Second Edition

Dictionary of DNA and GENOME TECHNOLOGY

Paul Singleton

WILEY-BLACKWELL

Dictionary of DNA and Genome Technology

Second Edition

Paul Singleton

WILEY-BLACKWELL
A John Wiley & Sons, Ltd., Publication

Dictionary of
DNA and Genome Technology

Second Edition

Dictionary of DNA and Genome Technology

Second Edition

Paul Singleton

WILEY-BLACKWELL
A John Wiley & Sons, Ltd., Publication

© 2010 Paul Singleton

First edition 2008
Second edition 2010

Wiley-Blackwell is an imprint of John Wiley & Sons, formed by the merger of Wiley's global Scientific, Technical and Medical business with Blackwell Publishing.

Registered office: John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK

Other Editorial Offices:
9600 Garsington Road, Oxford, OX4 2DQ, UK

111 River Street, Hoboken, NJ 07030-5774, USA

For details of our global editorial offices, for customer services and for information about how to apply for permission to reuse the copyright material in this book please see our website at www.wiley.com/wiley-blackwell

The right of the author to be identified as the author of this work has been asserted in accordance with the Copyright, Designs and Patents Act 1988.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, except as permitted by the UK Copyright, Designs and Patents Act 1988, without the prior permission of the publisher.

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic books.

Designations used by companies to distinguish their products are often claimed as trademarks. All brand names and product names used in this book are trade names, service marks, trademarks or registered trademarks of their respective owners. The publisher is not associated with any product or vendor mentioned in this book. This publication is designed to provide accurate and authoritative information in regard to the subject matter covered. It is sold on the understanding that the publisher is not engaged in rendering professional services. If professional advice or other expert assistance is required, the services of a competent professional should be sought.

Library of Congress Cataloguing-in-Publication Data

Record on file

ISBN 978-0-470-74732-2 (HB) 978-0-470-74731-5 (PB)

A catalogue record for this book is available from the British Library.

Set in 8/9 Times by Integra Software Services Pvt. Ltd., Pondicherry, India.

Printed in Great Britain by Antony Rowe Ltd., Chippenham, Wiltshire.

First Impression 2010

For Mobby

Contents

Preface	ix
Notes for the user	xi
Ready reference	xv
Dictionary	1
Appendix: alphabetical list of genera	405

Preface

This edition has been written to accomodate the recent spate of innovations and developments in DNA technology. It covers a wide range of new methods and adaptations, as well as terminology, and reflects current thinking in both *in vitro* and *in vivo* studies. The opportunity was also taken to update and extend entries from the 1st edition. Every effort has been made to present up-to-date information – much of it newly published in mainstream journals within the last 6–12 months.

The dictionary is designed for use throughout the biomedical sciences, particularly in areas such as:

- biotechnology
- diagnosis (hereditary and infectious diseases)
- drug development
- epidemiology
- forensic science
- gene therapy
- genetically modified (GM) foods
- genomics
- industrial enzymes
- microbiology
- molecular biology
- oncology
- systems biology
- taxonomy
- vaccine development

References to published papers and reviews are cited throughout the dictionary. Some of these are the source(s) of information on which particular entries are based; they can provide additional details (e.g. protocols), and they also permit the reader to make his or her own assessment of a given source. Other references are included in order to indicate further information which is relevant to a given entry.

Some pervasive topics – for example, PCR, gene fusion, forensics, phages, overexpression, retroviruses, typing, microarrays, stem cells – are treated more extensively. These essay-style entries are intended to offer the newcomer a broad working knowledge of the area. Such entries may also be of use to the over-specialized researcher.

Commercial systems and materials are widely used in DNA technology, and a range of entries describe these products. This kind of entry may be useful, for example, when a paper offers no information about

a given product other than "used according to the manufacturer's instructions" – leaving unanswered questions on the product, the protocol or the principle. Accordingly, these entries give brief overviews of products – and their uses – and in many cases they also cite relevant papers describing work in which a given product has been used. Where the names of products and systems are known to be trademarks this has been indicated.

Notes for the user on the following pages will facilitate use of the dictionary. The attention of the reader is drawn, in particular, to the item on alphabetization.

Paul Singleton
Clannaborough (UK), September 1st 2009

Notes for the user

Alphabetization

The headwords in any dictionary can be listed in either of two distinct ways:

1. According to the way in which the terms are actually written. This approach, which is called the 'word-by-word' approach, is used in this dictionary.
2. According to the way in which the terms *would* appear if all the spaces and hyphens etc. were deleted. This is the 'letter-by-letter' approach.

The ***order*** in which headwords are listed in any dictionary depends on the particular approach used; for example:

Approach 1
A site
A-tract
AAS
abacavir
abasic site
ABC excinuclease

bla gene
black–white screening
blue–white screening
Bluescript®
branch migration
branched DNA assay

col plasmid
colchicine
cos site
cosmid
Cre–*loxP* system
CREB

Approach 2
AAS
abacavir
abasic site
ABC excinuclease
A site
A-tract

black–white screening
bla gene
Bluescript®
blue–white screening
branched DNA assay
branch migration

colchicine
col plasmid
cosmid
cos site
CREB
Cre–*loxP* system

Note that a one-letter descriptor (such as the 'A' in 'A site') is treated as a word for the purposes of alphabetization in approach 1.

In practice, neither approach is foolproof. For example, in the first approach, the order in which a given term is listed may depend on whether or not a hyphen is regarded as a necessary part of the term. Thus, on deleting a hyphen – and closing-up the intervening space – the characters on either side of the hyphen become contiguous; in this case, the character which followed the hyphen is important as a primary determinant of alphabetical order. In the second approach, strict adherence to the basic 'letter-by-letter' rule would lead – for example – to the following order:

factor I	(1)
factor II	(2)
factor III	(3)
factor IV	(4)
factor IX	(9)
factor V	(5)
factor VI	(6)
factor VII	(7)
factor VIII	(8)
factor X	(10)

To some, this order may seem reasonable, even preferable. To others, it runs counter to common sense: Roman numerals are generally seen (and used) as the equivalent of numbers – and ought therefore to be arranged in numerical order.

When a Greek letter is a significant component of an entry heading – as e.g. in λ phage – it is treated as a word and is listed in the relevant alphabetical position indicated by the English name (i.e. alpha, lambda etc.). However, β-galactosidase and β-lactamases are listed under G and L, respectively; this rule applies also to entry headings starting with letters such as L-, *p*-, *N*-, *O*- etc. which precede the names of certain chemicals. In some cases a headword is given in *both* possible locations, with suitable cross-referencing; this has been done simply in order to assist readers.

Cross references

Words in SMALL CAPITALS refer the reader to entries elsewhere in the dictionary. Such cross references are included e.g. to extend the reader's knowledge into related fields or topics. Cross references may be particularly useful for directing the reader to allied, or parallel, subjects whose relationship to the entry being read may not be immediately obvious.

In some cases a complete understanding of a given entry, or a full appreciation of its context, depends on information contained in other entries – which are indicated by cross reference(s). Dictionaries are often arranged in this way because it avoids the need to repeat information. If it is especially important to follow-up a cross reference, then the cross reference is followed by '(q.v.)'. In other cases, in which the purpose of a cross reference is simply to link one topic with another, the cross reference may be preceded by 'See also. . ..' or 'cf.'.

External references

References to papers, articles or reviews in journals are given in square brackets. The names of journals are abbreviated to save space. The abbreviated journal name is followed by the year of publication, the volume number (and frequently the issue number), and page number(s). This information is sufficient to enable the reader to obtain any given reference. (When a paper has been cited as an advance publication it may be referred to by its digital object identifier (doi) number; this number allows the reference to be followed-up.)

Commercial products

Many of the commercial products are listed under their trade names. In general, these products are widely used in studies on all aspects of DNA-based technology and are cited in many research papers. It should be noted that the inclusion of any given product in the dictionary is not based on an evaluation of that product, and implies no comparison of that product with any similar product(s) marketed by other companies. It is obviously not possible to include every product currently on the market. Importantly, any details of a product given in the dictionary are those details which are to hand at the time of writing; companies are continually modifying and updating their products, so that the reader should refer to the manufacturer's literature for details of any modifications.

Ready reference

The Greek alphabet

A	α	alpha	N	ν	nu
B	β	beta	Ξ	ξ	xi
Γ	γ	gamma	O	o	omicron
Δ	δ	delta	Π	π	pi
E	ϵ	epsilon	P	ρ	rho
Z	ζ	zeta	Σ	σ	sigma
H	η	eta	T	τ	tau
Θ	θ	theta	Y	υ	upsilon
I	ι	iota	Φ	φ	phi
K	κ	kappa	X	χ	chi
Λ	λ	lambda	Ψ	ψ	psi
M	μ	mu	Ω	ω	omega

Amino acids

alanine	Ala	A	A	Ala	alanine
arginine	Arg	R	C	Cys	cysteine
asparagine	Asn	N	D	Asp	aspartic acid
aspartic acid	Asp	D	E	Glu	glutamic acid
cysteine	Cys	C	F	Phe	phenylalanine
glutamic acid	Glu	E	G	Gly	glycine
glutamine	Gln	Q	H	His	histidine
glycine	Gly	G	I	Ile	isoleucine
histidine	His	H	K	Lys	lysine
isoleucine	Ile	I	L	Leu	leucine
leucine	Leu	L	M	Met	methionine
lysine	Lys	K	N	Asn	asparagine
methionine	Met	M	P	Pro	proline
phenylalanine	Phe	F	Q	Gln	glutamine
proline	Pro	P	R	Arg	arginine
serine	Ser	S	S	Ser	serine
threonine	Thr	T	T	Thr	threonine
tryptophan	Trp	W	V	Val	valine
tyrosine	Tyr	Y	W	Trp	tryptophan
unknown	Xaa	X	X	Xaa	unknown
valine	Val	V	Y	Tyr	tyrosine

Prefixes used with SI (Système International) units

(prefix,	value,	symbol)	(value,	prefix,	symbol)
		FRACTIONS			
atto	10^{-18}	a	10^{-1}	deci	d
centi	10^{-2}	c	10^{-2}	centi	c
deci	10^{-1}	d	10^{-3}	milli	m
femto	10^{-15}	f	10^{-6}	micro	μ
micro	10^{-6}	μ	10^{-9}	nano	n
milli	10^{-3}	m	10^{-12}	pico	p
nano	10^{-9}	n	10^{-15}	femto	f
pico	10^{-12}	p	10^{-18}	atto	a
yocto	10^{-24}	y	10^{-21}	zepto	z
zepto	10^{-21}	z	10^{-24}	yocto	y
		MULTIPLES			
deca	10	da	10	deca	da
exa	10^{18}	E	10^{2}	hecto	h
giga	10^{9}	G	10^{3}	kilo	k
hecto	10^{2}	h	10^{6}	mega	M
kilo	10^{3}	k	10^{9}	giga	G
mega	10^{6}	M	10^{12}	tera	T
peta	10^{15}	P	10^{15}	peta	P
tera	10^{12}	T	10^{18}	exa	E
yotta	10^{24}	Y	10^{21}	zetta	Z
zetta	10^{21}	Z	10^{24}	yotta	Y

Micro-measurements

1 Å (Ångström unit) = 10^{-1} nm = 10^{-4} μm = 10^{-10} m
1 nm (nanometer) = 10^{-3} μm = 10^{-6} mm = 10^{-9} m
1 μm (micrometer, formerly micron) = 10^{-3} mm = 10^{-6} m
1 mm = 10^{-1} cm = 10^{-3} m = 10^{-6} km

A

A (1) Adenine (as a base, or the corresponding nucleoside or nucleotide).
(2) L-Alanine (alternative to Ala).

Å Ångström unit, 10^{-10} m; a unit of length used e.g. to indicate intermolecular distances.

A_{260} See the entry ULTRAVIOLET ABSORBANCE.

A box The adenine riboswitch aptamer (see RIBOSWITCH).

A-DNA One of the conformations adopted by dsDNA: a right-handed helix with ∼11 base-pairs per turn.
(Note that aDNA is used to refer to ANCIENT DNA.)
(cf. B-DNA and Z-DNA.)

A family (of DNA polymerases) A group of DNA-DEPENDENT DNA POLYMERASES that include prokaryotic, eukaryotic and viral enzymes. Members of the A family include some phage polymerases (although not those from phages φ29 or T4) and the *Escherichia coli* pol I (involved e.g. in the maturation of Okazaki fragments and in BASE EXCISION REPAIR).
Also included in this family is POLQ (= pol θ; pol theta), an enzyme found in human and other eukaryotic cells. POLQ is able to carry out translesion synthesis of DNA. It may participate in base excision repair, a suggestion supported by the *in vitro* demonstration of its 5′-deoxyribose phosphate lyase activity, a role apparently involved in single-nucleotide base excision repair [Nucleic Acids Res (2009) 37(6):1868–1877].
(See also B FAMILY, X FAMILY and Y FAMILY.)

A site (of a ribosome) The aminoacyl or 'acceptor' site at which tRNA molecules carrying the second and subsequent amino acids bind during translation. (cf. P SITE.)

A-tract In genomic DNA: a nucleotide motif that is reported to be associated with regions of the most pronounced curvature of the molecule; an A-tract is a poly(A) (i.e. poly-adenosine) sequence. In the genome of *Escherichia coli*, A-tracts were reported to be distributed 'quasi-regularly', in both coding and non-coding sequences; the A-tracts occur in clusters ∼100 bp long, with consecutive A-tracts exhibiting a periodicity of 10 to 12 bp. It was suggested that the clusters of A-tracts may constitute a form of 'structural code' for DNA compaction in the NUCLEOID [Nucleic Acids Res (2005) 33:3907–3918].
Studies on the mechanics and dynamics of DNA suggested a rationale, incorporating A-tracts, for the stable bending of DNA [Nucleic Acids Res (2008) 3:2268–2283].
Studies on *eukaryotic* genomes have reported that A-tracts are absent specifically in those coding sequences (exons) that correspond to the locations of nucleosomes. It was concluded that the pattern of absence/presence of A-tracts in the genome constitutes a code for the presence/absence – respectively – of nucleosome locations. [The coexistence of the nucleosome positioning code with the GENETIC CODE on eukaryotic genomes: Nucleic Acids Res (2009) doi: 10.1093/nar/gkp689.]
(cf. CLASS A FLEXIBLE PATTERNS.)

AAA ATPases 'ATPases associated with diverse cellular activities' – ATPases which are found in various locations, such as proteasomes and peroxisomes. They have been categorized as AAA+ PROTEINS.

AAA+ proteins A family of NTPases whose members include proteins with diverse functions; AAA ATPASES are examples of this group.
[Review of AAA+ proteins: Genome Biol (2008) 9(4): 216.]

AAAVs Avian adeno-associated viruses (see the entry AAVS).

AAS Aminoalkylsilane (3-aminopropyltriethoxysilane; APES): a reagent used e.g. to bind tissue sections to glass (for *in situ* hybridization etc.).
[Uses (e.g.): Am J Pathol (2006) 169(1):258–267; Nucleic Acids Res (2008) 36(16):5335–5349.]

***aat* gene** In *Escherichia coli*: a gene encoding the enzyme that catalyzes addition of a leucine or phenylalanine residue to the N-terminal of proteins that are synthesized with either an N-terminal arginine or a lysine residue; such addition facilitates degradation of the protein.
(See also N-END RULE.)

AatII A RESTRICTION ENDONUCLEASE from *Acetobacter aceti*. Recognition sequence/cutting site: GACGT↓C.

AAUAAA In a pre-mRNA: a polyadenylation signal upstream of the site at which the molecule is cut and polyadenylated; the polyadenylation sequence is similar in various organisms, although there are variations.
Other *cis*-acting elements may have roles in regulating the polyadenylation of human mRNAs – including upstream U-rich sequences similar to those which have been identified in yeast and plants.
As well as acting as a polyadenylation signal, this sequence was reported to affect the *rate* of transcription [RNA (2006) 12(8):1534–1544].

AAV Adeno-associated virus: see the entry AAVS.

AAV Helper-Free System A commercial gene-delivery system (Stratagene, La Jolla CA) in which the genes in two plasmids provide functions necessary for production of infective AAV virions (see AAVS) without the need for a helper virus; these virions are used to deliver genes to target cells within which viral DNA – containing the gene of interest – integrates in the host cell's DNA.
Essentially, the gene/fragment of interest is first cloned in a plasmid cloning vector in which the insert is bracketed by a pair of inverted terminal repeats (ITRs) which are necessary for subsequent viral packaging. This plasmid is then used to transfect PACKAGING CELLS – which are *co-transfected* with two other plasmids: (i) a plasmid containing the genes that encode viral capsid and replication functions, (ii) a plasmid containing genes that encode the lytic phase of AAV. The resulting infective (*but still replication-deficient*) virions that are produced in the packaging cells can then be used to infect the required target cells (in which the gene of interest can be expressed).
This GENE-DELIVERY SYSTEM has been used e.g. to express

siRNAs [Mitochondrion (2007) 7(4):253–259]; to deliver an anti-angiogenic gene (for investigating age-related macular degeneration) [Mol Vision (2008) 14:471–480]; and to study some features of food/energy metabolism [J Neurosci (2009) 29(1):179–190].

(See also: VIRAPORT RETROVIRAL GENE EXPRESSION SYSTEM and VIRAPOWER LENTIVIRAL EXPRESSION SYSTEM.)

AAVs Adeno-associated viruses (also known as: adeno-satellite viruses): defective viruses that are able to replicate only when certain functions are provided by a co-infecting *helper virus* (adenovirus or herpesvirus) – or, in certain *in vitro* systems, when these functions are provided by plasmid-borne genes (as e.g. in the AAV HELPER-FREE SYSTEM).

Functions provided by adenovirus type 5 (for AAV type 5) include both positive and negative effects. For example, the E4Orf6 function (involved in replication of AAV5 genomic DNA) – together with E1b – degrades AAV5 capsid proteins and Rep52 [J Virol (2007) 81(5):2205–2212]. The functions provided by herpes simplex virus type 1 (for the early stages of AAV replication) were reported to involve nine proteins from the helper virus [PLoS Pathog (2009) 5(3):e1000340].

The AAVs are parvoviruses in which the genome is linear ssDNA. Positive and negative strands of the viral DNA are encapsidated in separate virions.

The AAVs infect a wide range of vertebrates. Initial stages of infection, including internalization of DNA, occur without a helper virus. [Cloning an *avian* AAV (an AAAV) and the generation of recombinant AAAVs: J Virol (2003) 77:6799–6810.]

AAVs are used, for example, in GENE THERAPY. Efforts are being made to increase the efficacy of AAV vectors in gene therapy by designing the CAPSID on the basis of e.g. information obtained from studies on the naturally occurring capsid variants of AAVs in mammals [see: Gene Therapy (2009) 16: 311–319]. (See also KU70.)

An inducible and highly efficient system was reported for the production of recombinant AAV vectors in insect (Sf9) cells [Proc Natl Acad Sci USA (2009) 106(13):5059–5064].

AAV vectors, encoding genes of the α and the β subunits of hexosaminidase, were inoculated, *intracranially*, into mice in order to assess the potential of gene therapy for treatment of the human GM2 gangliosidoses such as Tay–Sachs disease and Sandhoff disease [Proc Natl Acad Sci USA (2006) 103 (27):10373–10378]. A simpler method for delivering genes to brain cells was reported later (see below).

AAV9 has been used, in mice, for gene delivery to cells of the central nervous system (brain and spinal cord) *by intravenous injection*. It was thought that this approach may allow the development of gene therapy for e.g. some human neurodegenerative diseases [Nature Biotechnol (2008) 27:59–65].

AAV vectors were also used for the genetic manipulation of cultured neurons [Brain Res (2008) 1190:15–22].

It was reported earlier that, in human cells, AAV DNA (in the absence of helper virus) integrates in the genome with an apparent preference for CPG ISLANDS. More recently, AAVs have been reported to integrate, site-specifically, into a locus on chromosome 19, and the occurrence of such integration is apparently influenced by the TRP-185 protein [J Virol (2007) 81(4):1990–2001]. Palindromes of length greater than about 40 bp are reported to be significant targets for the integration of recombinant AAV vectors [J Virol (2007) 81(20):11290–11303].

The site of insertion of AAVs within chromosome 19 was reported to contain a 347-bp sequence capable of enhancing the promoter and transcriptional functions of AAV vectors in liver cells; inclusion of this small fragment in AAV vectors may thus facilitate their use for the delivery and expression of transgenes in liver cells [Gene Therapy (2009) 16:43–51].

AB1380 A strain of the yeast *Saccharomyces cerevisiae* (see the entry SACCHAROMYCES for some details).

(See also YEAST ARTIFICIAL CHROMOSOME.)

abacavir A NUCLEOSIDE REVERSE TRANSCRIPTASE INHIBITOR used e.g in antiretroviral therapy; CSF–plasma ratios indicate that it may reach therapeutic levels in the cerebrospinal fluid (CSF).

A trial that compared abacavir with nevirapine (as part of a combined therapy) reported that abacavir tended to produce a lower rate of serious adverse effects, suggesting a wider use of this drug in resource-limited settings [Trop Med Int Health (2008) 13(1):6–16].

abasic site *Syn*. AP SITE.

abasic-site mimic See the entry RPA.

ABC excinuclease See UVRABC-MEDIATED REPAIR.

Abelson murine leukemia virus See the entry ABL.

aberrant RNA (aRNA) See the entry ARNA (sense 2).

abl (*ABL*) An ONCOGENE first identified in the Abelson murine leukemia virus. The v-*abl* product has TYROSINE KINASE activity. The human homolog of v-*abl*, c-*abl*, is usually present on chromosome 9; however, in the majority of patients with CHRONIC MYELOGENOUS LEUKEMIA it has been translocated to chromosome 22, forming a chimeric gene, known as *bcr-abl*, that encodes a tumor-specific tyrosine kinase (designated P210). Chromosome 22 containing the chimeric *bcr-abl* gene is called the Philadelphia chromosome (also called Ph^1).

Subcellular localization of c-Abl protein at an early stage in myogenic differentiation was reported to be influenced by its acetylation [EMBO Rep (2006) 7(7):727–733].

abortive transduction TRANSDUCTION in which the transduced DNA persists in a recipient cell as a stable, extrachromosomal but non-replicating molecule; when the recipient divides only one daughter cell receives the DNA fragment.

absorbance (ultraviolet) See ULTRAVIOLET ABSORBANCE.

abzyme *Syn*. CATALYTIC ANTIBODY.

Abzyme® A reagent kit (Abbott Laboratories) used for detecting antibodies in the context of hepatitis B.

acceptor splice site (acceptor splice junction) In a pre-mRNA: the splice site (consensus AG) at the 3′ end of an intron.

(cf. DONOR SPLICE SITE.)

accession number A number which refers to a unique database entry for a given sequence or gene. Some examples include:

(i) GenBank® accession number X17012, referring to data on the gene for rat insulin-like growth factor II (IGF II); (ii) GenBank® accession number AY024353, referring to data on the *ftsZ* gene of the bacterium *Sodalis glossinidius;* (iii) GenBank® accession number AM160602, referring to data on mRNA of the gene for cinnamyl alcohol dehydrogenase in a species of oak (*Quercus ilex*).

(See also ANNOTATION.)

AccuPrime^{TM} GC-rich DNA polymerase A DNA polymerase (Invitrogen, Carlsbad CA) optimized for DNA synthesis on 'difficult-to-amplify' templates, including those with a GC content >65%. Targets up to 5 kb may be amplified with this polymerase.

[Uses (e.g.): J Bacteriol (2008) 190(24):8096–8105; J Exp Clin Cancer Res (2008) 27:54; FEMS Microbiol Lett (2009) 294(1)32–36.]

AccuProbe® A family of PROBES (Gen-Probe, San Diego CA) used for identifying certain medically important bacteria by detecting specific sequences of nucleotides from lysed cells. The method involves a *hybridization protection assay*. In this assay, an added reagent cleaves the acridinium ester label on all *unbound* probes. Labels on the *bound* probes (which are protected from cleavage by virtue of their position in the probe–target duplex) react with a second reagent, producing a chemiluminescent (light) signal. The light produced by this reaction is measured in RLUs (i.e. relative light units). The threshold value (in RLUs) for a positive result must be carefully examined [see for example: J Clin Microbiol (2005) 43:3474–3478].

[Use for *Staphylococcus aureus*: J Clin Microbiol (2008) 46(6):1989–1995. Use for *Streptococcus pneumoniae* (as a reference): J Clin Microbiol (2008) 46(7):2184–2188. Use for *Mycobacterium avium*: J Clin Microbiol (2008) 46 (8):2790–2793. Use for identifying *Mycobacterium* spp: Emerg Infect Dis (2009) 15(1):53–55, and Emerg Infect Dis (2009) 15(2):242–249.]

(See also PACE 2C and TMA.)

acetosyringone A phenolic substance which promotes activity of the *vir* operon in species of the plant-pathogenic bacterium *Agrobacterium* (see CROWN GALL).

(See also AGROINFILTRATION.)

Acetosyringone has been used e.g. for studies on terpenoid metabolism in the tomato plant [Plant Physiol (2009) 149(1): 499–514], and studies on the transformation of wheat [Plant Cell Rep (2009) 28(6):903–913].

Agrobacterium can also transfer T-DNA to other types of cell, including e.g. human and fungal cells; acetosyringone was used to promote transfer of T-DNA from *Agrobacterium* to the fungus *Aspergillus fumigatus* for (random) insertional mutagenesis [PLoS ONE (2009) 4(1):e4224].

***N*-acetyl-L-cysteine** See MUCOLYTIC AGENT.

acetylation (of histones) HISTONE acetylation is regulated e.g. by the opposing effects of histone acetyltransferases (HATs) and histone deacetylases (see HDAC); the (de)acetylation of histones can affect CHROMATIN structure, and may therefore alter the accessibility of DNA for events such as transcription and repair.

The acetylation of histones can be studied/manipulated e.g. by using HDAC *inhibitors* (e.g. TRICHOSTATIN A).

[Genome-wide analysis of histone acetylation and its effect on gene expression in (the protozoan) *Entamoeba histolytica*: BMC Genomics (2007) 8:216.]

A general perception is that transcription of genes requires – as a pre-condition – an 'open' form of CHROMATIN (the so-called *euchromatin*) in the vicinity of the given genes; acetylation of vicinal histone(s) is usually regarded as an important factor associated with the presence of euchromatin. (In some types of chemotherapy – EPIGENETIC THERAPY – an inhibitor of HDACs is sometimes included in order to promote 'open' chromatin in the vicinity of specific gene(s) with the object of contributing to de-repression of the genes.) However, histone acetylation is only one factor that regulates gene expression; for example, it was reported that the drug-induced formation of 'open' chromatin (involving hyper-acetylation of vicinal histone(s)) was not, on its own, sufficient to de-repress lytic-cycle genes in the Epstein–Barr virus [J Virol (2008) 82(10): 4706–4719]. Nevertheless, the complexity of this issue may be indicated by a study in which histone acetylation – but not DNA demethylation – *was* found to be sufficient to break the latency of gamma-herpesvirus 68 in a mouse cell line [PLoS ONE (2009) 4 (2):e4556].

In a genomewide study of HDACs in *Schizosaccharomyces pombe* (a fission yeast), the patterns of histone acetylation, HDAC binding and nucleosome density were compared with gene expression profiles; it was found that different HDACs may have different roles in repression and activation of genes [EMBO J (2005) 24(16):2906–2918]. Following *damage* to DNA in *S. pombe*, the restoration of chromatin structure was reported to involve deacetylation of histone H3 by Hst4 (a putative HDAC) [Eukaryotic Cell (2008) 7:800–813], while recovery from DNA damage was reported to involve Mst1 (a histone acetyltransferase) [Genetics (2008) 179(2):757–771].

In (human) nucleosomes, the acetylation of certain lysine residues depends primarily on HATs, but the effect of these enzymes appears to be promoted by binding protein HMGN1 [EMBO J (2005) 24(17):3038–3048].

Acetylation of the histone *chaperone* NUCLEOPHOSMIN, as well as histone acetylation, apparently promotes transcription [Mol Cell Biol (2005) 25(17):7534–7545], while chaperone-stimulated, histone-acetylation-independent transcription has also been reported [Nucleic Acids Res (2007) 35:705–715].

HDACs in (human) development and physiology have been reviewed within the context of implications for disease and therapy [Nature Rev Genetics (2009) 10:32–42].

The c-Abl protein (see ABL) was reported to be a substrate for the p300 and other histone acetyltransferases.

***N*-acetylmuramidase** See LYSOZYME.

***N*-acetylneuraminic acid** (NANA) See NEURAMINIDASE.

ACF APOBEC-1 complementation factor: see RNA EDITING.

Achilles' heel technique A technique in which a RESTRICTION ENDONUCLEASE is targeted to one *particular* recognition site when multiple copies of that site are freely available. In one method, a triplex-forming oligonucleotide (see TRIPLEX DNA) is used to mask the required cleavage site. While this site is masked, the remaining sites are methylated in order to inhibit subsequent cleavage; the triplex is then removed and specific cleavage can be carried out.

(See also PROGRAMABLE ENDONUCLEASE.)

aciclovir Alternative spelling for ACYCLOVIR.

acid-fast bacilli Those bacilli (i.e. rod-shaped bacteria) which, when stained with the Ziehl–Neelsen (or similar) stain, resist decolorization with mineral acid or an acid–alcohol mixture. This kind of staining method is used for screening respiratory specimens, e.g. samples of sputum, and for examining other types of specimen, for *Mycobacterium tuberculosis* (an acid-fast species).

Ac*M*NPV *Autographa californica* NPV: see NUCLEAR POLYHEDROSIS VIRUSES.

AcNPV *Syn*. Ac*M*NPV – see entry NUCLEAR POLYHEDROSIS VIRUSES.

acridines Heterocyclic, fluorescent compounds which bind to dsDNA (primarily as an INTERCALATING AGENT) and also to single-stranded nucleic acids (and to the backbone chains of double-stranded nucleic acids). Acridines have antimicrobial activity and they are mutagenic; they are also used as stains for nucleic acids and can be used for CURING plasmids.

acridinium ester label (on probes) See ACCUPROBE.

acrocentric Refers to a CHROMOSOME in which the CENTROMERE is located close to one end.

acrydite hybridization assay An assay in which molecules of labeled ssDNA or ssRNA, passing through a polyacrylamide gel by electrophoresis, are captured (bound) by complementary oligonucleotides immobilized in a (central) 'capture zone' within the gel; all the molecules of nucleic acid that are *not* complementary to the capture oligos pass through the central capture zone and continue their migration to the end of the gel strip. The complementary oligos are synthesized with a 5′ terminal acrydite group which binds them to the polyacrylamide matrix so that they are immobilized in the gel. (Note that the central region of the gel strip is prepared separately.)

acrylamide A toxic, water-soluble agent (CH_2=CH—$CONH_2$) which can be polymerized to POLYACRYLAMIDE by catalysts such as *N*,*N*′-methylene-*bis*-acrylamide ('Bis') which promote cross-linking.

actinomycin C_1 *Syn*. ACTINOMYCIN D.

actinomycin D An antibiotic (a substituted phenoxazone linked to two pentapeptide lactone rings) produced by some species of *Streptomyces;* it acts as an INTERCALATING AGENT, binding to DNA and inhibiting DNA-dependent RNA polymerase. The drug has low affinity for AT-rich promoter regions; hence, *initiation* of transcription from such promoters may be little affected by the antibiotic.

activation domain (AD) See YEAST TWO-HYBRID SYSTEM.

activation-induced cytidine deaminase (AID) An enzyme that occurs in germinal center B lymphocytes (B cells) and which is an absolute requirement for affinity maturation and class switching in normal development of antibodies.

The (autosomal recessive) form of HYPER-IGM SYNDROME has been linked to a deficiency of AID (see the table in entry GENETIC DISEASE).

AID also inhibits retrotransposition of L1 – suggesting that it has function(s) in addition to its role of creating antibody diversity [Nucleic Acids Res (2009) 37(6):1854–1867].

(See also CYTIDINE DEAMINASE AND RNA EDITING.)

activation/regulation of genes (*DNA technol*.) See e.g. entry CONDITIONAL GENE ACTIVATION/REGULATION.

activity-based probe A type of probe used for the real-time study of APOPTOSIS (q.v.).

acyclonucleotide Any analog of a deoxyribonucleotide or ribonucleotide in which a non-cyclic moiety carries the base. One example is a monomer of glycerol nucleic acid (see the entry GNA). Polymerization of certain acyclonucleotides on a DNA template has been achieved with THERMINATOR DNA POLYMERASE.

acyclovir (alternative spelling: aciclovir) 9-(2-hydroxyethoxymethyl)guanine: an antiviral agent which is active against a number of herpesviruses, including herpes simplex. In cells, acyclovir is phosphorylated to the monophosphate by (viral) thymidine kinase; subsequently it is converted to the (active) triphosphate form via host-encoded enzymes. The active drug inhibits *viral* DNA polymerase; the host cell's polymerase is much less sensitive.

In cells which are not virally infected, acyclovir appears not to be significantly phosphorylated.

Acyclovir has been used topically and systemically.

***N*-acyl-homocysteine thiolactone** See QUORUM SENSING.

***N*-acyl-L-homoserine lactone** (AHL) See QUORUM SENSING.

acylneuraminyl hydrolase *Syn*. NEURAMINIDASE.

AD primer (arbitrary degenerate primer) See TAIL-PCR.

Ada protein (in *Escherichia coli*) See DNA REPAIR.

adaptamer See ORFMER SETS.

adaptive response (to alkylating agents) See DNA REPAIR.

adaptor A short, synthetic, double-stranded fragment of DNA which is similar, in principle, to a LINKER but which generally offers more flexibility. Thus, for example, the two ends of a given adaptor may consist of dissimilar STICKY ENDS – one end able to bind to a (complementary) sticky end on a DNA fragment and the other able to bind to a different sticky end on a vector molecule, facilitating the integration of fragments and vectors which were cut by different restriction enzymes. An adaptor may also include one or more internal restriction sites, offering the chance to select alternative sticky ends at a later stage in the work, and there may also be primer-binding sites and/or a promoter sequence etc.

(See also NOTI.)

ADAR1 A dsRNA ADENOSINE DEAMINASE which is involved e.g. in RNA EDITING.

(See also Zα in the entry Z-DNA.)

AdEasy™ XL adenoviral vector system A GENE-DELIVERY

SYSTEM (marketed by Stratagene, La Jolla CA, USA) which uses ADENOVIRUS-based vectors. Essentially, it facilitates the preparation of a recombinant adenoviral vector containing the gene/fragment of interest.

Initially, the gene/insert is cloned in a small shuttle vector (~7 kb) which includes: (i) the left and right ITRs (inverted terminal repeats) of the adenovirus genome; (ii) two regions homologous to two sequences in another plasmid, pAdEasy-1 (see later), (iii) a gene encoding resistance to kanamycin; and (iv) a recognition site for the restriction endonuclease PmeI.

After cloning, the shuttle vector is linearized (by cleavage with PmeI); linearization leaves the two homologous regions (see above) in terminal positions.

The linearized shuttle vector is inserted, by transformation, into a strain of *Escherichia coli*, BJ5183-AD-1, that already contains the (circular) plasmid vector pAdEasy-1. pAdEasy-1 (~33 kb) includes modified genomic DNA of human adenovirus serotype 5 (containing deletions in both the E1 and E3 regions). Homologous recombination occurs (within *E. coli*) between the linearized shuttle vector and homologous regions in pAdEasy-1. Cells containing the recombinant plasmids are selected on kanamycin-containing media.

The recombinant plasmids are isolated, and the next stage is conducted *in vitro*.

Recombinant plasmids are cleaved by a restriction enzyme, PacI, at selected sites, yielding a linear construct with adenoviral terminal sequences (i.e. ITRs). This construct is used to transfect specialized, competent AD-293 cells – within which infective adenovirus virions (containing the gene/fragment of interest) are produced; the virions are released for subsequent use in gene-expression studies in mammalian cells.

The principle of the AdEasy™ system was exploited in the production of oncolytic adenovirus [BMC Biotechnol (2006) 6:36].

The AdEasy™ XL vector system has been used in a range of studies [see e.g. Nucleic Acids Res (2006) 34(12):3577–3584; Vaccine (2007) 25(52):8687–8701; Mol Ther (2008) 16(5):886–892; J Clin Invest (2009) 119:976–985; Cell Cycle (2009) 8(2):257–267; Biochim Biophys Acta (2009) 1793(8): 1379–1386; J Exp Clin Cancer Res (2009) 28(1):75; BMC Immunol (2009) 10:14].

adefovir A NUCLEOSIDE REVERSE TRANSCRIPTASE INHIBITOR.

adenine phosphoribosyltransferase An enzyme (EC 2.4.2.7) which forms adenosine monophosphate (AMP) from adenine and 5-phosphoribosyl-1-diphosphate.

In humans, a deficiency of adenine phosphoribosyltransferase (an autosomal recessive disorder) can cause excretion of adenine (in the urine) and the formation of a highly insoluble product, 2,8-dihydroxyadenine, which can give rise to kidney stones and renal failure.

adeno-associated viruses See AAVS.

adeno-satellite viruses See AAVS.

adenosine A RIBONUCLEOSIDE.

adenosine deaminase An enzyme (EC 3.5.4.4) which catalyzes the conversion of adenosine to inosine.

(See also ADAR1.)

adenosine deaminase deficiency The congenital deficiency of a functional adenosine deaminase (EC 3.5.4.4) characterized by lack of normal development of T cells and a (consequent) marked immunodeficiency in which the patient is susceptible to infection by opportunist pathogens. This disorder has been treated successfully by GENE THERAPY.

(See also GENETIC DISEASE (table).)

***S*-adenosyl-L-methionine** A methyl group (CH_3–) donor that is used in various reactions.

Adenovirus A genus of icosahedral, non-enveloped viruses of the family Adenoviridae; the genome is linear dsDNA. These viruses infect mammals and birds; each type of adenovirus is usually specific for one, or a limited range, of closely related host species.

Adenovirus pathogenesis (in humans) commonly involves respiratory-tract infections, while some adenoviruses are able to induce tumors in rodents (rats). The *oncogenic* potential of adenoviruses has been investigated by using Ad5 adenovirus *dl520* to infect human U2OS cells, and monitoring levels of the Myc protein (see MYC); it was reported that the (virus-encoded) E1A protein interacts with the cell's p400, resulting in the stabilization of Myc and induction of Myc target genes [Proc Natl Acad Sci USA (2008) 105(16):6103–6108)]. (See also HEPATOCELLULAR CARCINOMA SUPPRESSOR 1.)

The adenovirus virion is ~70–90 nm in diameter; the capsid encloses a core containing genomic DNA (which is closely associated with an arginine-rich polypeptide). The 5′ end of each strand of the DNA is covalently linked to a hydrophobic 'terminal protein' (TP).

Both ends of the genomic DNA are characterized by an inverted terminal repeat (ITR) – which varies in length in the different types of adenovirus; the 5′ end residue is commonly dCMP.

During infection, the core enters the nucleus, releasing viral DNA. Replication of viral DNA involves TP and also a virus-encoded DNA polymerase, as well as other virus- and host-encoded proteins. TP, synthesized in precursor form, binds covalently to DNA during replication and is later cleaved to the mature (DNA-bound) TP. A TP-mediated form of DNA replication also occurs in PHAGE φ29 (q.v. for details).

Expression of late viral genes, encoding structural proteins, is accompanied by the cessation of cellular protein synthesis.

Some 10^5 virions may be formed within a single host cell.

Adenoviruses as vectors

Adenoviruses are employed as VECTORS in various types of investigation, including GENE THERAPY. (See also the entry ADEASY XL ADENOVIRAL VECTOR SYSTEM.) Re-targeting of adenovirus type 5 vectors to cell-surface $\alpha_v\beta 6$ integrin molecules resulted in reduced hepatotoxicity and improved uptake by tumor cells following systemic delivery [J Virol (2009) 83 (13):6416–6428].

Titers of recombinant adenoviruses in packaging cells may be optimized in various ways: see e.g. PACKAGING CELL.

Variable genome size in replication-deficient vectors may

affect viral stability [J Virol (2009) 83(4):2025–2028].

[Vaccine vectors: Mol Therapy (2009) 178:1333–1339.]

Adenovirus dodecahedron base

A construct, which consists of twelve copies of a pentameric adenoviral capsid protein, involved in the penetration of cells, has been used as a vehicle for increasing the uptake, by cells, of the antitumor agent bleomycin (which was tethered to the construct) [PLoS ONE (2009) 4(5):e5569].

adenoviruses Viruses of (i) the genus ADENOVIRUS – *or* (ii) the family Adenoviridae.

adenylate cyclase An enzyme (EC 4.6.1.1) which catalyzes the conversion of ATP to CYCLIC AMP.

In *Escherichia coli*, the activity of adenylate cyclase (the *cya* gene product) is regulated e.g. in association with the CATABOLITE REPRESSION system.

(See also BACTERIAL TWO-HYBRID SYSTEM.)

In mammals, the enzyme forms part of a plasma membrane complex and is regulated e.g. via certain G proteins; it is activated by some bacterial exotoxins (e.g. PERTUSSIS TOXIN).

Anthrax toxin (EF component) and *cyclolysin* (a virulence factor synthesized by the Gram-negative bacterial pathogen *Bordetella pertussis*) both exhibit adenylate cyclase activity, which is stimulated by CALMODULIN.

adenylate kinase An enzyme (EC 2.7.4.3) which catalyzes the (reversible) conversion: 2ADP⟷ATP+AMP.

aDNA *Ancient* DNA – see the entry ANCIENT DNA.

ADO ALLELE DROP-OUT.

AdoMet Abbreviation for *S*-adenosyl-L-methionine.

ADP-ribosylation The transfer, to a protein, of an ADP-ribosyl group from NAD^+, mediated by ADP-ribosyltransferase (EC 2.4.2.30).

In eukaryotes ADP-ribosylation can (for example) regulate the properties of HISTONES.

In *Escherichia coli*, RNA polymerase is ADP-ribosylated (with change in activity) following infection with bacteriophage T4.

ADP-ribosylation is an intracellular effect of some bacterial exotoxins (e.g. cholera toxin and PERTUSSIS TOXIN).

Polymerized ADP-ribosyl subunits (up to 50) may be found on certain eukaryotic proteins.

adult stem cell See STEM CELL.

Aedes aegypti A species of mosquito which can act as a vector in the transmission of e.g. dengue and yellow fever [genome: Science (2007) 316(5832):1718–1723].

(See also ARBOVIRUSES.)

It was reported that NUMTS are prevalent in the genome of *A. aegypti*, suggesting that future phylogenetic or population genetic studies should be carried out with nuclear, rather than mitochondrial, DNA markers [BMC Genet (2009) 10:11].

affinity capture electrophoresis Electrophoresis in a medium containing immobilized capture probes; it is used e.g. for the isolation of a given fragment of ssDNA, or a fragment of triplex-forming dsDNA.

(See also ACRYDITE HYBRIDIZATION ASSAY.)

affinity chromatography Chromatography in which particular molecules are isolated (adsorbed) owing to their affinity for an immobilized ligand – any non-specific unbound molecules being removed from the immobilized matrix. This procedure may be used e.g. for isolating/purifing a given type of molecule (see e.g. GENE FUSION (uses)).

Affinity® protein expression and purification A product of Stratagene (La Jolla CA) designed to facilitate the expression and purification of proteins expressed in prokaryotic systems; the product includes various pCAL plasmid vectors, shown in the diagram – each encoding a CALMODULIN-binding affinity tag.

Protein expression is maximized by a vector that includes a T7/LacO promoter system. In suitable strains of *Escherichia coli* (e.g. BL21(DE3)), T7 RNA polymerase is expressed in the presence of the inducer IPTG and drives expression from the T7/LacO promoter system on the plasmid. Tight control of expression is achieved with a plasmid-borne copy of $lacI^q$. Efficient translation of the protein of interest is promoted by using the strong ribosome-binding site (RBS) of T7 gene 10.

pCAL vectors contain a ColE1 origin of replication and an ampicillin-resistance gene.

All pCAL vectors encode a CALMODULIN-binding peptide (CBP) tag which forms a fusion product with the expressed protein and permits high-level purification following a single passage through CALMODULIN-AFFINITY RESIN. The (small) size of the CBP tag (about 4 kDa) may be expected to have a smaller effect on the protein of interest compared with larger tags – such as the (26-kDa) glutathione *S*-transferase (GST) affinity tag.

One of the pCAL vectors includes a KEMPTIDE SEQUENCE which can be used e.g. for *in vitro* labeling of the expressed protein with protein kinase A (PKA) and ^{32}P.

All of the pCAL vectors include a cleavage site (for enterokinase or thrombin proteinase) for removal of the CBT tag.

One of the pCAL vectors includes a FLAG sequence.

Affinity® vectors were used e.g. in studies on a fluorescent reporter protein [Proc Natl Acad Sci USA (2008) 105:20038–20043]; a light-activated DNA-binding protein [Proc Natl Acad Sci USA (2008) 105(31):10709–10714; and HMG box proteins [Mol Endocrinol (2008) 22(5):1141–1153].

affinity resin See e.g. NICKEL-CHARGED AFFINITY RESIN.

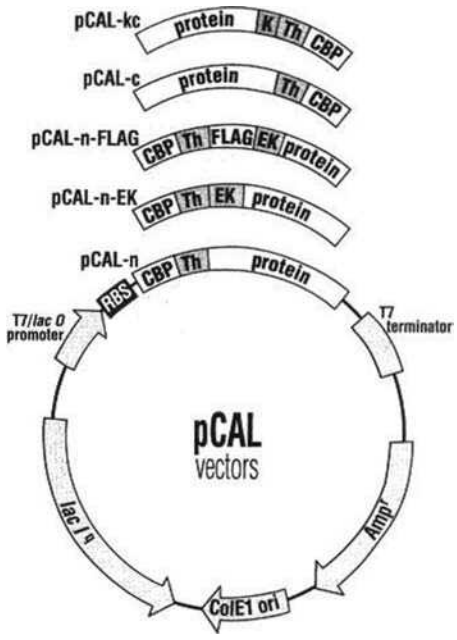
affinity tag *Syn*. AFFINITY TAIL.

affinity tail (affinity tag) Part of a FUSION PRODUCT which can facilitate detection/isolation of a (recombinant) protein e.g. by AFFINITY CHROMATOGRAPHY or by the use of an affinity resin (such as a NICKEL-CHARGED AFFINITY RESIN).

Some affinity tails are small peptides. One advantage of a small affinity tail is that it is less likely to interfere with the function of the fusion protein – so that its removal may not be necessary.

(See also FLAG, PESC VECTORS and SIX-HISTIDINE TAG.)

Large (protein) tails, for example glutathione *S*-transferase, may improve the solubility of the fusion protein but they may need subsequent removal in order to avoid interference with the function of the recombinant target protein.



Affinity® PROTEIN EXPRESSION AND PURIFICATION The range of pCAL vectors (see entry for details of the method).CBP (in each vector) refers to calmodulin-binding peptide. EK = enterokinase; K = Kemptide sequence; Th = thrombin proteinase. (See also entry for FLAG.)

Courtesy of Stratagene, La Jolla CA, USA.

(See also CHAMPION PET SUMO VECTOR.)

A highly temperature-stable affinity tail (a lectin, stable up to 80°C) may be useful for proteins originating from thermophilic organisms; the fusion proteins bind specifically to an agarose matrix that contains D-mannose, and the affinity tail can by cleaved by an enterokinase [BioTechniques (2006) 41 (3):327–332].

[Affinity as a tool in life science (a review): BioTechniques (2008) 44(5):649–654.]

(See also AFFINITY PROTEIN EXPRESSION AND PURIFICATION.)

aflatoxins Heat-stable toxins produced by certain fungi (strains of *Aspergillus flavus* and *A. parasiticus*); the molecule of an aflatoxin contains a bifuran moiety fused with a (substituted) coumarin.

[Genes for aflatoxin biosynthesis: Appl Environ Microbiol (2005) 71:3192–3198.]

Aflatoxins damage DNA (producing e.g. mutations, strand breakage and inhibition of repair) and they are also highly carcinogenic in mammalian species; the aflatoxins are associated e.g. with cases of hepatocellular carcinoma. Different species may be affected in different ways by these toxins.

In mouse hepatocytes, it was found that α-mannan protects against DNA damage from aflatoxin B_1 [Int J Mol Sci (2009) 10(2):395–406].

AFLP Either of *two distinct* PCR-based approaches for TYPING bacteria.

One approach ('amplified fragment length polymorphism') includes a number of variant forms of arbitrarily primed PCR (AP-PCR), including e.g. RAPD analysis.

The other approach, outlined here, involves initial digestion of genomic DNA with two types of RESTRICTION ENDONUCLEASE; it is sometimes called 'amplified restriction fragment length polymorphism' – but this is not recommended by the original authors [see comments in: Nucleic Acids Res (1995) 23:4407–4414].

In the digested genome each fragment is flanked by STICKY ENDS produced by one or other of the two types of restriction enzyme. Two types of adaptor molecule (A, B) are added; the A molecules have *one* sticky end which binds to sites cleaved by one of the restriction enzymes, and B molecules have *one* sticky end that binds to sites cleaved by the other enzyme. A site cleaved by EcoRI (left) and a matching adaptor (right) is shown below:

```
5′-----NNG          AATTGNNNNN-3′
  -----NNCTTAA            CNNNNN
```

Fragment–adaptor binding is followed by ligation – but the cleavage site of EcoRI is *not* regenerated, avoiding repeated restriction. Fragments (with their adaptors) are amplified by PCR. Each primer is complementary to a sequence covering part of an adaptor and the (contiguous) restriction site of the fragment. Each primer's 3′ end extends beyond the restriction site for one (or several) nucleotides; thus, a given primer will be extended only if the primer's 3′ terminal 'selective' nucleotide(s) align with *complementary* base(s) in the fragment. In the example given above, one template strand is:

```
5′-----NNGAATTGNNNNN-3′
```

and a primer with a deoxycytidine (C) selective 3′ nucleotide will bind as follows:

```
5′-----NNGAATTGNNNNN-3′
       CCTTAACNNNNN-5′   ← primer
```

The same principle holds for each primer-binding site: only some of the primers will be extended owing to their selective 3′ end.

The PCR products undergo gel electrophoresis, and bands of products (made visible e.g. by the use of labeled primers) form the *fingerprint*.

AFLP discriminated easily between *Clostridium botulinum* types A, B, E and F, but only limited intraspecies differences

were reported in isolates of *Brucella*.

AFM ATOMIC FORCE MICROSCOPY.

agar A complex mixture of galactans obtained from certain red algae (e.g. *Gelidium*); these compounds form part of the cell wall and/or intercellular matrix.

Agar comprises two main components: AGAROSE and agaropectin. Agaropectin is a mixture of sulfated galactans, some of which contain glucuronic acid and/or other constituents.

An agar *gel* is used as a matrix in many types of solid and semi-solid microbiological medium. This translucent, jelly-like material is prepared by heating a mixture of agar – e.g. 1.5% w/v – and water to >100°C and then cooling it to room temperature; gelling begins at ~40–45°C.

Agar can be an inhibitory factor in PCR (cf. GELLAN GUM).

One report indicated that agar promotes transformation in *Escherichia coli* [J Bacteriol (2009) 191(3):713–719].

agarose The major constituent of AGAR: a non-sulfated linear polymer consisting of alternating residues of D-galactose and 3,6-anhydro-L-galactose. Agarose is used e.g. as a medium in GEL ELECTROPHORESIS for the separation of large fragments of nucleic acid.

(cf. POLYACRYLAMIDE.)

age-related macular degeneration See the entry AMD.

agnoprotein A regulatory protein, encoded by JC VIRUS, which plays a major role in the infective cycle; it may be involved in facilitating the transport of virions from the nucleus (where virus assembly occurs) to the cytoplasm.

Mutant, phosphorylated forms of agnoprotein are unable to sustain the infective cycle [J Virol (2006) 80(8):3893–3903].

Expression of a transfected agnoprotein-encoding gene was reported to give rise to inhibition of the low-fidelity pathway of DSB (double-strand break) repair, involving protein KU70.

Ago protein See ARGONAUTE.

Agrobacterium A genus of motile, Gram-negative bacteria that are found primarily in the rhizosphere (root environment of plants).

GC% of genomic DNA: 57–63.

Optimal temperature for growth: 25–28°C. Various mono- and disaccharides can be metabolized; glucose is metabolized e.g. via the Entner–Doudoroff pathway and the hexose monophosphate pathway.

Colonies which develop on media containing carbohydrates are generally mucilaginous, copious slime being formed.

Most species are pathogenic to plants – *A. tumefaciens* and *A. vitis* are causal agents of CROWN GALL.

(Some relevant entries: AGROINFECTION, BINARY VECTOR SYSTEM, FLORAL DIP METHOD AND TRIFOLITOXIN.)

Agrobacterium was used for random insertional mutagenesis in the fungus *Aspergillus fumigatus* (by the insertion of T-DNA: see CROWN GALL) [PLoS ONE (2009) 4(1):e4224].

Taxonomic note

One taxonomic study has classified all species of *Agrobacterium* in the genus *Rhizobium* [Int J Syst Evol Microbiol (2001) 51: 89–103]. According to this paper, *A. tumefaciens* is renamed *Rhizobium radiobacter*, and *A. vitis* is renamed as *Rhizobium vitis*. (The name *Agrobacterium tumefaciens* has been – and continues to be – widely used in publications.)

agroinfection A method for introducing viral DNA (or cDNA) into plant cells.

In the original procedure, viral DNA is inserted into the T-DNA region of the Ti plasmid of bacterium *Agrobacterium tumefaciens* (see the entry CROWN GALL for details of the Ti plasmid); thus, when infecting a plant, *A. tumefaciens* injects viral DNA (within the T-DNA) into plant cells.

Currently, the commonly used procedure involves the use of a *binary vector* system. In this approach the gene/sequence of interest is inserted into a small (binary) vector in which it is flanked, on each side, by the left border and right border of T-DNA. The binary vector, containing the gene, is inserted into an engineered strain of *Agrobacterium tumefaciens* that contains the *vir* (virulence) region of the Ti plasmid, which is concerned with transfer of DNA into plant cells. This strain is then used to infect a plant. When the *vir* genes are activated, DNA from the *binary vector* (specifically, the DNA from the section bracketed by the left and right borders of T-DNA) is transferred into the plant cells – such transfer being mediated by factors encoded by the *vir* region acting in *trans*.

For developing transgenic plants it is desirable to minimize the content of extraneous DNA which is transferred to plant cells through the vector system, particularly when such plants are to be made available in a general agricultural setting; the content of extraneous DNA is covered by certain laws which relate to transgenic plants. To this end, a number of minimal T-DNA vectors have been suggested [BioTechniques (2006) 41(6):708–710].

Agroinfection was used e.g. to study resistance to disease in the potato and the avirulence genes in *Phytophthora infestans* (the late blight pathogen) [PLoS ONE (2008) 3(8):e2875].

[High-efficiency expression of proteins in plants from agroinfection-compatible TMV (tobacco mosaic virus) expression vectors: BMC Biotechnol (2007) 7:52.]

(See also FLORAL DIP METHOD; cf. AGROINFILTRATION.)

agroinfiltration A procedure, used for inserting a vector into plant cells, in which cells of the plant-pathogenic bacterium *Agrobacterium tumefaciens* (containing the vector, and in the presence of ACETOSYRINGONE) are injected, with a syringe, through leaf stomata.

[Uses (e.g.): Plant Physiol (2007) 145(1):5–16 and (2009) 149(2):1005–1016.]

(cf. AGROINFECTION.)

AGT O^6-alkylguanine-DNA alkyltransferase: see *Uses of gene fusion* in the entry GENE FUSION.

AHL (*N*-acyl-L-homoserine lactone) See QUORUM SENSING.

***ahpC* gene** See ISONIAZID.

AHT (*N*-acyl-homocysteine thiolactone) See QUORUM SENSING.

AID ACTIVATION-INDUCED CYTIDINE DEAMINASE.

AIDS Acquired immune deficiency syndrome. An HIV^+ person

with AIDS has counts of $CD4^+$ T cells below a certain level and, additionally, the presence of one or more types of AIDS-defining disease (such as: candidiasis of the *lower* respiratory tract, retinitis with CMV (cytomegalovirus), *extra*pulmonary infection with *Mycobacterium tuberculosis*, pneumonia due to *Pneumocystis carinii* etc.).

The causal agent is generally HIV-1 (q.v.); the disease can also be caused by HIV-2. (See RETROVIRUSES for a general background to HIV-1 and other, related viruses.)

Chemotherapy used in the treatment of AIDS may include various ANTIRETROVIRAL AGENTS (q.v.). (See also HAART.)

Quiescent HIV-1

The chemotherapeutic agents currently in use aim primarily to break the replicative cycle of the virus (e.g. by inhibiting maturation of the virions). However, this does not address the covert problem of viral *persistence* resulting from the 'silent' *integrated* HIV genomes that pose an ongoing threat (as they escape the immune system as well as therapeutic agents) [see e.g. Nature Rev Microbiol (2009) 7:798–812].

In one approach to this problem, siRNAs have been used to investigate those particular histone deacetylases (see HDAC) that contribute to HIV-1 repression in latently infected cells. From the results of these studies it was suggested that HDAC inhibitors, specific for particular HDACs, may be useful for the targeted disruption of persistent HIV-1 infection [J Virol (2009) 83(10):4749–4756]. (In an analogous (murine) study, histone acetylation, but not DNA demethylation, was found to reactivate a latent gammaherpesvirus (MHV-68) in a cell line [PLoS ONE (2009) 4(2):e4556].)

In a different approach to the disruption of latency in HIV, an alternatively spliced transcription factor (ΔVII-Ets-1) was found to activate transcription in latent HIV-1 in cells derived from patients on a HAART chemotherapeutic regime. Unlike some other approaches, this method did not have the major disadvantage of causing significant T cell activation [Proc Natl Acad Sci USA (2009) 106(15):6321–6326].

aiRNA Asymmetric interfering RNA: see the entry SIRNA.

alanine scan mutagenesis A method that is used for studying the binding properties (or other characteristics) of particular residues in a protein by replacing them with alanine residues. Replacement of residue(s) may be achieved e.g. by synthesizing the protein from a modified (engineered) mRNA. After insertion of the alanine residues the protein is examined for specific properties/functions.

alarmone Any of various low-molecular-weight molecules that are able to mediate some change in cellular metabolism as a response to a particular type of stress. One example is ppGpp which is formed in the stringent response in *Escherichia coli*.

albamycin *Syn*. NOVOBIOCIN.

(See also ANTIBIOTIC (table).)

Alexa Fluor® 488 See XERODERMA PIGMENTOSUM (diagnosis by the EdU method).

AlgR In *Pseudomonas aeruginosa*: the response regulator of a TWO-COMPONENT REGULATORY SYSTEM (q.v.), FimS/AlgR, which controls biosynthesis of type IV fimbriae (responsible for so-called 'twitching motility'); this regulation involves activating the pilVWXY1Y2E operon [J Bacteriol (2008) 190(6):2023–2030].

AlgR also has roles in regulating type III protein secretion genes and alginate biosynthesis.

AlgZ *Syn*. FimS: see entry ALGR.

AlkA protein See DNA REPAIR.

alkaline phosphatase (AP) An enzyme (EC 3.1.3.1; maximum stability at pH ~7.5–9.5) which is used e.g. for removing 5′ terminal phosphate groups from nucleic acids.

AP is also used in reporter systems. For example, when it is bound to antibodies, AP can act as a label to any molecule of the relevant antigen to which the antibodies bind; the AP is then detectable by using any of various types of substrate that yield a colored product or light. Substrates that yield colored products include NBT (nitroblue tetrazolium) + BCIP (5-bromo-4-chloro-3-indolyl phosphate). ATTOPHOS is cleaved by AP to yield a fluorescent molecule (fluorophore). Several 1,2-dioxetanes, such as CSPD® and AMPPD®, give rise to CHEMILUMINESCENCE when cleaved by AP.

AP is used as a reporter e.g. in ELISA.

alkaline stripping Stripping of hybridized RNA probes from a DNA MICROARRAY by degradation with buffers that contain NaOH (sodium hydroxide) under carefully regulated conditions (e.g. a temperature of 60–62°C).

A stripped microarray can be re-used – reducing costs – but a microarray cannot be stripped twice without loss of quality. Microarrays which had been stripped once were reported to give results similar to those from virgin (non-stripped) arrays.

***O^6*-alkylguanine-DNA alkyltransferase**(AGT) See *Uses of gene fusion* in the entry GENE FUSION.

allele (allelomorph) Any one of two or more different versions of a particular GENE; the product or function of a given allele may exhibit qualitative and/or quantitative difference(s) from the product or function of other alleles of that gene.

In a diploid cell or organism, if an *allelic pair* (i.e. the two alleles of a given gene) consists of two identical alleles then the cell/organism is said to be *homozygous* for that particular gene; if different, the cell/organism is said to be *heterozygous* for the gene.

(See also ALLELE-SPECIFIC GENE EXPRESSION.)

allele drop-out (ADO) The failure of PCR to amplify one of the two alleles of a given gene.

ADO can be problematical e.g. for preimplantation genetic diagnosis (see PGD) in which the sample cell is heterozygous for the relevant gene; in some cases a PGD misdiagnosis has been due to ADO.

allele-specific DNA methylation analysis A method designed for determining allele-specific methylation. It involves PCR-amplification of BISULFITE-treated DNA followed by PYROSEQUENCING of the amplicons of each allele using primers that include allele-specific single-nucleotide polymorphisms (see SNP). Use of the method was demonstrated by analyzing methylation of the *H19* gene – in which the paternal allele is usually imprinted [BioTechniques (2006) 41(6):734–739].

allele-specific gene expression The two alleles of a given gene (in a diploid cell) may be expressed at different levels for a number of reasons – e.g. one allele may be transcriptionally silent owing to X-INACTIVATION, or it may be expressed at a different level owing to the presence of one or more polymorphisms (see SNP) which affect transcriptional regulation. Such differential allelic expression may be detected e.g. by a method in which the levels of mRNA from each of the two alleles are compared – the transcripts from one allele being distinguished from those of the other allele by the presence of a specific polymorphism (acting as a marker) [PLoS Genetics (2008) 4(2):e1000006]. Other workers screened 11500 SNPs to identify differential allelic expression and concluded that allele-specific gene expression is widespread across the genome [PLoS ONE (2009) 4(1):e4150].

Allele-specific gene expression patterns were investigated by quantitative genotyping of 2529 genes from patients with acute lymphoblastic leukemia – revealing regulation of gene expression by methylation at CpG sites [Genome Res (2009) 19(1):1–11].

(See also EPIGENETIC ALLELIC RATIO ASSAY.)

allele-specific PCR A variant form of PCR designed to amplify a particular allele of a given gene – but no other allele(s) of that gene. One of the two primers is designed with a 3′ terminal nucleotide that pairs with a *specific* base in the required allele – a base which is known to be different from that in the other allele(s) at this location; this primer can be extended on the required allele, which will be amplified. Other alleles are not amplified as the mismatch at the primer's 3′ terminal will inhibit extension.

An essential requirement is the use of a polymerase which *lacks* proofreading ability (i.e. one which lacks 3′-to-5′ exonuclease activity); such an enzyme (e.g. the *Taq* polymerase) is not able to remove the primer's terminal nucleotide (i.e. it cannot correct the mismatch) and does not amplify unwanted alleles.

It is also important to carry out the reaction with an appropriate level of stringency.

A modified form of this method is found e.g. in SNP GENOTYPING.

allelic pair See ALLELE.

allelic ratio assay See e.g. EPIGENETIC ALLELIC RATIO ASSAY.

allelomorph *Syn.* ALLELE.

allolactose β-D-Galactopyranosyl-(1→6)-D-glucopyranose: the natural inducer of the LAC OPERON in *Escherichia coli;* it is formed as a minor product during the cleavage of lactose by β-galactosidase.

(See also IPTG.)

allosteric effect The effect which is produced when the binding of a ligand to a given target molecule affects the properties of other site(s) on the same molecule; allosteric effects are due to conformational changes that result from the binding of the ligand.

allosteric nucleic acid enzymes Nucleic acid enzymes (see e.g. APTAZYME) whose function can be subject to regulation e.g. by the activity of a low-molecular-weight molecule – or the activity of an oligonucleotide.

(See BINARY DEOXYRIBOZYME LIGASE and MAXIZYME.)

alpha (α, *Lk*) A symbol for LINKING NUMBER.

alpha peptide (α-peptide) See the entry α-PEPTIDE.

Alphanodavirus See NODAVIRUSES.

Alphavirus A genus of viruses that include many which cause disease (involving e.g. encephalitis and arthralgia) in humans and animals. The (enveloped) virion is spherical, ∼50–70 nm diam., and the (monopartite) genome is linear, positive-sense ssRNA.

Alphaviruses include e.g. the SINDBIS VIRUS and Western equine encephalomyelitis virus.

alternation of generations See PLOIDY.

alternative polyadenylation See POLYADENYLATION.

alternative splicing SPLICING of a given pre-mRNA which can proceed in at least two different ways, the different modes of splicing producing mRNAs with different overall sequences.

Alternative splicing is a common (i.e. natural) phenomenon [Nucleic Acids Res (2004) 32(13):3977–3983] which allows a given gene to encode extra information. Thus, e.g. a given gene may encode two mutually antagonistic messages (whose expression must be under appropriate control). Exceptionally, it was reported that a single gene may potentially give rise to >100 different transcripts.

The regulation of (normal) alternative splicing involves a balance between certain factors which promote or inhibit the use of specific splicing sites in pre-mRNA.

Events that may occur during alternative splicing include the splicing out (loss) of an exon (*exon skipping*), inclusion of an intron (*intron retention*) and/or splicing at a site *within* an exon.

Differential, SNP-associated alternative splicing may occur in different tissues [PLoS Biol (2008) 6(12):e1000001].

A comparative study on eight organisms found alternative splicing to be more extensive in vertebrates than in invertebrates [Nucleic Acids Res (2007) 35(1):125–131].

Atypical (U1-*independent*) splicing may be a characteristic feature of some types of alternative splicing event [Nucleic Acids Res (2009) 37(6):1907–1914].

The occurrence of alternative splicing, in respect of a given gene, may be investigated (e.g.) by initially isolating mRNAs from an extract of the relevant (living) cells. For eukaryotic cells, this can be done most easily by exploiting the poly(A) tail which characteristically forms the 3′ end of mRNAs; the cell extract is passed through a column containing the ligand oligo(dT)-cellulose that binds mRNAs (by their poly(A) tails) and allows other types of RNA (e.g. ribosomal RNA) to be eliminated. When eluted from the column, the population of mRNAs is subjected to GEL ELECTROPHORESIS and the gel is then used for NORTHERN BLOTTING; individual mRNAs can be identified (on the membrane) by highly specific, labeled PROBES. This procedure can thus reveal mRNAs *of different sizes* that were derived from a given gene.

Alternative splicing may also be investigated by the use of

an EXON ARRAY (q.v.).

The choice of splice site in a transcript can be modulated, *in vitro* and *in vivo*, by using a specific antisense molecule. Thus, splicing of pre-mRNA of gene *bcl-x* normally produces a long anti-apoptotic mRNA as well as a short pro-apoptotic mRNA, and the ratio of these two mRNAs in the cell is a key determinant of cancer progression. By using a short antisense PNA conjugated to an oligopeptide containing eight serine–arginine repeats (cf. SR PROTEINS), it was possible to modify the splicing pattern so that the short (pro-apoptotic) mRNA became the dominant influence; this procedure was able to produce APOPTOSIS in HeLa (cancer) cells [Nucleic Acids Res (2005) 33(20):6547–6554].

siRNA-controlled alternative splicing has been reported in which the siRNAs targeted appropriate sequences in introns or exons [Nature Struct Mol Biol (2009) 16:717–724].

Aberrant splicing causes many types of disease. Correction of aberrant splicing has been demonstrated both *in vitro* and *in vivo* by using specifically designed *splice-switching oligonucleotides* (SSOs) that can re-direct the splicing machinery to normal splice sites. Positive results were reported with the use of SSOs in patients with Duchenne muscular dystrophy. [Therapeutic potential of splice-switching oligonucleotides (a review): Oligonucleotides (2009) 19(1):1–14.]

An alternatively spliced transcription factor was reported to disrupt latency (i.e. activate transcription) of HIV-1, offering a potential approach to the problem of quiescent infection in AIDS [Proc Natl Acad Sci USA (2009) 106(15):6321–6326].

(See also the entries CRYPTIC SPLICING, EXON TRAPPING, and *alternative polyadenylation* in POLYADENYLATION.)

Alu sequences (*Alu* sequences) In (at least) some mammalian genomes: a family of related sequences, each typically about 300 nt long and commonly having a recognition site for the RESTRICTION ENDONUCLEASE AluI (AG↓CT); the human genome may contain about one million copies. Alu sequences are RETROTRANSPOSONS; owing to their ability to generate insertional mutations they are regarded as potential factors in genetic disorders.

Most Alu sequences in the human genome (those from subfamilies S and J) seem to be largely inactive – transpositional activity being essentially confined to a much smaller group of 'younger' Alu sequences in the Y family [Genome Res (2009) 19:545–555].

Some Alu sequences with regions similar to splice sites are found in certain genes and are recognized as exons. (See also EXONIZATION.) Splicing patterns of exonized Alu sequences have been described in (human) tissues [PLoS Genet (2008) 4(10):e1000225].

A study of exonizing and non-exonizing Alu sequences has reported various features required for precise recognition of exons by the splicing machinery [PLoS Comput Biol (2009) 5(3):e1000300].

(See also SINE and LINE.)

AluI A RESTRICTION ENDONUCLEASE from *Arthrobacter luteus;* recognition site: AG↓CT.

α-amanitin A complex, substituted cyclic peptide (one member of a family of toxins produced e.g. by the 'death cap fungus', *Amanita phalloides*) which, at low concentrations, can inhibit (eukaryotic) DNA-dependent RNA polymerase II.

This agent was used e.g. for studies on RNA polymerase II [PLoS ONE (2008) 3(2):e1661; Acta Crystallogr (Section D) (2009) 65(2):112–120].

amantadine (1-adamantanamine hydrochloride) A polycyclic ANTIVIRAL AGENT used (e.g.) for prophylaxis and early treatment of infection by type A INFLUENZAVIRUS; *rimantadine*, α-methyl-1-adamantane methylamine hydrochloride, has similar activity.

Amantadine inhibits viral replication by blocking a proton channel formed by the M2 protein of influenza virus A [Proc Natl Acad Sci USA (2008) 105(5):1483–1488].

(See also *swine flu* in the entry INFLUENZAVIRUS.)

Amaxa Nucleofector® See NUCLEOFECTION.

amber codon See NONSENSE CODON.

amber mutation A mutation that creates an amber codon (see NONSENSE CODON).

amber suppressor See SUPPRESSOR MUTATION.

ambisense RNA Viral ssRNA in which some gene(s) occur in positive-sense form and other(s) occur in negative-sense form (in the same strand).

Ambisense RNA occurs e.g. in the genome in viruses of the Arenaviridae, including Lassa fever virus, and in the genome in viruses of the genus *Phlebovirus*, e.g. the Rift Valley fever virus; in Rift Valley fever virus the tripartite genome consists of two fragments of negative-sense ssRNA and one fragment of ambisense ssRNA.

In many ambisense viruses, the junction between opposite-sense ORFs in the genome exhibits a hairpin structure which acts as a termination signal. The hairpin structures appear to be absent in the small (S) ambisense fragment in Rift Valley fever virus; instead, a conserved motif was reported to act as a termination signal in both the negative-sense and ambisense fragments [J Virol (2007) 81(10):5246–5256].

AMD Age-related macular degeneration: any of certain forms of pathologic condition affecting the eyes in the elderly. One form involves the development of new blood vessels (angiogenesis) in retinal tissue; the potential for GENE THERAPY in this condition has been investigated – see e.g. AAV HELPER-FREE SYSTEM.

amelanotic melanoma See MELANOMA.

amelogenin A protein associated with dental development. A gene encoding human amelogenin (*AMELX*) occurs on the X chromosome (at location Xp22.3–Xp22.1) and another gene encoding amelogenin (*AMELY*) occurs on the Y chromosome (location Yp11.2).

The amelogenin genes are exploited e.g. in a gender identification assay: see DNA SEX DETERMINATION ASSAY.

***AMELX* gene** See AMELOGENIN.

***AMELY* gene** See AMELOGENIN.

Ames strain (of *Bacillus anthracis*) See BACTERIA (table).

Ames test (Mutatest; *Salmonella*/microsome assay) A test used

to determine whether a given agent is mutagenic (and therefore possibly carcinogenic) by investigating its ability to *reverse* an auxotrophic mutation in *Salmonella typhimurium;* the mutation in *S. typhimurium* makes the organism dependent on an exogenous source of histidine, and reversal of the mutation would allow the organism to synthesize its own histidine (i.e. it would revert to prototrophy). Different strains of *S. typhimurium* may be used, each with a different type of mutation in the histidine operon. Some of these test strains may also contain mutations which make them more permeable to certain chemicals and/or which prevent them from carrying out DNA repair. Moreover, the test strains may contain the plasmid pKM101 which includes genes for so-called error-prone repair and which therefore promotes the mutagenic effects of any DNA-damaging agents present in the reaction.

Because some chemical agents exhibit mutagenic activity only after their metabolic activation, the test system generally includes microsomal enzymes from a liver homogenate (the 9000 ***g*** supernatant, fraction 'S9') from rats pre-treated with a carcinogen to induce production of the appropriate enzymes.

When performed as a 'plate incorporation test', a culture of *S. typhimurium*, an S9 preparation and the substance under test are mixed with soft agar (which includes a low level of histidine) and this is poured onto a plate of *minimal agar* – which is then incubated at 37°C in the dark. (Minimal agar permits the growth of prototrophs but does not permit growth of the (auxotrophic) test strain of *S. typhimurium*.) The low level of histidine in the soft agar allows (only) limited growth of the (auxotrophic) *S. typhimurium* and this results in a light, confluent growth of this organism in the upper layer of agar (the 'top agar').

If the test substance had caused reversion to prototrophy in any cells of the test strain, those cells (whose growth would not be limited) can grow and form visible colonies.

When interpreting the results, several factors must be borne in mind. An absence of growth in the 'top agar' would suggest that the substance under test has general antibacterial activity, and that any colonies which develop on the plate are unlikely to be true revertants. Again, before drawing conclusions from a number of *apparent* revertant colonies, it is necessary to take into account the known *spontaneous* reversion rate of the particular mutation in the given test strain.

Various modifications of the basic Ames test are used for specific purposes.

[Ames test with a derivative of carbamic acid: Antimicrob Agents Chemother (2005) 49:1160–1168.]

An Ames test was employed in an assay of mutagenicity on certain sulfur analogs of polycyclic aromatic hydrocarbons [J Org Chem (2007) 72(22):8383–8393].

[Comparison of Ames II and traditional Ames test responses with respect to mutagenicity, strain specificities, need for metabolism and correlation with rodent carcinogenicity: Mutagenesis (2009) 24(4):359–366.]

(See also SOS CHROMOTEST.)

amethopterin (methotrexate) 4-Amino-10-methylfolic acid: an agent which blocks nucleotide synthesis; it is used e.g. in the chemotherapy of leukemia.

(cf. AMINOPTERIN.)

amikacin See AMINOGLYCOSIDE ANTIBIOTIC.

amino acid A term which (in the context of DNA technology) generally refers to one of 20 compounds (listed in the table) whose residues are common components of oligopeptides, polypeptides and proteins. Each amino acid is represented by at least one CODON (see the table, and see GENETIC CODE).

Ornithine is just one example of an amino acid which is not represented by a CODON but which is nevertheless found e.g. in certain oligopeptide antibiotics; this kind of oligopeptide is synthesized by a ribosome-free enzyme system (rather than by translation) (see NON-RIBOSOMAL PEPTIDE SYNTHETASE).

4-amino-10-methylfolic acid See AMETHOPTERIN.

aminoalkylsilane See AAS.

4-aminofolic acid See AMINOPTERIN.

aminoglycoside antibiotic Any of a group of (broad-spectrum) antibiotics in which the molecular structure typically includes an aminosugar and either 2-deoxystreptamine or streptidine; these antibiotics bind to the bacterial 30S ribosomal subunit and inhibit protein synthesis.

The aminoglycosides include amikacin, framycetin, gentamicin, hygromycin B, kanamycin, neomycin, streptomycin and tobramycin; they are typically bactericidal (at appropriate concentrations) and are active against a wide range of Gram-positive and Gram-negative species.

Resistance to aminoglycosides can arise e.g. by (i) mutation in proteins of the ribosomal 30S subunit (affecting the binding of antibiotics); (ii) inactivation of antibiotics by bacterial enzymes which e.g. carry out *N*-acetylation or *O*-phosphorylation; (iii) decreased uptake by the cell.

(See also G418 SULFATE.)

aminopterin 4-Aminofolic acid: an agent that blocks the pathway of nucleotide synthesis. Aminopterin is used e.g. in HAT medium for preparing a HYBRIDOMA (q.v.).

[Example of the use of aminopterin (for the preparation of a hybridoma): BMC Immunol (2009) 10:16.]

(cf. AMETHOPTERIN.)

AMO Anti-miRNA oligonucleotide: see e.g. MICRORNAS.

AMP CT Amplified *Chlamydia trachomatis* test: a TMA-based assay used for detecting the pathogen *Chlamydia trachomatis* in clinical specimens (Gen-Probe, San Diego CA). One early study [J Clin Microbiol (1997) 35:676–678] examined urine specimens from female patients as a non-invasive method of diagnosing chlamydial infection; both the AMP CT assay and a PCR-based method were found to be sensitive and specific methods for detecting *C. trachomatis*, and it was concluded that both methods were suitable screening procedures.

ampholyte Any electrolyte with both acidic and basic groups.

ampicillin 6(α-aminobenzylamido)-penicillanic acid: a semi-synthetic PENICILLIN used in media e.g. as a selective agent for bacteria containing a vector with an ampicillin-resistance marker gene.

amplicon (1) A specific (precise) sequence of nucleotides, part

AMINO ACIDS: symbols, molecular weights and codons

Amino acid	1-letter symbol	3-letter symbol	Molecular weight	Codons
Alanine	A	Ala	89	GCA, GCC, GCG, GCU
Arginine	R	Arg	174	AGA, AGG, CGA, CGC, CGG, CGU
Asparagine	N	Asn	150	AAC, AAU
Aspartic acid	D	Asp	133	GAC, GAU
Cysteine	C	Cys	121	UGC, UGU
Glutamic acid	E	Glu	147	GAA, GAG
Glutamine	Q	Gln	146	CAA, CAG
Glycine	G	Gly	75	GGA, GGC, GGG, GGU
Histidine	H	His	155	CAC, CAU
Isoleucine	I	Ile	131	AUA, AUC, AUU
Leucine	L	Leu	131	CUA, CUC, CUG, CUU, UUA, UUG
Lysine	K	Lys	146	AAA, AAG
Methionine	M	Met	149	AUG
Phenylalanine	F	Phe	165	UUC, UUU
Proline	P	Pro	115	CCA, CCC, CCG, CCU
Serine	S	Ser	105	AGC, AGU, UCA, UCC, UCG, UCU
Threonine	T	Thr	119	ACA, ACC, ACG, ACU
Tryptophan	W	Trp	204	UGG
Tyrosine	Y	Tyr	181	UAC, UAU
Valine	V	Val	117	GUA, GUC, GUG, GUU

of a larger nucleic acid molecule, which is copied (amplified) by an *in vitro* amplification process such as NASBA or PCR.
(2) One of the copies of a sequence of nucleotides which has been copied (amplified) by methods such as NASBA or PCR.
(3) One of a number of elements of linear DNA (~100 kb) formed in studies on the *JBP1* gene of *Leishmania tarentolae* [term used in: Nucleic Acids Res (2005) 33(5):1699–1709].
(4) *Formerly* [Proc Natl Acad Sci USA (1985) 82:694–698]: a defective virus vector.
(5) Within a Y chromosome: a segment with >99% sequence identity to other region(s) in the given chromosome (hence the adjective *ampliconic*). [Use of term: BMC Genet (2007) 8:11.]
(6) A given sequence of nucleotides *within which* a specific sub-region may be identified as being of special interest and which may be treated as an amplicon (sense 1).

amplicon containment One approach to the minimization of contamination by amplicons from previous assays in methods such as PCR. Essentially, this involves division of the working environment into several dedicated areas, each of which is used for only certain specific stage(s) of the procedure. For example, when working with PCR, it is usual to carry out the thermal cycling and the analysis of products (e.g. electrophoresis) in separate areas, and separate areas may also be specified for extracting target nucleic acid and for preparing reagents.

(See also AMPLICON INACTIVATION.)

amplicon inactivation In PCR: any method which avoids contamination by destroying carry-over amplicons from previous assays. Such contamination can be a major problem e.g. in those clinical laboratories in which specimens are examined routinely for only a small number of target sequences; under these conditions new specimens may risk contamination if amplicons are allowed to build up in the laboratory environment (e.g. in/on the equipment or in reagents).

Methods for amplicon inactivation

The uracil-N-glycosylase method. In this method, *all* assays are conducted in the normal way except that deoxythymidine triphosphate (dTTP) is replaced by deoxyuridine triphosphate

(dUTP) in the reaction mixture. All amplicons produced in each assay therefore contain dUMP instead of dTMP. These amplicons can be analysed in the normal way by gel electrophoresis etc.

In addition to the use of dUTP, the reaction mixture also includes the enzyme URACIL-N-GLYCOSYLASE (UNG). Thus, if an assay is contaminated with amplicons from a previous assay, these amplicons will act as substrates for the enzyme, uracil being cleaved from each dUMP; this, in itself, does not bring about strand breakage, but the amplicons are degraded to non-amplifiable pieces by the high temperature used for the initial denaturation of target DNA. The high temperature also inactivates UNG; this is necessary in order to avoid degradation of amplicons from the current assay. Note that the *target* DNA in the reaction is not affected by UNG as it contains dTMP.

Normally, this method cannot be used in PCR-based studies of DNA methylation in which the sample DNA is treated with BISULFITE; this is because bisulfite treatment converts non-methylated cytosines to uracil, so that the template DNA itself would be subjected to degradation. However, unlike the usual form of bisulfite treatment (in which DNA is sulfonated and is subsequently desulfonated), it has been found that non-desulfonated DNA can be amplified by PCR with the UNG method of decontamination because this (sulfonated) form of DNA is resistant to UNG. Desulfonation of DNA is achieved by a prolonged (30-minute) initial stage of denaturation (at 95°C) [Nucleic Acids Res (2007) 35(1):e4].

The isopsoralen method. In this method isopsoralen is added to the reaction mixture. Isopsoralen is a heterocyclic compound which, when bound to DNA, can form covalent interstrand crosslinks when photoactivated by ultraviolet radiation (e.g. 365 nm/15 min/4°C); activation of isopsoralen at low temperatures was reported to be more efficient than at room temperature. As the double-stranded amplicons are *covalently* crosslinked they cannot be denatured to single strands; hence, they cannot serve as templates, so that, if they contaminate a subsequent assay the outcome would not be affected.

Amplicons produced by this method are suitable for examination by processes such as gel electrophoresis and staining (e.g. for confirming the presence of a given target sequence in the sample DNA). However, they cannot be used for any process (such as SSCP analysis) that requires single-stranded samples.

APSR. In another approach to amplicon inactivation, all the primers have a 5′ tag which incorporates a binding site for a type IIS restriction endonuclease; this enzyme, in the reaction mixture, cleaves any contaminating amplicons – and is itself inactivated at the initial high-temperature stage (see APSR).

amplicon primer site restriction See APSR.

ampliconic See AMPLICON (sense 5).

amplification (of DNA *in vitro*) See DNA AMPLIFICATION.

amplification (of RNA *in vitro*) See RNA AMPLIFICATION.

amplification-refractory mutation system See ARMS.

amplified fragment length polymorphism (AFLP) See AFLP.

amplified restriction fragment length polymorphism A non-recommended name which has been used for a method for TYPING bacteria – see the entry AFLP.

Ampligase® See DNA LIGASE.

amplimer Any primer used in PCR.

AmpliTaq^{TM} Gold DNA polymerase A 94-kDa thermostable, recombinant DNA polymerase (Applied Biosystems) used in PCR; its optimal extension temperature is reported to be 72–80°C, with an extension rate of 2 to 4 kb per minute at 72°C.

[Uses (e.g.): Malaria J (2009) 8:154; Mol Vision (2009) 15: 1620–1630; BMC Cancer (2009) 9:258.]

AmpliWax^{TM} See HOT-START PCR.

AMPPD® A 1,2-dioxetane substrate that emits light when dephosphorylated by ALKALINE PHOSPHATASE (AP). It is used e.g. for detecting AP-labeled probes.

(See also CHEMILUMINESCENCE.)

amprenavir See PROTEASE INHIBITORS.

AMTDT Amplified *Mycobacterium tuberculosis* direct test: a TMA-based assay for detecting *Mycobacterium tuberculosis* in clinical specimens (Gen-Probe, San Diego CA).

The AMTDT was approved in 1995 by the American FDA (Food and Drug Administration) for use with *smear-positive* respiratory specimens; a smear-positive specimen is one from which a smear showing ACID-FAST BACILLI can be prepared.

In the original test, the specimen was initially treated with a MUCOLYTIC AGENT, decontaminated with sodium hydroxide, and finally sonicated (to lyse organisms and release nucleic acids). The specimen was then heated (95°C/15 minutes) to remove intra-strand base-pairing in the rRNA.

The reaction mixture contained 45 μL or 50 μL of sample, and amplification (at 42°C) was conducted for 2 hours. The amplification product was detected by the addition of target-specific probes (for details see entry ACCUPROBE).

An attempt was made to adapt AMTDT for the detection of *Mycobacterium tuberculosis* in *non*-respiratory specimens [J Clin Microbiol (1997) 35:307–310], and studies were made to compare the original and subsequent (improved) versions of AMTDT for the detection of *M. tuberculosis* in respiratory and non-respiratory specimens [J Clin Microbiol (1998) 36: 684–689].

The new-format 'enhanced' AMTDT was approved by the FDA in 1998. Among other changes, this version involved the use of a 450 μL aliquot of the sample (instead of a 40–50 μL sample).

One study reported false-positive results in tests on sputa from patients infected with *Mycobacterium kansasii* and *M. avium;* these species of *Mycobacterium* are not infrequently isolated from infections in immunocompromised patients. The authors of this study suggested a change in the threshold value of luminometer readings considered to be an indication of a positive result [J Clin Microbiol (1999) 37:175–178].

anaerobic respiration RESPIRATION (q.v.) which occurs in the absence of oxygen.

analyte In a test system: that component whose properties are being studied/measured.

anchor primer A primer that binds to an ANCHOR SEQUENCE.

anchor sequence Commonly, a sequence of nucleotides, with a known composition, which is present in a given molecule or which is ligated to another sequence (or added by tailing) in order to serve a particular function – e.g. as a primer-binding site. For example, in ANCHORED PCR an anchor sequence is used to provide an otherwise unavailable site for priming the amplification of an unknown sequence.

Certain natural sequences, such as the poly(A) tail on many mRNA molecules, have been described as anchor sequences.

anchored PCR A form of PCR used for amplifying an unknown sequence of nucleotides adjacent to a known sequence on a fragment of DNA; this approach addresses the problem of the lack of a primer-binding site in the unknown sequence.

To each end of the fragment is ligated a short segment of DNA of known sequence (e.g. a linker). (If the fragment has 3′ or 5′ overhangs then these can be eliminated, enzymically, in order to prepare the fragment for blunt-end ligation to the linkers.) Following ligation, that linker which is contiguous with the unknown region provides an ANCHOR SEQUENCE which is able to serve as a primer-binding site for one of the PCR primers. The second primer is designed to bind at a site within the known sequence. If PCR is primed in this way, the resulting amplicons will include the unknown sequence and at least part of the known sequence. If the known sequence occurs in the *center* of the fragment, then amplification, as described above, can be carried out for both of the unknown flanking regions.

(See also TAIL-PCR and VECTORETTE PCR.)

anchoring enzyme In SAGE (q.v.): a name sometimes given to the enzyme used for initial cleavage of the cDNAs.

ancient DNA (aDNA) DNA recovered from specimens which are thousands or millions of years old.

[Isolation of ancient DNA (optimization of procedures): BioTechniques (2007) 42(3):343–352.]

(See also SPEX.)

Studies on (permafrost-derived) mammoth bones provided some indication that mitochondrial DNA is preserved better than nuclear DNA in permafrost [Nucleic Acids Res (2009) doi: 10.1093/nar/gkp159].

ANDENYALAA The (10-amino-acid) tag which functions as a proteolysis signal on an aborted protein in the TRANS TRANSLATION process (q.v.).

aneuploid Refers to a genome that has one or more chromosomes in excess of, or less than, the number characteristic of the species (see e.g. DOWN'S SYNDROME).

Angelman syndrome A disorder that may be caused by any of various genetic mechanisms – see the table in entry GENETIC DISEASE for further details.

angiogenesis (neovascularization) Development of (new) blood vessels – e.g. in an embryo or in a tumor.

(See also CANCER THERAPY.)

annealing The hybridization of two complementary (or near-complementary) sequences of nucleotides to form a double-stranded molecule or a double-stranded region within a larger molecule (e.g. the binding of primers to primer-binding sites in PCR).

annotation Information supplied with the database entry of a given sequence or gene.

For an uncharacterized gene, the annotation may include a predicted function and/or an attempted characterization of the gene's product based on sequence homology with a gene of known function present in another organism.

(See also ACCESSION NUMBER.)

antagomir Any of a range of synthetic, chemically engineered oligonucleotides that bind to, and antagonize, specific types of MICRORNA molecule. Intravenous administration (in mice) of antagomirs directed against particular miRNAs resulted in efficient and specific silencing of the given miRNAs [Nature (2005) 438:685–689].

[Lentivirus-mediated antagomir expression for the specific inhibition of miRNA function: Nucleic Acids Res (2007) 35: e149.]

Antagomirs have been used to study the role of microRNAs miR-27a and miR-451 in regulating the expression of *MDR*1/P-glycoprotein in ovarian cancer cells [Biochem Pharmacol (2008) 76(5):582–588], and in a study on the TAR miRNA of HIV-1 [Retrovirology (2009) 6:18].

anthrax toxin A toxin, produced by the Gram-positive pathogen *Bacillus anthracis*, which gives rise to the symptoms of anthrax. It comprises three protein components – each, alone, being unable to function as a toxin; these three proteins are encoded by the plasmid pXO1. (Another plasmid, pXO2, is needed for the pathogenicity of *B. anthracis;* this encodes an essential anti-phagocytic capsule which protects the organism from the host's immune system.)

One component of the toxin localizes in the cell membrane and permits internalization of the other two components – a zinc protease (which disrupts intracellular signaling) and an ADENYLATE CYCLASE (which e.g. promotes edema).

anti Abbreviation for ANTICLINAL.

antibiotic Any of an extensive range of natural, semi-synthetic and fully synthetic compounds which, in low concentrations, are able, selectively, to inhibit or kill specific types of microorganism and, in some cases, other types of cell – e.g. tumor cells; an antibiotic acts at specific site(s) in a susceptible cell. (Compounds that are active against *viruses* are usually called 'antiviral agents' rather than antibiotics.)

Natural antibiotics (see e.g. BACTERIOCIN) have ecological roles. Some types of antibiotic have medical/veterinary uses in the prevention and/or treatment of infectious diseases, and some (e.g. nisin) are used as food preservatives.

(For examples of antibiotics see: AMINOGLYCOSIDE ANTIBIOTIC; CHLORAMPHENICOL; β-LACTAM ANTIBIOTICS; OXAZOLIDINONE ANTIBIOTICS, QUINOLONE ANTIBIOTICS, QUINOXALINE ANTIBIOTICS, RIFAMYCINS and VANCOMYCIN.)

In nucleic-acid-based technology, antibiotics are used in a variety of *selective* procedures. The presence of an antibiotic

ANTIBIOTIC: some antibiotics used in DNA technology (e.g. for marker selection)

Antibiotic	Group	Target organisms	Antibiotic action
Actinomycin D	–	Prokaryotic and eukaryotic	Intercalating agent; inhibits DNA-dependent RNA polymerase
Ampicillin	β-Lactams	Bacteria	Blocks synthesis of cell-wall polymer peptidoglycan
Blasticidin S	Nucleoside	Prokaryotic and eukaryotic	Inhibits protein synthesis by inhibiting the peptidyltransferase-mediated reaction at the ribosome
Carbenicillin	β-Lactams	Bacteria	Blocks synthesis of cell-wall polymer peptidoglycan
Cefotaxime	β-Lactams	Bacteria (mainly Gram-negative)	Blocks synthesis of cell-wall polymer peptidoglycan
Chloramphenicol (= chloromycetin)	–	Bacteria (broad spectrum), some fungi	Binds to prokaryotic and mitochondrial ribosomes; inhibits peptidyltransferase and (hence) protein synthesis
G418 sulfate	Related to gentamicin	Prokaryotes, yeasts, plants, mammalian cells	Inhibits protein synthesis
Gentamicin	Aminoglycosides	Bacteria	Binds to 30S subunit of bacterial ribosomes and inhibits protein synthesis
Hygromycin B	Aminoglycosides	Mammalian and plant cells	Inhibits protein synthesis
Kanamycin	Aminoglycosides	Bacteria	Binds to 30S subunit of bacterial ribosomes and inhibits protein synthesis
Kasugamycin	Aminoglycosides	Bacteria, some fungi	Inhibits polypeptide chain initiation and, hence, protein synthesis
Mycophenolic acid	–	Bacteria (also antitumor agent)	Inhibits synthesis of guanosine monophosphate, inhibiting synthesis of nucleic acids
Nalidixic acid	Quinolones	Bacteria (mainly Gram-negative)	Inhibits function of the A subunit of gyrase; inhibits DNA synthesis
Novobiocin	–	Bacteria	Inhibits binding of ATP to the B subunit of gyrase, inhibiting DNA synthesis
Penicillin G (= benzylpenicillin)	β-Lactams	Bacteria (mainly Gram-positive)	Blocks synthesis of cell-wall polymer peptidoglycan
Polymyxin B	Polymyxins	Bacteria (mainly Gram-negative)	Increases permeability of the cytoplasmic membrane and affects the integrity of the outer membrane
Puromycin	Nucleoside	Prokaryotic and eukaryotic	Inhibits protein synthesis by acting as an analog of part of an aminoacyl-tRNA
Rifampicin	Rifamycins	Bacteria (mainly Gram-positive)	Inhibits DNA-dependent RNA polymerase by binding to the β subunit of the enzyme
Streptomycin	Amino glycosides	Bacteria, some fungi	Interacts with ribosomes and inhibits protein synthesis
Tetracycline	Tetracyclines	Bacteria	Binds to ribosomes; inhibits protein synthesis by blocking the binding of aminoacyl-tRNAs to the A site
Zeocin™	Bleomycin/ phleomycin	Bacteria, yeast, and mammalian cells	Binds to, and cleaves, DNA

in a growth medium selects for those cells that are expressing a gene conferring resistance to that antibiotic; those cells that are susceptible to the antibiotic – i.e. which do not contain a relevant antibiotic-resistance gene – are growth-inhibited or killed by the antibiotic. An antibiotic-resistance gene may be included e.g. in a vector used for transfection of a population of cells; cells which internalize the vector (and which express the antibiotic-resistance gene) are then selected by growth on a medium containing the relevant antibiotic.

An antibiotic may be MICROBICIDAL or MICROBISTATIC; at lower concentrations, a microbicidal antibiotic may behave as a microbistatic antibiotic.

A mixture of antibiotics may behave synergistically or antagonistically (or may not display either effect).

Synergism is shown when different antibiotics, that are acting (simultaneously) on a given organism, produce an effect which is greater than the sum of their individual effects. For example, the antibiotics sulfamethoxazole and trimethoprim block different reactions in the same major metabolic pathway: sulfamethoxazole inhibits the formation of dihydrofolic acid (DHF), and trimethoprim inhibits the conversion of DHF to the important coenzyme tetrahydrofolate (THF); these two antibiotics act synergistically and they are used, together, in the therapeutic agent *cotrimoxazole*.

Antagonism, the converse of synergism, can occur in different ways. In one form, an antibiotic that inhibits growth (e.g. CHLORAMPHENICOL) antagonizes those antibiotics (such as the penicillins and other β-lactam antibiotics) whose activity depends on growth in the target cell. In a different form of antagonism, certain antibiotics stimulate cells to produce enzymes that inactivate *other* antibiotics; in one example, the β-lactam imipenem (or cefoxitin) induces the synthesis of β-lactamases – enzymes which can inactivate certain other β-lactam antibiotics.

Modes of action

To be effective at all, an antibiotic must be able to enter, or pass through, the cell envelope in order to reach the relevant target site(s). Moreover, an antibiotic can be effective against a given population of cells only if its concentration is above the appropriate minimum level for that agent under the given conditions.

Modes of action include:

- Interference with DNA gyrase (a topoisomerase), with consequent inhibition of DNA synthesis (e.g. novobiocin, quinolone antibiotics).
- Depletion of guanine nucleotides (by inhibiting synthesis of GMP), affecting the synthesis of nucleic acids (e.g. MYCOPHENOLIC ACID).
- Binding to ribosomes and inhibiting protein synthesis (e.g. aminoglycoside antibiotics, chloramphenicol, macrolide antibiotics (such as erythromycin), tetracyclines, viomycin).
- Binding to RNA polymerase, inhibiting transcription (e.g. rifamycins).
- Disruption of the bacterial cytoplasmic membrane, altered permeability of the membrane affecting the cell's integrity/survival (e.g. gramicidins, polymyxins).
- Inhibition of synthesis of the bacterial cell wall polymer peptidoglycan, leading to cell lysis (e.g. β-lactam antibiotics, vancomycin).
- Inhibition of the enzyme dihydrofolate reductase, thereby inhibiting tetrahydrofolate-dependent reactions (that include synthesis of deoxythymidine – and, hence, synthesis of DNA) (e.g. pyrimethamine, trimethoprim).
- Interference with DNA function by intercalating agents (e.g. actinomycin D, quinoxaline antibiotics).
- Interaction with sterols in the cytoplasmic membrane (e.g. in yeasts and other fungi), causing leakage (e.g. polyene antibiotics).
- Inhibition of the enzyme chitin synthase (in certain fungi), affecting cell wall synthesis (e.g. polyoxins).

Mechanisms of bacterial resistance to antibiotics

Resistance to a particular type of antibiotic is *constitutive* in cells which (i) lack the antibiotic's specific target, (ii) have a variant form of the target which is not susceptible to the antibiotic, and (iii) are impermeable to the antibiotic. Examples: (i) *Mycoplasma* is resistant to β-lactam antibiotics because it lacks a cell wall; (ii) strains of *Staphylococcus aureus* known as MRSA (methicillin-resistant *S. aureus*) generally contain a modified target PBP that is not susceptible to methicillin; (iii) typically, Gram-negative bacteria are insensitive to penicillin G (a β-lactam antibiotic) because this antibiotic is not able to penetrate the outer membrane of Gram-negative bacteria.

As well as constitutive resistance (see above) resistance can be acquired e.g. by mutation, or by the acquisition of plasmid (or transposon) gene(s) specifying resistance to one or more types of antibiotic.

Examples of resistance mechanisms include:

- Due to mutation, the target of a given antibiotic may be altered so that it fails to bind the antibiotic; consequently, the target (e.g. an enzyme) is not affected by otherwise inhibitory concentrations of that antibiotic. Thus e.g. a mutant form of the ribosomal protein L22 in *Staphylococcus aureus* confers resistance to quinupristin/dalfopristin (Synercid®), a streptogramin, and in *Mycobacterium tuberculosis* point mutations in the *rpoB* gene (encoding the β subunit of RNA polymerase) can confer resistance to rifamycins (such as rifampin) for which RNA polymerase is the target. A novel mechanism for antibiotic resistance has been reported in the Gram-positive pathogen *Streptococcus pneumoniae;* in this organism, resistance to linezolid (an oxazolidinone) is linked to mutation in a gene encoding a methyltransferase which, in *sensitive* strains, methylates a site in the 23S rRNA (the target for linezolid) [Genome Res (2009) 19:1214–1223].
- Transposon Tn*10* encodes an inducible efflux system that enables certain Gram-negative bacteria to externalize tetracycline via an 'efflux pump' located in the cell envelope.
- Mutant forms of certain envelope proteins are associated with decreased permeability. For example, alteration in outer membrane porins in *Enterobacter aerogenes* increases resistance to certain antibiotics, and in *Pseudomonas aeruginosa*

resistance to the aminoglycosides and other antibiotics can be determined through membrane permeability controlled by a TWO-COMPONENT REGULATORY SYSTEM.

- Degradation of antibiotics by plasmid-encoded or chromosome-encoded enzymes. Such enzymes include the inducible and constitutive β-LACTAMASES that cleave the β-lactam ring in, and inactivate, β-lactam antibiotics such as penicillins and cephalosporins. The enzyme chloramphenicol acetyltransferase, which degrades CHLORAMPHENICOL, is another example – as are acetyltransferases, adenylyltransferases and phosphotransferases that inactivate aminoglycoside antibiotics.
- Increased production of an affected metabolite. Thus, for example, synthesis of higher levels of *p*-aminobenzoic acid (PABA) may overcome the effect of competitive inhibition by sulfonamides.
- Gene amplification. At least some bacteria respond to antibiotics by a gene amplification mechanism which can e.g. upregulate enzymes that hydrolyze antibiotic(s). [Bacterial gene amplification: Nature Rev Microbiol (2009) 7:578–588.]

[Antibiotic resistance genes (database): Nucleic Acids Res (2009) 37(Database issue):D443–D447.]

Nucleic-acid-based tests for antibiotic resistance

See the entry ANTIBIOTIC RESISTANCE TESTING.

antibiotic resistance testing (nucleic-acid-based approaches for bacteria) The traditional, culture-based approaches for detecting antibiotic resistance in bacteria have been appropriate mainly for the species (e.g. *Escherichia coli*, *Salmonella* spp) that grow rapidly on laboratory media; for these species, sensitivity to a given antibiotic is shown by growth inhibition in the presence of the given antibiotic. Resistance to particular antibiotics in such organisms may be determined e.g. by growing the organisms on agar plates that contain antibiotic-impregnated paper disks (on the surface); antibiotics diffuse out from the disks, and organisms that are sensitive to any given antibiotic form a zone of growth-inhibition around the relevant disk.

Growth-based methods are less satisfactory for some slow-growing species (such as *Mycobacterium tuberculosis*) – and for pathogens such as *Chlamydia trachomatis* (which have a complex life cycle, and which fail to grow on ordinary laboratory media). Moreover, while culture-based methods can, in some cases, indicate the MIC (i.e. the minimum inhibitory concentration) of a given antibiotic (using e.g. the so-called E test) they give no information regarding the precise causation (i.e. mechanism) of resistance.

Nucleic-acid-based ('genotypic') methods for detecting antibiotic resistance (in various organisms) have developed over the last few decades. Below is a brief glimpse at some of the methods used for two important pathogens: *Staphylococcus aureus* and *Mycobacterium tuberculosis*.

Staphylococcus aureus

Among early studies, a BDNA ASSAY was used for detecting the MECA GENE of *S. aureus* in samples from blood-culture bottles (without the usual prior subculture to solid media) [J Clin Microbiol (1999) 37:4192–4193]; it was suggested that the results of this approach could help the clinician to choose an appropriate antibiotic (β-lactam or vancomycin) to match the actual requirement of an infection.

MULTIPLEX PCR has also been used for detecting the *mecA gene* (and the COA GENE) in *S. aureus*.

A pentaplex assay (targeting five genes) has been designed (a) to detect *S. aureus*, (b) to identify the *mecA* gene, and (c) to help distinguish between *community-acquired* MRSA and hospital-acquired MRSA (see the entry MRSA); for the latter purpose, one of the targeted genes was *lukS* – which encodes the PANTON–VALENTINE LEUCOCIDIN (q.v.) [BMC Microbiol (2009) 9:13].

Mycobacterium tuberculosis

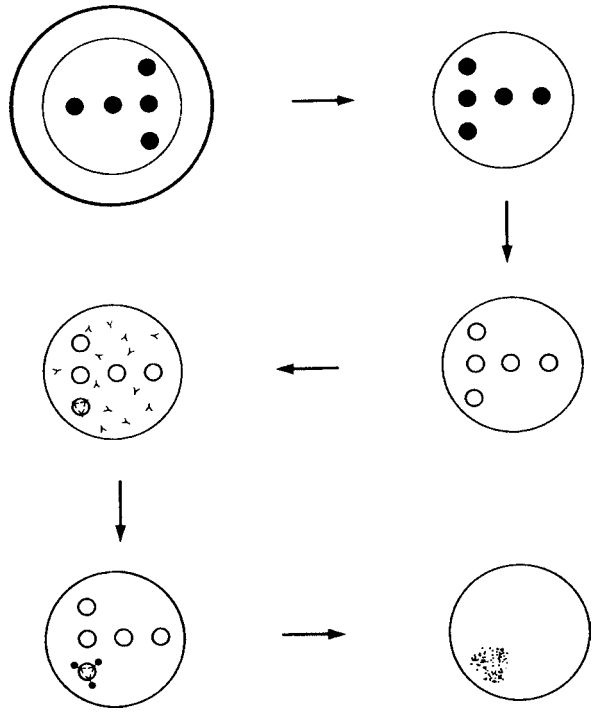
The resurgence of *M. tuberculosis* as a major threat, and the appearance of multidrug-resistant (MDR) strains, has given a strong impetus to the search for suitable nucleic-acid-based methods for detecting antibiotic resistance in this organism. Unlike resistance to β-lactam antibiotics in *S. aureus*, which is commonly due to specific gene products (e.g. PBP 2a or β-lactamase enzymes), the resistance of *M. tuberculosis* to anti-tuberculosis drugs usually arises through mutation(s); these mutations affect the genes whose products are the targets of anti-tuberculosis drugs. While resistance to some drugs (e.g. isoniazid, streptomycin) can be due to mutation in more than one gene, resistance to rifampin (see RIFAMYCINS) is usually due to mutation(s) in one gene, *rpoB*, which encodes the β-subunit of the RNA polymerase. Hence, detecting resistance to rifampin is technically simpler; moreover, strains that are resistant to rifampin are frequently resistant to various other anti-tuberculosis drugs, so that rifampin resistance may be a useful marker for multi-resistant strains.

Mutations in the *rpoB* gene are detected in various ways. In some methods the initial step is to use PCR to copy the *rpoB* gene – or a specific region of *rpoB* in which the resistance-associated mutations commonly occur. The PCR products are then examined e.g. by SSCP ANALYSIS or by a LINE PROBE ASSAY. Both methods are able to indicate *specific* mutations in the *rpoB* gene. (By identifying specific mutations, methods such as these are valuable e.g. for tracking resistance in the pathogen population and for epidemiological studies.)

In a different approach, *M. tuberculosis* was infected with mycophages carrying a fluorescent reporter gene (encoding either green fluorescent protein or ZsYellow); infection of these cells was carried out in the presence of either streptomycin or rifampin. The cells that were susceptible (sensitive) to the drug failed to exhibit fluorescence because (owing to inhibition of their RNA polymerase) the reporter gene was not transcribed; cells that were resistant to the drug were able to transcribe the reporter gene and (hence) exhibited fluorescence [PLoS ONE (2009) 4(3):e4870].

antibody-based library screening An alternative to screening with a nucleic acid probe (cf. COLONY HYBRIDIZATION). The method is shown diagrammatically in the figure.

antibody-labeling reagents See e.g. ZENON ANTIBODY LABELING REAGENTS.



ANTIBODY-BASED LIBRARY SCREENING: screening a cDNA expression library (diagrammatic). The cDNA library – consisting of protein-encoding cDNAs within expression vectors – was inserted into a population of bacteria; these bacteria were then inoculated onto agar plates in such a way that each bacterium gave rise to an individual colony. The object now is to determine *which* colony contains the cDNA that encodes (and expresses) the particular protein of interest. While each plate may contain >100 colonies, the plate considered here has only five colonies (for clarity).

The cDNA-containing colonies (*top*, *left*) are lightly overlaid with a nitrocellulose filter; when lifted off, the filter contains a mirror-image replica of the colonies (*top*, *right*). The cells on the filter are lysed, and the proteins they contain (including any encoded by the cDNAs) are bound to the filter (*center*, *right*). The filter is then treated with antibodies (Y) that are specific to the protein of interest (*center*, *left*); antibodies have bound to the given protein (present in *one* of the colonies). Unbound antibodies are washed away. The filter is then treated with *protein A* – a protein (derived from the bacterium *Staphylococcus aureus*) which binds to antibodies. Unbound protein A is washed away, leaving some protein A (•) bound to the specific antibody–protein complex (*bottom*, *left*). Before use, the protein A is labeled with a radioactive source so that its presence can be detected by *autoradiography;* this involves exposing a photographic film to the nitrocellulose filter for an appropriate time and then developing the film. When the film is developed (*bottom*, *right*) it identifies the particular colony on the original plate which contains the cDNA of interest. This colony can then be used as an inoculum to grow more of the cells which contain the specific cDNA.

This procedure clearly depends on efficient and specific protein–antibody binding. A protein synthesized in bacteria from a *eukaryotic* gene may lack the three-dimensional structure and/or post-translational modification that characterizes the native eukaryotic protein; it may therefore fail to bind antibodies raised against the native protein. (Moreover, even if the protein *is* synthesized with normal conformation, the process used to bind it to the filter may alter that conformation.) However, it is possible to produce antibodies that recognise, and bind to, a short, specific run of amino acid residues, even when the target protein lacks the conformation and/or post-translational modification of the native protein.

Finally, a colony containing the *particular* cDNA of interest may not be present on the plate being examined. Hence, it may be necessary to carry out the procedure on more than one plate.

Figure reproduced from *Bacteria in Biology, Biotechnology and Medicine*, 6th edition, Figure 8.10, page 222, Paul Singleton (2004) John Wiley & Sons Ltd, UK [ISBN 0-470-09027-8] with permission from the publisher.

anti-cancer therapy (*DNA technol.*) See CANCER THERAPY.

anticlinal (*anti*) Of a nucleotide: the conformation in which the oxygen atom within the sugar ring (–O–) is at the maximum distance from the 6-position of a purine (or the 2-position of a pyrimidine).

(cf. SYNCLINAL.)

anticoagulant (*DNA technol.*) A term that usually refers to an agent which inhibits the coagulation (i.e. clotting) of blood.

The anticoagulants include sodium citrate, sodium oxalate, heparin and SODIUM POLYANETHOLESULFONATE (SPS); the latter two agents can inhibit PCR.

anticodon Three consecutive bases in a tRNA molecule which are complementary to a CODON which specifies the particular amino acid carried by that tRNA.

An anticodon is written in the 5′-to-3′ direction – as is a codon.

(See also WOBBLE HYPOTHESIS.)

anti-downstream box (*or* antidownstream box) See DOWNSTREAM BOX.

antigenic drift (in viruses) See the entry GENETIC DRIFT.

antigenic variation Successive changes in cell-surface antigens which are exhibited by certain types of microorganism (e.g. *Trypanosoma* – see e.g. VSG).

At least two (distinct) mechanisms are known: (i) the alternative antigens are transcribed from specific, *pre-existing* genes (see PHASE VARIATION), and (ii) the alternative antigens arise by ongoing recombinational events – i.e. they are encoded by newly formed, rather than pre-existing, genes. An example of the second mechanism is the formation of new versions of a subunit in the fimbriae of *Neisseria gonorrhoeae;* variant forms of this subunit arise through repeated recombination between the chromosomal subunit gene, *pilE*, and another chromosomal gene, *pilS*, as well as between *pilE* and any homologous DNA received e.g. by transformation.

Some surface proteins of the fungus *Aspergillus fumigatus* may be hypervariable owing to recombination among internal mini- or microsatellite tandem repeat sequences in the genes encoding them [Eukaryotic Cell (2007) 6(8):1380–1391].

Antigenic variation apparently helps a pathogen to evade a host's immunologic defense mechanisms.

anti-miRNA oligonucleotide See e.g. the entry MICRORNAS.

antimutator gene Any gene whose activity reduces the rate of spontaneous mutation in a cell. For example, some strains of *Escherichia coli*, with a mutant form of DNA polymerase, have mutation rates below those of wild-type strains; in this case the antimutator activity presumably involves improved fidelity/proof-reading.

antiparallel (of strands in dsDNA) The (usual) arrangement in which a 5′-to-3′ strand is hybridized to a 3′-to-5′ strand, i.e. each end of a double-stranded DNA molecule has a 5′ terminal and a 3′ terminal.

(See also DNA.)

antiretroviral agents Agents with activity against retroviruses; some are useful e.g. in chemotherapy against AIDS – see e.g. the NON-NUCLEOSIDE REVERSE TRANSCRIPTASE INHIBITORS, NUCLEOSIDE REVERSE TRANSCRIPTASE INHIBITORS and PROTEASE INHIBITORS. Other agents that are useful against AIDS include the INTEGRASE INHIBITORS, THE FUSION INHIBITORS and the CHEMOKINE CORECEPTOR ANTAGONISTS.

(See also HAART, HIV-1 and RETROVIRUSES.)

The ability of certain indole derivatives to inhibit synthesis of a retroviral glycoprotein, protecting mice from a challenge with MLV (murine leukemia virus), has suggested a potential use for these compounds as antiretroviral therapeutic agents [PLoS ONE (2009) 4(2):e4533].

Specific HDAC inhibitors (see HDAC), as well as genome-specified agents, have been suggested as a means to address the problem of *latency* in HIV-1: see *Quiescent HIV-1* in the entry AIDS.

anti-reverse cap analog (ARCA) A chemically modified form of CAP ANALOG (Ambion, Austin TX) designed to maximize the efficiency of *in vitro* translation by ensuring that the cap analog is incorporated in the transcript in the correct orientation (a cap analog incorporated in the reverse orientation does not support translation).

The use of an anti-reverse cap analog enhanced translation by a factor of 1.5–1.8 in studies on translational efficiency of mRNA encoding the HIV-1 *gag* region [PLoS ONE (2008) 3(6):e2356].

antisense gene In a genetically engineered cell: any gene, inserted into that cell, whose presence is intended to inhibit or block the expression of another, endogenous gene.

antisense RNA Any natural or synthetic RNA whose sequence permits interaction with a given sense sequence – in RNA or in DNA – and which can affect the activity/expression of the given target molecule: see e.g. FINOP SYSTEM, MICRORNAS, MULTICOPY INHIBITION, POST-SEGREGATIONAL KILLING, the R1 PLASMID, RNA INTERFERENCE.

Other antisense systems

Engineered antisense systems have been devised using PNA or MORPHOLINO ANTISENSE OLIGOMERS.

(See also GENE SILENCING.)

antisense strand (of DNA) The non-CODING STRAND.

antisense transcriptome See TRANSCRIPTOME.

antiviral agent Any agent which has activity against viruses – *in vivo* and/or *in vitro*. (cf. ANTIBIOTIC.)

(See e.g. ANTIRETROVIRAL AGENTS; see also ACYCLOVIR, AMANTADINE, GANCICLOVIR, LAMIVUDINE, OSELTAMIVIR, RIBAVIRIN AND ZANAMIVIR.)

antizyme Any of a group of proteins associated with regulation of ornithine decarboxylase (ODC) – which is involved in biosynthesis of polyamines. In eukaryotes, functional antizyme is expressed in the presence of increased levels of polyamines – antizyme 1 promoting UBIQUITIN-dependent degradation of ODC via the 26S PROTEASOME. Antizymes are regulated by antizyme inhibitor.

[Antizyme in prokaryotes: BMC Biochem (2007) 8:1.]

AOX1 In the (methylotrophic) yeast *Pichia pastoris*: a highly regulated, inducible gene which encodes alcohol oxidase – a peroxisomal enzyme involved in metabolism of methanol.

(See also PPICZ VECTOR.)

AP ALKALINE PHOSPHATASE.

AP endonuclease Any enzyme with endonuclease activity that is involved in the excision of apurinic/apyrimidinic nucleotide residues (see e.g. BASE EXCISION REPAIR).

AP-PCR Arbitrarily primed PCR: any form of PCR which uses primers of arbitrary sequence and which amplifies random, but discrete, sequences of chromosomal DNA; AP-PCR has been used for TYPING bacteria.

PCR is initially carried out under low stringency, and the primers bind at various sites to each strand of heat-denatured chromosomal DNA; the binding of primers occurs at 'best-fit' sequences, and may include mismatches. In some cases two primers bind with relative efficiency, on opposite strands, at locations separated by a few hundred bases. If synthesis can occur normally from these two primers, a further round of cycling under low-stringency conditions, followed by many cycles under high-stringency conditions, may produce copies of an amplicon delimited by the two best-fit sequences. In the phase of high-stringency cycling not all the primers will bind to their best-fit sequences – so that only a proportion of the amplicons produced under low-stringency conditions will be amplified in the high-stringency phase.

The amplicons from a given sample are subjected to gel electrophoresis, and the stained bands of amplicons form the fingerprint. Strains are compared and classified on the basis of their fingerprints.

One advantage of this approach is that there is no need for prior knowledge of the genome sequence; there is no need to design specific primers, and any isolate is potentially typable.

Results are generally reproducible under standardized conditions in a given laboratory, but comparable results will not necessarily be obtained in other laboratories unless the procedures are *identical;* reproducibility of results depends not only on the primer sequence but also e.g. on the particular type of polymerase used and on the initial procedure used for preparing the sample DNA.

Some other named methods are based on the same principle – e.g. RAPD (random amplified polymorphic DNA) analysis and DAF (direct amplification fingerprinting). Such methods may differ e.g. in the length of primers used, annealing temperature for the primers, and the type of gel used for electrophoresis. The original AP-PCR employed primers of 20 to 50 nucleotides, an annealing temperature of ~40°C, and agarose gel. In RAPD the primers are often 10–20 nt, with annealing at ~36°C, and products separated in an agarose gel. DAF uses short primers (5–8 nt) at an annealing temperature of ~30°C; as there are many more (smaller) products, electrophoresis is carried out in a polyacrylamide gel and silver staining is used to detect bands in the fingerprint.

AP site (abasic site) In a nucleotide sequence: a site at which the base (purine or pyrimidine) is missing – the remainder of the nucleotide (sugar, phosphate) being present. (cf. GAP.)

APES See the entry AAS.

aphidicolin A tetracyclic diterpenoid, isolated from a fungus, which strongly inhibits the eukaryotic α DNA polymerase. Bacterial DNA polymerases are unaffected, but aphidicolin was reported to inhibit DNA synthesis in at least some members of the domain ARCHAEA (e.g. some methanogens).

[Example of use: Proc Natl Acad Sci USA (2009) 106(3): 803–807.]

Aphthovirus A genus of PICORNAVIRUSES which includes the causal agent of foot-and-mouth disease.

[Inhibition of replication of *Aphthovirus* by siRNA: Virol J (2008) 5:86. Diagnosis of foot-and-mouth disease in Brazil by RT-PCR: BMC Vet Res (2008) 4:53.]

APO-1 (CD95) The cell-surface Fas receptor: see APOPTOSIS.

APOBEC-1 See CYTIDINE DEAMINASE and RNA EDITING.

apolipoprotein B See RNA EDITING.

apoptosis In various types of eukaryotic cell: death that occurs in a regulated ('programed') manner – an organized process in which fragmentation of the genome is a characteristic feature.

DNA fragmentation can be assessed by the TUNEL ASSAY.

In the vertebrates, apoptosis occurs naturally, for example, during embryogenesis. It occurs during metamorphosis in the invertebrates.

Apoptosis can be a response to damaged DNA in the cell: see e.g. the entry P53.

Apoptosis can be induced by physical agents, such as heat and radiation, and by certain chemical agents; for example, in osteoblasts, it may be stimulated by glucocorticoids (leading to osteoporosis).

Major players in apoptosis are certain intracellular cysteine proteases (*caspases*) that are synthesized in an inactive form (zymogen) and activated specifically during apoptosis; these enzymes cleave specific protein substrates at a site next to an aspartic acid residue. Cleavage by a caspase may inactivate or activate the substrate – e.g. caspase-mediated activation of the nuclease *caspase-activated DNase* (CAD) results in DNA fragmentation. (CAD is normally present in an inactive form, complexed with an inhibitory partner known as ICAD.)

Caspase activation apparently occurs primarily by cleavage of the zymogen. In at least some cases, one caspase may be activated by another, 'upstream' caspase in a caspase cascade.

In macrophages that have taken up the bacterium *Shigella flexneri* (by phagocytosis), apoptosis in the macrophage can be promoted by a plasmid-encoded protein, IpaB, secreted by *S. flexneri;* IpaB activates the host's IL-1β-converting enzyme (ICE, a cysteine protease) which then initiates apoptosis.

Other factors that may promote apoptosis include binding of certain cytokines (e.g. TNFα) and cytotoxic ($CD8^+$) T cells to specific cell-surface receptors; in the latter case, Fas ligand (on the T cell) binds to the target cell's Fas receptor (which may also be referred to as CD95 or APO-1).

Inhibition of apoptosis is seen, for example, in B lymphocytes latently infected with EPSTEIN–BARR VIRUS; in these B cells apoptosis is inhibited by small EBV-encoded RNAs: see EBER.

Fluorophore-labeled probes, which label *active* caspases *in vivo* ('activity-based probes'), allowed researchers to monitor

the kinetics of apoptosis in live mice [Nature Med (2009) 15: 967–973].

Ongoing study of apoptosis has enabled the development of pro-apoptotic receptor agonists that may be used to promote apoptosis in cancer cells [Nature Rev Drug Discovery (2008) 7:1001–1012].

(See also ALTERNATIVE SPLICING (gene *bcl-x*), BARNASE, BINASE, CISPLATIN, MDA-7/IL-24, MYB, MYC and SURVIVIN.)

APSR (amplicon primer site restriction) In PCR, a method for preventing the contamination of a reaction mixture with amplicons from previous assays. In APSR, *all* the assays are conducted with primers whose 5′ ends carry a recognition site for a (type IIS) RESTRICTION ENDONUCLEASE which cleaves both strands 3′ of its binding site. In a reaction mixture, the (added) restriction enzyme will cut any carry-over amplicons but will not cleave the *template* DNA (unless, by chance, it contains the given recognition site); the restriction enzyme is inactivated during PCR temperature cycling.

(See also AMPLICON INACTIVATION.)

APT paper See SOUTHERN BLOTTING.

aptabody A construct, made from multiple RNA APTAMERS, which can function as a protein. The prototype aptabody was a tetravalent construct that was able to mimic an antibody by simultaneously binding two molecules of the *Drosophila* B52 protein and two molecules of streptavidin. The performance of the aptabody was reported to rival that of a monoclonal antibody against the B52 protein [Nucleic Acids Res (2009) doi: 10.1093/nar/gkp243].

aptamer (1) Any of a large number of synthetic oligonucleotides that can adopt (at least one) three-dimensional structure and bind, with high specificity, to a given ligand; both RNA and DNA aptamers can be synthesized.

(See also APTABODY and INTRAMER.)

Selection of an RNA aptamer for a given target molecule can be achieved by SELEX (systematic evolution of ligands by exponential enrichment). Briefly, in SELEX, the immobilized target ligand (for example, a protein) is exposed to a large and diverse population of oligonucleotides synthesized with random sequences. After removal of unbound oligos, the relatively few bound oligos are eluted and then converted to cDNAs. The cDNAs are amplified by PCR using primers that incorporate a promoter; the amplicons are transcribed, and the transcripts are used in another round of selection with the target molecule. This cycle is repeated, target–RNA binding becoming more specific at each round; the range and affinity of the selected aptamers can be determined e.g. by regulating buffer conditions at the binding stage. (A principle similar to this is found in PHAGE DISPLAY.)

In general, SELEX is useful for analyzing protein–nucleic acid binding and e.g. interactions between RNA and various low-molecular-weight molecules.

[Combining SELEX and yeast 3-hybrid system for *in vivo* selection and classification of aptamers: RNA (2007) 13(4): 614–622.]

Uses/applications of aptamers

The high-level binding specificity of the aptamers, and their ability to be linked to other molecules, has suggested various uses.

An aptamer–shRNA fusion transcript was used to regulate gene expression within mammalian cells. The activity of the shRNA moiety was controlled by an interaction between the aptamer and its ligand, theophylline (1,3-dimethylxanthine), the latter being supplied exogenously.

An aptamer – known to bind to prostate tumor cells – was conjugated to an siRNA via a streptavidin bridge; on addition to cells, this conjugate was internalized, and the siRNA was able to inhibit gene expression as efficiently as when it was transfected by a lipid-based procedure.

[Magnetically mediated selection of DNA aptamers: BioTechniques (2007) 43(3):344–353.]

[Cellulose-binding aptamers: Nucleic Acids Res (2007) 35 (19):6378–6388.]

[An *in silico* directed evolution study of DNA aptamers that bind allophycocyanin: Nucleic Acids Res (2009) 37(1):e6.]

[*In silico* selection of (RNA) aptamers: Nucleic Acids Res (2009) doi: 10.1093/nar/gkp408.]

A 58-nt aptamer sequence (which was developed by the use of SELEX) can – following transcription – adopt *two* distinct structures with different functions; both of the structures are required (jointly) to inhibit a particular receptor. These two structures cannot be interconverted (by unfolding and then re-folding) apparently owing to their inherent structural stability [Nucleic Acids Res (2009) doi:10.1093/nar/gkp284].

Using a SELEX-based method, five aptamers were selected for their high specificity towards ligands at the surface of the (pathogenic) Gram-positive bacterium *Staphylococcus aureus*– each of the five aptamers recognizing a different molecular target. When used in combination, this panel of five aptamers constitutes a system for the detection and identification of *S. aureus* [Nucleic Acids Res (2009) doi: 10.1093/nar/gkp489].

(See also APTAZYME and PROXIMITY LIGATION ASSAY.)

(2) A *natural* sequence within a RIBOSWITCH.

aptazyme (*DNA technol.*) An APTAMER–RIBOZYME construct in which the aptamer – on binding a certain ligand – typically undergoes a conformational change resulting in modification/regulation of the ribozyme.

In one application, ligand-binding by the aptamer promotes the ability of the ribozyme moiety to polymerize ribonucleotides – using another copy of the ribozyme as template. Thus, in the reaction mixture, the ribozymes synthesize copies of each other, and exponential amplification of the RNA occurs in the presence (although not in the absence) of the particular ligand. As the rate of increase in copies of RNA is governed by the ligand's *concentration*, the concentration of the ligand in a given sample can be assessed from the increase in RNA. Thus, systems can be set up in which an aptamer, in responding to a particular ligand, provides a sensitive assay for that

ligand. The ligand may be any of a variety of small molecules or proteins – e.g. those relevant in medical or environmental investigations. [Nature Biotechnol (2009) 27:288–292.]

One study has examined the performance of an engineered drug-sensing aptazyme made from a theophylline-responsive aptamer fused to a HAMMERHEAD RIBOZYME [RNA (2009) 15(1):76–84].

apurinic Lacking a purine residue: see AP SITE.

apyrase A nucleotide-degrading enzyme (EC 3.6.1.5) with various applications in technology (see e.g. PYROSEQUENCING).

apyrimidinic Lacking a pyrimidine residue: see AP SITE.

***araBAD* operon** See OPERON.

Arabidopsis thaliana A small cruciferous plant commonly used in plant genetics because of its simple genome and its short generation time.

[Genome of *A. thaliana*: Nature (2000) 408:791–826.]

(See also FLORAL DIP METHOD.)

araC See OPERON.

Aranesp® See BIOPHARMACEUTICAL (table).

arbitrarily primed PCR See AP-PCR.

arbitrary degenerate primer (AD primer) See TAIL-PCR.

arboviruses Arthropod-borne viruses: a (non-taxonomic) group which includes those viruses that are able to replicate within a vertebrate host *and* in arthropods (such as mosquitoes and ticks). Many of the arboviruses can cause disease in humans or animals; they include bluetongue virus (genus *Orbivirus*), Lassa fever virus (genus *Arenavirus*), West Nile virus (genus *Flavivirus*) and yellow fever virus (genus *Flavivirus*).

Arboviruses that infect mosquitoes can cause damage to the *mosquito* host – damage which may be limited by the insect's antiviral response: RNA INTERFERENCE (q.v.). An experiment to study the role of RNAi in protecting the mosquito used a recombinant SINDBIS VIRUS that expressed a protein (the B2 protein of Flock House virus) which inhibits RNAi. In this experiment it was found that expression of the B2 protein in mosquito cells inhibits RNAi. The recombinant Sindbis virus, inoculated intrathoracically, was lethal in different mosquito species, and the recombinant virus caused mortality in *Aedes aegypti* in a dose-dependent manner [BMC Microbiol (2009) 9:49].

ARCA See ANTI-REVERSE CAP ANALOG.

Archaea One of the two domains of prokaryotic organisms, the other being BACTERIA. Organisms in the two domains differ e.g. in 16S rRNA sequences, composition of cell wall macromolecules, composition of cytoplasmic membrane lipids and flagellar structure. The general features of gene expression in the two domains are also dissimilar.

(See also PROKARYOTE.)

archaean (*syn*. archaeon) An organism within the (prokaryotic) domain ARCHAEA.

Archaebacteria A now-obsolete kingdom of prokaryotes; the organisms formerly placed in this taxon are currently classified in the domain ARCHAEA.

ArchaeMaxx™ A polymerase-enhancing factor, marketed by Stratagene (La Jolla CA), which was designed to overcome the so-called DUTP POISONING effect.

This factor is used e.g. in association with the *PfuTurbo*® and Herculase® DNA polymerases (also marketed by Stratagene).

[Example of use: PLoS Biol (2006) 4(3):e73.]

archaeon See the entry ARCHAEAN.

ARES™ See PROBE LABELING.

Argonaute Any of a family of proteins which occur in higher eukaryotes and which are involved e.g. as effector molecules in the gene-silencing activities of certain small RNAs (see e.g. MICRORNAS and RNA INTERFERENCE).

Two subfamilies of Argonaute proteins are recognized. The proteins of the Ago subfamily are ubiquitous, being involved e.g. in post-transcriptional gene silencing (by miRNAs and siRNAs) in which mRNAs may be either physically degraded or functionally repressed. (See also BIM.) Proteins of the Piwi subfamily (see PIWI PROTEIN) apparently occur primarily in germ line cells – in which they act in conjunction with small RNAs (piRNAs) e.g. to silence TRANSPOSABLE ELEMENTS.

Various species are known to have several (or many) types of Argonaute protein, and, in a given species, these proteins may carry out different functions. Some Argonaute proteins have endonucleolytic activity, and these, in association with miRNAs or siRNAs, are able to cleave specific mRNAs; this cleavage is reported to occur when there is perfect complementarity between the small RNA and its target sequence on an mRNA. In human cells, only the Ago2 was reported to have nucleolytic capability [Genome Biol (2008) 9(2):210]. Both Ago1 and Ago2 apparently have nucleolytic capability in the cells of *Drosophila* – in which miRNA-directed cleavage of mRNAs appears to involve Ago 1, while Ago 2 works mainly in association with siRNAs.

Other types of Argonaute protein may be involved in the repression of translation when miRNAs bind to imperfectly complementary target sites.

Ago proteins may also be involved in *transcriptional gene silencing* and *transcriptional gene activation* (see SIRNA).

(See also ARNA (sense 2).)

ARMS Amplification-refractory mutation system: a procedure used e.g. for demonstrating or detecting a point mutation *at a specific site* in DNA whose wild-type (non-mutant) sequence is known.

Essentially, use is made of a primer in which the 3′-terminal nucleotide is complementary to the given *mutant* base at the specific site. After hybridization, extension of the primer by a polymerase signals the presence of the mutation at that site, while the absence of extension indicates the presence of a wild-type (or other) nucleotide.

armyworm The insect *Spodoptera frugiperda*. Cell cultures of this organism are used e.g. for the synthesis of recombinant proteins in BACULOVIRUS EXPRESSION SYSTEMS.

(See also SF9 CELLS.)

aRNA (1) Antisense RNA – see the entry MESSAGEAMP ARNA AMPLIFICATION KIT.

(2) Aberrant RNA: a type of RNA which is produced in cells (under certain conditions) and which can be converted to an effector molecule by modification involving the activity of an RNA-dependent RNA polymerase.

In the fungus *Neurospora crassa*, an aRNA that is induced by damage to DNA acts as a precursor to a small interfering RNA (∼20 nt long) referred to as qiRNA. The production of qiRNA (so named because of its interaction with Argonaute protein QDE-2) depends on the activity of an RNA-dependent RNA polymerase – and also requires a helicase and the Dicer enzyme [Nature (2009) 459:274–277].

Note. aDNA is used to refer to *ancient* DNA.

array A shortened version of MICROARRAY – but also used to refer to other oligonucleotide- or tissue-based arrangements etc. with analogous or distinct uses.

ARS Autonomously replicating sequence: a genomic sequence which, if isolated and linked to a non-replicative fragment of DNA, promotes independent (extrachromosomal) replication of that fragment within the cell.

ARSs were first identified in *Saccharomyces cerevisiae*, an organism which contains, on average, one ARS in every ∼40 kb of genomic DNA. ARSs also occur in at least some yeast plasmids.

It appears that some ARSs are active chromosomal origins while others are *silent origins;* some of the silent origins (and some active ones) may function as transcription silencers.

[Genome-wide hierarchy of replication origin usage in the yeast *Saccharomyces cerevisiae*: PLoS Genetics (2006) 2(9): e141.]

Factors reported to contribute to the efficient replication of ARS-containing plasmids in yeast cells include (i) the CEN (centromere) element and (ii) minichromosome maintenance protein 1 (Mcm1).

The circular dsDNA genome of human papillomavirus type 16 can replicate stably in *S. cerevisiae* independently of ARS or CEN; sequences in the viral DNA apparently substitute for both ARS and CEN [J Virol (2005) 79(10):5933–5942].

ARSs were also reported in species of the ARCHAEA.

(See also YEAST ARTIFICIAL CHROMOSOME.)

Artemis A DNA sequence viewer and annotation tool (Sanger Institute) used for displaying features of a given sequence of nucleotides and the results of analyses of that sequence (and of its six-reading-frame translation).

Artemis nuclease A multifunctional enzyme involved e.g. in a DNA repair mechanism – *non-homologous DNA end-joining*, NHEJ – used e.g. for repair of double-stranded breaks caused by ionizing radiation; it trims the ends of a double-stranded break, making them suitable for subsequent ligation by other components of the NHEJ system [see e.g. Nucleic Acids Res (2008) 36(10):3354–3365].

arthropod-borne viruses See the entry ARBOVIRUSES.

artificial *trans*-encoded sRNAs See the entry SRNAS.

ascospores See SACCHAROMYCES.

ascus See SACCHAROMYCES.

aseptic technique Measures which avoid the contamination of cultures, sterile media etc. – and/or contamination of persons, animals or plants – by microorganisms which are present in the environment (e.g. in the air) or which may be associated with particular source(s).

In this approach, the vessels used for media etc. must be sterile before use (e.g. pre-sterilized Petri dishes), and sterile material should not be exposed to any non-sterile conditions before use (see also STERILIZATION).

The working surfaces of forceps and other types of metal instrument (such as bacteriological loops etc.) are sterilized by 'flaming' before use, and the rims of bottles etc. used for dispensing sterile (non-flammable) materials are also flamed.

Benches are regularly treated with disinfectants and/or with ULTRAVIOLET RADIATION (UVR). The so-called 'germicidal' lamps, which may emit UVR at ∼254 nm, are used e.g. for the disinfection of air and exposed surfaces in enclosed areas. In general, UVR has rather poor powers of penetration, and its effects on microorganisms may be reversible by certain DNA repair processes (see e.g. UVRABC-MEDIATED REPAIR).

Some procedures, e.g. handling specimens likely to contain certain pathogens (such as *Mycobacterium tuberculosis*, or certain highly hazardous viruses such as Ebola virus or Lassa fever virus), are carried out in a SAFETY CABINET.

Asian flu See the entry INFLUENZAVIRUS.

ASLV vectors Avian sarcoma and leukosis virus-based vectors – see e.g. GENE THERAPY (*Viral vectors*).

ASP APOBEC-1-stimulating protein: see RNA EDITING.

aspart See INSULIN ASPART.

Aspergillus A genus of (mycelial) fungi (class Hyphomycetes) which are common in nature. *A. fumigatus* is pathogenic (see also AFLATOXINS), while many species are saprotrophic.

[*Aspergillus* genome database: Nucleic Acids Res (2009) doi: 10.1093/nar/gkp751.]

(See also INSERTIONAL MUTAGENESIS.)

ASR/GMP oligonucleotides Primers that comply with specifications of the FDA (Food and Drug Administration) in the USA and which are used (in restricted circumstances, and under stated conditions) for specific laboratory tests designed to provide analytical results.

assembly of DNA (*in vitro*) See the entry DNA CUTTING AND ASSEMBLY.

assisted reproductive technology See e.g. SCNT (for cloning of domestic animals) and IVF (in humans).

association studies See e.g. COPY NUMBER VARIANT and SNP GENOTYPING.

asymmetric PCR A form of PCR in which the concentration of *one* of the primers is much lower than that of the other (e.g. a ratio of 1:50); during temperature cycling, this primer will be quickly used up – so that only one strand of the target sequence will be significantly amplified.

Uses of asymmetric PCR include the preparation of probes and the preparation of single-stranded DNA for sequencing.

ssDNA products from PCR can also be obtained in a different way. One of the two types of primer can be labeled with BIOTIN and the reaction carried out with both primers in their

normal concentrations. When the reaction finishes, STREPTAVIDIN is added; this binds only to the biotin-labeled strands. Subsequent gel electrophoresis (in a denaturing gel) separates the two types of strand: the mobility of a streptavidin-bound strand is much lower. In this approach the biotinylated primer gives rise to the strand which is *not* required.

AT type See the entry BASE RATIO.

ATMS *p*-Aminophenyltrimethoxysilane: a reagent used for covalently binding DNA probes to a solid support when preparing a MICROARRAY. (In an earlier procedure, DNA was bound *non*-covalently to glass slides by the reagent poly-L-lysine.) [Method: Nucleic Acids Res (2001) 29:e107.]

(See also DENDRICHIP.)

atomic force microscopy (AFM; or scanning force microscopy) A method for imaging *surfaces*, including those of molecules and of (living) cells, in e.g. air or liquid, at nanometer-scale resolution. Essentially, the object's surface is scanned in a raster pattern with a fine probe located underneath a traveling cantilever; a laser, reflected from the cantilever – and thus providing information on the movements of the cantilever – is detected by a photodiode system, and the incoming signals are converted, by computer, into a surface profile.

Atomic force microscopy was used e.g. to study chromatin structure and nucleosome remodeling [Methods (2007) 41(3): 333–341]; to analyze repetitive α-satellite DNA [Eur Biophys J (2007) 37(1):81–93]; and for the characterization of an antibody scFv [Nanomedicine (2008) 4(1):1–7].

Atomic force microscopy has been used for direct visualization of G-quadruplexes in DNA [Nucleic Acids Res (2009) doi: 10.1093/nar/gkp679].

atsRNAs Artificial *trans*-encoded sRNAs: see the entry SRNAS.

***att* sites** Sites involved e.g. in SITE-SPECIFIC RECOMBINATION that occurs when the PHAGE LAMBDA genome integrates into a bacterial chromosome.

Lambda *att* sites are used in commercial DNA technology systems: see e.g. the GATEWAY SITE-SPECIFIC RECOMBINATION SYSTEM, MULTISITE GATEWAY TECHNOLOGY, BP CLONASE and LR CLONASE.

(See also *att*Tn*7* in the entry TN7.)

attaching and effacing lesion See PATHOGENICITY ISLAND.

***att*B, *att*P** See the entry PHAGE LAMBDA. (See also ATT SITES and GATEWAY SITE-SPECIFIC RECOMBINATION SYSTEM.)

attenuator control See OPERON.

***att*L, *att*R** See the entry PHAGE LAMBDA. (See also ATT SITES AND GATEWAY SITE-SPECIFIC RECOMBINATION SYSTEM.)

atto- A prefix meaning 10^{-18}.

AttoPhos™ A reagent (Promega, Madison WI) which can be cleaved (by ALKALINE PHOSPHATASE) to yield a fluorophore.

Attophos™ has been used e.g. in studies on the staging of trypanosomiasis [PLoS Negl Trop Dis (2009) 3(6):e459] and in immunoblotting in studies on prion proteins [Mol Biol Cell (2009) 20(1):233–244].

att*Tn*7 See the entry TN7.

Aubergine A member of the Piwi subfamily of ARGONAUTE proteins which is found in *Drosophila*. It is reported to be essential e.g. for germ cell formation.

(See also PIWI PROTEIN.)

Augmentin A combined formulation of the β-lactam antibiotics clavulanic acid (see β-LACTAMASES) and amoxycillin; it is used e.g. for the treatment of urinary tract infections.

autoactivation (*syn*. self-activation) (two-hybrid systems) Activation of a reporter system without prior interaction between the bait and prey proteins.

autocatalytic aptazyme See APTAZYME.

autocatalytic splicing See SPLICING.

autoclave An apparatus within which objects and/or materials are sterilized by saturated (air-free) steam under pressure; the conditions in a working autoclave are commonly within the range 115°C (∼69 kPa; 10 lb/inch^2) to 134°C (∼207 kPa; 30 lb/inch^2).

STERILIZATION in an autoclave is carried out e.g. when preparing certain types of media. Heat-labile constituents of a medium (e.g. a solution of an antibiotic) may be membrane-filtered before being added to a sterile (autoclaved) medium.

Some steam-impermeable items that cannot be sterilized by autoclaving may be sterilized in a hot-air oven at 160–170°C for ∼1 hour.

autoclave tape A paper strip (usually self-adhesive) which is included with the objects being sterilized in an autoclave; it exhibits a visible change (e.g. in color) when it is subjected to appropriate sterilizing conditions, and can therefore act as a check on the correct operation of the autoclave.

***Autographa californica* NPV** See NUCLEAR POLYHEDROSIS VIRUSES.

autoinducer (in quorum sensing) See QUORUM SENSING.

automated sequencing (of DNA) A method used for rapidly sequencing DNA fragments of up to ∼800 nt. Essentially, the process involves conventional chain-termination (i.e. Sanger) sequencing (see DIDEOXY METHOD) with fluorophore-labeled ddNTPs – each type of ddNTP (A, G, C, T) being labeled with a fluorophor that emits a distinctive color on excitation. The sequencing products are separated by polyacrylamide gel electrophoresis and the bands of products are scanned by a laser; the positions of individual nucleotides, identified by the color of the fluorescence, are recorded automatically.

(See also DNA SEQUENCING AND PYROSEQUENCING.)

autonomously replicating sequence See ARS.

autoplast A PROTOPLAST OR SPHEROPLAST which develops as a result of activity of an organism's own autolytic enzymes.

autoradiography A procedure in which a radioactive source is detected or quantitated by its effect on a photographic film; a film is exposed to the radioactive source, in the dark, and for an appropriate period of time, and is subsequently processed.

Autoradiography is used e.g. for investigating intracellular processes (radioactive isotopes being incorporated into biomolecules), and is also used for detecting bands of products (fragments of DNA), following gel electrophoresis, by means of radioactively labeled probes (see PROBE LABELING).

In general, optimal resolution may require the use of those isotopes which have relatively low-energy emission (such as

tritium, ^{3}H) rather than those (such as ^{32}P) which have high-energy emission.

autosomal dominant disorder Refers to a genetic disorder in which the phenotypic manifestation arising from expression of an abnormal autosomic allele occurs in the presence of the corresponding normal allele – that is, the influence of the abnormal allele overrides that of the normal allele. This type of disorder tends to exhibit a so-called *vertical* pattern of transmission from one generation to the next; in such cases, an abnormal trait is more likely (than in an autosomal recessive condition) to affect each successive generation.

Both males and females can be affected – and, unlike the situation in X-linked dominant disorders, father-to-son transmission can occur.

Examples of this disorder: AXENFELD–RIEGER SYNDROME and PEUTZ–JEGHERS SYNDROME (and see also CHARCOT–MARIE–TOOTH DISEASE).

(cf. AUTOSOMAL RECESSIVE DISORDER and X-LINKED DISORDER.)

autosomal recessive disorder Refers to a genetic disorder in which manifestation of the abnormal phenotype is exhibited when an abnormal allele is not accompanied by the presence of the corresponding normal, wild-type allele; heterozygous individuals with one normal allele do not usually exhibit the abnormal phenotype.

Both males and females can be affected.

Only when mating occurs between two (homozygously) affected individuals will the trait necessarily appear in all the offspring; mating between two heterozygous individuals, or between one heterozygous and one normal individual, tends to spare some of the offspring, so that an autosomal recessive disorder may miss generation(s) and is said to exhibit a *horizontal* mode of transmission.

(cf. AUTOSOMAL DOMINANT DISORDER and X-LINKED DISORDER.)

autosome Any chromosome other than a HETEROSOME.

autotransporter See the entry OMPT GENE.

auxins Phytohormones (plant hormones) which promote stem elongation and other aspects of plant development; the auxins are derivatives of tryptophan. Indole 3-acetic acid (IAA; also called 'auxin' or 'heteroauxin') is a major auxin; it is synthesized from the precursor indole 3-acetonitrile (IAN).

Abnormally high levels of auxins (*hyperauxiny*) are found in some plant diseases (e.g. CROWN GALL).

auxotrophic mutant Any microorganism which, as a result of a mutation, is unable to synthesize an essential nutrient and which therefore can grow only if provided with an exogenous source of that nutrient. (An organism which does *not* contain such a mutation, and which can synthesize all of its essential nutrients, is called a *prototroph*.)

One example of an auxotrophic mutant is mentioned in the entry AMES TEST.

Isolation of auxotrophic bacteria

The usual selective procedures are not suitable for isolating auxotrophic bacteria from a mixture of prototrophs and auxotrophs: a medium that supports the growth of auxotrophs will also permit the growth of the corresponding prototrophs.

The *limited enrichment* method uses an agar-based *minimal medium* enriched with *small* amounts of nutrients. A minimal medium is one which supports the growth of prototrophs but – because it lacks one or more essential nutrients – does not support the growth of auxotrophs. Any colony of auxotrophic cells will quickly exhaust the small amount of nutrients in its vicinity – so that the colony will remain small; however, the unrestricted growth of prototrophs on this medium means that colonies of prototrophs will be larger than those of the auxotrophs. Hence, the small colonies indicate presumptive auxotrophs.

The *delayed enrichment* technique employs an agar-based minimal medium for initial growth, so that prototrophs (only) form colonies in the initial incubation. Complete medium is then poured onto the plate and allowed to set; the nutrients in this medium diffuse into the minimal medium below, allowing the growth of auxotrophs. Again, small colonies indicate presumptive auxotrophs.

For (penicillin-sensitive) bacteria, auxotrophic mutants can be isolated by virtue of their inability to grow in a (penicillin-containing) minimal medium; penicillin is an antibiotic that acts only on *growing* cells. In this technique, a well-washed population of bacteria (that includes auxotrophs) is exposed to penicillin in a minimal medium; the prototrophs (which are able to grow) are killed by the penicillin. The remaining cells are washed and re-plated on a complete medium in order to recover any auxotrophs that may be present. It's important to note that, if the auxotrophs had developed through an *in vitro* process of mutagenization, the cells must be allowed to grow for several generations in complete medium prior to exposure to penicillin; this is because newly mutated cells will contain a full complement of (prototrophic) enzymes – auxotrophy developing only after several rounds of cell division (during which the prototrophic enzymes are 'diluted out'). A failure to observe this requirement would cause the death of any newly mutagenized auxotrophs on exposure to penicillin. A further requirement in this method is that only a low concentration of cells be used; the reason for this is that auxotrophs should not be allowed to grow on nutrients released by lysed prototrophs as this would render them susceptible to lysis by penicillin. For this reason, STREPTOZOTOCIN (q.v.) may be employed in place of penicillin.

Auxotrophs may also be isolated by REPLICA PLATING.

avian erythroblastosis virus See the entry ERB.

avirulence gene See GENE-FOR-GENE CONCEPT.

Axenfeld–Rieger syndrome An autosomal dominant disorder involving eye defects and certain systemic abnormalities. The syndrome has been associated with mutations in the *PITX2* gene – which encodes a transcription factor – or mutation in other genes (e.g. *FOXC1*, *PAX6*). [A novel *PITX2* mutation: Mol Vision (2008) 14:2205–2210.]

In some patients the disorder is reported to involve aberrant splicing of pre-mRNA; it was suggested that variability in the

extent of the splicing fault may be reflected in the variability of phenotypic manifestations [BMC Med Genet (2006) 7:59].

5-aza-2′-deoxycytidine (decitabine) A nucleoside analog used for studying DEMETHYLATION and as a therapeutic agent in the treatment of e.g. myelodysplastic syndrome and myeloid leukemia. In cells, 5-aza-2′-deoxycytidine is phosphorylated and incorporated into DNA; it then inhibits methyltransferase activity – for example, it has been shown to reactivate tumor suppressor genes silenced by methylation of promoter DNA. While effective against leukemias, this agent is generally less effective against solid tumors. Studies on MELANOMA cells suggest that the activity against this type of cancer could be made more effective given a better knowledge of the epigenetic and genetic background of each individual tumor [PLoS ONE (2009) 4(2):e4563]. (These studies have also suggested that HDAC and proteasomal inhibitors might act synergistically with the methyltransferase inhibitor in some patients.)

Studies to determine optimal therapeutic dosing of 5-aza-2′-deoxycytidine have been carried out in cultured cells [BMC Cancer (2008) 8:128], and 5-aza-2′-deoxycytidine (combined with various HDACs) has been studied in Ewing's sarcoma cells [Cancer Cell Int (2008) 8:16]. The agent has also been used in studies on promoter methylation in breast tumor cell lines [BMC Cancer (2009) 9:80] and for the identification of hypermethylated genes [BMC Med Genomics (2009) 2:11].

5-aza-2dC The nucleoside analog 5-AZA-2′-DEOXYCYTIDINE.

5-aza-CdR The nucleoside analog 5-AZA-2′-DEOXYCYTIDINE.

5-azacytidine A nucleoside analog that is used e.g. for inhibiting DNA methyltransferases; its activity is similar to that of 5-AZA-2′-DEOXYCYTIDINE (q.v.).

azaserine (*O*-diazoacetyl-L-serine) An agent with antimicrobial and antitumor activity produced by *Streptomyces* sp (a Gram-positive bacterium). Azaserine inhibits the activity of certain enzymes, including phosphoribosylformylglycinamidine synthetase – thus inhibiting biosynthesis of purines and, hence, nucleotides.

(cf. DON; see also HADACIDIN.)

AZT See ZIDOVUDINE.

B

B A specific indicator of ambiguity in the recognition site of a RESTRICTION ENDONUCLEASE or in another nucleic acid sequence; 'B' indicates C or G or T in DNA, C or G or U in RNA.

B-DNA The form of DNA that appears to be the most common *in vivo* – a RIGHT-HANDED HELIX of ∼10.4 base pairs/turn, diameter ∼20Å. [Flexibility: Nucleic Acids Res (2009) doi: 10.1093/nar/gkp962.] This is the basis of the Watson–Crick model of DNA (the 'double helix').

(cf. A-DNA and Z-DNA.)

B family (of DNA polymerases) A group of DNA-DEPENDENT DNA POLYMERASES that include prokaryotic, eukaryotic and viral enzymes. Members of the B family include DNA polymerases of phages φ29 and T4; the eukaryotic enzymes pol α, pol δ and pol ϵ (important for genome replication and DNA repair functions e.g. in a human context); and pol ζ (see DNA POLYMERASE ζ).

(See also A FAMILY, X FAMILY and Y FAMILY.)

B2 protein(of Flock House virus) See FLOCK HOUSE VIRUS.

BAC BACTERIAL ARTIFICIAL CHROMOSOME.

Bac-to-Bac® A BACULOVIRUS EXPRESSION SYSTEM, marketed by Invitrogen (Carlsbad CA), in which a recombinant baculovirus genome – containing the gene of interest – is generated within specialized cells of *Escherichia coli* by a method that involves site-specific transposition.

The strain of *E. coli* used in this system (DH10Bac™) contains a BACMID (q.v.) that includes the specific target site of transposon Tn*7* (see TN7). During this procedure, the gene of interest (carried in a small plasmid) is inserted into the Tn*7* target site of the bacmid by site-specific transposition (forming an *expression bacmid*).

Initially, the gene of interest is inserted into the polylinker of a small (4.8 kb) pFastBac™ plasmid (page 30). Within this vector, the inserted gene is positioned downstream of the (strong) promoter of the baculovirus polyhedrin gene. The expression cassette within the pFastBac vector is flanked by the terminal sequences, Tn*7*L and Tn*7*R, of transposon Tn*7*. The vector, containing the gene of interest, is then transfected into cells of the DH10Bac strain of *E. coli*, within which the expression cassette is transferred from the vector to the bacmid by site-specific transposition.

The recombinant bacmid, containing the gene of interest, is isolated from *E. coli* and is then transfected into insect cells for production (amplification) of recombinant baculovirus – which can be used to infect fresh insect cells for production of the required recombinant protein.

The Bac-to-Bac system has been used e.g. in studies on the capsid protein of the beak-and-feather-disease virus (BFDV) [J Virol (2006) 80(14):7219–7225]; for studying conformational states of a SARS virus protein [J Virol (2006) 80(14): 6794–6800]; for studying a processive microtubule polymerase [Cell (2008) 132(1):79–88]; for studies on the regulation of *Drosophila melanogaster* genes [BMC Genomics (2008) 9:424]; for studies on the nuclear egress of cytomegalovirus [PLoS Pathogens (2009) 5(1):e1000275]; for studies on Orc6 involvement in septin complex activity [Mol Biol Cell (2009) 20(1):270–281]; and a study on STAT1 signaling [Genes Dev (2009) 23(1):118–132].

(See also PFASTBAC DUAL VECTOR KIT.)

Bacillus anthracis A species of Gram-positive, spore-forming, typically square-ended rod-shaped bacteria, virulent strains of which form anthrax toxin (encoded by plasmid pXO1) and a poly-D-glutamic acid capsule (encoded by plasmid pXO2). *B. anthracis* grows on nutient agar; in biochemical tests it gives results similar to those from *Bacillus cereus*. *B. anthracis* is susceptible to the γ phage (gamma phage).

The Sterne strain of *B. anthracis* forms anthrax toxin but lacks the capsule – lack of the capsule permitting elimination of the pathogen by phagocytosis.

Ames strain: see entry BACTERIA (table).

Disease similar to anthrax has been reported to be caused e.g. by strains of *Bacillus cereus* [see PLoS Pathogens (2006) 2(11):e122].

(See also SORTASE.)

back mutation (reverse mutation) (1) Following a given (primary) mutation: a mutation that restores the original nucleotide sequence.

(2) An *intragenic* SUPPRESSOR MUTATION.

back translation (reverse translation) The process of *deducing* or *inferring* the probable nucleic acid sequence that encodes a given sequence of amino acids in a protein or polypeptide. One problem in doing this is that, given the degeneracy of the GENETIC CODE, most of the amino acids can be encoded by more than one codon (see the table in the entry AMINO ACID) – so that a number of different nucleic acid sequences can be postulated to encode the given sequence of amino acids. To estimate which of these nucleic acid sequences is likely to the the right one, it may be possible e.g. to obtain clues from the codon usage pattern of the particular organism from which the protein/polypeptide originated (see CODON BIAS, senses 2 and 3).

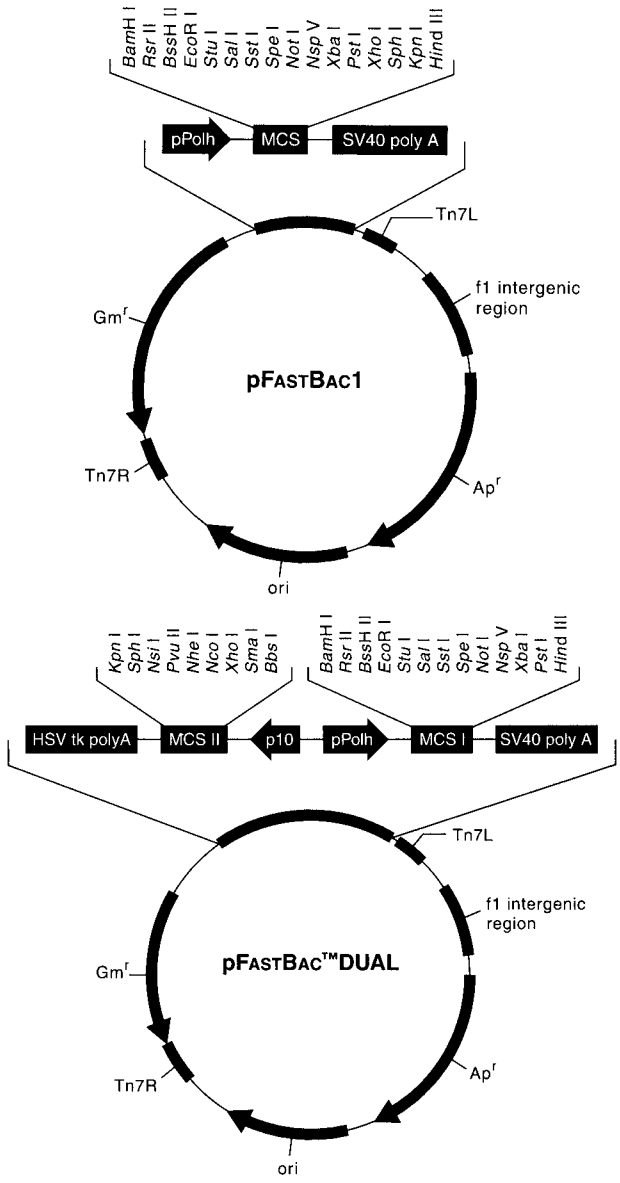
***Bac*Light™** See LIVE BACLIGHT BACTERIAL GRAM STAIN KIT and LIVE/DEAD BACLIGHT BACTERIAL VIABILITY KIT.

bacmid A replicon, first constructed in the 1990s, which can replicate as a plasmid in *Escherichia coli* and can be used to transfect (susceptible) insect cells.

A bacmid consists of a recombinant baculovirus genome (see BACULOVIRIDAE) which includes a plasmid origin of replication and also a target (insertion) site for the transposon Tn*7* (*att*Tn*7*).

If a bacmid is present in *E. coli*, a gene cassette – flanked by the two terminal parts of Tn*7* – can be transposed into the *att*Tn*7* site in the bacmid by site-specific transposition, given that the functions for transposition are provided e.g. from a helper plasmid in the bacterium.

[Use of bacmids (e.g.): Insect Biochem Mol Biol (2008) 38 (1):89–98.]



Bac-to-Bac® VECTORS *Above.* The basic pFastBac^{TM}1 vector. See entry for details of the method. *Below.* The pFastBac^{TM} Dual vector into which two DNA sequences can be inserted and expressed simultaneously.

Courtesy of Invitrogen, Carlsbad CA, USA.

BACTERIAL GENOMES: examples of size and sources of information

Species	Genome (∼Mbp)[a]	Reference
Gram-positive		
Bacillus anthracis (Ames strain)	5.2	Nature (2003) 423:81–86
Bacillus subtilis	4.2	Nature (1997) 390:249–256
Bacteroides thetaiotaomicron	6.3	Science (2003) 299:2074–2076
Clostridium perfringens	3.0	PNAS[b] (2002) 99(2):996–1001
Lactobacillus acidophilus NCFM	2.0	PNAS[b] (2005) 102(11):3906–3912
Lactobacillus plantarum	3.3	PNAS[b] (2003) 100:1900–1995
Listeria welshimeri	2.8	J Bacteriol (2006) 188:7405–7415
Mycobacterium avium subsp *paratuberculosis*	4.8	PNAS[b] (2005) 102(35):12344–12349
Mycobacterium leprae	3.3	Nature (2001) 409:1007–1011
Mycobacterium tuberculosis	4.4	Nature (1998) 393:537–544
Rhodococcus sp RHA1	9.7	PNAS[b] (2006) 103:15582–15587
Streptococcus pneumoniae	2.1	Science (2001) 293:498–506
Streptococcus pyogenes	1.9	PNAS[b] (2001) 98:4656–4663
Tropheryma whipplei	0.9	Lancet (2003) 361:637–644
Gram-negative		
Aeromonas hydrophila	4.7	J Bacteriol (2006) 188(3):8272–8282
Borrelia burgdorferi	0.9	Nature (1997) 390:580–586
Caulobacter crescentus	4.0	PNAS[b] (2001) 98(7):4136–4141
Chlamydia trachomatis	1.0	Science (1998) 282:754–759
Coxiella burnetii	2.0	PNAS[b] (2003) 100(9):5455–5460
Escherichia coli	5.0[c]	Science (1997) 277:1453–1474
Escherichia coli O157:H7	5.5	Nature (2001) 409:529–533; erratum (2001) 410:240
Helicobacter pylori	1.6	Nature (1997) 388:539–547
Helicobacter pylori (HPAG1)	1.6	PNAS[b] (2006) 103(26):9999–10004
Neisseria meningitidis	2.2	Nature (2000) 404:502–506
Pasteurella multocida	2.2	PNAS[b] (2001) 98:3460–3465
Pseudomonas syringae pv. tomato	6.5	PNAS[b] (2003) 100(18):10181–10186
Rhizobium leguminosarum	7.7	Genome Biol (2006) 7(4):R34
Rickettsia typhi	1.1	J Bacteriol (2004) 186(17):5842–5855
Shigella flexneri 5b	4.5	BMC Genomics (2006) 7:62
Thiomicrospira crunogena	2.4	PLoS Biol (2006) 4(12):e383
Vibrio cholerae	4.0[d]	Nature (2000) 406:477–483

[a]Approximate genome size in Mbp (bp × 10^6). Variation in genome size can occur within a given species.
[b]Proc Natl Acad Sci USA.
[c]Values of ∼4.7–5.5 have been reported for different strains.
[d]*Vibrio cholerae* chromosome I ∼3.0 Mbp, chromosome II ∼1.0 Mbp (genome size: ∼4.0 Mbp).

(See also BAC-TO-BAC.)

bacteria A term used to refer to some or all organisms of the domain BACTERIA. Note that this term has a lower-case 'b'; an upper-case 'B' is used for the *domain* Bacteria because this is the name of a *taxon* (a specific taxonomic category).

Bacteria One of the two domains of prokaryotic organisms – the other being ARCHAEA. Organisms in these two domains differ e.g. in their 16S rRNA, in the composition of their cell-wall macromolecules, in the composition of their membrane lipids and also in flagellar structure; moreover, in general, gene expression in the Archaea (in terms of transcription and translation) appears to resemble the eukaryotic mode of expression more closely than the bacterial mode.

Genome and chromosomes

The *typical* bacterial genome is a circular chromosome (i.e. ccc dsDNA). Cells may contain a single chromosome or may contain more than one copy per cell (the number depending on species and on growth conditions).

In some species of bacteria the genome is *linear* dsDNA: see e.g. BORRELIA BURGDORFERI. (A linear chromosome is also present e.g. in a species with one of the largest genomes sequenced thus far in bacteria (∼9.7 Mbp): *Rhodococcus* sp RHA1 [Proc Natl Acad Sci USA (2006) 103:15582–15587].)

In general, only one *type* of chromosome is found in the cells of a given species. However, in *Vibrio cholerae* each of the cells has *two* types of chromosome – one (chromosome I) being larger than the other (chromosome II); jointly, these two chromosomes form the genome of *V. cholerae*.

Bacterial chromosomes are extensively folded into compact bodies (see NUCLEOID).

The size of the genome varies among different species. In a number of cases the genome has been completely sequenced. The table lists the approximate size of the genome in various species and includes sources of information on sequences.

Plasmids

In addition to chromosome(s), many bacteria contain one or more types of PLASMID; however, plasmids are not normally present in certain bacteria (e.g. *Brucella* and *Rickettsia*), and, even when a plasmid is present, its presence is commonly not essential to the host cell. In *Rhodococcus* sp RHA1, there are three (linear) plasmids as well as the (linear) chromosome.

Susceptibility of bacteria to bacteriophages

Bacteria are commonly susceptible to one or more types of BACTERIOPHAGE, and a lysogenic strain of bacteria contains the phage's genome (the *prophage*) as well as the cell's own genome. Commonly, the prophage integrates in the bacterial chromosome; this is not the case, however, in e.g. phage P1.

Sometimes the prophage confers important characteristic(s) on the host cell – e.g. the CTXϕ prophage (which encodes cholera toxin) is the major virulence factor in cholera-causing strains of *Vibrio cholerae*.

Bacterial life style etc.

Some bacteria are *obligate* intracellular parasites/pathogens – i.e. they grow *only* within certain types of host cell; among these bacteria, CHLAMYDIA undergoes a developmental cycle (which is atypical in bacteria). Many of these obligately intracellular bacteria cause disease in their hosts. Most bacteria, however, are free-living organisms, and many can be grown on inanimate laboratory media. Even so, the laboratory media (including provision of various specific requirements) and conditions (such as temperature) must be appropriate in order to achieve growth; thus, while some species of bacteria, e.g. *Escherichia coli*, do grow well on simple media (such as NUTRIENT AGAR), others will not grow if the medium is not supplemented with ('enriched' with) substances such as serum or blood.

Some bacteria are obligate aerobes, meaning that they need oxygen (as an essential requirement) for growth, while some are obligate anaerobes – organisms which will not grow (and which may be killed) in the presence of oxygen. However, many species (including *E. coli*) are facultatively anaerobic – i.e. they can grow in the presence or absence of oxygen; such organisms may adopt a respiratory or fermentative mode of metabolism according to the presence or absence of oxygen, respectively.

Bacterial taxonomy and typing

Before the 1980s, the taxonomy of bacteria was based largely on various phenotypic characteristics (including e.g. reaction to certain dyes, particularly the Gram stain). Modern bacterial taxonomy is based primarily on comparison of nucleic acids, and most TYPING methods are also genotypic – an important exception being PHAGE TYPING.

Bacteria in DNA technology

Bacteria are used frequently in recombinant technology for a variety of reasons which include: (i) typically, rapid growth and division, (ii) ease of manipulation and mutagenesis, (iii) ease of storage, and (iv) (in comparison with eukaryotic cells) cellular processes which are relatively easy to manipulate. In particular, *Escherichia coli (E. coli)* is extensively used as an experimental organism. (Other bacteria which have been widely used as experimental organisms include e.g. *Bacillus subtilis* (Gram-positive), *Pseudomonas aeruginosa* (Gram-negative), and *Salmonella typhimurium* (Gram-negative).) Moreover, in addition to bacterial *cells*, certain bacterial *components* (e.g. plasmids and RESTRICTION ENDONUCLEASES) are invaluable in this area of study.

bacterial artificial chromosome (BAC) An *in vitro* construct, based on the low-copy-number F PLASMID, which is used e.g. as a vector for CLONING large fragments of DNA (∼300 kb).

The use of a low-copy-number plasmid vector may appear a strange choice when the object of cloning is to replicate the target sequence with high efficiency. However, one problem with small, high-copy-number plasmids is that use of these vectors can often result in instability of the cloned DNA; the instability, manifested e.g. in deletions/rearrangements, may occur because of the ease with which recombination can take place between the fragments carried by different copies of the plasmid. Fragments of eukaryotic DNA, which often include repetitive sequences, are particularly prone to instability in the small high-copy-number vectors.

BACs, which exist stably in one or two copies per cell, are widely used for making genomic libraries. They are inserted into bacteria by methods such as ELECTROPORATION.

[BACs in immunology: Immunology (2007) 121(3):308–313.]

BACs containing a cloned herpesvirus genome were stably maintained in bacteria; virions were produced when CCB (permissive) cells were transfected with these BACs [J Virol (2008) 82(10):4955–4964].

(Compare with the HUMAN ARTIFICIAL CHROMOSOME and YEAST ARTIFICIAL CHROMOSOME.)

bacterial DNase See entry DNASE for staphylococcal thermonuclease; see also STREPTODORNASE.

bacterial two-hybrid system An approach used for studying protein–protein interactions within bacterial cells. (cf. YEAST TWO-HYBRID SYSTEM.)

There are at least two distinct forms of methodology.

The (commercial) BACTERIOMATCH TWO-HYBRID SYSTEM resembles the yeast two-hybrid system in that interaction of the two proteins of interest is revealed by the transcription of a reporter gene – i.e. the genes of the two proteins of interest are fused, separately, to the two parts of a transcription activator.

In a different approach (that does not involve transcription activation), the gene of each of the two proteins of interest is fused, separately, to a fragment of the gene (*cya*) encoding the enzyme ADENYLATE CYCLASE (see also GENE FUSION) – each gene being fused to a different fragment of the *cya* gene. The two fusion genes are then expressed in bacterial cells. In these cells, the two proteins of interest may interact in such a way that the two fragments of adenylate cyclase are brought together to form a functional enzyme. Whether or not this occurs can be investigated by carrying out the test in cells that are deficient in adenylate cyclase (e.g. the *cya*-99 strain of *Escherichia coli* BTH101). To test for interaction between the two proteins of interest (as detected by the formation of active adenylate cyclase), cells are inoculated into a minimal medium with lactose as sole carbon source and the indicator X-GAL. If the two proteins interact, and adenylate cyclase is formed, cyclic AMP (cAMP, produced by adenylate cyclase) permits activation of the LAC OPERON by the lactose (see the entry CATABOLITE REPRESSION); the *lac* operon produces β-galactosidase – which cleaves the X-gal with the formation of a blue coloration. (If the two proteins do *not* interact, cAMP is not produced and the use of lactose by the cells is blocked.) This method has been used for studying protein interactions in *Mycobacterium tuberculosis* [BMC Mol Biol (2009) 10:3].

bacterial vectors (in gene therapy) See the section *Vectors in gene therapy* in the entry GENE THERAPY.

bacteriocin Any of various typically protein or peptide factors, produced by certain bacteria (including both Gram-negative and Gram-positive species), that kill or inhibit other bacteria – frequently closely related species; bacteriocins range from high-molecular-weight proteins and quite short peptides to a simple modified amino acid (microcin A15). Some bacteriocins require post-translational modification for their activity.

Analogous agents are produced by members of the Archaea (e.g. *halocins*) but their amino acid sequences are apparently not homologous to those of their bacterial counterparts.

Bacteriocins are of various distinct types: see e.g. COLICIN, LANTIBIOTIC and MICROCIN.

Most bacteriocins are encoded by genes in conjugative or non-conjugative plasmids. Chromosomally encoded bacteriocins include some of those formed by lactic acid bacteria and e.g. 'bacteriocin 28b' (a colicin from *Serratia marcescens*).

Bacteriocins act in characteristic ways. Some (e.g. microcin B17) inhibit DNA gyrase; some (e.g. cloacin DF13) cleave 16S rRNA; and some form pores in the bacterial cytoplasmic membrane.

bacteriocyte See e.g. BUCHNERA and WIGGLESWORTHIA.

BacterioMatch® two-hybrid system A system (Stratagene, La Jolla CA) for detecting interaction between two proteins in a bacterial cell. The proteins (*bait* and *target*) correspond to the 'bait' and 'prey' proteins in the YEAST TWO-HYBRID SYSTEM (q.v. for basic information on two-hybrid systems). Each of these proteins is encoded in a separate expression vector – in which its sequence is part of a fusion gene; the two vectors, pBT and pTRG (see diagram on page 34), are introduced into a bacterial cell and are expressed together in the cell.

The *bait vector* contains the gene of the bait protein fused to the gene of the cI protein of PHAGE LAMBDA. Expression of this vector within the cell therefore produces a (bait + cI) fusion protein; this fusion protein has a region that can bind to a DNA sequence (the λ operator sequence) located a short distance upstream of the promoter of a reporter gene. (Thus, the bait + cI fusion protein is the 'DNA-binding' entity in this two-hybrid system.)

In the *target vector* the gene of the target protein is fused to a sequence encoding the N-terminal region of the (bacterial) RNA polymerase α-subunit; the fusion protein is expressed in the cell.

Within the (engineered) reporter cell, interaction between bait and target proteins results in the recruitment of a functional RNA polymerase to the (nearby) downstream promoter, thus initiating transcription of the reporter gene.

The reporter gene encodes resistance to ampicillin – which allows the positive selection of cells in which protein–protein interaction has occurred (because these cells are able to grow on ampicillin-containing media). Another reporter gene, *lacZ* (encoding β-galactosidase), is transcribed in series with the Amp^r gene; it may be used to validate interaction between the two proteins.

The above scheme describes only one approach to detecting protein–protein interaction in bacteria: see BACTERIAL TWO-HYBRID SYSTEM for an alternative approach.

[Use of the BacterioMatch® system (examples): J Bacteriol (2008) 190:4971–4978; BMC Genomics (2009) 10:118; and Neural Dev (2009) 4:1.]

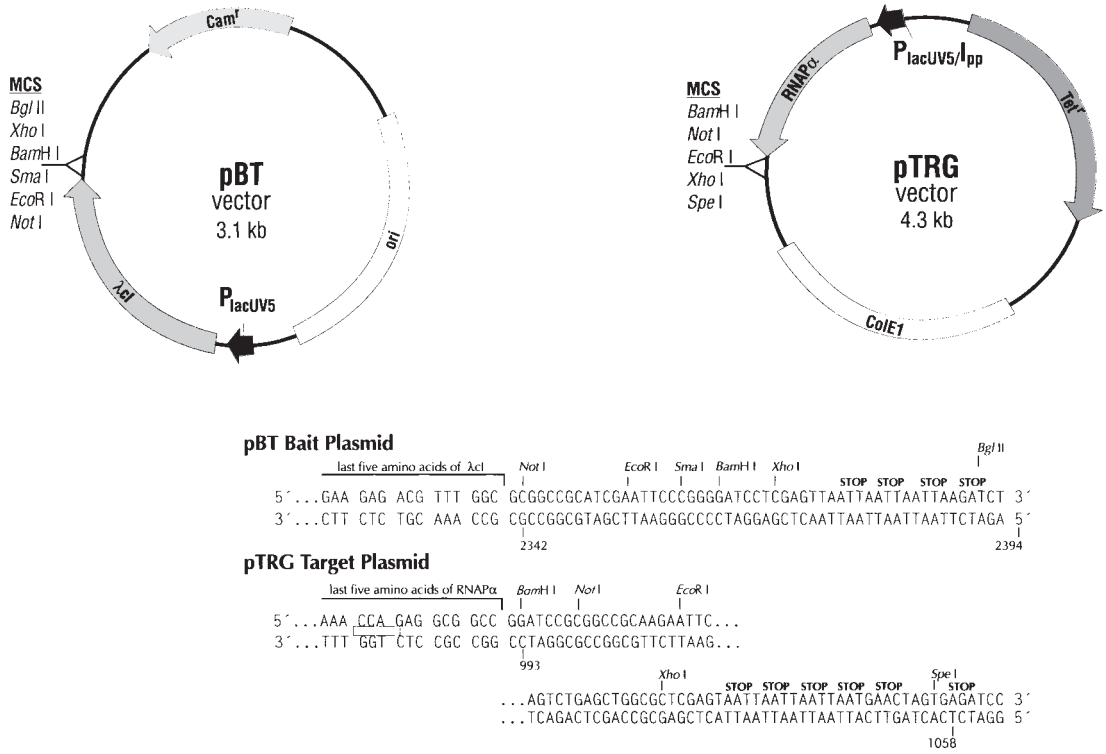
bacteriophage (phage) Any virus that infects *bacteria;* many (perhaps most, or all) bacteria are susceptible to one or more types of phage. (See entries under PHAGE for some specific examples.)

A given phage may specifically infect bacteria of only one species or strain – or it may have a wider host range. Certain phages are specific for conjugative *donor* cells (i.e. 'male' cells), in which the PILUS forms the site of initial attachment. (See also FEMALE-SPECIFIC PHAGES.)

Virulent and temperate phages

A phage is often described as *virulent* or *temperate*. Virulent phages lyse (kill) the bacterial host cell. A temperate phage can form a more or less stable relationship with a bacterium; in most cases the phage genome (the *prophage*) integrates in the bacterial chromosome, but in some cases (e.g. phage P1) it exists as a circular, extrachromosomal 'plasmid' (see also PHAGE N15).

Under certain conditions a temperate phage may be induced to enter the lytic cycle – in which case virions are formed and



BacterioMatch® VECTORS The 'bait' and 'target' vectors (pBT and pTRG, respectively) of the BacterioMatch® two-hybrid system. See entry for details of the method. λcI refers to the gene of the cI protein of bacteriophage lambda (λ). MCS = multiple cloning site (polylinker); P = promoter; ColE1 = origin of replication of the ColE1 plasmid; RNAPα = a region encoding the N-terminal sequence of the α subunit of (bacterial) RNA polymerase; Tet^r = resistance to the antibiotic tetracycline; Cam^r = resistance to the antibiotic chloramphenicol.

Courtesy of Stratagene, La Jolla CA, USA.

cell lysis occurs. Thus, the prophage may retain the potential for virulence, and in a population of *lysogenic* bacteria (i.e. bacteria which are hosts to a temperate phage), spontaneous induction may occur in a small number of cells.

Genome (within the phage particle)

Depending on phage, the genome may be linear dsDNA (e.g. phages λ, P1, P2, P22, φ29, T4, T7); ccc dsDNA (e.g. PM2); ccc ssDNA (e.g. f1, φX174); ssRNA (e.g. MS2, Qβ) – or dsRNA (e.g. φ6).

Phage φ6 is unusual in that it has a SEGMENTED GENOME.

Composition, morphology and size of phages

Phages commonly consist of nucleic acid enclosed within a protein coat called a *capsid*. However, certain phages contain lipid, either within the virion itself (e.g. phage PM2 in *Alteromonas espejiana*) or as an outer *envelope* (for example, phage φ6 in *Pseudomonas syringae* pv. *phaseolicola*), and certain phages have other components such as fucose (e.g. phage MV-L3) or spermidine (e.g. phage MS2).

Phages of the family Inoviridae (e.g. phages f1, fd, M13 in enterobacteria; Pf1 in *Pseudomonas;* Xf in *Xanthomonas;* v6 in *Vibrio*) are long FILAMENTOUS PHAGES which, in most or all cases, infect only so-called 'male' (i.e. conjugative donor) cells (see e.g. PHAGE M13). While these phages replicate (and form virions) in their host cells, they do not lyse the cells: virions are extruded through the cell envelope of the (living) host cells.

Other phages vary greatly in shape and size. Many phages have a 'spherical' or icosahedral protein *head*, containing the genome, which carries a thin, elongated, contractile or non-contractile *tail* involved in attachment to the host cell. Some

BACTERIOPHAGE: some uses of phages in DNA technology

Bacteriophage	Genome	Uses (e.g.)
f1	ccc ssDNA	• Replication origin used in pBluescript® phagemids • PHAGE DISPLAY technology
λ (lambda)	linear dsDNA	• cI protein gene used in bacterial two-hybrid system (see BACTERIOMATCH TWO-HYBRID SYSTEM) • *cos* site used in cosmids and phasmids (see COSMID, PHASMID) • Derivatives used as cloning vectors (see e.g. ZAP EXPRESS VECTOR) • λ Red recombination system used for *in vivo* recombination (see the entry LAMBDA (λ) RED RECOMBINATION) • Constituents used in a range of GATEWAY SITE-SPECIFIC RECOMBINATION systems
M13	ccc ssDNA	• Used for making single-stranded copies of DNA target sequences for site-specific mutagenesis or sequencing etc. • Helper phage for phagemids, including e.g. gene-delivery phagemid particle vectors for eukaryotic cells (see entry GENE THERAPY)
N15	linear dsDNA	• Derivative used as a linear cloning vector in the BigEasy™ linear cloning system (q.v.)
P1	linear dsDNA	• Cre–*loxP* site-specific recombination system used e.g. in cassette-exchange mechanisms (e.g. RECOMBINASE-MEDIATED CASSETTE EXCHANGE)
φ29 (phi 29)	linear dsDNA	• MULTIPLE DISPLACEMENT AMPLIFICATION
T3	linear dsDNA	• RNA polymerase used for *in vitro* transcription • RNA polymerase used for incorporation of labeled nucleotides
T4	linear dsDNA	• DNA polymerase used for incorporation of digoxigenin-labeled nucleotides • DNA ligase used for closing nicks • UvsX recombinase used for the (isothermal) recombinase polymerase amplification of DNA (see entry RPA, sense 1)
T7	linear dsDNA	• As for T3 • Eberwine technique • DNA polymerase (Sequenase®), helicase and single-strand binding protein (gp2.5) used in circular helicase-dependent amplification of plasmids (see entry CHDA)

are small, icosahedral virions which lack a tail. Some have a complex, irregular shape.

In size, phages range from ~25 nm in diameter (e.g. φX174, Qβ – both tail-less, icosahedral phages) to ~200 nm (e.g. λ, T4 – both tailed phages), while filamentous phages (e.g. f1, M13) can be >750 nm in length, although they are only ~6 nm in width.

Phages in DNA technology

Phages (and their components/enzymes) are widely used: see e.g. the table (and PHAGE MU). [Phage gene regulation (database): Nucleic Acids Res (2009) doi: 10.1093/nar/gkp911.]

Phages as potential therapeutic agents

Many studies have been carried out to investigate the use of phages against the bacteria that cause certain human diseases [see e.g. Antimicrob Agents Chemother (2001) 45:649–659]. Phage-encoded endolysins (enzymes that cleave the bacterial cell-wall polymer peptidoglycan) have also been considered as therapeutic agents [Exp Biol Med (2006) 231:366–377].

In a different approach, a number of studies have assessed phages, or phagemid particles, for use in gene delivery in the genetic modification of mammalian cells (and, hence, their possible use in GENE THERAPY); for example, an engineered helper phage was developed to enable phagemid particles to incorporate a specific, helper-encoded peptide ligand in order to facilitate binding to mammalian target cells.

Phage delivery vehicles have also been used in the context of DNA VACCINES [see e.g. Infect Immun (2006) 74(1):167–174].

bacteriophage conversion See LYSOGENY.

bacteroid A term used to refer to certain forms of bacteria that are distinct (e.g. morphologically) from related, normal wild-type cells – e.g. the bacterial nitrogen-fixing cells found in root nodules of leguminous plants.

BaculoDirect™ A BACULOVIRUS EXPRESSION SYSTEM (Invitrogen, Carlsbad CA) in which a gene/fragment of interest is inserted into a modified, linearized baculovirus genome in a 1-hour *in vitro* step – following which the recombinant genome can be transfected directly into insect cells (such as

Sf21 cells). After 72 hours, the supernatant (containing extracellular recombinant virions) can be either stored or used to infect a new batch of insect cells for expression of the gene of interest.

Insertion of a gene of interest into the baculovirus genome is carried out by a technique involving the GATEWAY SITE-SPECIFIC RECOMBINATION SYSTEM: the baculovirus genome includes a pair of *att*R sites into which the *att*L-flanked target DNA is inserted by the activity of LR Clonase.

During the LR Clonase reaction – in which the target DNA is inserted – part of the linearized baculovirus genome, which contains a *lacZ* gene, is excised and forms a by-product. Subsequent staining of the baculovirus-infected cells to detect the *absence* of *lacZ* enables the researcher to check that the cells do not contain non-recombinant baculovirus vectors.

In the recombinant baculovirus vector, the target DNA is driven by a strong promoter (the promoter of the polyhedrin gene), and the vector encodes a V5-His tag (see also NICKEL-CHARGED AFFINITY RESIN) to facilitate product purification.

One version of this vector contains a sequence encoding the MELITTIN peptide which is intended to promote secretion of the expressed protein.

BaculoDirect™ was used e.g. for studying monooxygenase transcripts [Biochem Pharmacol (2008) 75(2):570–579]; for studying the regulation of Golgi structure/function [Mol Biol Cell (2009) 20(5):1388–1399]; and for studying the activity of the Werner DNA helicase [PLoS ONE (2009) 4(3):e4673].

Baculoviridae A family of enveloped viruses which (in nature) infect a wide range of arthropods – including members of the Coleoptera (beetles), Diptera (flies), Hymenoptera (ants, bees etc.) and Lepidoptera (butterflies, moths). (Baculoviruses can also infect mammalian cells under experimental conditions – and have been considered as gene-delivery vehicles for gene therapy.)

The baculovirus genome consists of ccc dsDNA and ranges from ∼80 kbp to ∼200 kbp, according to virus.

In DNA technology, these viruses are exploited e.g. in the BACULOVIRUS EXPRESSION SYSTEMS used for the expression of recombinant proteins.

The baculoviruses include two major groups: the NUCLEAR POLYHEDROSIS VIRUSES (NPVs) and the GRANULOSIS VIRUSES (GVs).

OBs*, *ODVs

Late in the phase of infection, within the cell's nucleus, many of the replicated viral nucleocapsids become enveloped by a lipid-containing membrane (that apparently derives from the nuclear membrane). Subsequently, these (enveloped) nucleocapsids are embedded in a crystalline matrix composed of a virus-encoded protein, forming the so-called *occlusion bodies* (OBs); the viruses that are occluded within these bodies are called *occlusion-derived virions* (ODVs).

The OBs of the nuclear polyhedrosis viruses (also referred to as *polyhedra*, owing to their shape) contain multiple enveloped particles, each of which contains either a single nucleocapsid or multiple nucleocapsids, according to virus. On the other hand, in granulosis viruses, each of the OBs contains a single enveloped particle containing a single nucleocapsid or, rarely, two nucleocapsids.

In the NPVs the crystalline matrix of the occlusion body is composed of the protein *polyhedrin*. In the granulosis viruses the matrix protein is *granulin*.

The OBs serve as disseminative units: when the host dies, and OBs are ingested by a new, susceptible host, infectious virions are released in the midgut of the new host and initiate infection.

BVs

Baculoviruses also form so-called *budded virions* (BVs). The BVs are formed when a proportion of the nucleocapsids are transported from the nucleus to the cell's plasma membrane; here, they bud from the cell's surface into the extracellular environment to produce enveloped virions (BVs). BVs spread the infection to other cells within the same (infected) host.

In *Autographa californica* NPV, the *me53* gene is reported to be required for efficient production of BVs, deletion of this gene resulting in a dramatic reduction in yield [J Virol (2009) 83(15):7440–7448].

Baculovirus genes/gene products

Baculovirus genes/gene products referred to in the literature include e.g. *ie-1* (encoding a transregulatory protein activator of transcription of both early and late genes); the *lef* genes (late expression factor genes, encoding e.g. a primase and its accessory factor); *gp64* (encoding a protein, GP64, involved in virion attachment to the host cell membrane); and *p143* (a helicase).

The gene encoding very late expression factor 1 (VLF-1) is apparently present in the genomes of all baculoviruses. The product, VLF-1, is reported to have homology with members of the tyrosine recombinase family; however, a function unrelated to recombinase action (in which VLF-1 is involved in the production of normal capsid particles) has been reported for this product [J Virol (2006) 80(4):1724–1733].

Biological control

In addition to their uses in DNA technology, baculoviruses have been widely used for the biological control of various insect pests; they are regarded as particularly suitable for this purpose because of their apparent lack of relationship to the animal and plant viruses. The baculoviruses have been used e.g. to control the European pine sawfly (*Neodiprion sertifer*) and the European spruce sawfly (*Gilpinia hercyniae*) as well as the Small White Butterfly (*Artogeia rapae*).

baculovirus Any virus of the family BACULOVIRIDAE.

baculovirus expression systems Systems in which vectors that are based on the baculovirus genome (see BACULOVIRIDAE) are used for the expression of recombinant proteins in insect cells (or, more recently, in the cells of vertebrates). The use of baculovirus vectors for this purpose was suggested by the large quantities of matrix protein that are produced in insect cells infected with these viruses; as early as the 1980s, human interferon was produced in this kind of system.

(cf. OVEREXPRESSION.)

Baculovirus expression systems include e.g. BAC-TO-BAC and BACULODIRECT.

Advantages of baculovirus expression systems

The advantages claimed for these systems include:

- High-level expression of recombinant proteins. Given the use of strong promoters to drive recombinant genes, yields of recombinant protein have been reported to reach up to ~500 mg/L (up to ~50% of total cell protein).
- Unlike prokaryotic cells, insect cells can carry out various post-translational modifications of expressed proteins similar to those carried out in mammalian cells.
- Procedures for growing (culturing) insect cells are simpler than those involved in the growth of mammalian cells – e.g. relatively simple, serum-free media can be used, and there is no requirement for CO_2.
- High-density suspension culture can be employed.
- Protein expression can be easily scaled-up.
- Insect cells are characteristically refractory to infection by human viruses – so that recombinant proteins are unlikely to be contaminated by viral pathogens.

Baculovirus expression systems have been used e.g. for the production of veterinary products (which include a vaccine against swine fever).

Disadvantages of baculovirus expression systems

These expression systems have some disadvantages:

- Some of the required post-translational modifications are poorly represented, or lacking, in insect cells. For example, patterns of glycosylation may be significantly different from those that characterize human glycoproteins. To combat this, genetically modified cells (see e.g. MIMIC SF9 INSECT CELLS) can be used; these cells are designed to express stably various mammalian glycosyltransferases.
- Recombinant proteins expressed in insect cells may remain largely intracellular. This is contrary to the general requirements of a production process – in which the formation of extracellular (i.e. secreted) protein facilitates the downstream processing of the product. In some systems, the gene of the recombinant protein is fused with a tag (e.g. a tag encoding multiple histidine residues – see NICKEL-CHARGED AFFINITY RESIN) which can facilitate detection/isolation/purification of the product following cell lysis.

Development of baculovirus expression systems

Early procedures used for inserting a foreign gene into the baculovirus genome depended on homologous recombination – insect cells were co-transfected with (i) baculovirus DNA and (ii) the gene of interest, present in a separate vector and flanked by (non-essential) baculovirus DNA (e.g. sequences from the polyhedrin gene). If the gene of interest inserts into the polyhedrin gene (homologous recombination), the plaque morphology of the resulting recombinant baculovirus will be altered, allowing identification of recombinant viruses (and, hence, separation from the background of non-recombinant viruses). This approach, however, was found to yield only a low proportion (<1%) of recombinant viruses.

Improved recovery of recombinant viruses was achieved by using linearized viral genomes, and a further improvement resulted from the prior removal from the viral genome of an essential sequence, ORF 1629; in this case, a functional viral genome is produced only if recombination occurs with the experimental DNA – which includes ORF 1629 as well as the target sequence.

A rapid, commercial system produces recombinant baculovirus genomes by a 1-hour procedure involving *in vitro* site-specific (*att*-mediated) recombination between a linearized, modified baculovirus genome and the cloned gene of interest: see BACULODIRECT.

(See also BAC-TO-BAC and PFASTBAC DUAL VECTOR KIT.)

In some cases, expression of recombinant proteins has been achieved in whole (live) insects. For example, production of the veterinary product Vibragen Omega (used therapeutically for canine parvovirus infection) is carried out in silkworms; in this process, insects are infected with a recombinant polyhedrosis virus containing a sequence encoding recombinant feline interferon omega.

Baculovirus-independent gene expression in insect cells

Expression of heterologous proteins in insect cells can also be carried out in a baculovirus-independent manner by using a simpler vector system (including Gateway® and topo-type systems) in *Drosophila* Schneider S2 cells: see DES.

BaeI A type IIB RESTRICTION ENDONUCLEASE from *Bacillus sphaericus*.

bait protein See YEAST TWO-HYBRID SYSTEM and BACTERIO-MATCH TWO-HYBRID SYSTEM.

BALB/c mice An inbred, laboratory strain of mice prone to the development of myelomas following certain stimuli.

barcoding (1) See entry MOLECULAR INVERSION PROBE GENOTYPING.

(2) See entry DNA BARCODING.

(See also ZIPCODE ARRAY.)

Bardet–Biedl syndrome A genetically heterogeneous disorder reported to involve a range of chromosomes (2, 3, 4, 7, 9, 11, 12, 14, 15, 16, 20); the syndrome is characterized by obesity and various other manifestations, including renal and retinal abnormalities.

Protein(s) affected in the Bardet–Biedl syndrome appear to be required for *leptin* receptor signaling in the hypothalamus; mice with knock-out mutations in some of the relevant genes did not respond to leptin e.g. by failing to reduce intake. This suggested that reduced leptin signaling is causally connected with obesity in this syndrome [Hum Mol Genet (2009) 18(7): 1323–1331].

(See also OB GENE.)

barnase A bacterial endoribonuclease, produced by *Bacillus amyloliquefaciens*, that forms cyclic products. (A protein of the barnase family was reported to be encoded by *Geobacter metallireducens* [BMC Microbiol (2009) 9:109].)

Barnase has been reported to trigger APOPTOSIS in human cancer cells [PLoS ONE (2008) 3(6):e2434].

(See also BARSTAR and BINASE.)

Barr body (chromatin body; sex chromatin; X chromatin) A small, discrete body of CHROMATIN which can be seen in the nucleus in stained, interphase cells from female mammals. A Barr body is derived from an inactivated X CHROMOSOME (see X-INACTIVATION); hence, cells from a (normal) female contain one Barr body per nucleus, while cells from a normal male contain none.

In those individuals with atypical genotypes – such as XXX (superfemale) or XO (Turner's syndrome) – the number of Barr bodies per nucleus is also atypical (two and zero in the superfemale and in those with Turner's syndrome, respectively).

(See also DNA SEX DETERMINATION ASSAY.)

barstar An (intracellular) protein that binds with high affinity (and inhibits) the ribonuclease BARNASE.

BART BamHI rightward transcript: any of a family of virus-encoded RNAs found in cells latently infected with EPSTEIN–BARR VIRUS.

(See also miR-BART2 in the entry MICRORNAS (in section: *miRNAs encoded by viruses*.)

base composition (1) *Syn*. BASE RATIO.

(2) The relative amounts of A, C, G and T in a given sample of nucleic acid.

base excision repair A DNA repair system, found in organisms ranging from bacteria to humans, that deals with chemically induced damage, including damage caused by ROS (reactive oxygen species).

Excision of an aberrant, chemically modified base is catalyzed by a DNA glycosylase. (See also entry 8-OXOG.) The phosphodiester bond on one side of the AP site is then cut by an AP endonuclease – e.g. exonuclease III in *Escherichia coli* (an enzyme which has exonuclease and apurinic/apyrimidinic endonuclease activity). Replacement of the damaged nucleotide (and apparently of several adjacent nucleotides) is then effected by a DNA polymerase.

(See also A FAMILY and MISMATCH REPAIR.)

base J β-D-Glucosylhydroxymethyluracil: an unusual base in the nuclear DNA of kinetoplastid protozoans and in *Euglena*. In trypanosomatids, base J appears to replace a proportion of thymine residues and is found e.g. in the telomere region.

Thymidine hydrolases JBP1 and JBP2 were reported to be involved in the synthesis of base J [Nucleic Acids Res (2009) 37(5):1452–1462].

base pairing (in double-stranded DNA and RNA) The pairing (i.e. binding, annealing, hybridization) of each base in one of the strands with a particular, *complementary* base in the other strand; such *complementarity* depends on the specificity of binding between cytosine and guanine, and between thymine (or uracil in RNA) and adenine.

Base-pairing involves hydrogen-bonding between each pair of complementary bases.

base ratio (base composition; dissymmetry ratio) In a sample of DNA, the ratio:

$$[A+T]/[G+C]$$

in which A, T, G and C are relative molar amounts of the four types of nucleotide. Microbes with a base ratio of >1 are sometimes called *AT types;* those with a base ratio <1 are *GC types*.

(See also the entry GC%.)

base-unpaired region See MATRIX-ATTACHMENT REGION.

BAT-26 A region of MICROSATELLITE DNA which contains 26 consecutive adenines. Instability of BAT-26 is associated e.g. with certain cancers (see CANCER DETECTION). This is one of the sequences usually monitored in microsatellite instability (MSI) tests (see MSI ANALYSIS).

Bc12 protein See e.g. BIM.

BCIP 5-Bromo-4-chloro-3-indolyl phosphate: a substrate for the enzyme alkaline phosphatase which is used e.g. in PROBE LABELING (q.v.).

BCIP has been used e.g. in optimization of staining protocols for specific types of cell [PLoS ONE (2009) 4(5): e5536], and for identifying genes by EST analysis [BMC Genomics (2009) 10:109].

bcr-abl (*BCR-ABL*) See ABL.

***t*-BDMS** See BUTYLDIMETHYLSILYL.

bDNA assay Branched DNA assay: a method for quantitation of a specific DNA or RNA target sequence. Based on SIGNAL AMPLIFICATION, it avoids a major problem associated with TARGET AMPLIFICATION methods such as PCR, i.e. potential contamination of apparatus and environment with amplicons from previous assays.

Essentially, the bDNA assay involves: (i) a target-specific PROBE that immobilizes target sequences on a solid support, and (ii) a second type of probe which links each immobilized target to a synthetic, *multi-branched* molecule of DNA (so-called bDNA), which, in turn, can bind many enzyme-linked probes; using a dioxetane substrate, the enzyme generates an amplified chemiluminescent signal whose intensity is related to the number of target sequences in the sample.

bDNA kits (Quantiplex™) have been used e.g. for assaying HIV (the human immunodeficiency virus), hepatitis C virus (HCV), and simian immunodeficiency virus.

Uses (e.g.): HIV [BMC Public Health (2008) 8:169], [PLoS ONE (2008) 3(12):e3918]; HCV [Hum Immunol (2008) 69 (3):158–164].

Beijing/W See W-BEIJING STRAIN.

BENA435 A cell-permeant DNA-staining fluorophore that has absorption/emission maxima of 435/484 nm when complexed with DNA; this agent was described as non-toxic and capable of photoactivation *in vivo*. It appears to be an intercalating agent. Fluorescence was reported to be stronger with dA/dT homopolymers (in comparison with dG/dC homopolymers).

It appears that BENA435 is not related to any other DNA-staining agent.

[BENA435: Nucleic Acids Res (2006) 34(5):e43.]

(See also DNA STAINING.)

Benefix® A recombinant form of blood-clotting factor IX used for the management of hemophilia B.

benzene-1,3,5-triacetic acid See the entry BTA.

O^6-benzylguanine labels See GENE FUSION (uses).

benzylpenicillin *Syn*. Penicillin G (see also table in the entry ANTIBIOTIC).

Bernard–Soulier syndrome (hemorrhagiparous thrombocytic dystrophy) A rare autosomal recessive condition characterized by a tendency for abnormal/excessive bleeding. Genetic lesions associated with this condition, on chromosomes 3, 17 and 22, affect genes that encode subunits of a multi-subunit receptor for the von Willebrand factor.

[Bernard–Soulier syndrome: Orphanet J Rare Dis (2006) 1: 46.]

Betanodavirus See NODAVIRUSES.

bevirimat (3-*O*-(3′,3′-dimethylsuccinyl)betulinic acid; PA-457) A potent inhibitor of HIV-1 (human immunodeficiency virus) which blocks the activity of the viral protease during normal maturation of the virus (and is commonly called a 'maturation inhibitor'). [Maturation as a therapeutic target: Expert Opin Ther Targets (2009) 13(8):895–908.] (See also VLP.) Antiretrovirals which inhibit the viral protease by *binding* to it are considered under PROTEASE INHIBITOR.

BEVS Baculovirus expression vector system(s) (see BACULOVIRUS EXPRESSION SYSTEMS).

***bgl* operon** See CATABOLITE REPRESSION.

BglF See CATABOLITE REPRESSION.

BglG See CATABOLITE REPRESSION.

BglI A type IIP RESTRICTION ENDONUCLEASE which has been reported to yield an increased number of ribotypes of *Vibrio cholerae* (see the entry RIBOTYPING).

BHK cells Cells derived from baby hamster kidney.

bi-allelic mutagenesis See ZINC-FINGER NUCLEASE.

BiFC BIMOLECULAR FLUORESCENCE COMPLEMENTATION.

bifunctional intercalating agent See INTERCALATING AGENT.

bifunctional vector *Syn*. SHUTTLE VECTOR.

BigEasy™ linear cloning system A cloning system (Lucigen Corporation, Middleton WI) which involves a *linear* plasmid vector (pJAZZ-KA) derived from the linear dsDNA genome of coliphage N15. (See PHAGE N15.) The prophage of phage N15 does not integrate into the bacterial chromosome – it replicates as an extrachromosomal 'plasmid'. The mechanism of replication in N15 has features in common with replication in some other (linear genome) phages (e.g. φKO2) and with replication in the bacterium *Borrelia burgdorferi* (which also has a linear genome); see BORRELIA BURGDORFERI for more details.

Advantages of a linear vector system include the ability to clone sequences (e.g. inverted repeats, AT-rich regions) that may exhibit instability if cloned within circular, supercoiled plasmid vectors.

[Use of BigEasy™ (e.g.): BMC Plant Biol (2008) 8:124.]

Bim In e.g. mouse embryonic stem cells: a protein which acts as a mediator in signaling pathways. Bim's role is affected e.g. by phosphorylation; thus, phosphorylation at the Thr-112 residue reduces the binding of Bim to anti-apoptotic protein Bc12 (thus increasing cell survival), while phosphorylation at some other sites promotes its apoptotic role by reducing its susceptibility to proteasomal degradation [Mol Cell (2008) 30(4):415–425]. Bim's mRNA is normally a target for microRNAs (miRNAs), and the upregulation of Bim in cells which are deficient in ARGONAUTE proteins (Ago 1–4) (cells which are, as a consequence, deficient in miRNA activity) promotes APOPTOSIS [Genes Dev (2009) 23:304–317].

bimolecular fluorescence complementation (BiFC) A method used for detecting protein–protein interaction *in vivo*. In this method, the two proteins of interest are each tagged e.g. with a *fragment* of YELLOW FLUORESCENT PROTEIN, each of the fragments (alone) being non-fluorescent; interaction between the two proteins – *in vivo* – may result in appropriate juxtaposition of the two parts of the fluorophore, with consequent production of a fluorescent signal on suitable excitation.

In the procedure called extended BiFC (exBiFC), the pair of proteins whose interaction is being studied are each tagged with a *constitutively* fluorescent domain – a different fluorophore for each protein – in addition to the fragment of fluorophore used for the basic BiFC assay. This approach enables the researcher to follow (i) the *expression* of each protein and (ii) the *localization* of each protein – and it permits normalization of the BiFC signal in relation to the levels of expression of the two proteins.

[exBiFC: BioTechniques (2006) 41(6):688–692.]

[BiFC fragments: BioTechniques (2006) 40(1):61–66.]

BiFC for demonstrating protein interaction in *Arabidopsis* [Plant Physiol (2007) 143(2):650–660] and in a fungus, *Saccharomyces cerevisiae* [Eukaryotic Cell (2007) 6(3): 378–387]; for studying the *Arabidopsis* calcium sensor [Plant Physiol (2008) 148(4):1883–1896]; for studying rice kinase–protein interactions PLoS Physiol (2009) 149(3):1478–1492] and promoter activation by NF-κB p65 [PLoS Biol (2009) 7(3):e1000073].

(See also COMPLEMENTATION.)

binary deoxyribozyme ligase A (synthetic) DEOXYRIBOZYME which can be used to re-code sequence information [Nucleic Acids Res (2006) 34(8):2166–2172].

Based on the MAXIZYME principle, this ligase consists of two (ssDNA) 'half-deoxyribozymes' which can be activated to become a functional enzyme by the binding of a sequence-specific 'bridging oligonucleotide'. In the active enzyme, the two 'half-deoxyribozymes' are base-paired – in their *central* regions – to form a dsDNA stem; at *each* end of this stem the pair of single strands – one from each 'half-deoxyribozyme' – can act as recognition sequences. At each end of the stem, i.e. at the junction between double-stranded and single-stranded regions, is a catalytic (ligase) domain.

When a single *effector* strand of DNA is hybridized to both of the recognition sequences at *one* end of the stem (acting as a 'bridging oligonucleotide' across the two sequences), the enzyme is stimulated to carry out ligation of two separate DNA strands which have bound to the recognition sequences at the *other* end of the stem. Hence, it is possible to ligate two *specified* sequences to form a new sequence – i.e. to 'write' a new sequence.

binary probe A composite PROBE that consists of two single-stranded oligonucleotides designed to hybridize to the target sequence in an end-to-end fashion – with the (labeled) 5′ end of one oligo juxtaposed to the (labeled) 3′ end of the other. *Correct* hybridization of *both* probes to the target sequence can be indicated by a reporter system in several ways.

In one approach, the 5′ and 3′ labels are a compatible pair of fluorophores which, when closely juxtaposed in a correctly hybridized probe, produce so-called fluorescence resonance energy transfer (see FRET; see also LIGHTCYCLER).

In a different approach, both of the oligos are labeled with a pyrene molecule. The rationale is that the emission maxima of pyrene on free (unbound) oligos are significantly different from those of two juxtaposed pyrene molecules on a bound probe (an excited dimer, or so-called *excimer*).

(See also PYRENE BINARY PROBE.)

In addition to the above schemes, the correct juxtaposition of the oligonucleotides on both halves of the target sequence can be indicated by their susceptibility to *ligation;* this basic idea is exploited in a number of techniques (for example, the LIGASE CHAIN REACTION).

binary vector system A vector system in which a given event or function requires the co-ordinated activity of two separate components.

A binary vector system is used, for example, in the method with which plants are genetically modified by the activity of the plant-pathogenic bacterium *Agrobacterium tumefaciens*. This method exploits the ability of *A. tumefaciens* to transfer DNA to plant cells – as occurs when it causes CROWN GALL (q.v.). Genetic modification of plants by this method involves specialized, engineered strains of *A. tumefaciens* which retain the *vir* (virulence) function, i.e. they are able to promote the transfer of T-DNA – or of DNA which is flanked by the end-sequences of T-DNA – into plant cells. The given DNA to be inserted into plant cells (e.g. viral DNA) is first inserted, *in vitro*, into a small vector molecule at a location flanked by T-DNA border sequences. This vector is then inserted into one of the engineered strains of *A. tumefaciens*, forming a binary vector system; thus, when the (Ti-dependent) *vir* function is activated in the bacterium, the T-DNA-flanked target DNA is transferred from the small vector into plant cells.

The design of the small vector (i.e. the one carrying the T-DNA-flanked target sequence) is particularly important if the genetically modified plant is subsequently to enter the general agricultural environment. These vectors should exclude extraneous DNA, and various minimal T-DNA vectors have been proposed [e.g. BioTechniques (2006) 41(6):708–710].

binase An RNase formed (as an extracellular product) by the Gram-positive bacterium *Bacillus intermedius* (and also synthesized from expression plasmids in *Escherichia coli*). It has been used (e.g.) to demonstrate susceptibility of unprotected and unhybridized oligoribonucleotides (in a hydrogel microchip) to RNase activity [BioTechniques (2008) 44(1):77–83].

Binase can promote APOPTOSIS in some cells, and it is also active against tumor cells.

[Binase and other microbial RNases as potential anticancer agents: Bioessays (2008) 30(8):781–790.]

(See also BARNASE.)

biodegradable nanoparticles See NANOPARTICLES.

BioEase™ The designation of several expression vectors (from Invitrogen, Carlsbad CA) which encode a 72-amino-acid tag that directs IN VIVO BIOTINYLATION of a protein–tag fusion product; biotinylation can be achieved in bacterial, insect and mammalian cells.

Biotinylated proteins can be detected and purified *in vitro* with STREPTAVIDIN-based systems.

[Use (e.g.): J Biol Chem (2008) 283(23):15577–15588.]

(See also pcDNA™6/BioEase™-DEST in entry GATEWAY SITE-SPECIFIC RECOMBINATION SYSTEM (table).)

biofilm (extraction of *extra*cellular DNA) See DNA ISOLATION.

biogenesis The process of development by which a given entity is formed in nature. For example, a hormone is the end result of a series of chemical reactions, and mature MICRORNAS are formed from primary transcripts by multi-stage processing.

bioinformatics A computer-based approach to storing, organizing, making accessible and analyzing the vast amounts of information derived primarily from studies on genomes and proteins.

Much of the information consists of sequences of nucleotide residues (from DNA and RNA) and amino acid residues (from proteins) that are stored in databases (see DATABASE). Such stored information can be used, for example, to infer the function of a newly identified gene, or any protein, by comparing its sequence with the sequences of existing (known) genes or proteins.

Analysis of the sequence data (by a process referred to as *computational biology*) has a number of aims. In the context of nucleic acids, one of the requirements is to identify genes among the long, continuous sequences of nucleotide residues. A further aim is to develop approaches that enable prediction of the three-dimensional structure and function of molecules of RNA and protein.

An overarching aim is to classify data into categories which reflect evolutionary relationships.

Desktop facilities for studies in bioinformatics include e.g. the VECTOR NTI ADVANCE 9.1.

biolistic method A form of 'bombardment' used for inserting DNA into certain types of cell; the DNA is coated onto μm-scale gold (or other) particles and fired into the cells.

The biolistic method was used e.g. for inserting constructs into epidermal cells of the onion (*Allium cepa*) [Plant Physiol (2009) 149(1):181–194].

(See also GENE GUN.)

biological containment A safety-orientated approach to experimentation with genetically engineered, or other, organisms or entities in which e.g. a 'limited life' or 'self-destruct' mechanism is brought into operation automatically when the function of the organism or entity has been completed.

(See also AMPLICON CONTAINMENT.)

bioluminescence Light generated within living organisms, or in

systems obtained from living organisms, that is derived from a chemical reaction. (See also entry for (engineered) IN VIVO BIOLUMINESCENCE.) (cf. CHEMILUMINESCENCE.)

Bioluminescence is exhibited by some bacteria (e.g. species of *Alteromonas*, *Vibrio* (*Photobacterium*) and *Xenorhabdus*), fungi (e.g. *Armillaria mellea*, species of *Mycena*), dinoflagellates (e.g. *Gonyaulax*, *Noctiluca*, *Pyrocystis*, *Pyrodinium*) and fireflies (*Photinus pyralis*).

All organisms that exhibit bioluminescence use a (species-specific) thiol-containing oxidoreductase involved in excitation of the light-emitting entity; the generic term for this enzyme is LUCIFERASE (q.v.).

The light-emitting entity from different species is referred to by the generic term *luciferin*. In the firefly, luciferin is 4,5-dihydro-2-(6-hydroxy-2-benzothiazolyl)-4-thiazolecarboxylic acid.

Production of bioluminescence depends on the presence of free oxygen – although only very low levels may be required. The source of energy in bioluminescence depends on species; thus, for example, the firefly uses ATP, but bacteria use NAD and FMN.

Different organisms emit light of different wavelengths; for example, the light from fireflies is ∼560 nm while that from bacteria is ∼475–505 nm. Some dinoflagellates emit light of ∼480 nm in pulses of ∼0.1 second duration.

Applications of bioluminescence

The luciferase–luciferin system of the firefly is employed e.g. for detecting ATP in PYROSEQUENCING.

A bioluminescent system has been used e.g. to detect ATP in certain products (such as UHT milk), the presence of ATP in samples of these products being used as an indication of metabolism by contaminating organisms.

Systemic administration of luminol in animal models was reported to enable *in vivo* imaging of myeloperoxidase activity by bioluminescence [Nature Med (2009) 15:455–461].

(See also BRET, and *Vibrio* in the entry QUORUM SENSING.)

bioluminescence resonance energy transfer See BRET.

biomagnetic separation See DYNABEADS.

biopanning See PHAGE DISPLAY.

biopharmaceutical Any of a wide category of pharmaceutical products (including nucleic-acid- and protein-based agents) which are used therapeutically, or as *in vivo* diagnostics, and which are prepared by methods other than simple extraction of compounds from non-engineered biological sources.

The biopharmaceuticals include a number of products that depend on recombinant DNA technology. In some cases (see e.g. INSULINS) a recombinant protein is largely based on, and is designed to mimic, a natural protein – but may act over a different time-scale. In other cases (for example, the fusion protein etanercept: see accompanying table) the product is an artificial construct which has no natural counterpart.

A number of allied products are manufactured specifically for research, and are not intended for therapeutic, diagnostic or other clinical purposes. They include various products that are used in cell/tissue-culture studies – e.g. basic fibroblast growth factor (bFGF), acidic fibroblast growth factor (aFGF), glial-derived neurotrophic factor (GDNF) and insulin-like growth factor-1 (IGF-1); all of these are recombinant human proteins which are synthesized in *Escherichia coli*.

The manufacture of proteins that are active in a medical or veterinary setting requires not only the correct formulation of nucleic acid coding sequences but also synthesis of proteins which, for correct activity, have the appropriate type of post-translational modification, e.g. specific patterns of glycosylation. This requirement may preclude the use of cells which are unable to carry out a specific type of post-translational modification; for example, bacteria typically do not carry out the type of glycosylation required in a number of mammalian proteins. Nevertheless, in some cases, such a facility may be incorporated in cells in which it is normally lacking (see e.g. GLYCOSYLATION (for 'glycoengineered' *Pichia pastoris*) and MIMIC SF9 INSECT CELLS).

The table shows only a limited selection of the wide range of products that have been produced on a commercial basis.

bioterrorism See e.g. under FORENSIC APPLICATIONS.

biotin (coenzyme R, vitamin H) A cofactor which, *in vivo*, is involved in various types of carboxylation reaction; in these reactions biotin binds (via its carboxyl group) to the ϵ-amino group of a lysine residue in the enzyme.

(See also BIOTIN INTERFERENCE ASSAY and IN VIVO BIOTINYLATION.)

Biotin is used e.g. for labeling PROBES. For this purpose it can be linked, covalently, to nucleoside triphosphate molecules (e.g. dATP) through an allylamine *spacer arm* that may contain e.g. a chain of 10 or more carbon atoms. Biotinylated nucleoside triphosphates can be incorporated into DNA e.g. by 3′ END LABELING or by NICK TRANSLATION.

Biotin can be used to label DNA or RNA. For example, in one commercial product, biotin is linked via a 14-C spacer arm to the 6-position of dATP; this labeled nucleotide can be incorporated efficiently into DNA by nick translation in the presence of dCTP, dGTP and dTTP. CTP labeled at the N4-position via a 14-C spacer arm can be incorporated efficiently into transcripts (e.g. with T7 or T3 RNA polymerase) in the presence of the template and ATP, GTP and UTP.

Site-specific biotinylation of RNA may be achieved during transcription (with T7 RNA polymerase) by the use of special engineered bases, one of which can carry a biotin (or other) residue [Nucleic Acids Res (2005) 33(15):e129].

A biotinylated label may be detected e.g. with a STREPTAVIDIN-bound fluorophore or enzyme; as the biotin molecule is tethered to its nucleotide via the (long) allylamine chain, biotin–streptavidin binding can occur freely (i.e. unhindered by the nucleic acid molecule).

biotin interference assay An assay which has been used e.g. to determine whether or not a given protein needs to interact with a specific DNA locus, at any time during the course of a reaction, in order for the reaction to proceed normally. This is achieved by tethering a molecule of BIOTIN to a nucleotide in the given DNA locus; while the biotin does not significantly

BIOPHARMACEUTICAL: examples of some commercial products whose production involves recombinant DNA technology

Product (Company)	Function/use
Aranesp® (Amgen) (darbepoietin-α)	An erythropoietic protein which stimulates production of erythrocytes (red blood cells). This recombinant protein is produced in a mammalian cell line. Aranesp® is used e.g. for the treatment of anemia associated with chronic renal disease, a condition in which the body's natural erythropoietic system is impaired
Benefix® (Genetics Institute)	A recombinant form of blood-clotting factor IX, produced in Chinese hamster ovary (CHO) cells, which is used in the management of hemophilia B
Enbrel® (Amgen) (etanercept)	A soluble fusion protein combining the extracellular domain of the p75 cell-surface receptor of tumor necrosis factor (TNF) with the Fc portion of immunoglobulin IgG. By sequestering TNF, this agent inhibits the binding of TNF (a pro-inflammatory cytokine) to its natural cell-surface receptor. (The Fc moiety, which has a binding site on macrophages, may facilitate the removal of complexed TNF via phagocytosis.) The product has been used successfully in rheumatoid arthritis to reduce pain and inflammation, an effect which is associated with a fall in the level of TNF
EPOGEN® (Amgen)	A recombinant form of the natural agent erythropoietin (involved in stimulating production of red blood cells) which is produced in a mammalian cell line. It is used in some cases of anemia
Fabrazyme® (Genzyme)	A recombinant form of the enzyme α-galactosidase A (prepared in Chinese hamster ovary (CHO) cells). The product is used for treating Fabry disease (see entry), a condition characterized by a deficiency in α-galactosidase A activity
INTRON® A (Schering-Plough)	Recombinant interferon α-2b. This agent is intended for use in a range of conditions, including e.g. malignant melanoma, chronic hepatitis C and AIDS-related Kaposi sarcoma
Kepivance® (Amgen) (palifermin)	A recombinant form of the human keratinocyte growth factor (KGF). This agent is used in severe oral mucositis, in which cells of the mouth and throat surfaces have been damaged by cancer chemotherapy or by other treatments; it helps to protect the existing cells and stimulates growth/development of new cells
Kineret® (Amgen)	A recombinant form of the interleukin-1 (IL-1) receptor antagonist (IL-1ra) which is produced in *Escherichia coli*. IL-1 is formed in excess in patients with rheumatoid arthritis; by inhibiting the activity of IL-1 in the body, this agent is able to reduce the inflammatory response in the disease. (Other products synthesized in *E. coli* include e.g. Neupogen® (see below); see also entry GLYCOSYLATION.)
Kogenate® (Bayer)	A recombinant form of human factor VIII which is produced in BHK (baby hamster kidney) cells. It is used therapeutically in cases of hemophilia A
Lantus® (Sanofi-Aventis)	INSULIN GLARGINE: a long-acting recombinant form of human insulin used for the treatment of diabetics.(See also NovoRapid®, below.)
Nespo® (Dompé Biotec)	A product in the same category as Aranesp® (see above)
Neulasta® (Amgen) (PEGfilgrastim)	A longer-acting form of Neupogen® (see below)
Neupogen® (Amgen) (filgrastim)	A recombinant form of G-CSF (granulocyte colony-stimulating factor) (produced in *Escherichia coli*). This agent stimulates the production of neutrophils (infection-fighting white blood cells) which are depleted in neutropenia. The recombinant product is not glycosylated (unlike the natural protein) but its biological activity is equivalent to that of the natural G-CSF
NovoRapid® (Novo Nordisk)	A short-acting, recombinant form of insulin: insulin aspart (see entry INSULINS) used for the treatment of diabetics. (See also Lantus®, above.)
NovoSeven® (Novo Nordisk)	A recombinant form of factor VIIa (produced in baby hamster kidney (BHK) cells). This agent permits the coagulation of blood in the absence of factors VIII and IX (although it does not replace factors VIII and IX). It is used e.g. for the prevention of bleeding during surgery in hemophiliac patients
Refludan® (Hoechst) (lepirudin)	A recombinant form of the thrombin-inhibitor hirudin (produced in the yeast *Saccharomyces cerevisiae*). This product is used e.g. as an anticoagulant
Vitravene® (ISIS)	An antisense oligonucleotide (21 nucleotides in length) used for the treatment of CMV (cytomegalovirus) retinitis in AIDS patients. This agent is injected directly into the eye and inhibits viral replication by interfering with the function of certain viral mRNAs

inhibit duplex formation it does inhibit protein binding to the biotin-modified site.

A biotin interference assay has been used e.g. to investigate interactions at recombinase-binding sites; in this study, biotin was attached to a thymidine residue (at the 5-position) via a 16-atom linker, the biotin being ~25Å from the major groove [J Biol Chem (2008) 283(18):12402–12414].

(See also DNA–PROTEIN INTERACTIONS.)

biotin ligase See IN VIVO BIOTINYLATION (engineered).

biotin ligase recognition peptide (BLRP) See IN VIVO BIOTINYLATION (engineered).

biotinylation *in vivo* (engineered) See entry IN VIVO BIOTINYLATION (engineered).

bipartite 'Two-piece' – refers e.g. to the genome of a virus that consists of two pieces of nucleic acid. Viruses with a bipartite genome include the arenaviruses (ssRNA) – e.g. Lassa fever virus.

(cf. MONOPARTITE and MULTIPARTITE.)

BirA The protein–biotin ligase of *Escherichia coli* (encoded by gene *birA*). (See IN VIVO BIOTINYLATION (engineered).)

bird flu See INFLUENZAVIRUS.

Bis A shorthand form for *N,N′*-methylene-bis-acrylamide, a catalyst used for cross-linking ACRYLAMIDE in the preparation of POLYACRYLAMIDE.

bisulfite (action on DNA) Bisulfite can deaminate cytosine to uracil in single-stranded DNA; subsequent replication leads to a GC-to-AT change. *Methylated* cytosines are refractory to this treatment.

Bisulfite can also cause cross-linking between nucleic acids and proteins.

(See also COBRA, HEADLOOP SUPPRESSION PCR, METHYLATION-SPECIFIC PCR, METHYLATION-SPECIFIC SINGLE-BASE EXTENSION.)

(See also the note on bisulfite-treated (sulfonated) DNA in the entry AMPLICON INACTIVATION.)

BK virus A polyomavirus (genome: ccc dsDNA) which infects humans and other mammals. It is related to JC VIRUS (q.v.).

BL21-CodonPlus® See CODON BIAS.

BL21 Star™ Strains of *Escherichia coli* (Invitrogen, Carlsbad CA) used e.g. for T7-promoter-based protein expression. The strains have a mutation, *rne131*, in the gene encoding RNase E; the resulting decrease in RNase activity in these strains is reported to result in an improvement in the stability of transcripts and an increase in the yield of protein.

[Examples of use of BL21 Star cells: PLoS Pathog (2009) 5(4):e1000390; Nucleic Acids Res (2009) 37(4):1269–1279; J Biol Chem (2009) 284(13):8866–8876.]

***bla* gene** A gene encoding a β-LACTAMASE.

black–white screening A method of screening for recombinants in which the host cells contain the same indicator system (β-galactosidase and α-peptide genes) as that used in *blue–white screening* (see PBLUESCRIPT for details) but in which a different chromogenic reagent (Sgal) is used. In this system, the presence of an insert (recombinants: no β-galactosidase activity) is indicated by the development of white colonies, while cells lacking an insert (β-galactosidase activity) cleave the Sgal to products that form a black compound with added ferric chloride (and hence form black colonies).

BLAST Basic local alignment search tool (a program used for comparing sequences):
http://www.ncbi.nlm.nih.gov/BLAST/

Using this search tool, it is possible to make the following comparisons:

1. Nucleotide query sequence compared with nucleotide database.
2. Products of the six possible reading frames of a nucleotide query sequence (three reading frames per strand) compared with a protein database.
3. Six possible reading frames of a nucleotide query sequence compared with six reading frames in a nucleotide database.
4. A polypeptide or protein query sequence compared with a protein database.
5. A polypeptide/protein query sequence compared with the products of all six reading frames of a nucleotide database.

Analysis of BLAST alignment reports may be facilitated by web-based interface SEQUEROME [BioTechniques (2005) 39(2):186–188].

While BLAST is often used to search for *fragments* (i.e. incomplete sequences) of genes etc., a search for the entire sequence of a homologous gene may be facilitated by the use of the graph-based algorithm genBlastA [Genome Res (2009) 19:143–149].

blasticidin S An antibiotic which inhibits protein synthesis by inhibiting the peptidyltransferase reaction during translation in both prokaryotic and eukaryotic cells.

Examples of use of blasticidin S in mammalian cells (at 6 μg/mL) [PLoS Medicine (2006) 3(10):e420]; in fungal cells (at 300 μg/mL) [Eukaryotic Cell (2006) 5(6):896–904]; and for studies on *Plasmodium* (at 2 μg/mL) PLoS Pathog (2009) 5(1):e1000256].

blastocyst A structure formed from those cells (*blastomeres*) produced by division of the fertilized egg; it is a hollow ball of cells that consists of two main parts: (i) an outer layer of cells (the *trophectoderm*), which forms the wall of the sac, and which encloses the blastocyst cavity, and (ii) the *inner cell mass* (ICM), which is found within the cavity attached to the inner wall of the trophectoderm (and which will develop to form the embryo). (The trophectoderm which is in contact with the ICM is referred to as *polar* trophectoderm, while the rest of the trophectoderm, which forms the boundary of the blastocyst cavity, is called the *mural* trophectoderm.) Shortly after its formation, the blastocyst attaches to the wall of the uterus (*implantation*) and undergoes further development.

In genetic experimentation, individual blastomeres can be subjected to the required manipulation; they are then inserted into a blastocyst – which is inserted into a female animal for further development. The fate of specific, individual blastomeres within the developing blastocyst and embryo may be followed by a fluorescence-based tracking system – see e.g. CELL LINEAGE TRACKING.

A blastocyst stage develops during the process of whole-animal cloning following the SCNT procedure.

In vitro, the blastocyst can be a source of embryonic STEM CELLS, which are derived from the ICM cells (see above). In culture, trophoblast cells give rise to a monolayer while the ICM cells form a three-dimensional colony.

In cultures of mouse embryonic stem cells the LEUKEMIA INHIBITORY FACTOR can be used to inhibit differentiation of the ICM cells, i.e. it promotes pluripotent growth. (See also FEEDER CELLS.)

blastomeres Cells formed by cell division of a fertilized egg – see BLASTOCYST.

blastulation The process in which cells (*blastomeres*) give rise to a BLASTOCYST (q.v.).

bleomycin Any one of a group of structurally complex, DNA-binding, glycoprotein antitumor agents (produced by a Gram-positive bacterium, *Streptomyces verticillus*) which can cause both single-strand and double-strand cleavage of DNA in the presence of Fe^{2+} and oxygen.

[Mechanistic studies on bleomycin-mediated DNA damage (multiple binding modes can result in double-stranded DNA cleavage): Nucleic Acids Res (2008) 36(11):3781–3790.]

In the yeast *Saccharomyces cerevisiae*, susceptibility to the recombinogenic/mutagenic effects of bleomycin have been reported to be enhanced by intercalation of ACRIDINES in the DNA [Mutagenesis (2009) 24(4):317–329].

Delivery of bleomycin (uptake by cells) has been improved (giving increased bioavailability) by tethering this agent to a protein construct referred to as the adenovirus *dodecahedron base*: a complex consisting of twelve copies of a pentameric adenovirus protein which is involved in viral penetration of the cell [PLoS ONE (2009) 4(5):e5569].

***Blm* gene** See BLOOM'S SYNDROME.

***BLM* gene** See BLOOM'S SYNDROME.

blocked reading frame See the entry OPEN READING FRAME.

blood–brain barrier (in gene transfer) See GENE THERAPY and GENE-DELIVERY SYSTEM.

blood RNA isolation kit See e.g. PAXGENE BLOOD RNA KIT.

***Bloom* gene** See BLOOM'S SYNDROME.

Bloom's syndrome A syndrome, characterized by an increased incidence of various types of cancer, which is associated with mutation (functional deficiency) in a gene (variously referred to as *Bloom*, *Blm* and *BLM*) that encodes a HELICASE of the RecQ family (see RECQ PROTEIN). Deficiency in this helicase is linked to genomic instability.

Bloom's syndrome protein A RecQ helicase whose deficiency is associated with BLOOM'S SYNDROME.

blotting Any of various procedures in which molecules or fragments are transferred from a gel (following electrophoresis) to a nitrocellulose or other matrix.

(See SOUTHERN BLOTTING, NORTHERN BLOTTING and WESTERN BLOTTING, IMMUNOBLOTTING.)

BLRP Biotin ligase recognition peptide – see IN VIVO BIOTINYLATION.

blue-green algae The former name of the CYANOBACTERIA; they were called 'blue-green algae' because of their ability to carry out photosynthesis in the oxygenic manner, i.e. a mode similar to that found in algae (and in higher plants).

blue–white screening See PBLUESCRIPT.

(cf. BLACK–WHITE SCREENING.)

Bluescript® See the entry PBLUESCRIPT.

blunt-end ligation Ligation of two blunt-ended molecules of nucleic acid using e.g. the (ATP-dependent) T4 DNA ligase. (See also POLISHING and ZERO BLUNT.)

blunt-ended DNA Double-stranded DNA which has no 5′ or 3′ single-stranded overhangs. Blunt-ended termini are generated e.g. by cleavage with certain RESTRICTION ENDONUCLEASES (e.g. HpaI) which make a 'blunt-ended cut'. (Compare with STICKY ENDS.)

(See also entries END-IT DNA END-REPAIR KIT, POLISHING, and ZERO BLUNT.)

Bluo-gal See X-GAL.

Bm*S*NPV See NUCLEAR POLYHEDROSIS VIRUSES.

BODIPY The trade name of a range of fluorophores based on a substituted tricyclic indacene core. Some of the dyes are used e.g. in the automated sequencing of DNA and for the labeling of probes etc.

BODIPY dyes are designated systematically on the basis of their approximate absorption/emission maxima (nm), as determined in methanol.

***Bombyx mori* NPV** See NUCLEAR POLYHEDROSIS VIRUSES.

Borrelia burgdorferi A species of Gram-negative bacteria in which the genome consists of *linear* double-stranded DNA. In this organism the terminal regions of the chromosome are closed (hairpin) structures called *telomeres*.

Chromosomal replication in *B. burgdorferi* is initiated bi-directionally from an internal origin. Replication involves an intermediate circular molecule which consists of a chromosome *dimer* in a head-to-head, tail-to-tail arrangement. The junctions between the individual chromosomes are processed by a breakage-and-reunion mechanism (*telomere resolution*) to form the closed (hairpin) termini in each daughter chromosome. Telomere resolution is catalysed by an enzyme called PROTELOMERASE.

Features of this type of genome replication are also shown by certain phages (e.g. N15, φKO2).

B. burgdorferi (*sensu lato*) is the causal agent of the (tick-borne) infection Lyme disease.

bortezomib (PS-341) A dipeptidyl boronic acid which inhibits 26S proteasomes (see PROTEASOME) and inhibits the activation of transcription factor NF-κB.

Bortezomib has been approved for the treatment of multiple myeloma (although it has been associated e.g. with thrombocytopenia).

Bortezomib was reported to be more effective at killing B lymphocytes infected with EPSTEIN–BARR VIRUS in the type III phase of latency than B cells infected with EBV in type I latency [J Virol (2007) 81(18):10029–10036].

Treatment with bortezomib is reported to enhance antigen-specific $CD8^+$ T-cell-mediated antitumor immunity induced

by DNA vaccination [J Mol Med (2008) 86(8):899–908].

[Using small-molecule proteasome inhibitors to explore the proteasome's role in cell biology: J Med Chem (2008) 51(9): 2600–2605.]

bovine syncytial virus See SPUMAVIRINAE.

BOX element A type of repetitive intergenic sequence which occurs in the genomes of certain bacteria (e.g. *Streptococcus pneumoniae*) and which has been used in one form of REP-PCR.

In *S. pneumoniae*, the genome is reported to contain more than 100 BOX elements; these elements have been found to affect the expression of neighboring genes, and there is also evidence that they are mobile elements [J Bacteriol (2006) 188(23):8307–8312].

BOXTO A fluorescent cyanine dye which (bound to dsDNA) has an emission maximum at 552 nm on excitation at 515 nm; it was reported to show preferential binding in the minor groove [Nucleic Acids Res (2003) 31(21):6227–6234].

A combination of BOXTO with a target-specific probe has been used in real-time PCR [BioTechniques (2006) 40(3): 315–319].

bp Base pairs: see the entry KB.

BP Clonase™ An enzyme used in the GATEWAY SITE-SPECIFIC RECOMBINATION SYSTEM. BP Clonase is a recombinase that can mediate *in vitro* exchange between a segment flanked by *att*B sites (e.g. a linear dsDNA PCR-generated amplicon) and a segment flanked by *att*P sites (e.g. a site in a circular vector molecule). (PCR amplicons that are suitable for this reaction can be generated by using primers that incorporate a 5′ *att*B tag.)

Insertion of an *att*B-flanked segment into an *att*P-flanked region within a vector molecule (by BP Clonase) produces a vector containing the insert flanked by hybrid *att*B/*att*P sites which are called *att*L sites. This vector can then be used as an *entry clone* in the GATEWAY SITE-SPECIFIC RECOMBINATION SYSTEM – i.e. the *att*L-flanked insert can be inserted into the *att*R-flanked region in any required destination vector.

(cf. LR CLONASE.)

***BRAF* pseudogene** See PSEUDOGENE.

branch migration See HOLLIDAY JUNCTION.

branched DNA assay See BDNA ASSAY.

BRCA Breast cancer gene: either of two genes, referred to as *BRCA1* and *BRCA2*. Inherited mutations in these genes are associated with a risk for breast cancer and ovarian cancer.

BrdU (BUdR) 5-Bromo-2′-deoxyuridine: a mutagenic analog of thymidine which can be phosphorylated by thymidine kinase and incorporated into DNA during replication.

BrdU is used e.g. in a CELL PROLIFERATION ASSAY, and in a method for the diagnosis of XERODERMA PIGMENTOSUM.

BRE Transcription factor IIB recognition element: a eukaryotic promoter sequence that recognizes transcription factor IIB – see also entry CORE PROMOTER.

breast cancer (relevant topics) See (for example) the entries BRCA, CANCER, EPSTEIN–BARR VIRUS, EPIDERMAL GROWTH FACTOR RECEPTOR FAMILY.

breast cancer genes See e.g. BRCA.

BRET Bioluminescence resonance energy transfer: the phenomenon, analogous to FRET, in which a donor entity is capable of exhibiting CHEMILUMINESCENCE given the presence of a suitable substrate; if an appropriate *acceptor* molecule is in close proximity to the donor (within ∼100Å), and in the right orientation, energy is transferred from the donor (without the production of light) to the acceptor – which is thus activated to produce a *fluorescent* signal.

BRET has been used e.g. for monitoring protein–protein interactions within living cells – the donor molecule being linked to one of the proteins and the acceptor molecule being linked to the other.

One advantage of BRET, compared with FRET, is that, owing to the absence of external excitation (and, therefore, the avoidance of autofluorescence), it offers a higher signal-to-background ratio and is well suited to those conditions in which low-level signals are created *within* experimental cells/animals.

[Method for enhancing the image: Proc Natl Acad Sci USA (2007) 104(24):10264–10269.]

[Subcellular imaging of dynamic protein interactions: Biophys J (2008) 94(3):1001–1009.]

[BRET used for studying the mechanism of antipsychotics: Proc Natl Acad Sci USA (2008) 105(36):13656–13661.]

[Use of BRET for studying autophagy in mammalian cells: Mol Biol Cell (2009) 20(3):870–881.]

***brlf-1* gene** See ZEBRA.

broad-range primer ('universal' primer) Any PRIMER whose target sequence is a conserved sequence of nucleotides that occurs in a wide range of species. Thus, for example, certain sequences in the bacterial 16S rRNA gene are common to all species of bacteria. Primers that are complementary to these sequences have various uses. Broad-range primers have been used e.g. as a control in PCR-based assays of species-specific targets in bacterial genomic DNA; in such an assay, failure to amplify a conserved sequence with the broad-range primers may indicate e.g. a failure of the DNA extraction process or failure of the particular PCR protocol used to amplify target sequences.

Another use of broad-range primers in PCR is the detection of PCR-inhibitory substances in test samples; thus, inhibitors may be suspected if a sample containing bacterial genomic DNA yields no products when assayed by PCR with broad-range primers.

broad-spectrum antibiotic An ANTIBIOTIC which is effective against a wide range of species of bacteria; those susceptible may include both Gram-positive and Gram-negative species.

5-bromouracil (BU) A mutagenic analog of thymine. The *keto* tautomer pairs with adenine, but the *enol* tautomer pairs with cytosine; if present in the *enol* form during DNA replication it can cause an AT→GC mutation.

BSA Bovine serum albumin (used e.g. as a facilitator of PCR).

***Bs*PolX** See X FAMILY.

BssHII A type IIP RESTRICTION ENDONUCLEASE produced by

Bacillus stearothermophilus; recognition site: G↓CGCGC. It forms fragments with a 5′ overhang. BssHII is useful e.g. for the preferential cutting of genomic DNA at (GC-rich) sites in CPG ISLANDS.

BTA Benzene-1,3,5-triacetic acid: an agent used e.g. for attaching 5′-aminated oligonucleotides to an aminosilanized glass surface [Nucleic Acids Res (2006) 34(3):e22].

BU BROMOURACIL.

bubble-linker PCR *Syn.* VECTORETTE PCR.

buccal cell sampling Collection of cells from inside the cheek as a source of DNA for analysis (e.g. DNA profiling); a swab or filter-paper device is commonly used. The yield of DNA per sample varies considerably between individuals – as does the ratio of human to non-human (e.g. bacterial) DNA in the sample. Using a filter-paper device, the (total) DNA yield per sample may be within the range 20–500 ng [BioTechniques (2005) 39:257–261].

Double-stranded DNA in solution can be quantified e.g. by the PICOGREEN ASSAY.

(See also FORENSIC APPLICATIONS.)

Analogous procedures are used for collecting the DNA samples from animals.

Among *non*-mammalian species, efficient buccal swabbing for DNA samples has been reported for the sunfish (*Lepomis*) [BioTechniques (2005) 38:188–192].

Buchnera A genus of bacteria found within specialized cells (*bacteriocytes*) in aphids; reproduction of the aphids appears to be wholly dependent on the *Buchnera* endosymbionts.

Buchnera has one of the smallest genomes among bacteria (640681 bp). [Genome sequence of *Buchnera* sp APS: Nature (2000) 407:81–86.]

(See also WIGGLESWORTHIA and WOLBACHIA.)

budded virions (BVs) See BACULOVIRIDAE.

BUdR *Syn.* BRDU.

buffy coat The layer of white cells above the packed red cells in a sample of centrifuged unclotted blood.

bulged duplex A phrase which is sometimes used to refer to a region in a HETERODUPLEX (sense 2) in which the strands are separated ('bulge') owing to non-complementarity.

bumper primer See the entry STRAND DISPLACEMENT.

Bunyaviridae A family of enveloped ssRNA viruses that have a SEGMENTED GENOME; in viruses of the genus *Bunyavirus*, all three genome segments are negative sense. Hosts include vertebrates, insects and plants. Some bunyaviruses are major human pathogens – e.g. those causing arthropod-borne viral hemorrhagic fevers. Although many bunyaviruses are maintained in arthropods (e.g. mosquitoes, sandflies, ticks) viruses of the genus *Hantavirus* occur in populations of persistently infected rodents; infection of human hosts usually occurs via arthropod vectors or (in the case of *Hantavirus*) by the inhalation of virus-containing aerosols.

(See also RIFT VALLEY FEVER VIRUS.)

bunyaviruses Viruses of the family BUNYAVIRIDAE.

BUR (in genomic DNA) Base-unpaired region: see the entry MATRIX-ATTACHMENT REGION.

BUR nucleation sequence A nucleotide sequence that is often found in a base-unpaired region in a MATRIX-ATTACHMENT REGION.

***t*-butyldimethylsilyl-** A group which has been used at the 2′-O-position of oligoribonucleotides (in a hydrogel microchip) to inhibit enzymic degradation of the oligoribonucleotides; such modification can be removed, as required, prior to use of the oligoribonucleotides in hybridization studies [BioTechniques (2008) 44(1):77–82].

BVs (budded virions) See BACULOVIRIDAE.

Bxb1 See PHAGE BXB1.

bypass Refers to the action of a 'translesion' DNA polymerase (see e.g. Y FAMILY) when, on encountering certain types of DNA lesion (e.g. 8-OXOG) in the template strand, it continues to synthesize the new strand of DNA. Typically, an incorrect nucleotide is inserted opposite an 8-oxoG lesion.

(See also BYPASS DNA POLYMERASE.)

bypass DNA polymerase Any DNA polymerase that continues to synthesize the new strand of DNA even when the template contains certain types of lesion (e.g. 8-OXOG), such synthesis (typically) resulting in the incorporation of incorrect nucleotides opposite the lesions. (Other, routine, 'high-fidelity' DNA polymerases are normally inhibited/blocked by such lesions.) Bypass DNA polymerases occur e.g. in the Y FAMILY.

A Y-family archaeal DNA polymerase, Dpo4, was reported to make an error-free BYPASS at an 8-oxoG lesion – correctly inserting a cytosine opposite 8-oxoG [PLoS Biol (2006) 4(1): e11]. A more characteristic reaction of the bypass DNA polymerases is exhibited by the yeast polymerases δ (delta) and η (eta) – which incorporate adenine opposite an 8-oxoG lesion [Nucleic Acids Res (2009) 37(9):2830–2840].

(See also MISMATCH EXTENSION.)

bZIP Basic zipper: a feature of certain proteins consisting of a basic region and a LEUCINE ZIPPER.

***bzlf-1* gene** See ZEBRA.

C

C L-Cysteine (alternative to Cys).

c-*myc* See MYC. (See also GELDANAMYCIN.)

c-*onc* A *cellular* ONCOGENE.

C value (*C*-value etc.) The quantity of DNA in a *haploid* genome measured in e.g. picograms or number of kilobase-pairs.

C19MC A cluster of 46 miRNA-encoding genes that is present on (human) chromosome 19 (see MICRORNAS).

CA-MRSA Community-acquired MRSA: see the entry MRSA.

CAD Caspase-activated DNase: see APOPTOSIS.

cadang-cadang viroid COCONUT CADANG-CADANG VIROID.

Caenorhabditis elegans A small (~1 mm) transparent nematode worm used in genetic studies.

The genome (~97 Mbp) is not typically eukaryotic because many of the genes occur in *operons* (clusters of contiguous genes which are typical of bacterial genomic organization); the polycistronic transcripts from these operons are processed by TRANS SPLICING.

[Genome sequence: Science (1998) 282:2012–2018.]

[Transcriptome analysis: BMC Biol (2008) 6:30.]

[Massively parallel sequencing of the polyadenylated transcriptome of *C elegans*: Genome Res (2009) 19:657–666.]

Caf1 A factor involved in deadenylation of the poly(A) tail of mRNAs in eukaryotes.

[Caf1 in mRNA deadenylation in *Trypanosoma brucei* and human cells: Nucleic Acids Res (2008) 36(10):3374–3388.]

(See also PARN.)

CAGE Cap analysis gene expression: an approach used e.g. for mapping the location of transcriptional start sites throughout a given genome. Essentially, this involves sequencing a short (~20 nt) tag originating from the 5′ end of each full-length mRNA and then using the data from the whole population of transcripts to identify those sites in the genome from which transcription had occurred.

Each transcript is initially reverse transcribed with an oligo-dT primer (see FIRST STRAND) – or with a random primer in order to include non-polyadenylated transcripts. RNA/DNA hybrids that include a *complete* 5′ end are selected by biotinylating the CAP structure and isolating by streptavidin-coated DYNABEADS (see the entry CAP TRAPPER). Subsequently, a LINKER which incorporates the recognition site of the restriction enzyme MmeI (see table in entry RESTRICTION ENDONUCLEASE) is ligated to the cDNA – producing a *cutting* site for MmeI ~20 nt proximal to the point of ligation, so that MmeI can excise a short tag from the cDNA. Tags derived from various RNAs are sequenced, and the data used to determine transcription start sites in the genome.

Using high-throughput sequencing of the tags, it is possible to quantify the different types of tag in a given preparation and, in this way, to estimate the intracellular concentration of the transcript corresponding to each of the CAGE tags; by this means, it is possible to estimate the level of expression of a given transcript within the cell (as well as enabling its start site to be determined).

['Deep CAGE' (genome-wide identification of promoters, quantification of their expression, and network inference) – a mini-review: BioTechniques (2008) 44(5):627–632.]

caged DNA DNA that is transcriptionally inactivated (or partly so) by covalent modification with photolabile compounds; it can be inserted into cells and then 'uncaged' (i.e. made transcriptionally active) by a short burst of ultraviolet radiation (at e.g. λ 360 nm) – which does not damage the cells. A similar approach is used with individual nuleotides, e.g. ATP.

Caging is used e.g. for studying intracellular events with a precise control of timing.

[Use (e.g.): Nucleic Acids Res (2009) 37(8):e58.]

caged luciferin A form of inactive luciferin that can be inserted into cells (across the cytoplasmic membrane) and then activated intracellularly as and when required.

(See also BIOLUMINESCENCE.)

Within cells, luciferin can be activated instantaneously by a pulse of ultraviolet radiation, or alternatively, active luciferin can be released – over time – by the activity of (endogenous) esterases on the caged luciferin, thus allowing measurements to be made over extended periods.

Caged luciferin is available commercially.

(See also LUCIFERASE.)

caged nucleotide See CAGED DNA.

Cairns' mechanism (theta (θ) replication) A mechanism for the replication of a ds (i.e. double-stranded) cccDNA molecule in which replication begins at a fixed point, progresses – either unidirectionally or bidirectionally (according to organism) – and finishes at a defined terminus; both strands are replicated more or less simultaneously. A partly replicated intermediate form has a 'θ'-like structure. Bidirectional replication occurs e.g. in the chromosome of *Escherichia coli*.

Cajal body An organelle found within the nucleus in a range of organisms, including animals, plants and yeasts; it has been associated with roles such as the assembly of protein/siRNA silencing complexes and the maturation of snRNPs.

[Poly(ADP-ribose) polymerase 1 in localization of proteins to the Cajal body: PLoS Genet (2009) 5(2):e1000387.]

calmodulin A ~17-kDa heat-stable, acid-stable protein which, when activated by calcium ions, stimulates various enzymes, including certain bacterial exotoxins which have ADENYLATE CYCLASE activity. (See also AFFINITY PROTEIN EXPRESSION AND PURIFICATION.)

calmodulin affinity resin A product that binds CALMODULIN-binding peptide tag (see AFFINITY PROTEIN EXPRESSION AND PURIFICATION) in the presence of low concentrations of Ca^{2+}; elution can be achieved in the presence of 2 mM EGTA at pH 7.

[Use (e.g.): studies on DNA ligase 1-deficient plants: BMC Plant Biol (2009) 9:79.]

calreticulin A 46-kDa Ca^{2+}-binding protein, associated (e.g.) with the endoplasmic reticulum, which has role(s) in antigen

presentation. Complexes formed with peptides, *in vitro*, can induce a peptide-specific T cell response.

A DNA VACCINE that also encoded calreticulin was found to elicit a better immune response.

CAM (1) CALMODULIN.
(2) Cell adhesion molecule.
(3) Chorioallantoic membrane.

CAM plasmid A large *Pseudomonas* plasmid which encodes functions for the degradation of camphor.

Cambridge Reference Sequence A sequence of mitochondrial DNA (mtDNA) used as a standard against which individual mtDNA profiles are defined in the US government's mtDNA Popstats Population Database (see FORENSIC APPLICATIONS).

cAMP CYCLIC AMP.

cAMP phosphodiesterase See CYCLIC AMP.

cAMP receptor protein (CRP) See CATABOLITE REPRESSION.

Campbell model An early model, proposed by Campbell in the 1960s, for the integration of a circular plasmid or phage into a bacterial chromosome by a single cross-over.

Campbell-type integration The mechanism also referred to as INSERTION–DUPLICATION RECOMBINATION (q.v.).

Campylobacter A genus of Gram-negative, motile, respiratory (oxidative), oxidase-positive bacteria that metabolize amino acids and TCA cycle intermediates; the cells, spirally curved rods, are often 1–5 μm in length.

The GC% of the genomic DNA is 30–38. The type species is *C. fetus*.

[Complete genome of *C. jejuni* (NCTC11828): J Bacteriol (2007) 189(22):8402–8403.]

Campylobacter spp are found as parasites and pathogens in man and other animals. *C. jejuni* and *C. coli* can cause e.g. food-borne infections in man. *C. fetus* can cause abortion in sheep. *C. sputorum* can cause enteric disease in hogs (pigs).

cancer Any disease of man or animals, involving unrestricted growth and multiplication of a subset of cells, in which a malignant tumor is formed at a given location; a *carcinoma* is a type of cancer which develops from epithelial cells, while a *sarcoma* develops from connective tissue. A malignant tumor may invade and destroy adjacent tissues, and it may also shed cancer cells which disseminate (*metastasize*) to other regions of the body (via the bloodstream or lymph), resulting in the formation of secondary tumors at close or distant sites.

(See also the comparison of gene expression in cancer cells, *embryonic* stem (ES) cells and adult tissue stem cells in the entry STEM CELL.)

Leukemia is a malignant disease, also due to abnormal cell proliferation, in which the bloodstream contains an increased number of white cells (of abnormal/immature type); it can be caused e.g. by some RETROVIRUSES. Characteristically, there is anemia, a tendency to hemorrhage, and increased susceptibility to infections. (In some patients, leukemia has resulted from retrovirus-mediated gene therapy [J Clin Invest (2008) 118:3132–3142].)

(See also CANCER DETECTION and CANCER THERAPY.)

In *plants*, a cancer is referred to as a *gall* (see e.g. CROWN GALL).

Causation of cancer in man and animals

In man and animals, the cause of a given cancer may involve one or more of the following factors: (i) a CARCINOGEN (see e.g. ULTRAVIOLET RADIATION and MELANOMA); (ii) loss of function of a tumor suppressor (see e.g. P53); (iii) the activity of an ONCOGENE; (iv) the activity of certain viruses (see e.g. EPSTEIN–BARR VIRUS); (v) GENE AMPLIFICATION and/or the overexpression of products that promote growth and division of cells. (vi) an ONCOMIR.

Some individuals have a genetic pre-disposition to develop certain types of cancer (e.g. breast cancer: see BRCA; see also BLOOM'S SYNDROME, WERNER SYNDROME and XERODERMA PIGMENTOSUM).

(See also: CISPLATIN; DIGENE HYBRID CAPTURE 2 ASSAY; EPIDERMAL GROWTH FACTOR RECEPTOR FAMILY; ERLOTINIB; GEFITINIB; GENE THERAPY; IMATINIB; KAISO PROTEIN; KIT; MSI ANALYSIS; MUTAGEN; PAPILLOMAVIRUSES; enzyme RPase III in the entry RNA POLYMERASE.)

cancer detection Laboratory-based methods for the detection of cancer include histology of biopsies and assays for cancer-associated antigens; an example of a cancer-associated antigen is PSA (prostate-specific antigen), levels of which (in the blood) are significantly elevated in prostate cancer patients.

DNA-based methods include monitoring gene expression patterns in order to reveal dysregulated cells. In one approach the messenger RNAs (mRNAs) are examined. More recently, interest has been shown in monitoring patterns of expression of microRNAs (miRNAs) in cancer. For example, miRNAs from prostate cancer tissue have been reported to be present in the circulation, and to permit blood samples from patients with prostate cancer to be distinguished from those of healthy patients [Proc Natl Acad Sci USA (2008) 105:10513–10518]. A similarly distinctive pattern of miRNA expression has been reported in ovarian cancer patients [Gynecol Oncol (2009) 112:55–59]. Using a microarray-based approach, expression patterns of miRNAs in serum were reported to be able to discriminate between the samples from cancer patients and samples from normal individuals (without amplification techniques) [PLoS ONE (2009) 4(7):e6229].

***DNA-based screening for colorectal cancer* (*CRC*)**

Although it is a major cause of morbidity and mortality, colorectal cancer can be successfully treated – if diagnosed early – and attention has been focused on identifying effective biomarkers of the disease for timely detection.

Various DNA-based markers for CRC have been described. These include:

- Point mutations in certain genes, including the K-*ras* oncogene (see RAS) and the tumor suppressor gene *TP53*.
- Instability of MICROSATELLITE DNA – e.g. in the BAT-26 sequence (a region of 26 consecutive adenines).
- Hypermethylation (silencing) of the *TFPI2* gene (the tissue factor pathway inhibitor-2 gene); aberrant methylation in this gene has been reported in almost all CRC adenomas. [Cancer Res (2009) 69(11):4691–4699.]

• Long-fragment DNA in (exfoliated) cells in stool. This has been regarded as an indication of the presence of APOPTOSIS-resistant tumor cells – and has been assayed by a fecal DNA test: the *DNA integrity assay* (DIA).

[Genetic and epigenetic marker-based DNA test of stool is a promising approach for colorectal cancer screening: Yonsei Med J (2009) 50(3):331–334.]

cancer terminator virus See MELANOMA.

(See also HEPATOCELLULAR CARCINOMA SUPPRESSOR 1.)

cancer therapy (*DNA technol.*) Therapy designed specifically to combat CANCER. In a context of DNA technology, current approaches include: use of ONCOLYTIC VIRUSES; other oncolytic systems, e.g. specifically targeted pro-apoptotic genes; angiogenesis-blocking methods; inactivation of ONCOGENES; anti-metastasis (see e.g. NM23-H1); use of cancer-cell-specific cytotoxic systems; and the combined use of oncolytic viruses with SHORT HAIRPIN RNA (see the entry ONCOLYTIC VIRUS). For therapy against solid tumors, work is now being carried out to develop (more effective) *replicating* retroviral vectors (see the section on RCR vectors in the entry RETROVIRUSES).

(*Chemotherapy* using thiopurine base analogs is associated with a risk of skin cancer: see *Mutagenic effects of ultraviolet radiation* in the entry ULTRAVIOLET RADIATION.)

The targets in cancer therapy also include tumor suppressor genes which, due to methylation-dependent inactivation, have failed to prevent tumor development (see the entry DEMETHYLATION).

Yet another target is the enzyme topoisomerase II; here, the idea is selectively to interfere with the enzyme's normal role in a manner that creates damage to the DNA in cancer cells. [Targeting DNA topoisomerase II in cancer chemotherapy: Nature Rev Cancer (2009) 9:338–350.]

In one scheme, oncolytic viruses were adsorbed onto silica implants and used in an *extended release* program for treating gastric and pancreatic tumors in mice [Gene Therapy (2009) 16:103–110].

Inhibition of neovascularization (angiogenesis) in human tumors – developed in nude mice – was reported with the use of a DNA enzyme (DNAzyme) active against the mRNA of a β_1 INTEGRIN; downregulation of the expression of β_1 integrin in *endothelial* cells appears to inhibit the ability of these cells to form microcapillaries. On injection into tumors, the DNAzyme significantly reduced both tumor size and number of microvessels in the 3 weeks of experimentation [Cancer Gene Therapy (2009) doi: 10.1038/cgt.2009.13].

Multitarget therapy, using multiple shRNAs, was suggested as a potentially effective strategy for use in cancer therapy [Cancer Gene Therapy (2009) doi: 10.1038/cgt.2008.102].

Induction of senescence has been suggested as a possible anti-cancer therapy [Mol Cancer (2009) 8:3].

(See also MDA-7/IL-24.)

ONCOGENE-related transcription factors have been assessed for their potential as targets in individualized therapy [Cancer Gene Therapy (2009) 16:103–112].

Genes that are delivered for therapeutic purposes must be adequately controlled in order that: (i) optimal efficacy can be obtained, and (ii) the gene's activity can be switched off, if necessary, given adverse effects. One review has looked at engineered transcriptional switches which could be employed within those gene-delivery systems involved in anti-cancer gene therapy [Mol Cancer Therapy (2008) 7(3):439–448].

Cannabis sativa (forensic analysis) See FORENSIC APPLICATIONS (*Plant pathogen forensics*/*plant forensics*).

cap (in mRNA) In most eukaryotic (and some viral) mRNAs: the 5′ terminal region consisting of:

$$m^7G^{5'}ppp^{5'}N^{3'}p^{5'}N^{3'}p\ldots\ldots 3'$$

which is added shortly after the start of transcription. The guanosine is linked 5′-to-5′ to the 5′-triphosphate end of the transcript by guanylyl transferase and is then methylated by a specific 7-methyltransferase. (Adjacent nucleotide(s) are also methylated in some caps.) The methylguanosine appears to facilitate translation by binding to *cap-binding proteins* in the 40S ribosomal subunit.

Cap-independent translation of proteins can be achieved, *in vitro* and *in vivo*, e.g. by using an internal ribosomal entry site (see IRES).

The cap may serve to inhibit 5′ exonucleolytic degradation of mRNA. Interaction between the cap and poly(A)-specific ribonuclease (see PARN) was reported to promote degradation of the (3′) poly(A) tail.

The cap structure is recognized by (eukaryotic) eIF4E (part of the translation initiation complex).

A (5′) 2,2,7-trimethylguanosine cap is found on snRNAs.

(See also CAP TRAPPER, CAPFINDER, CAPSELECT.)

(cf. CAP ANALOG.)

CAP Catabolite activator protein in CATABOLITE REPRESSION.

cap analog Any of various synthetic sequences of nucleotides – based on the CAP sequence and its natural variant forms – that are used for preparing transcripts for *in vitro* translation reactions. (See also IRES.)

Various types of cap analog can be obtained commercially (from e.g. Ambion, Austin TX, USA).

[Novel cap analogs for *in vitro* synthesis of mRNAs with high translational efficiency: RNA (2004) 10(9):1479–1487.]

(See also ANTI-REVERSE CAP ANALOG.)

cap analysis gene expression See CAGE.

cap-binding proteins See the entry CAP (in mRNA).

cap-independent translation See e.g. IRES.

cap trapper A technique used for obtaining full-length cDNAs from mRNAs. The FIRST STRAND is synthesized. Then, those RNA/DNA hybrid structures containing a full-length mRNA (including the CAP region) undergo chemical modification in which a BIOTIN molecule is introduced into the diol residue of the cap; such (full-length) structures can then be selected by STREPTAVIDIN-coated magnetic beads. (Incomplete RNA/DNA hybrids that lack the cap structure are not biotinylated and are thus eliminated.)

First reported some time ago [Genomics (1996) 37(3):327–

336], this approach has been widely used – e.g. for studies on the plant *Thellungiella halophila* [BMC Plant Biol (2008) 8: 115]; studies on barley [DNA Res (2009) 16(2):81–89]; and studies on adaptive evolution between wheat and rice [BMC Genomics (2009) 10:271].

(See also CAGE.)

CapFinder A technique used for facilitating the preparation of *full-length* cDNAs. Essentially, when synthesis of the FIRST STRAND has copied the 5′ CAP region on the mRNA template strand, the MMLV reverse transcriptase adds several deoxycytidines to the 3′ end of the newly formed first strand; this addition appears to occur preferentially when the first strand has *completely* copied the capped mRNA. The reaction mixture includes copies of a short oligonucleotide (the so-called *T-S oligo*) whose 3′ end consists of several guanosine residues. This oligo binds to the short chain of terminal deoxycytidines in the first strand. The enzyme reverse transcriptase then *switches template* from the mRNA template to the T-S oligo template and, as synthesis continues, a complementary copy of the 5′ end of the T-S oligo is added to the 3′ end of the first strand. The T-S oligo primes synthesis of the second strand.

One problem with this scheme is that the 3′ end of the T-S oligo may bind to any complementary sequence(s) available in the mixture; hence, were the T-S oligo to be used as a PCR primer in a subsequent round of amplification there might be heavy contamination with random products.

The problem of contamination was addressed by a protocol which combines the template-switching effect of CapFinder with INVERSE PCR. In this approach, the cDNA, prepared as described above, was first amplified by PCR using the oligo d(T) and T-S oligo primers. Products from the PCR reaction were phosphorylated and then self-ligated (i.e. circularized). The circularized products were used as templates for inverse PCR – in which both primers were complementary to *known* (internal) sequences within the double-stranded cDNA; any random (non-specific) products would not be amplified by these primers because such products would lack the specific primer-binding sites. In this method, therefore, the products of inverse PCR include both the 5′ and the 3′ end-sequences of the (full-length) cDNA [BioTechniques (2006) 40(2):187–189].

(See also CAP TRAPPER and CAPSELECT.)

caprine (*adj*.) Of, or pertaining to, goats.

CAPS Cleaved amplified polymorphic sequences: a technique that has been used e.g. for TYPING isolates of bacteria and for detecting polymorphisms. The underlying principle relates to the ability of a polymorphism (for example, an SNP) to create or destroy the recognition sequence of a given RESTRICTION ENDONUCLEASE; in either case (i.e. whether a restriction site is created or destroyed) the relevant site may be scored as a CAPS marker. An SNP that gives rise to a CAPS marker has been termed a SNIP-SNP.

In one study, a CAPS approach was examined for its ability to type strains of *Staphylococcus epidermidis*. A selection of genomic target sequences were chosen and the corresponding sequences from each strain were amplified by PCR; each type of amplicon from each strain was then subjected to a panel of restriction enzymes. The results showed that most strains (33 out of 35) could be classified into five categories (referred to as CAPS types A–E) on the basis of CAPS markers detected in the amplicons.

As SNPs may produce CAPS markers it has been suggested that the latter may be useful as indicators for detecting SNPs.

capsduction Transfer of random sequences of (chromosomal or plasmid) linear dsDNA between strains of *Rhodobacter capsulatus* (*Rhodopseudomonas capsulata*) by a phage-like particle (*gene transfer agent*; GTA). No phage-specific DNA has been identified. Some strains of *R. capsulatus* can apparently receive DNA by capsduction without themselves becoming donors of GTAs. Capsduction has also been reported in the archaean *Methanococcus voltae*.

CapSelect A technique used for facilitating the preparation of *full-length* cDNAs from mRNAs [first described in: Nucleic Acids Res (1999) 27(21):e31]. As in the CAPFINDER method (q.v.), a short chain of deoxycytidines is added to the 3′ end of the growing FIRST STRAND following synthesis of the cap region on the mRNA template. However, template-switching (as in CapFinder) is not involved. Instead, the 3′ end of the first strand (the short chain of deoxycytidines) is extended by controlled ribonucleotide tailing with rATP; hence, the 3′ end of the first strand is modified to:

3′-rA.rA. . .dC.dC. . .

The modified 3′ end of the first strand is then ligated to a dsDNA adaptor which contains the 3′ overhang:

dT.dT. . .dG.dG. . .3′

in which the strand containing the overhang forms the primer for second-strand synthesis.

(See also CAP TRAPPER.)

capsid The protein *coat* or *shell* surrounding the nucleic acid genome (or surrounding the nucleoprotein core) of a virion; in some viruses the proteins of the capsid are arranged with icosahedral symmetry. In some viruses there is an outer (lipid or lipoprotein) *envelope* (*peplos*) that surrounds the capsid; the lipid components are apparently always derived from the host cell's membranes. (In viruses of the BACULOVIRIDAE the envelope of the BVs derives from the cytoplasmic membrane while the envelope of the ODVs is derived from the nuclear membrane.)

carbon nanotubes (in PCR) Carbon nanotubes – single-walled type (outside diameter 1–2 nm) and multi-walled type (outside diameter <8–30 nm) – have been reported to enhance the specificity and overall efficiency of LONG PCR using a 14-kb template of phage λ DNA. The mechanism of enhancement was not known, but suggestions included the promotion of heat equilibrium in the system and direct interaction between the nanotubes and components of the PCR reaction mixture

[BioTechniques (2008) 44(4):537–545].
(See also NANOPARTICLES.)

carbon source responsive element See CSRE.

carcinogen Any CANCER-inducing agent. Carcinogens include various types of low-molecular-weight chemical MUTAGEN; complex substances (e.g. asbestos in mesothelioma); physical agents, e.g. ULTRAVIOLET RADIATION and X-rays; and some types of virus (e.g. Epstein–Barr virus in Burkitt's lymphoma, hepatitis B virus in hepatocellular carcinoma).
(See also AMES TEST.)

carcinoma A type of CANCER which develops from epithelial cells (cf. SARCOMA).

CARD Catalyzed reporter deposition: see TYRAMIDE SIGNAL AMPLIFICATION.

carrier (*immunol.*) See HAPTEN.

carrier DNA DNA included in TRANSFORMATION experiments (i.e. in addition to 'experimental' DNA) in order to improve the efficiency of uptake.

carrier gene See GENE FUSION.

Cascade™ expression system A system (Active Motif) which is designed to provide tightly regulated expression of proteins in *Escherichia coli*, with low basal levels of expression prior to induction; the latter feature is important e.g. for producing toxic proteins (i.e. proteins that are toxic to the cells in which they are synthesized).

The system involves *two* transcriptional regulators, NahR and XylS2, which act sequentially. The initial induction of the system is achieved with salicylate – NahR then initiating transcription of XylS2. When produced, XylS2 is activated by the salicylate and it then induces transcription of the gene of interest. (See also SALICYLATE.)

In this system, the genes for NahR and XylS2 are located in the chromosome of an *ad hoc* expression strain of *E. coli*; the gene of interest can be either plasmid-borne or chromosomal.

The advantages of this system include the tightly controlled expression, minimal background expression, and an inducer (salicylate) which is much cheaper than e.g. IPTG (used for inducing *lac*-based systems) – so that large-scale work can be carried out more economically.

caspase See APOPTOSIS.

caspase-activated DNase See APOPTOSIS.

cassette (gene cassette) A DNA segment (usually encoding one or more genes) which behaves as a single unit when inserted into, or excised from, a larger molecule.
(See also INTEGRON.)

CAT Chloramphenicol acetyltransferase (see REPORTER GENE).

catabolite activator protein See CATABOLITE REPRESSION.

catabolite control protein A See CATABOLITE REPRESSION.

catabolite repression In bacteria – Gram-negative and Gram-positive: a phenomenon in which certain substrates (sources of carbon/energy) are used in preference to other substrates, even when the latter are freely available. For example, in the bacterium *Escherichia coli*, the LAC OPERON is not induced (i.e. lactose is not utilized), even in the presence of lactose, if glucose is available; this so-called *glucose effect* is also seen in other operons (e.g. the *ara* operon, involved in metabolism of the sugar arabinose).

Catabolite repression can operate by at least two distinct mechanisms – one characteristic of *E. coli* (and other Gram-negative, enteric bacteria) and one that has been found almost exclusively in Gram-positive species of bacteria. Both of the mechanisms are regulated by signals from the phosphoenol-pyruvate-dependent phosphotransferase system (see PTS), a transport system used by various Gram-positive and Gram-negative bacteria for the uptake (internalization) of a range of substrates.

Catabolite repression in Gram-negative, enteric bacteria

In these bacteria the key regulator molecule is a cytoplasmic protein: the IIA component of the *glucose permease* complex (see PTS). When glucose is *absent*, IIA∼P, the phosphorylated form of IIA, remains phosphorylated and – in this state – it activates the enzyme adenylate cyclase; this stimulates the synthesis of cyclic AMP (cAMP). cAMP forms a complex with the *cAMP-receptor protein* (CRP; = *catabolite activator protein*, CAP). This complex (cAMP-CRP) acts as a transcriptional activator, binding to the promoters of e.g. the *ara*, *lac* and other operons, permitting expression of these operons in the presence of their respective inducers; thus, e.g. in the presence of lactose (and in the absence of glucose), lactose can induce the *lac* operon.

When glucose is *present*, the *un*phosphorylated form of IIA binds to the (membrane-associated) permeases of lactose and certain other sugars; this inhibits uptake of the corresponding sugars. Because these sugars act as inducers of their respective operons, this phenomenon has been referred to as *inducer exclusion*.

Catabolite repression in Gram-positive bacteria

Certain operons in these bacteria are controlled by *operon-specific* regulator proteins. According to one scheme, the activity of these proteins depends not only on whether or not they are phosphorylated but also on *which* PTS protein was involved in their phosphorylation. In each regulator protein there are two copies of a specific phosphorylation site, each copy being designated a PRD (for PTS regulation domain).

A given PRD may be phosphorylated either by the phosphorylated IIB (IIB∼P) component *of the relevant permease* or by HPr∼P (both intermediates in the PTS). Phosphorylation by IIB∼P is inhibitory (causing the regulator to inhibit expression of the corresponding operon). Phosphorylation by HPr∼P tends to promote expression of the relevant operon.

In this model, an operon is expressed when its regulator is phosphorylated *only* by HPr∼P, i.e. in the absence of glucose (and in the presence of the specific inducer). An operon stays inactive if one of the PRD sites is phosphorylated by HPr∼P and the other is phosphorylated by IIB∼P (i.e. in the absence of both glucose and inducer). The rationale for this is that, for the permease controlling the uptake of a given substrate, the *absence* of the substrate (i.e. inducer) leaves IIB phosphorylated, and therefore able to phosphorylate a PRD site – thus inhibiting the corresponding operon; this is appropriate as, in

the absence of the inducer (substrate) there is no need for the given operon to be active. When the inducer is present, IIB is dephosphorylated by the inducer and so cannot phosphorylate a PRD site; if, concurrently, *glucose is absent*, HPr∼P will be available to phosphorylate a PRD, thus promoting expression of the relevant operon.

In certain Gram-positive bacteria catabolite repression may involve a mechanism in which an (ATP-dependent) enzyme, *HPr kinase*, phosphorylates HPr. HPr∼P, produced by HPr kinase-mediated phosphorylation, can form a complex with the *catabolite control protein A* (CcpA); this complex is able to control transcription in the target operons by binding to the appropriate regulatory site. In *Staphylococcus xylosus*, inactivation of *hprK* (the gene that encodes HPr kinase) was found to abolish this regulation in three catabolic enzyme systems.

The *bgl* operon in (Gram-negative) *E. coli*, which encodes a catabolic system involved e.g. in the metabolism of salicin, is regulated by a PRD-containing protein designated BglG. In the absence of a β-glucoside (e.g. salicin) BglG is phosphorylated by the IIB component of β-glucoside permease (BglF), thus blocking transcription of the *bgl* operon.

(See also CSRE.)

catalytic antibody (abzyme) An antibody which has catalytic activity.

catalyzed reporter deposition See TYRAMIDE SIGNAL AMPLIFICATION.

catenane A complex consisting of two or more circular molecules of nucleic acid interlocked like the links in a chain.

(cf. CONCATEMER, CONCATENATE.)

CBP (1) CRE-binding protein: see CREB PROTEIN.

(2) CALMODULIN-binding peptide: a feature of the AFFINITY PROTEIN EXPRESSION AND PURIFICATION system (q.v.).

CBPs Cap-binding proteins: see CAP.

cccDNA Covalently closed circular DNA – single-stranded or double-stranded DNA which has no free 5′ or 3′ ends.

CCD Charge-coupled device.

***ccd* mechanism** One form of POST-SEGREGATIONAL KILLING, encoded by the F PLASMID, that promotes stable maintenance of the plasmid in a bacterial population. The genes *ccdB* (previously named *letB*) and *ccdA* (*letA*) encode a lethal toxin (CcdB) and its antidote (CcdA), respectively. CcdB is stable, but CcdA is slowly degraded by the Lon protease of the host cell; hence, CcdA can protect against CcdB only when it is being synthesized in (plasmid-containing) cells. Plasmid-free daughter cells are killed by CcdB (derived from the parent cell) because they lack the (plasmid-borne) *ccdA* gene from which the antidote can be synthesized.

The designation *ccd* is derived from 'coupled cell division' or from 'control of cell death' (according to different authors).

If used *in vitro* for ensuring the continued presence of the F plasmid in a bacterial strain, the *ccd* mechanism is reported to break down eventually, yielding plasmid-free cells. To avoid this problem of plasmid instability, the *ccd* system has been engineered so that the gene encoding the toxin resides on the *bacterial chromosome* and the antidote-encoding gene stays in the plasmid. With this arrangement, plasmid-less cells are automatically killed by the toxin [BioTechniques (2005) 38 (5):775–781].

CCF2 substrate See FRET.

CcpA See CATABOLITE REPRESSION.

***ccr* genes** (in *Staphylococcus aureus*) See SCCMEC.

CCR3 (in HIV-1 infection) See HIV-1.

CCR5 (in HIV-1 infection) See HIV-1.

(See also CHEMOKINE CORECEPTOR ANTAGONISTS.)

[Overexpression of CCR5 in *Escherichia coli*: PLoS ONE (2009) 4(2):e4509].

CD11a/CD18 (leukocyte function-associated antigen, LFA-1) A cell-surface molecule of the β_2 INTEGRIN group which occurs on leukocytes. On activation, during inflammation, this molecule binds strongly to ligands, such as ICAM-1, on vascular endothelium. Individuals with leukocyte adhesion deficiency, an autosomal recessive disorder, are characterized by inadequate β_2 integrins and an abnormal susceptibility to certain infectious diseases.

CD11b/CD18 (CR3; Mac-1) A cell-surface molecule of the β_2 INTEGRIN group which occurs on certain leukocytes and is involved e.g. in adhesion reactions.

CD11c/CD18 (CR4) A cell-surface protein of the β_2 INTEGRIN group found e.g. on macrophages and neutrophils; it binds components of the complement system.

CD11d/CD18 A protein of the β_2 INTEGRIN group, found e.g. on macrophage foam cells, which is reported to bind multiple ligands [Blood (2006) 107(4):1643–1650].

CD21 A cell-surface molecule found e.g. on epithelial cells of the human oropharynx and on B lymphocytes (B cells). The molecule acts as a receptor for the EPSTEIN–BARR VIRUS.

CD27 A cell-surface molecule which is used as a marker for (human) memory B lymphocytes (memory B cells).

CD34 A cell-surface antigen on hemopoietic stem cells.

CD38 A transmembrane glycoprotein (ecto)enzyme expressed on various types of cell, including B cells and T cells; as well as being a cell-surface receptor, it has ADP-ribosyl cyclase activity. In chronic lymphocytic leukemia it is expressed on a subset of leukemic cells; high levels of CD38 expression are associated with a poor clinical outcome.

(See also ONCOLYTIC VIRUS.)

CD40 A cell-surface molecule on B lymphocytes (B cells) that interacts with the CD40L ligand on T cells, thus triggering an important phase of development in the B cell.

(See also CD40L and HYPER-IGM SYNDROME.)

CD40L CD40 ligand: a cell-surface molecule on T lymphocytes (T cells) that interacts with CD40 on B cells; such interaction promotes development of a normal antibody response in the B cells and is also a factor in the development of the T cells.

(See also CD40 and HYPER-IGM SYNDROME.)

CD81 A cell-surface protein (of the tetraspanin family) found in human cells. CD81 acts as a receptor for the E2 envelope glycoprotein of hepatitis C virus.

Factors other than CD81 are apparently also required for

infection by the hepatitis C virus [J Virol 2006) 80(10):4940–4948].

CD95 (APO-1) The cell-surface Fas receptor: see APOPTOSIS.

CD209 *Syn*. DC-SIGN (q.v.).

CD209L (L-SIGN) A C-type lectin reported to act as a receptor for the SARS virus.

Cdc2 kinase (lamin kinase) See NUCLEAR LAMINA.

Cdc6 protein (in archaeans) See DNAA GENE.

cdc25 (in *Saccharomyces cerevisiae*) See e.g. CYTOTRAP TWO-HYBRID SYSTEM.

cDNA Copy DNA (sometimes called complementary DNA): a dsDNA or ssDNA copy of an RNA molecule – commonly a mature, poly(A)-tailed mRNA (see also FIRST STRAND).

Note. C-DNA refers to a particular conformation of DNA; it is unrelated to cDNA.

Some methods for producing cDNA can be inefficient. One problem is that reverse transcriptase may have exonuclease activity (i.e. it may degrade a newly synthesized strand); this problem may be overcome e.g. by using a recombinant form of Moloney murine leukemia virus (MMLV) reverse transcriptase which has no exonuclease activity.

In some methods, priming of the second strand depends on the 3′ end of the first strand folding back on itself (to form a hairpin structure); the 3′ end is then extended – using the first strand as template. Because the hairpin's (single-stranded) loop has to be enzymically removed (e.g. with endonuclease S_1), this inevitably leads to a loss of part of the 5′ sequence of the mRNA.

Full-length cDNAs (i.e. those containing the *entire* coding sequence) are an important requirement for the preparation of adequate cDNA libraries. Because the (5′) CAP sequence of a eukaryotic mRNA is a good marker for the presence of the 5′ end of the mRNA, various methods have been devised which depend on copying this structure during the synthesis of (full-length) cDNAs: see e.g. CAPFINDER and CAPSELECT.

Full-length cDNAs are often obtained by the CAP TRAPPER method. [Development of 5006 full-length cDNAs in barley: DNA Res (2009) 16(2):81–89; studies on adaptive evolution between wheat and rice: BMC Genomics (2009) 10:271.]

cDNA library See LIBRARY.

CDP™ A CHEMILUMINESCENCE-generating substrate which is cleaved by ALKALINE PHOSPHATASE.

CECF protein synthesis Continuous-exchange cell-free protein synthesis: see CELL-FREE PROTEIN SYNTHESIS.

cefepime See CEPHALOSPORINS.

cefoxitin See CEPHAMYCINS.

ceftazidime See CEPHALOSPORINS.

ceftriaxone See CEPHALOSPORINS.

CEL I A member of the SINGLE-STRAND-SPECIFIC NUCLEASE group.

(See also CEL I NUCLEASE MISMATCH ASSAY.)

CEL I nuclease mismatch assay An assay used e.g. to indicate the frequency of mutation achieved in a pool of cells which has been subjected to targeted (gene-specific) mutagenesis/knock-out. Genomic DNA is isolated from the mutagenized cells, and the targeted sequence is then amplified by PCR. The resulting amplicons will contain a mixture of wild-type (i.e. non-mutagenized) and mutant sequences. On melting and re-annealing these sequences, mismatches occur between wild-type and mutant strands, forming heteroduplex DNA; hence, localized unpaired (single) strands will occur at the site of the mutation in the heteroduplex molecules. The mixture is then subjected to the single-strand-specific nuclease CEL I which cuts a single strand at the 3′ end of the single-stranded region in each heteroduplex molecule. The resulting products (the cleaved + intact strands) are subjected to gel electrophoresis and quantitated densitometrically. Comparison of the relative intensity of the cleavage products with that of the uncleaved products gives an indication of the frequency of mutagenesis in the given gene/sequence [Proc Natl Acad Sci USA (2008) 105(15):5809–5814].

cell-free protein synthesis ('coupled transcription–translation') Any of various forms of *in vitro* (i.e. cell-free) system used for the transcription and translation of a given recombinant protein from a specific template introduced into the system.

Commonly used commercial and non-commercial systems are often based on lysates of *Escherichia coli* that contain all the molecular machinery needed for transcription and translation of a linear or circular template added to the system.

A cell-free system necessarily includes components such as ribosomes, tRNAs and transcription factors, and it commonly includes RNA polymerase from bacteriophage T7; the gene/fragment of interest is accordingly flanked by a T7 promoter and a terminator sequence.

Quantitative studies on polysomes in cell-free protein synthesis systems have indicated that the rates of synthesis could be improved e.g. by adding purified elongation factors to the reaction mixture [Biotechnol Bioeng (2005) 91(4):425–435].

The activity (functionality) of proteins produced in cell-free systems may be suboptimal, and it was hypothesized that this may be caused e.g. by poor coupling between transcription and translation due to the rapid activity of the T7 polymerase (as compared with the native bacterial enzyme). The situation was found to be improved by using a slower T7 enzyme or carrying out the process at 20°C [Nucleic Acids Res (2006) 34(19):e135].

Studies on translation of luciferase mRNA in a *eukaryotic* cell-free system found that a rise in the initiation rate results from addition of extra ribosomes to the polysome [Nucleic Acids Res (2007) 35(19):6547–6559].

Yields of up to several grams/liter of protein are reported to be obtained under optimal conditions from cell-free systems.

The so-called 'P-gel' system of cell-free protein production is based on a DNA hydrogel in which genes form part of the gel scaffolding. The yield was reported to be up to 5 mg/mL, the high efficiency being attributed to e.g. improved stability of the genes and higher local enzyme concentrations [Nature Materials (2009) 8:432–437].

Advantages of cell-free protein synthesis

- Speed. There is no requirement for initial transfection and

selection, cell culture etc.

- Potential to focus the synthetic capacity of the system on the target template.
- Ability to experiment more freely with conditions and e.g. to use amino acid analogs.
- Ability to synthesize proteins which would be toxic/lethal if produced intracellularly.

[Studies on translation of polysomal mRNA: Nucleic Acids Res (2008) 36(8):2476–2488.]

[Cell-free production of (human) olfactory receptors: Proc Natl Acad Sci USA (2008) 105(41):15726–15731.]

Commercial systems for cell-free protein synthesis include the Expressway™ Plus Expression System from Invitrogen (Carlsbad CA) that uses pEXP-DEST vectors; these plasmids contain *att* sites and are accordingly accessible as destination vectors within the GATEWAY SITE-SPECIFIC RECOMBINATION SYSTEM. The vectors include a T7 promoter and terminator; they also contain sequences encoding tags (e.g. XPRESS, SIX-HISTIDINE TAG) which facilitate purification of the expressed protein.

The pEXP3-DEST vector (4.6 kb) includes a sequence that encodes a LUMIO tag which (used with the Lumio™ detection reagent) permits real-time detection of protein synthesis in the cell-free system. The specialized lysate used with this vector is derived from a *slyD* mutant of *Escherichia coli*; this avoids *non-specific* binding of Lumio™ reagent to the cell's SlyD protein and, hence, optimizes background (by reducing background signal) for detecting the given tag–protein fusion product.

[Use of Expressway™ (example): BMC Biotechnol (2008) 8:58.]

In *continuous-exchange cell-free* (CECF) protein synthesis the reaction is maintained by a continuous influx of reagents and efflux of byproducts across a dialysis membrane. While expression systems commonly involve a circular plasmid, it was reported that a linear (PCR-generated) template can give satisfactory results with an RNase E-deficient and chaperone-enriched extract; it was suggested that deficiency of RNase E may enhance the protein yield by avoiding a major loss of mRNA [BioTechniques (2008) 44(3):387–391].

[Use of a CECF kit (Roche) for *in vitro* translation of transcription factors: Plant Physiol (2009) 150(3):1286–1296.]

(See also POLYSOME.)

cell fusion (somatic cell hybridization) *In vitro* coalescence of cells, forming a single hybrid cell in which the nuclear and cytoplasmic contents of both cells are enclosed by a single cytoplasmic membrane.

In one method, populations of both types of cell are mixed and centrifuged. The pellet is then exposed briefly to an agent (*fusogen*) such as polyethylene glycol (PEG) which promotes cell fusion. (Lysophosphatidylcholine or inactivated Sendai virus may be used in place of PEG.) Free fusogen is removed by dilution, and the cells are resuspended in growth medium and incubated.

Cell fusion is used e.g. in the preparation of a HYBRIDOMA.

cell lineage tracking (with reporter genes) Tracking (identifying) the descendants of *particular* blastomeres (see BLASTOCYST) may be achieved by a fluorescence-based method with a suitably transgenic animal.

One study prepared transgenic mice with a double-reporter system containing (i) an (expressed) *loxP*-flanked gene (see entry CRE–LOXP SYSTEM) encoding a red fluorescent protein, and (ii) (downstream) a (non-expressed) gene which encoded a green fluorescent protein. The embryos derived from this mouse line can be manipulated at the preimplantation stage. Thus, a particular blastomere can be transfected with a Cre-expression vector which, on expressing the Cre recombinase, causes excision of the gene encoding red fluorescent protein. This blastomere, *and its descendants*, thus fail (permanently) to express the red fluorescent protein but, after removal of the gene for red fluorescent protein, are able to express the gene for green fluorescent protein – thus distinguishing them from descendants of the non-manipulated (non-transfected) blastomeres.

The above study also developed an alternative procedure in which transfection of a given blastomere resulted in a switch from green to red fluorescence [J Biomed Biotechnol (2009) doi: 10:1155/2009/985140].

cell proliferation assay Any assay which monitors an increase in the numbers of specific type(s) of cell. Such assays are carried out e.g. in order to determine whether or not cells are proliferating at a normal rate and/or to check the effect of a toxic (e.g. growth-inhibitory) agent on the cell population.

In one form of assay the cells are grown in a medium which contains BRDU; the BrdU becomes incorporated in the cells' DNA. Cell proliferation during this period of growth is monitored by measuring the amount of BrdU incorporated in the DNA in a given time – the amount of BrdU increasing with increasing numbers of cells. (This approach is also used for monitoring DNA synthesis.) The BrdU is monitored by first fixing and permeabilizing the cells, denaturing the DNA, and then exposing the whole preparation to (fluorophore-labeled) anti-BrdU antibodies. The BrdU can then be quantitated by a fluorescence-based technique.

For an alternative method that is simpler and more sensitive see CLICK-IT EDU.

cell replacement therapy See STEM CELL.

cell sorter See entries FLOW CYTOMETRY and FLUORESCENCE-ACTIVATED CELL SORTER.

cell wall sorting signal See SORTASE.

CENP-A, CENP-B etc. See KINETOCHORE.

centiMorgan (cM) A unit of distance on a chromosome: the distance between two loci when there is a 1% probability of those loci being separated by recombination during meiosis.

(See also MAP UNIT.)

centromere In a chromosome: the region of attachment of two CHROMATIDS; it is essential for chromosome–spindle attachment during mitosis.

(Relevant entries: ACROCENTRIC, ARS, HUMAN ARTIFICIAL CHROMOSOME, KINETOCHORE, METACENTRIC, YEAST ARTIFICIAL CHROMOSOME,

YEAST CENTROMERE PLASMID.)

(See also ICF SYNDROME.)

cephalosporins A category of β-LACTAM ANTIBIOTICS which include e.g. cefepime, cefixime, ceftazidime, ceftriaxone and cephalothin; the molecular structure is characterized by the β-lactam ring fused to a dihydrothiazine ring.

Note. In the literature, the names of some cephalosporins are spelt with either an 'f' or a 'ph'.

(cf. CEPHAMYCINS.)

cephalothin See CEPHALOSPORINS.

cephamycins A category of β-LACTAM ANTIBIOTICS: the 7-α-methoxycephalosporins; compared with the cephalosporins, cephamycins (e.g. cefoxitin) tend to have better penetration of the Gram-negative outer membrane, and to have increased resistance to some β-lactamases.

c-*erbB* See EPIDERMAL GROWTH FACTOR RECEPTOR FAMILY.

CFLP analysis CLEAVASE FRAGMENT LENGTH POLYMORPHISM ANALYSIS.

CFP-10 See ESAT-6.

CFPS CELL-FREE PROTEIN SYNTHESIS.

CFTR CYSTIC FIBROSIS transmembrane conductance regulator.

cfu Colony-forming unit – often refers to a single bacterial cell which is capable of giving rise to a colony on an appropriate medium and under suitable conditions.

CGH COMPARATIVE GENOMIC HYBRIDIZATION.

chain of infection (*epidemiol*.) See the entry TYPING (section: *The uses of typing*).

chain-terminating codon See NONSENSE CODON.

chain-termination method (DNA sequencing) The DIDEOXY METHOD.

Champion™ pET SUMO vector A 5.6-kb expression vector (Invitrogen, Carlsbad CA) designed for high-level expression of recombinant proteins in *Escherichia coli*. It includes a sequence encoding an 11-kDa solubility-enhancing SUMO protein which forms a fusion product with the target protein; following expression, the fusion protein can be cleaved at the C-terminal of the SUMO moiety by a SUMO protease (a recombinant form of the Ulp1 SUMO protease). (See the entry SUMOYLATION for the background on SUMO technology.) Cleavage of the fusion protein with this SUMO protease is effective except for those proteins whose terminal residue is leucine, lysine, proline or valine.

The plasmid vector also includes a T7 promoter and a tag encoding a series of six (consecutive) histidine residues; this tag facilitates purification of the target protein (see NICKEL-CHARGED AFFINITY RESIN for a rationale). The (linearized) vector has single-nucleotide (thymidine) 3′ overhangs for use with *Taq* DNA polymerase-generated PCR fragments (see EXTENDASE ACTIVITY and TA CLONING KITS for a rationale).

[Uses (e.g.): RNA (2008) 14(6):1026–1036; PLoS Pathog (2009) 5(1):e1000259; and PLoS ONE (2009) 4(4):e5227.]

(See also SOLUBILITY ENHANCEMENT TAG.)

Champion™ pET104-DEST vector A destination vector used in the GATEWAY SITE-SPECIFIC RECOMBINATION SYSTEM – see the table in that entry.

chaotrope Any agent which affects interactions between water molecules and which is able to promote transfer of non-polar substances to a hydrophilic medium. The chaotropes include thiocyanide, iodide, bromide and chloride ions – and certain compounds (e.g. urea); they are often used for disrupting cell membranes, for solubilizing proteins, and also for denaturing nucleic acids.

chaperone (1) (of proteins) See OVEREXPRESSION (the section: *Inclusion bodies*).

(2) (RNA) See RNA CHAPERONE ACTIVITY; see also FINOP SYSTEM and Hfq protein in the entry SRNAS.

(3) (histone) See e.g. *nucleophosmin* in the entry ACETYLATION (of histones).

Charcot–Marie–Tooth disease (CMT disease) Any of a group of heterogeneous inherited neuropathies which include both demyelinating and axonal types; autosomal dominant, autosomal recessive and X-linked cases are recognized.

Type 1A demyelinating CMT is associated with abnormality of the gene encoding peripheral myelin protein-22 located at 17p11.2, while type 1B demyelinating disease is associated with mutation in the gene for myelin protein zero at 1q22. X-linked CMT disease (CMTX) is reported to exhibit features of both the demyelinating and axonal types.

The type 2A disease is linked to mutations in the mitofusin 2 gene (*MFN2*) [BMC Med Genet (2006) 7:53].

ChargeSwitch® technology A method used for isolating and purifying DNA (Invitrogen, Carlsbad CA, USA). On release from the sample, DNA binds to minute, charged beads. This binding occurs because DNA carries a negative charge and the beads develop a positive charge when the pH (buffer) is <6.5. After two washes in a water-based buffer (during which time the beads are held by magnetic force) DNA is released from the beads by elution buffer (pH 8.5) which abolishes the mutual attraction between beads and DNA.

A ChargeSwitch® forensic DNA purification kit has been reported to recover ~50 to 150 ng of DNA (measured by the PICOGREEN and fluorometric techniques) from 48 specimen cigarette butts. (See also FORENSIC APPLICATIONS.)

[Uses (e.g.): Appl Environ Microbiol (2008) 74(18):5741–5749; J Clin Microbiol (2008) 46(9):3116–3118; PLoS Genet (2009) 5(3):e1000438; Retrovirology (2009) 6:57.]

ChargeSwitch® technology is also used for isolating RNA [e.g. Vaccine (2008) 26(8):1136–1141; J Virol (2009) 83(1):140–149].

charomid A COSMID containing a 'spacer' sequence – which differs, in length, in different charomids. In each charomid, a POLYLINKER permits the insertion of target DNA such that the total length of {spacer+target DNA} makes the charomid a suitable length for *in vitro* packaging – as it is then between the upper and lower size limits for packaging in the phage head; hence, the choice of a particular charomid (containing a spacer of given length) depends on the length of the target DNA which is to be inserted in the charomid.

Charomids were first used in the 1980s [see e.g. Proc Natl Acad Sci USA (1986) 83(22):8664–8668].

[Use (e.g.): J Bacteriol (2008) 190(8):3083–3087 and BMC Genomics (2008) 9:129.]

Charon vector A CLONING VECTOR, derived from the phage λ genome, which lacks the ability to establish LYSOGENY, thus avoiding integration into the chromosome of the host cell.

cHDA Circular helicase-dependent amplification: an isothermal approach to NUCLEIC ACID AMPLIFICATION in which the strand separation in a *circular* dsDNA template is achieved by the highly processive helicase of bacteriophage T7 – extension of sequence-specific primers by a ROLLING CIRCLE mechanism resulting in concatemeric products. cHDA also exploits the DNA polymerase of phage T7 (Sequenase®) and the phage's single-strand binding protein (gp2.5).

Amplification, reported to be up to 10000-fold, is achieved at 25°C.

The procedure may be used e.g. with a purified preparation of plasmids or with a crude cell lysate.

[Method: Nucleic Acids Res (2006) 34(13):e98.]

CHEF See PFGE.

chemical cleavage method (of DNA sequencing) See MAXAM–GILBERT METHOD.

chemiluminescence Light produced from a chemical reaction. One example is BIOLUMINESCENCE; other examples include light generated from certain types of *in vitro* reaction which involve enzymic cleavage of chemicals of non-living origin.

Chemiluminescence can be exploited in reporter systems in place of fluorescence (in various applications). The principle underlying all of these chemiluminescent reporter systems is the enzyme-mediated decomposition of certain types of substrate which, when cleaved, emit light. A high signal-to-noise ratio can be achieved owing to the low background. (See also CHEMILUMINESCENCE ENHANCEMENT.)

On membranes, chemiluminescence can be used for detecting specific biomolecules. For example, a nucleic acid PROBE which is labeled with BIOTIN (q.v.) can be detected (following hybridization to its target sequence) by incubating the membrane with a conjugate which consists of STREPTAVIDIN bound to ALKALINE PHOSPHATASE; the addition of a chemiluminescent substrate then yields a light signal at the location of the probe, thus indicating the specific target biomolecule.

Alternatively, a probe can be labeled with DIGOXIGENIN; detection is then achieved with a conjugate consisting of anti-DIG (antibody to digoxigenin) bound to alkaline phosphatase and subsequent exposure of the membrane to a chemiluminescent substrate.

A probe may also be labeled (directly) with alkaline phosphatase.

In DNA SEQUENCING, use can be made of biotinylated sequencing primers; following gel electrophoresis, the fragments are transferred to a nylon membrane and then visualized by a streptavidin–AP conjugate and a chemiluminescent substrate.

Substrates that emit light when cleaved by AP include the 1,2-dioxetanes CSPD®, AMPPD® and CDP^{TM}.

Enzymes other than AP are also used in chemiluminescent reactions. The reporter enzyme β-galactosidase is detected in cell extracts by using the substrate Galacton^{TM}.

The reporter enzyme β-glucuronidase can be detected in cell extracts by using the substrate Glucuron^{TM}.

Chemiluminescent reporting can also be used for proteins on Western blots. For example, a given target protein can be initially bound by its antibody – which, in turn, is bound by an anti-Ig (anti-immunoglobulin) antibody conjugated to AP and exposed e.g. to CSPD®.

Examples of the use of chemiluminescent systems

Studies on varicella-zoster virus involving DIG-labeled riboprobes detected with an anti-DIG–AP conjugate and CSPD® [J Virol (2006) 80(7):3238–3248]; detection of target protein with primary antibody, an AP-conjugated secondary antibody and CSPD® [BMC Cancer (2006) 6:75]; studies on Archaea-specific chemotaxis proteins using DIG-labeled probes with detection by CSPD® [BMC Microbiol (2009) 9:56] and non-radioactive tracking assay for detecting chloroquine-resistant mutants of *Plasmodium falciparum* using DIG-labeled probes and CSPD® [Malaria J (2009) 8:47].

(See also FLASH CHEMILUMINESCENT GENE MAPPING KIT.)

chemiluminescence enhancement A reduction in the (water-mediated) quenching of a chemiluminescent signal by addition of certain macromolecules ('enhancers') that effectively shield the site of chemiluminescence from water molecules. Enhancers can also shift the wavelength of the emitted light.

[Use of Sapphire-II^{TM} chemiluminescence enhancer: BMC Genetics (2006) 7:18.]

chemokine coreceptor antagonists (anti-HIV-1) ANTIRETROVIRAL AGENTS which bind to certain cell-surface chemokine coreceptors, blocking infection by strains of HIV-1 for which those coreceptor(s) are essential binding sites (see the entry HIV-1).

The chemokine coreceptor antagonist maraviroc is used for treating infections involving CCR5-tropic strains of HIV-1.

chi (χ) site See the entry HOMOLOGOUS RECOMBINATION.

chimera (chimaera) Any *hybrid* individual, cell or construct – formed from at least two genetically distinct sources. *Mosaic* is sometimes used as a synonym.

(See also TRANSGENESIS.)

chimeraplast (RDO: RNA–DNA chimeric oligonucleotide) A double-stranded nucleic acid structure – containing two hairpin ends – which develops by the folding (self-hybridization) of a single strand that includes stretches of RNA and DNA.

Chimeraplasts have been used for inducing single-base-pair changes at specified sites in the genomes of e.g. mammalian and plant cells.

Each chimeraplast includes a sequence which – except for a single mismatch – is homologous to the target sequence in the given genome. When a chimeraplast aligns with its target sequence, an endogenous repair process takes place at the site of the mismatch, leading to a single base change in the genomic DNA.

It has been suggested that such a system may have potential use in GENE THERAPY for correcting point mutations.

[Sequence details of a chimeraplast (example): see Figure 4

in Reprod Biol Endocrinol (2003) 1:83.]

[RDO for editing the *DMD* gene: Hum Mol Genet (2005) 14(2):221–233.]

chip (DNA chip) A small piece of glass (or other material) on which the components of a MICROARRAY or other test system are immobilized; the immobilized components may consist of e.g. oligonucleotides, strands of PNA, or sections of tissue. (See MICROARRAY for a description of some immobilization procedures; see also DENDRICHIP.)

ChIP See CHROMATIN IMMUNOPRECIPITATION.

CHIP 'C-terminus of Hsp70-interacting protein' – referring to a UBIQUITIN ligase that is involved in the polyubiquitinylation and degradation of proteins (see the entry PROTEASOME).

The activity of CHIP appears to depend on homodimerization via the U-box domain.

CHIP was reported to have a role in regulating oncogenic pathways [Nature Cell Biol (2009) doi: 10.1038/ncb1839].

ChIP-chip See CHROMATIN IMMUNOPRECIPITATION.

ChIP-PET See CHROMATIN IMMUNOPRECIPITATION.

ChIP-seq (ChIP-sequencing) See CHROMATIN IMMUNOPRECIPITATION.

Chlamydia A genus of Gram-negative, obligately intracellular parasitic/pathogenic bacteria found in man and other animals; species of *Chlamydia* give rise to diseases such as inclusion conjunctivitis and trachoma in man, and to enteric diseases in young domestic animals. These bacteria are usually sensitive to tetracyclines.

The cells are non-motile and coccoid, ranging in size from about 0.2 μm to 1.5 μm – depending on the stage within the developmental cycle (see below). The cell envelope appears to contain little or no peptidoglycan but does include an outer membrane similar to that of other Gram-negative bacteria.

The GC% of the genomic DNA is 41–44. (See also pCT in the table for the entry PLASMID.)

The type species of the genus is *C. trachomatis*.

Chlamydia cannot be grown in the usual laboratory media; these organisms are grown in animals, in the chick embryo (yolk sac), and in cell cultures.

Chlamydia undergoes a development cycle. The form of the organism which is infectious for host cells is designated the *elementary body* (EB); this is a small (∼0.2–0.4 μm) cell that does not divide and which has few ribosomes. Much of the cell's interior consists of condensed nuclear material. The EB binds to a host cell and is taken up by endocytosis; phagosomes containing EBs do not fuse with lysosomes. EBs later differentiate to become so-called *reticulate bodies* (RBs); the RBs are larger (∼0.6–1.5 μm), are metabolically more active, and they contain more ribosomes. Some 10 hours after initial infection RBs undergo binary fission; cell division continues, and from ∼20 hours after the initial infection the RBs start to give rise to EBs. The cycle is complete by about 48–72 hours – when each infected host cell contains a population of EBs that constitutes a so-called *inclusion* that may occupy a major part of the cell's interior.

Chlamydophila pneumoniae See NASBA (uses).

chloramphenicol (chloromycetin) A broad-spectrum antibiotic produced by the bacterium *Streptomyces venezuelae* and also prepared synthetically; it binds to prokaryotic and mitochondrial ribosomes, inhibiting the activity of peptidyltransferase and, hence, inhibiting protein synthesis. Chloramphenicol is bacteriostatic for a wide range of bacteria (Gram-positive and Gram-negative), and this antibiotic can also inhibit vegetative growth in some fungi.

Bacterial resistance to chloramphenicol is often due to the expression of a (plasmid-encoded) enzyme, chloramphenicol acetyltransferase (CAT), which catalyzes acetylation of the antibiotic in an acetyl-CoA-dependent way. Plasmid-encoded CAT is typically produced constitutively in Gram-negative bacteria but is inducible e.g. in species of the (Gram-positive) bacteria *Staphylococcus* and *Streptococcus*.

The pathogen *Pseudomonas aeruginosa* is innately resistant to chloramphenicol.

Chloramphenicol is antagonistic to those antibiotics – such as penicillins – which act only against growing and dividing bacteria.

chloramphenicol acetyltransferase See CHLORAMPHENICOL; see also REPORTER GENE.

chloromycetin *Syn*. CHLORAMPHENICOL.

***p*-chlorophenylalanine** A lethal substrate which has been used e.g. for the elimination of a SUICIDE VECTOR (q.v.).

CHO cells Cells derived from Chinese hamster ovary.

Christmas factor Blood-clotting factor IX; a deficiency in this factor is a (less common) cause of hemophilia (hemophilia B) (see also 'Hemophilia' in the entry GENETIC DISEASE (table)).

The product Benefix® is a recombinant form of factor IX that is used for the management of this condition; this protein is produced in Chinese hamster ovary (CHO) cells.

(See also PYROPHOSPHOROLYSIS-ACTIVATED POLYMERIZATION.)

chromatid Either of the two products formed when a (eukaryotic) chromosome replicates. *Sister* chromatids, derived from the same chromosome, are connected at the centromere.

(See also KINETOCHORE.)

ChromaTide™ nucleotides A variety of (fluorophore-labeled) nucleotides (Molecular Probes, Eugene OR, USA) which are used e.g. for direct labeling of probes (see PROBE LABELING).

[Use (examples): Genetics (2008) 179(4):2173–2182; BMC Microbiol (2009) 9:1.]

chromatin In a eukaryotic nucleus: the nucleoprotein complex which includes DNA, histones and non-histone proteins (see also CHROMOSOME and MATRIX-ATTACHMENT REGION).

The following is a generalized account.

Fully unraveled, chromatin is seen as fibers of DNA (∼10 nm diam.) with a regular beaded appearance. The beads are *nucleosomes*. Each nucleosome has a *core* consisting of two of *each* of the following types of HISTONE: H2A, H2B, H3 and H4; this structure is encircled by several turns of DNA (∼146 bp). The DNA enters and exits each nucleosome via histone H1. The nucleosomes are connected by *linker DNA* of variable length. (If treated with an endonuclease, the

beaded fibers yield *chromatosomes*, each consisting of a nucleosome and some associated DNA.)

The location of nucleosomes within the genomic DNA may be signaled by the so-called A-TRACTS (q.v.).

In vivo, the beaded (nucleosomal) fibers form superhelical chains: 30 nm fibers (*solenoids*) whose stabilization involves histone H1. Looped structures of nucleosomal and solenoidal DNA are anchored at intranuclear attachment sites.

Two basic types of chromatin may be distinguished in the interphase nucleus: (i) the highly condensed *heterochromatin* and (ii) *euchromatin*, a more loosely packed and less dense form. The heterochromatin may be generally inactive, while euchromatin seems to equate with active, transcribed regions associated with gene expression.

For transcription, replication and repair, the DNA must be accessible to enzymes and binding proteins etc.; chromatin must therefore regularly undergo modification or *remodeling*. Reported mechanisms include: modification of the properties of histones, post-translationally, e.g. by ACETYLATION (q.v.) ADP-RIBOSYLATION, phosphorylation and/or ubiquitylation; the mechanical displacement of nucleosomes along the DNA (i.e. 'relocation') by (ATP-dependent) remodeling complexes; incorporation of *variant* forms of HISTONES, modulating the properties of the nucleosomes.

In some cases, genes repressed (i.e. transcriptionally silent) by methylation in their promoter regions have been activated by the joint action of an inhibitor of methyltransferase (such as the nucleoside analog 5-AZA-2′-DEOXYCYTIDINE) and an inhibitor of histone deacetylases (such as trichostatin A). (See also METHYLATION.)

chromatin body *Syn*. BARR BODY.

chromatin immunoprecipitation (ChIP) A procedure used for investigating protein–DNA interactions, e.g. the interactions between transcription factors and their binding sites in DNA. Initially, a formalin-based protein–DNA cross-linking step is carried out with the *intact* cells; chromatin is then extracted from the cells and the DNA is sonicated into fragments of several hundred base-pairs in length. Antibodies, specific to a given protein of interest, are then used in order to precipitate the given protein *together with its bound* (*i.e. cross-linked*) *DNA*. When the DNA is isolated from the protein, the DNA fragment can be e.g. amplified and sequenced (see below).

To ensure that precipitation of a given target protein occurs optimally it is essential to use antibodies that bind effectively to the target protein; antibodies which bind well to the target protein in other contexts (e.g. IMMUNOBLOTTING) may not do so in ChIP owing e.g. to changes that may occur in the protein through the formalin-based cross-linking procedure. Whether or not a particular antibody will bind efficiently to the target protein in the ChIP procedure can be assessed by experimentation. In one approach, the protein of interest was expressed, within appropriate cells, as a *fusion protein* with the DNA-binding domain of GAL4. ChIP was then carried out – first with anti-GAL4 antibodies, and then with a candidate antibody against the target protein. Two different candidate antibodies (that were both effective in immunoblotting) were tested separately by ChIP; one was found to be markedly less effective than the other [BioTechniques (2008) 44(1):66–68].

The sensitivity of ChIP is reported to be improved by biotinylating the target proteins *in vivo*, that is, prior to treatment with formalin; this procedure is outlined in the entry IN VIVO BIOTINYLATION. As the biotinylated proteins can be captured and held securely by a STREPTAVIDIN-based system, biotinylation allows stringent washing conditions – thus leading to a better signal-to-noise ratio [Nucleic Acids Res (2006) 34: e33]. [Optimal use of tandem biotin and V5 tags in chromatin immunoprecipitation: BMC Mol Biol (2009) 10:6.]

The commonly used agent for DNA–protein cross-linking – formaldehyde – is appropriate for investigating proteins that are bound *directly* to DNA (such as histones), but it may not be as effective for investigating proteins which are *indirectly* associated with DNA (such as transcriptional co-activators). Studies on a number of compounds have found that ethylene glycol-bis-(succinimidyl succinate) (EGS) – which has a long 'spacer arm' – is an effective cross-linking agent; a study using both EGS and formaldehyde reported sequential cross-linking by these two agents, enabling the analysis of multiple transcriptional cofactors that cannot be studied satisfactorily by the (conventional) use of formaldehyde as the sole agent [BioTechniques (2006) 41(6):694–698].

The availability of ChIP technology enabled genome-wide mapping of the binding sites of transcription factors; previous methods had allowed the detection of only a limited number of these binding sites, for a given transcription factor, in each study. The importance of identifying these binding sites is that it promotes an understanding of the way in which a cell's genes, and its regulatory pathways, are controlled.

There are essentially two ChIP-based methods for mapping the genomic binding sites of transcription factors.

In ChIP-chip, amplified copies of a fragment of immunoprecipitated DNA, which includes the binding site of a *known* transcription factor, are used to probe a MICROARRAY; in the microarray, the oligonucleotide probes cover specific part(s) or all of the genome. Copies of the immunoprecipitated DNA may bind to one, several or many probes in the microarray, depending on the number of binding sites of the given transcription factor within the genome. One early study [Genes Dev (2002) 16(2):235–244] used a microarray of human CpG probes (see CPG ISLAND) to reveal promoters that bind the factor E2F; some 68 unique target loci were reported. More recently, studies using the ChIP-chip approach reported that the product of the oncogene *ZNF217* has thousands of target sites, including a variety of promoters, within the genomes of three tumor cell lines; certain of the observations in this work support the earlier suggestion that this protein behaves acts as a transcriptional repressor [J Biol Chem (2007) 282(13): 9703–9712].

Genome-scale ChIP-chip analysis has been carried out with as few as 10000 (human) cells [BioTechniques (2007) 43(6): 791–797].

ChIP-chip was also used for examining transcription factor IRF1 in the response to DNA damage [Nucleic Acids Res (2009) 37(4):1073–1085].

(See also TILING ARRAY.)

The other approach (ChIP-sequencing) involves sequencing of the immunoprecipitated DNA. In some cases an individual ChIP DNA fragment is sequenced. In other cases, sequencing is carried out, concomitantly, on a range of fragments which have been ligated together to form a single linear molecule; each of these fragments is a 'signature' sequence copied from a given ChIP DNA fragment. A variant form of the latter procedure is to ligate together fragments obtained by cloning the 5′ and 3′ ends of each of the ChIP fragments; this method is called the ChIP-PET approach (PET derived from Paired-End diTags). [PET-Tool: software for processing and managing of PET sequence data: BMC Bioinformatics (2006) 7:390.]

ChIP-precipitated DNA is now often examined by one of the 'new-generation' DNA sequencing systems; this approach was used e.g. in a study of the relationship between histone H3 methylation at lysine-4 and transcription factor binding [Genome Res (2008) 18(12):1906–1917].

[Analysis of the mouse embryonic stem cell regulatory networks obtained by ChIP-chip and ChIP-PET: Genome Biol (2008) 9(8):R126.]

The various ChIP-based approaches have been compared [Genome Res (2007) 17:898–909].

[Optimization of experimental design parameters for high-throughput ChIP: Nucleic Acids Res (2008) 36(21):e144.]

[ChIP for preparing a histone map of human chromosome 20q13.12: PLoS ONE (2009) 4(2):e4479.]

(cf. SOUTHWESTERN BLOTTING and see also DNA–PROTEIN INTERACTIONS.)

chromatin insulator A sequence of DNA, present in diverse species, which acts as a boundary between chromosomal loci that are subject to different patterns of regulation; insulators prevent the regulatory influence(s) that affect a given locus from affecting adjacent loci – e.g. a chromatin insulator may block the effect of an adjacent enhancer, thereby inhibiting gene expression.

Studies on a major insulator-binding protein in vertebrates, CTCF, showed demarcation of active and repressive domains [Genome Res (2009) 19(1):24–32].

Chromatin insulators were inserted into lentivirus vectors (used in gene therapy) to prevent these vectors from affecting the adjacent, endogenous genes following integration into the host's genome; however, while the vectors were functional, the presence of insulators was associated with a lower titer of integrants [BMC Biotechnol (2009) 9:13].

(See also ICR in the entry IMPRINTING CONTROL REGION.)

chromatosome See CHROMATIN.

Chromobacterium violaceum (CV026) (as biosensor in quorum sensing) See QUORUM SENSING.

chromosome A structure which contains part, or all, of a cell's genetic information; chromosomes consist essentially of DNA associated with certain proteins.

(See also CHROMATIN; cf. GENOME.)

The term 'chromosome' is usually not applied to the genetic material of viruses.

In prokaryotes (i.e. bacteria, archaeans) chromosome(s) and cytoplasm are in direct contact. (See also NUCLEOID.) In cells which lack extrachromosomal elements (such as a PLASMID) the chromosome(s) contain the entire genetic complement of the cell.

In a eukaryotic cell, all the chromosomes occur within the space enclosed by the nuclear membrane, and they are thus separated from the extranuclear region by this (permeable) barrier. The nuclear membrane must be permeable and must be able to support two-way traffic; for example, transcripts of polypeptide-encoding genes must be able to leave the nucleus in order to undergo translation on ribosomes. (See NUCLEUS.)

Chromosomes contain most of the genetic complement in eukaryotic cells; some DNA occurs in the mitochondria (see MITOCHONDRIAL DNA) and – in photosynthetic species – in chloroplasts. (See also KINETOPLAST.)

Circular and linear chromosomes

Most (not all) of the prokaryotic chromosomes studied contain a covalently-closed circular (ccc) dsDNA molecule. Size varies with species; for example, *Buchnera* (an intracellular parasite of aphids) has a rather small chromosome (∼640 kb), while the *Escherichia coli* chromosome is ∼5 Mbp.

Linear DNA is found in the chromosomes of e.g. the Gram-negative bacterium BORRELIA BURGDORFERI and the Gram-positive bacterium *Streptomyces*. A large (∼9.7 Mbp) genome consisting of a linear chromosome and three linear plasmids is found in *Rhodococcus* sp RHA1 [Proc Natl Acad Sci USA (2006) 103:15582–15587]. A feature of the prokaryotic linear chromosomes (such as the chromosome in *B. burgdorferi*) is the presence of closed, hairpin terminal regions (the so-called *telomeres*); during DNA replication (involving the formation of a circular chromosome dimer) the telomeres are fashioned by the enzyme PROTELOMERASE.

Eukaryotic chromosomes are structurally distinct. The ends of each linear dsDNA molecule include specialized sites that are involved in maintaining the integrity of the molecule: see TELOMERE. Moreover, the involvement of processes such as mitosis in cell division requires that these chromosomes contain specialized and structurally complex regions in order to permit their mechanical movement within the cell: see e.g. KINETOCHORE.

Number of chromosomes per cell

In prokaryotic cells, the number of chromosomes per cell is determined by (i) species and also by (ii) growth conditions. For example, while (slow-growing) cells of *Escherichia coli* contain one chromosome, a minimum of about four occur in *Deinococcus radiodurans*, and more than ten may be found in *Desulfovibrio gigas*.

In prokaryotes, it is generally assumed that when chromosomes occur in multiple copies in a cell then those chromosomes will be similar or identical to each other. Atypically, in e.g. the Gram-negative pathogen *Vibrio cholerae*, each cell

contains *two dissimilar types* of chromosome – one ∼3 Mbp, the other ∼1 Mbp; it has been reported that the initiation of replication of these chromosomes occurs at different times in the cell cycle [EMBO J (2007) 26(13):3124–3131].

In *E. coli*, growth rate affects the chromosome complement of the cell; thus, according to the Helmstetter–Cooper model, during rapid growth, a new round of DNA replication (i.e. chromosome replication) begins before the existing round has finished, leading to the presence of more than one chromosome in each rapidly growing cell.

Eukaryotic cells contain a species-dependent complement of chromosomes. Thus, cells of the slime mold *Dictyostelium* have six chromosomes (chromosome 2 containing about 25% of the genome), while the (somatic) cells of humans normally contain 46: 22 homologous pairs of autosomes and two so-called sex chromosomes (XX in females and XY in males). (See also X CHROMOSOME and Y CHROMOSOME.)

The diploid (2n) number of chromosomes in some other animals: pigeon (16), cat (38), pig (38), mouse (40), Rhesus monkey (42), rabbit (44), sheep (54), horse (64), chicken (78) and dog (78).

In the human population, atypical, i.e. abnormal, numbers of chromosomes are found in some individuals – e.g. XXX in superfemales and XYY in some males (see also e.g. DOWN'S SYNDROME).

Variation in the normal number of X chromosomes may be determined e.g. by examining the number of Barr bodies (see BARR BODY) in a stained preparation.

Citing locations in* (*human*) *chromosomes

A particular location on a given chromosome can be indicated with reference to a chromosome map. In this system, the location is indicated by (i) the chromosome number (i.e. a number from 1 to 22), or a letter (X or Y), followed by (ii) 'p' (the *short* arm of the chromosome) or 'q' (the *long* arm), and then (iii) a number that refers to a particular *band*, the bands reflecting the reaction of chromosomes to certain dyes. Thus, using this system, a location may be given as e.g. 14q32.1 or Xp22.

Examples of some chromosomal locations are given in the table for the entry GENETIC DISEASE.

(See also BACTERIAL ARTIFICIAL CHROMOSOME, HUMAN ARTIFICIAL CHROMOSOME and YEAST ARTIFICIAL CHROMOSOME.)

chromosome 6 (human) The location of alleles that encode the major histocompatibility complex.

chromosome 8 (human) The normal location of the MYC oncogene.

chromosome 9 (human) The normal location of the oncogene c-*abl*. In most patients with chronic myelogenous leukemia this gene is translocated to chromosome 22 (which is then called the 'Philadelphia chromosome').

chromosome 13 (human) The location of the coding region of the ONCOMIR-1 family of oncogenic miRNAs.

chromosome 19 (human) Relevant entries: AAVS, C19MC and MICRORNAS.

chromosome 22 (human) See CHROMOSOME 9. (See also ABL.)

chromosome aberration Any abnormality in the structure or number of a cell's chromosomes. Thus, e.g. a HETEROPLOID cell may lack a single chromosome from a (normal) diploid set, i.e. one of the chromosomes is unpaired; such a cell is referred to as *monosomic*. A diploid cell which has one extra chromosome (i.e. three copies of one particular chromosome) is *trisomic*.

(See also EUPLOID.)

chromosome banding A banding pattern seen when chromosomes are stained with certain dyes that bind preferentially to AT-rich regions in DNA.

chromosome painting A procedure in which a library of fluorescently labeled fragments of chromosome-specific DNA are used as PROBES e.g. to examine a preparation of immobilized metaphase chromosomes – or particular individual chromosome(s); the fragments are differentially labeled such that the emission spectrum from a given fragment correlates with its target.

In the context of pathology, this procedure is able to reveal chromosome aberrations such as aneuploidy and the kinds of rearrangement/translocation which are found in the chromosomes of some tumor cells; thus, for example, a translocation would be indicated if fragments specific to a given chromosome (identified by the emission spectrum of the fluorescent label) bind to a different chromosome.

chromosome walking A procedure used e.g. to assemble sequences of DNA, from a genomic library, into a series of *overlapping* fragments (i.e. to prepare a 'clone contig'). This is a preliminary step in the preparation of a physical map of the given genome. Once a contig has been established it is then possible to sequence each of the fragments of cloned DNA and arrive at a defined sequence of nucleotides for the entire segment of genomic DNA covered by the contig.

Initially, one of the cloned sequences in the genomic library is chosen as the starting clone; this sequence may e.g. contain a known gene. An *end* fragment of this sequence is then prepared, labeled, and used as an *end probe* to probe the other sequences in the library. Sequence(s) that are bound by the end probe are taken to be sequences that overlap the starting clone; these seqences, in turn, are used for the preparation of end probes for detecting *their* overlapping sequences, and so on.

Chromosome walking can proceed in both directions from the starting clone.

(cf. COSMID WALKING.)

chronic myelogenous leukemia (chronic myeloid leukemia) A disease involving hematopoietic stem cells; the chronic phase of the disease can develop to blast crisis via an incompletely understood mode of progression. (See also ABL.)

chylomicron See the entry RNA EDITING.

CI Cytoplasmic incompatibility – see e.g. *Wolbachia* in GMMS.

cI protein (of phage λ) See PHAGE LAMBDA.

(See also BACTERIOMATCH TWO-HYBRID SYSTEM.)

***cI857, cI*$_{857}$** See PHAGE LAMBDA.

CIB protein (in platelets) See WISKOTT–ALDRICH SYNDROME.
CIMP See CPG ISLAND METHYLATOR PHENOTYPE.
cinoxacin See QUINOLONE ANTIBIOTICS.
CIPer See CONNECTOR INVERSION PROBE TECHNOLOGY.
ciprofloxacin A QUINOLONE ANTIBIOTIC which can induce the SOS SYSTEM (q.v.).
circular helicase-dependent amplification See CHDA.
circularization The formation of a covalent linkage between the two ends of a linear (single-stranded or double-stranded) molecule of nucleic acid – usually involving the joining of a 3′-hydroxyl to 5′- phosphate, which involves the activity of a LIGASE. (In TOPOISOMERASE I CLONING, circularization, i.e. the insertion of a fragment into a linear vector, is achieved by the joining of 3′-phosphate and 5′-hydroxyl termini.)
circularly permuted Refers to molecules of nucleic acid which have different terminal nucleotides but whose nucleotides are in the same consecutive order, for example 5′-ABCDEFGHIJ, 5′-GHIJABCDEF, 5′-FGHIJABCDE.
circulating recombinant form (CRF) (of HIV-1): see HIV-1.
***cis*-acting element** A discrete sequence of nucleotides (such as a promoter) whose influence is limited to other regions of the *same molecule* of nucleic acid.

A *cis*-acting protein is one which acts on the molecule that encoded the protein.

(cf. TRANS-ACTING ELEMENT.)
cisplatin *cis*-Diaminedichloroplatinum: an agent which forms DNA adducts; its use can cause cell-cycle arrest, blocking of DNA replication, and APOPTOSIS. Cisplatin has been reported to induce looping and (at higher concentrations) condensation of single DNA molecules [Nucleic Acids Res (2009) 37(5): 1400–1410].

Cisplatin has been used for treating some types of cancer (e.g. ovarian and testicular cancers).
citrullinemia A disorder caused by a deficiency in the enzyme arginosuccinate synthetase (EC 6.3.4.5) – which catalyzes the reaction between citrulline and aspartate (resulting in arginosuccinate). Patients suffering from this condition accumulate citrulline and ammonia in their plasma and they often have a damaged nervous system.

Classic (type I) citrullinemia results from mutation in the gene encoding arginosuccinate synthetase.

Adult-onset (type II) citrullinemia is an autosomal recessive disorder resulting from mutation in the *SLC25A13* gene (that encodes citrin); a deficiency in citrin leads to a deficiency in arginosuccinate synthetase.

(See also GENETIC DISEASE (table).)
cko Conditional knock-out (mutation).
clamp (in DNA replication) See PCNA.
clamping (in PCR) See PCR CLAMPING.
class I transposon See TRANSPOSON.
class II transposon See TRANSPOSON.
class A flexible patterns In genomes of prokaryotes and lower eukaryotes: a strong 11-bp periodic bias in the distribution of nucleotides [BMC Bioinformatics (2005) 6:206].

(cf. A-TRACT.)
class switching (in B cells) See e.g. HYPER-IGM SYNDROME.
clavulanic acid See β-LACTAMASES.
clean-up resin (for DNA) See e.g. STRATACLEAN RESIN.
cleared lysate The supernatant obtained after lysis of a suspension of bacteria and ultracentrifugation; given appropriate methodology, a cleared lysate may contain e.g. plasmids (if present in the original cells) but should be free of cell debris and of (unsheared) chromosomal DNA.
Cleavase® A STRUCTURE-SPECIFIC ENDONUCLEASE (q.v. for the details/mechanism) (Third Wave Technologies Inc., Madison WI) which cleaves at a junction between single- and double-stranded regions in a DNA molecule; this enzyme is used e.g. in a typing/subtyping procedure: see CLEAVASE FRAGMENT LENGTH POLYMORPHISM ANALYSIS.

(See also SINGLE-STRAND-SPECIFIC NUCLEASE.)
cleavase fragment length polymorphism analysis A form of TYPING/subtyping in which single-stranded samples of the DNA under test are first heat-denatured (to remove any intra-strand base-pairing) and are then cooled to a temperature at which secondary structures (e.g. hairpins) are formed. These samples are then subjected to the enzyme CLEAVASE, and the resulting fragments (which had been initially end-labeled) are examined by gel electrophoresis.

As well as typing/subtyping, this technique has been used e.g. for investigating mutations and polymorphisms in PCR products.
cleaved amplified polymorphic sequences See CAPS.
Cleland's reagent The reagent DITHIOTHREITOL.
Click-iT™ EdU A system (Invitrogen, Carlsbad CA, USA) for detecting *in vivo* DNA synthesis – or e.g. for carrying out a CELL PROLIFERATION ASSAY. Instead of using e.g. BRDU, the cells' DNA is labeled with the thymidine analog 5-ethynyl-2′-deoxyuridine (EdU); this reagent is incorporated into DNA during growth. Subsequently, EdU is detected with an azide-conjugated Alexa Fluor® dye that forms a covalent link with the EdU.

This assay avoids (i) the requirement for denaturing DNA, and (ii) the need to use antibodies (both requirements being associated with the use of BrdU).

[EdU-based fast and sensitive detection of DNA synthesis *in vivo*: Proc Natl Acad Sci USA (2008) 105(7):2415–2420.]

[Examples of use: J Clin Invest (2008) 118(11):3714–3724; BMC Cancer (2008) 8:189; Nucleic Acids Res (2009) 37(7): 2087–2095; Mol Biol Cell (2009) 20(7):1960–1969; PLoS Pathog (2009) 5(1):e1000269.]
cloacin DF13 A plasmid-encoded COLICIN produced e.g. by *Enterobacter cloacae*.

(See also IUTA GENE.)
Clonase™ An enzyme used in the GATEWAY SITE-SPECIFIC RECOMBINATION SYSTEM: see BP CLONASE and LR CLONASE (enzymes active at different types of *att* site).
clone contig See CONTIG.
cloning (DNA cloning, gene cloning, genomic cloning, molecular cloning) A procedure used for producing millions of copies of a given sequence of nucleic acid (commonly a gene

or restriction fragment) which can then be used for sequencing or for subsequently expressing a product encoded by the gene etc.

Note. This process is distinct from *whole-animal* cloning: see the next entry (and see also the entry SCNT).

The term 'cloning' is generally understood to refer specifically to procedures in which the gene or fragment of interest is inserted into a VECTOR molecule – see CLONING VECTOR – which is then replicated in bacteria, or in eukaryotic cells. In this process, the number of target sequences progressively increases as the vector molecules continue to replicate within the growing population of cells. (PCR and other nucleic acid amplification procedures can be regarded as *in vitro* forms of cloning. However, note that 'PCR cloning' has been used to refer to a procedure in which amplicons, produced by PCR, are inserted into vector molecules which are then replicated in cells.)

The type of vector used for cloning depends e.g. on the size of the gene/insert to be cloned. Inserts up to ∼10 kb may be cloned in small, high-copy-number bacterial plasmid vectors or e.g. in the INSERTION VECTORS based on phage lambda. Inserts of ∼9–23 kb can be cloned in lambda REPLACEMENT VECTORS such as the LAMBDA FIX II VECTOR. Inserts of ∼40 kb can be cloned in COSMIDS. Inserts of up to ∼300 kb can be cloned in BACTERIAL ARTIFICIAL CHROMOSOMES (which are based on the F PLASMID), and inserts of up to 1 Mb or so can be cloned in a YEAST ARTIFICIAL CHROMOSOME. (See also HUMAN ARTIFICIAL CHROMOSOME.)

The type of vector used for cloning may also be influenced by the source of the 'target' sequence (i.e. the sequence to be cloned). For example, when the target sequence is produced by a PCR reaction in which the *Taq* DNA polymerase is used, the amplicons commonly have 3′ single-nucleotide adenosine overhangs resulting from EXTENDASE ACTIVITY; this feature is exploited by using vectors which have 3′ single-nucleotide *thymidine* overhangs that facilitate amplicon–vector ligation (see e.g. TOPO TA CLONING KIT).

Various commercial products offer a range of facilities for cloning. Thus, e.g. the advantages of *phage-mediated* cloning – avoiding the need for transformation or electroporation – are available in lambda-based vectors (such as LAMBDA FIX II VECTOR). Cloning of AT-rich sequences (or other sequences which may be unstable within circular, supercoiled plasmid vectors) may be facilitated in linear vectors (see e.g. BIGEASY LINEAR CLONING SYSTEM). Transfer of a gene/insert between different vectors may be facilitated e.g. with the GATEWAY SITE-SPECIFIC RECOMBINATION SYSTEM, and cassette-based changing of markers can be facilitated with the EXCHANGER SYSTEM.

(See also entries SEAMLESS CLONING KIT, TOPOISOMERASE I CLONING, USER CLONING and ZAP EXPRESS VECTOR.)

Linguistic note

A gene can be *inserted into* a vector and then *cloned in* that vector. The commonly used phrase "cloned into" is a misuse of English.

cloning (whole-animal cloning) A procedure that involves the replacement of chromosomes in an (unfertilized) oocyte with chromosomes from a somatic cell, the engineered oocyte then being allowed to develop (with suitable methodology) into a live animal.

The initial part of this process – in which a gamete's DNA is replaced by DNA from a somatic cell – is called *somatic cell nuclear transfer* (see the entry SCNT for further details).

(See also RECLONING.)

cloning site In general, in a CLONING VECTOR: the recognition sequence of a given RESTRICTION ENDONUCLEASE; when cut by the particular restriction enzyme, the site may provide STICKY ENDS which can be used for the insertion of a sample fragment which is flanked by the corresponding sticky ends. This kind of procedure (the use of a single cloning site) does not permit control of the *orientation* of the insert within the vector.

To promote flexibility, and to permit control of the orientation of the insert, vectors commonly include a region which contains a multiple cloning site (MCS), i.e. a POLYLINKER.

In TOPO TA CLONING KITS the (linearized) vector contains 3′ single-nucleotide thymidine overhangs that allow insertion of the (PCR-generated) fragment by topoisomerase activity – i.e. without the involvement of a ligase.

cloning vector A VECTOR which can replicate in a host cell and which is used for CLONING a gene or a fragment of DNA. A typical cloning vector can replicate independently of the host cell's chromosome(s).

(cf. EXPRESSION VECTOR.)

Various types of vector are used for cloning fragments in the size range ∼10 kb to ∼1 Mb (see the entry CLONING for details).

Many cloning vectors are small multicopy plasmids that are designed for extrachromosomal replication in bacteria. Such plasmids usually include a POLYLINKER (or ample restriction sites for inserting target DNA) and at least one gene that is suitable for selection (e.g. an antibiotic-resistance gene – cf. REPRESSOR TITRATION). (See also PUC PLASMIDS.)

Various vectors are used for cloning in *yeasts*; they include the high-copy-number *yeast episomal plasmids* (YEps, which are based on the TWO-MICRON PLASMID), and YACs (YEAST ARTIFICIAL CHROMOSOMES).

In some mammalian cells (e.g. COS CELLS), cloning vectors carrying the origin of replication of simian virus 40 (see the entry SV40) are able to replicate extrachromosomally (i.e. as a plasmid). In general, however (in mammalian cells), vectors are more stable when they integrate with chromosomal DNA (a notable exception being the HUMAN ARTIFICIAL CHROMOSOME).

Cloning vectors include circular molecules (e.g. many types of plasmid) and linear molecules (see e.g. YEAST ARTIFICIAL CHROMOSOME and BIGEASY LINEAR CLONING SYSTEM).

A particular *strand* of the target DNA can be cloned e.g. in a PHAGEMID (see also PBLUESCRIPT).

Cloning vectors derived from bacteriophage λ (Stratagene,

La Jolla CA) include INSERTION VECTORS (inserts up to 12 kb) and also REPLACEMENT VECTORS (inserts ~9 to 23 kb) – see e.g. LAMBDA FIX II VECTOR. These vectors have several advantages; for example, they avoid the requirement for processes such as transformation and electroporation (which are normally used to insert plasmid vectors into cells), and they do not suffer from the kind of size bias that may be seen with plasmids. (See also ZAP EXPRESS VECTOR.)

Some vectors exhibit instability within certain types of host cell – for example, lentiviral vectors are unstable in wild-type cells of *Escherichia coli* (see also entry REDUCED-GENOME ESCHERICHIA COLI).

The (3128-bp) cloning vector pJET1/blunt – marketed by Fermentas – provides a positive selection system for cloning blunt-ended PCR products. A pJET1/blunt plasmid containing the required insert (ligated into the multiple cloning site) is replicated following insertion into *Escherichia coli* cells by transformation; however, any re-circularized plasmid lacking the insert expresses a restriction endonuclease which kills the host cell.

(See also SHUTTLE VECTOR.)

Checking for the presence of an insert in the vector

See the entry VECTOR.

cluster (of genes) Any group of genes (in a given genome) that have a similar or identical pattern of expression.

cM CENTIMORGAN.

CML CHRONIC MYELOGENOUS LEUKEMIA.

CMTX X-linked CHARCOT–MARIE–TOOTH DISEASE.

CMV Cytomegalovirus (subfamily Betaherpesvirinae).

CNTs CARBON NANOTUBES.

CNV COPY NUMBER VARIANT.

***coa* gene** A gene which encodes the (plasma-clotting) enzyme coagulase; it has been widely used as a marker for the Gram-positive bacterium *Staphylococcus aureus*.

The typing method MLST was used for studying the genetic diversity of the *coa* gene in strains of *S. aureus* [PLoS ONE (2009) 4(5):e5714].

coat protein (*virol.*) See CAPSID.

COBRA Combined bisulfite restriction analysis: a method for detecting and/or quantitating methylation of cytosine bases in a given (known) sequence of genomic DNA [seminal paper: Nucleic Acids Res (1997) 25(12):2532–2534]. An aliquot of the sample DNA is initially treated with BISULFITE, and the (bisulfite-modified) DNA is subsequently amplified by PCR. [Primer design for PCR on bisulfite-treated genomes: Nucleic Acids Res (2005) 33(1):e9.]

The products of PCR reflect any changes – due to bisulfite treatment – at positions occupied by cytosine in the (original) untreated sample. Thus, for example, if a particular cytosine had been converted to uracil (by the bisulfite), then the PCR product would contain *thymine* in place of the original cytosine.

The PCR products are digested with a RESTRICTION ENDONUCLEASE whose recognition sequence includes a particular cytosine of interest. If (due to bisulfite treatment) the given cytosine is represented as a thymine in the PCR amplicons, then restriction will not occur at this site – and this can be observed when the restriction fragments are examined by gel electrophoresis. Had the given cytosine *not* been converted to uracil when treated with bisulfite (because it was methylated) it would remain as cytosine in the PCR amplicons – so that restriction would occur at this site. The restriction fragments can therefore indicate whether or not a given cytosine was methylated or unmethylated in the original sample of DNA.

The relative amounts of digested and undigested products of PCR can indicate the overall level of methylation in the original sample DNA.

Instead of depending on visual inspection of the restriction patterns, greater accuracy in quantitation can be achieved by using appropriate instrumentation [Nucleic Acids Res (2006) 34(3):e17].

Uses of COBRA (e.g.): studies on the effects of promoter demethylation and histone acetylation on a latent viral gene [PLoS ONE (2009) 4(2):e4556], and studies on methylation of the *ABO* gene promoter [PLoS ONE (2009) 4(3):e4788].

Cockayne syndrome A disorder involving e.g. photosensitivity and poor development, often with failure to survive beyond early adulthood; two categories of the condition, A and B, are associated with mutations in the genes encoding DNA-repair proteins CSA and CSB – at chromosome locations 5q12 and 10q11, respectively.

Cockayne syndrome involves a deficiency in transcription-coupled repair of CPDs (i.e. cyclobutyl pyrimidine dimers) induced by UV (ultraviolet) radiation. CSB was reported to be degraded via the ubiquitin–proteasome pathway in a CSA-dependent way [Genes Dev (2006) 20(11):1429–1434].

CSA has separable roles when responding to DNA damage resulting from (i) UV radiation, and (ii) oxidative stress; this was indicated in studies in which cells from a Cockayne syndrome patient were exposed, separately, to UV radiation and to the reagents potassium bromate (which induces oxidative damage) and menadione, which induces the development of reactive oxygen species [Proc Natl Acad Sci USA (2009) 106 (15):6209–6214].

(See also XERODERMA PIGMENTOSUM.)

coconut cadang-cadang viroid (CCCV) A VIROID (q.v.) that has caused severe economic losses of coconut palm (e.g. in the Philippines).

CCCV is an atypical viroid in that it exists in four different forms. During the early stages of cadang-cadang disease, the infected palms contain two species of CCCV, one 246 nt and the other 492 nt; however, two other species of CCCV appear later in the disease and eventually predominate. The 246-nt viroid is apparently the unit CCCV RNA.

coding region (in dsDNA) (1) Any region encoding product(s), part(s) of products, and/or function(s).

(2) An exon or, collectively, all the exons, in a given gene.

coding strand (of a gene) That strand which is homologous to – rather than complementary to – the corresponding mRNA. This appears to be the current (consensus) view. Earlier, the

opposite view was held by many, i.e. the coding strand was regarded as the strand *on which* mRNA is transcribed, i.e. a strand complementary to mRNA.

The coding strand is also called the *plus strand* or the *sense strand*.

The *non-coding strand* (= *minus strand* or *antisense strand*) is the strand that is complementary to the coding strand, i.e. the *template strand* on which mRNA is transcribed.

(See also FORWARD PRIMER.)

CODIS Combined DNA Index System: a database of human DNA profiles used by the Federal Bureau of Investigation in the USA. DNA profiling may be achievable with only minute amounts of DNA and may be possible even with damaged or degraded samples.

An individual's DNA profile is obtained by analyzing each of 13 *short tandem repeats* (see STR), found at specific chromosomal loci, and determining the number of repeat units in each of these STRs. The number of repeats in a given STR is recorded, and this is compared with the number of repeats, in the same STR, in other members of the database population. When all 13 STRs are analyzed (completing the profile), the 'uniqueness' of a given sample of DNA can be calculated by combining the probabilities determined for each STR site; the *random match probability* of a given DNA sample may be as low as $\sim$1 in 10^{14} or $\sim$1 in 10^{15}.

In the analysis, MULTIPLEX PCR is used for simultaneous amplification of multiple loci. The number of repeats in a given STR is indicated by the length of the corresponding amplicons: for each extra repeat unit the length increases by a fixed number of bases.

(See also FORENSIC APPLICATIONS.)

codominant gene Any gene which, in combination with other gene(s), contributes to a phenotype distinct from that specified in its absence.

codon A triplet of consecutive bases in a nucleic acid molecule encoding a given amino acid – or one of the three termination signals (see NONSENSE CODON) – cf. *Dual-function codons* below.

(cf. ANTICODON; see also SECIS.)

Codons are written in the 5′-to-3′ direction.

There are 64 possible triplets (from the four bases). Each of the two amino acids methionine and tryptophan is specified by a single codon (AUG and UGG respectively). Each of the remaining amino acids is specified by more than one codon, and the existence of multiple (*synonymous*) codons for each of these amino acids is referred to by saying that the genetic code is *degenerate*.

(See also WOBBLE HYPOTHESIS.)

When comparing codon usage among the different species, it was found that, for a given amino acid, different synonymous codons may be used, preferentially, by different organisms. Even in a given organism, a particular amino acid may be specified by different codons in different genes – see the entry CODON BIAS (sense 3). [General rules for the optimal codon choice: PLoS Genet (2009) 5(7):e1000556.]

Dual-function codons

In the ciliate protozoan *Euplotes crassus* it has been reported that the UGA codon can specify either of two amino acids – selenocysteine or cysteine. At any given UGA codon within an mRNA, the choice of which amino acid to insert depends on the distance between the given UGA codon and a specific sequence in the 3′-UTR region: SECIS (the *selenocysteine insertion sequence*) [Science (2009) 323(5911):259–261].

codon bias (1) (*DNA technol.*) During expression of a heterologous ('foreign') gene, e.g. a mammalian gene in bacteria: an *effect* due to the presence of certain codon(s) in the gene for which the expressing organism has an insufficient number of tRNAs; for example, if a gene with a high frequency of the arginine codon AGG, or the proline codon CCC, is expressed in *Escherichia coli*, translation is likely to be inefficient, or aborted, because these codons are much less common in the genes of *E. coli*.

One solution is to express heterologous genes in cells that are supplemented with extra copies of the given tRNAs (for example, encoded by plasmid genes). *Escherichia coli* BL21-CodonPlus® (Stratagene, La Jolla CA) carries extra copies of *argU* and other tRNA-encoding genes.

Another solution is to adapt the target gene so that its tRNA requirements match more closely those of the expressing organism. JCat (Java codon adaptation tool) is a program designed to facilitate such adaptation [Nucleic Acids Res (2005) 33: W526–W531].

(See also CODON OPTIMIZATION.)

(2) An *attribute*, of a given organism, involving a difference in codon usage compared with other (specified) organism(s). [Example of use: J Virol (2004) 74(10):4839–4852.]

(3) (*general biol.*) The *observation* that a given amino acid can be specified, preferentially, by different codons in different species – and also by different codons in different genes within the same organism.

In the bacterial chromosome it has been reported that genes exhibiting similar codon usage tend to be located close to each other in the genome and that (in at least some species) a number of 'codon usage domains' can be identified in the chromosome [PLoS Comput Biol (2006) 2(4):e37].

codon harmonization (translational attenuation) In a heterologous gene (i.e. one expressed in an organism other than the gene's natural host organism): adjustment of codon usage, by re-engineering the nucleotide sequence, in such a manner that codons used with a given frequency in the natural host are used with the same frequency in the transgenic (non-natural) host. This is reported to allow the translational machinery of the transgenic cells to progress and pause at appropriate sites in the mRNA and – as a consequence – to promote correct folding of the protein (in terms of the secondary and tertiary structures).

[Codon optimization versus codon harmonization for the expression of plasmodial proteins: Malaria J (2008) 7:197.]

(cf. CODON OPTIMIZATION.)

codon optimization The (pre-use) modification of a transgene

in order e.g. to exclude codons that are rarely used by the expressing organism; such codons, which are normal for the species of origin, are replaced by codons that are commonly used in the transgenic cell or organism – leading to enhanced expression of the given gene.

(cf. CODON HARMONIZATION; see also CODON BIAS (sense 1).)

When codons in a mycobacterial gene were optimized for expression in mammalian cells, a DNA VACCINE encoding the given gene exhibited an enhanced immune response within BALB/c mice [Infect Immun (2005) 73(9):5666–5674].

Replacement of a rare codon for leucine by a common codon in the gene encoding mono-oxygenase resulted in greater enzymic activity in the bacterial strain expressing this gene [Appl Environ Microbiol (2005) 71(11):6977–6985].

In studies on translational efficiency of an mRNA encoding the HIV-1 *gag* region, codon optimization enhanced translation by a factor of 1.4–1.7 [PLoS ONE (2008) 3(6):e2356].

[Codon optimization for expression of plasmodial proteins: Malaria J (2008) 7:197.]

Optimization of codon usage for expressing red fluorescent protein mRuby in mammalian cells resulted in a fluorescent signal 5–8 times stronger [PLoS ONE (2009) 4(2):e4391].

codon usage domain See CODON BIAS (sense 3).

coenzyme I See NAD.

coenzyme R *Syn*. BIOTIN.

cohesive ends *Syn*. STICKY ENDS.

co-immunoprecipitation An early method, used for detecting protein–protein interaction, in which an epitope-tagged 'bait' protein is precipitated by an anti-tag antibody; if any protein interacts with (binds to) the bait protein it is co-precipitated with the bait–antibody complex.

(See also PROTEIN–PROTEIN INTERACTIONS.)

cointegrate The (circular) product formed by the fusion of two circular replicons – e.g. during *replicative* transposition: see TRANSPOSABLE ELEMENT (figure and figure legend).

col plasmid COLICIN PLASMID.

colchicine A tricyclic alkaloid, from e.g. *Colchicum autumnale*, whose molecule includes a tropolone ring. Colchicine binds to tubulin and inhibits polymerization – blocking mitosis and e.g. promoting polyploidy in some types of cell.

The *in vitro* effect of colchicine can be terminated e.g. by using ultraviolet radiation; radiation converts colchicine to an inactive derivative: *lumicolchicine*.

ColE1 plasmid A small, non-conjugative, multi-copy PLASMID which is present in cells (strains of enterobacteria) typically at a copy number of about 10–30; it encodes a pore-forming COLICIN of the E category.

The GC% of ColE1 DNA is ∼50.

The initiation of replication of ColE1 DNA is regulated at the level of primer formation. A plasmid-encoded molecule of RNA (RNA II) is transcribed on a particular strand from a site upstream of the origin of replication. This transcript has the potential to develop into a primer from which DNA synthesis may ensue. Another plasmid-encoded molecule (RNA I), also transcribed from a site upstream of the origin (but on the other strand), has a regulatory role: it represses RNA II's function by base-pairing with it. Thus, whether or not a given initiation event takes place will depend on the outcome of the interaction between RNA II and RNA I.

The concentration of RNA I may be limited by the activity of an RNase when the copy number of ColE1 is submaximal; this, in turn, affects the freedom of the RNA II transcript to initiate replication.

A further regulatory influence on RNA II is the (plasmid-encoded) Rop protein (sometimes called Rom) which appears to promote the hybridization of RNA I and RNA II.

In *Escherichia coli*, mutation in the *pcnB* gene can affect the frequency of initiation of replication of ColE1 DNA. This gene encodes poly(A) polymerase I, an enzyme which adds a poly(A) tail to certain transcripts, including RNA I. The poly-(A) tail which is normally present on RNA I facilitates its degradation. Thus, a non-functional poly(A) polymerase that prevents poly(A) tailing of RNA I will enhance the stability of this molecule and therefore serve to enhance its inhibitory action on RNA II.

(See also MULTICOPY PLASMID.)

The origin of replication of ColE1 is widely used in various types of vector molecule.

colicin Any of various high-molecular-weight, mostly plasmid-encoded BACTERIOCINS formed by enterobacteria. [Biology: Microbiol Mol Biol Rev (2007) 71(1):158–229.]

Three genes are associated with a given colicin – the genes encoding (i) the colicin itself; (ii) an *immunity protein* (which protects against the *same* colicin produced by nearby cells); and (iii) a *lysis protein* (which promotes release of the colicin from the producing cell); the immunity protein is synthesized constitutively.

Synthesis of colicins (in *colicinogenic* cells) is triggered by the so-called SOS response, which is activated by damage to DNA.

Some colicins (e.g. E6, DF13) have RNase activity; some (e.g. E2, E8) have DNase activity; some (e.g. E1, A, N, B, Ia, Ib) are pore-formers, and others (e.g. M) inhibit synthesis of the bacterial cell wall polymer peptidoglycan.

colicin factor An early name for a COLICIN PLASMID.

colicin plasmid Any plasmid encoding one or more COLICINS; they include small multicopy plasmids and large low-copy-number plasmids (and both conjugative and non-conjugative plasmids).

Repression of transcription of colicin genes by LexA (see the entry SOS SYSTEM) is abolished during the cell's response to damaged DNA, i.e. agents which trigger the SOS response promote colicin synthesis.

(See also COLE1 PLASMID.)

coliphage Any bacteriophage which can infect *Escherichia coli* (*E. coli* is a species of *coliform* bacteria).

colloidal silica (in gene delivery) See the entry GENE-DELIVERY SYSTEM.

colony (*bacteriol*.) A discrete mound or heap of bacterial cells

which has developed, *from a single cell*, by repeated division of the progeny cells on (or within) a solid medium such as an agar plate.

Colonies of a given species of bacteria have a characteristic size, shape, color and consistency when growth occurs on a given type of medium for a specified period of time (see e.g. MACCONKEY'S AGAR).

colony hybridization (modified Grunstein–Hogness procedure) A method for selecting colonies of recombinant bacteria that contain a *particular* sequence of DNA. When a population of bacteria is transformed with a genomic LIBRARY, and then plated on a growth medium, *each* colony formed on the plate typically consists of bacteria containing *one* of the fragments from the library. To select colonies that contain a *particular* fragment, the original plate (= *master plate*) is replica plated (see REPLICA PLATING) to a nitrocellulose filter; cells from each colony adhere to the filter – which is then treated with alkali to (i) lyse the cells and (ii) denature DNA to the single-stranded state. The alkali is neutralized, and proteins are digested with proteinase. The filter is baked (at 70–80°C) under a vacuum, to bind ssDNA to the filter, and is then exposed to labeled probes (see PROBE) to detect any colony that contains the required sequence; this information can then be used to identify relevant colonies on the master plate.

Other methods for screening a genomic library include the RECACTIVE system.

(Compare with ANTIBODY-BASED LIBRARY SCREENING.)

colorectal cancer (detection) See CANCER DETECTION.

combination probe test Any test in which a PROBE is used to detect two (or more) target sequences simultaneously, i.e. in the same assay – see e.g. PACE 2C.

combinatorial (of oligonucleotides, oligopeptides etc.) Refers to a collection or library of oligonucleotides or oligopeptides etc. in which the constituent subunits (nucleotides and amino acid residues in these cases) in the individual members of the collection are arranged in many different combinations, such combinations sometimes covering every possible permutation of the subunits.

combined bisulfite restriction analysis See COBRA.

Combined DNA Index System See CODIS.

combined gold standard See GOLD STANDARD.

combing (of DNA) See DNA COMBING.

community-acquired MRSA See MRSA.

comparative DNA cleavage See e.g. CPNPG SITES.

comparative genomic hybridization (CGH) A technique for detecting changes (e.g. losses or gains) in particular chromosomes, or regions of a chromosome, in a sample of genomic DNA. Essentially, fragmented whole-genome preparations of sample DNA and normal DNA, each labeled with a different fluorescent dye, are allowed to hybridize simultaneously with a spread of *normal* metaphase chromosomes. The loss or gain of a chromosome, or sequence, in the sample DNA is detected by measuring the ratio of fluorescence intensities of the two fluorophores at various locations along the length of the metaphase chromosomes.

If, for example, a given sequence in a particular metaphase chromosome hybridizes with normal DNA but not with the sample DNA (shown by the absence of fluorescence from the *sample's* dye) this may indicate a deletion in the corresponding sequence of that chromosome in the sample DNA.

CGH has been used e.g. to detect aneuploidy (as in DOWN'S SYNDROME).

(cf. MULTIPLEX LIGATION-DEPENDENT PROBE AMPLIFICATION.)

(See also GENE AMPLIFICATION.)

compatible sticky ends See STICKY ENDS.

competence (in transformation) See TRANSFORMATION.

competent cells (*DNA technol.*) Cells selected, or engineered, for specific uses. Some competent cells have 'extra' functions while others are mutant/defective in some functions(s).

Cells which have 'extra' functions include e.g. *Escherichia coli* supplemented with an inducible gene for tyrosine kinase; unlike wild-type *E.* coli, these cells can synthesize tyrosine-phosphorylated proteins.

Some cells have additional translation capacity (see CODON BIAS, sense 1).

Cells with defective functions include e.g. METHYLATION-DEFICIENT CELLS (q.v.). Cells which contain an error-prone DNA polymerase (such as Mutazyme®; Stratagene, La Jolla CA) are used for MUTAGENESIS. Extra-permeable cells (e.g. ElectroTenBlue®; Stratagene) are used for ELECTROPORATION. Cells of *E. coli* with a *lacZ*Δ*M15* mutation permit blue–white color screening for recombinant phagemids when used e.g. for replicating pBluescript® vectors (see PBLUESCRIPT). Cells of *E. coli* which are deficient in ADENYLATE CYCLASE (e.g. *E. coli* BTH101 *cya*-99) were used e.g. in a BACTERIAL TWO-HYBRID SYSTEM.

(See the table in the entry ESCHERICHIA COLI for examples of various other modifications made to host cells for specific uses in recombinant DNA technology.)

competimer A PRIMER which has a non-extendable 3′ terminus but which competes with another primer for binding sites – thereby reducing the level of synthesis from the given binding site.

Competimers are used e.g. in quantitative RT-PCR when an 18S rRNA target is used as an internal control. As 18S rRNA is usually present in relatively large quantities, amplification of this target may go beyond the exponential phase in only a small number of cycles of PCR so that the internal control is not able to fulfill its function. To maintain a balance between levels of amplification of the internal control and the RNA of interest, the reaction mixture may be supplemented with a proportion of competimers in order to reduce amplification of the internal control; with an appropriate ratio of competimers to primers it is possible to reduce synthesis of the control sequence to the required level.

Competimers are obtainable commercially (Competimer™, Ambion, TX, USA).

[Uses (e.g.): Proc Natl Acad Sci USA (2007) 104(6):1766–1770; Virol J (2008) 5:111; Mediators Inflamm (2009) doi:

10.1155/2009/535348; BMC Plant Biol (2009) 9:64 (primer–competimer ratio 3:7); and Mol Vision (2009) 15:1277–1293 (primer–competimer ratio 1:4).]

In an analogous context, dominant *DNA* templates may be dealt with by techniques such as PCR CLAMPING and SUICIDE POLYMERASE ENDONUCLEASE RESTRICTION.

competitive clamping See PCR CLAMPING.

competitive RT-PCR A form of QUANTITATIVE PCR used for assessing the absolute number of specific target sequences in a given sample.

Essentially, the target is co-amplified with an (exogenous) 'competitor RNA' – using a single pair of primers for both the target and competitor. The assay is carried out with a range of concentrations of the competitor.

If, in a given reaction, the amount of product formed from the target matches that from the competitor, then this is taken to indicate that the target and the competitor were present in equal amounts at the start of that reaction.

For meaningful results, the competitor should be amplified with an efficiency similar to that of the target, and it should also give rise to a product which is distinguishable from that of the target sequence.

complementarity See BASE PAIRING.

complementary DNA (cDNA) See CDNA.

complementation Combination of two or more entities, either physically or in effect, forming a complete, functional and/or wild-type product, system or organism.

In one form of (genetic) complemetation, a wild-type gene introduced into a cell containing a defective (mutant) copy of that gene may compensate for the functional defect of the mutant gene (i.e. it may 'complement' the mutant gene) and thus confer on that cell a wild-type phenotype (in respect of the given gene).

In a number of cases, fragments of a given protein, which are inactive on their own, regain at least some activity when they are combined (not necessarily covalently) into a single particle: see BIMOLECULAR FLUORESCENCE COMPLEMENTATION, ENZYME FRAGMENT COMPLEMENTATION, α-PEPTIDE, and *split intein* in the entry INTEIN.

This principle also underpins techniques such as blue–white screening (see entry PBLUESCRIPT) and TWO-HYBRID SYSTEMS.

α-complementation A phenomenon involved in the blue–white screening method: see α-PEPTIDE and PBLUESCRIPT.

complete medium Any medium that supports the growth of a given AUXOTROPHIC MUTANT (q.v.); such a mutant does not grow in a *minimal medium* unless it is supplemented with the relevant growth requirement(s).

complete transduction See TRANSDUCTION.

composite transposon (class I transposon) See TRANSPOSON.

computational biology See BIOINFORMATICS.

ComX See TRANSFORMATION.

concatemer Two or more identical, linear molecules of nucleic acid linked (covalently) end-to-end – in the same orientation. The term CONCATENATE has been used as a synonym.

One example of a concatemer is the repetitive sequence formed when the PHAGE LAMBDA genome is replicated by the ROLLING CIRCLE mechanism.

(cf. CATENANE.)

concatenate (*verb*) To join (e.g. nucleic acid fragments) end-to-end (covalently), forming a chain.

If the fragments are identical the result is a CONCATEMER.

If the fragments etc. are not identical, the result is called a concatenate (*noun*) or a concatenation; such a chain is formed for example, during ligation of the ditags in SAGE, and in an analogous process in CHROMATIN IMMUNOPRECIPITATION.

(cf. CATENANE.)

conditional gene activation/regulation (*DNA technol.*) *In vivo* control of genes (e.g. starting and/or stopping transcription or translation) can be achieved in various ways. In the simplest case, a gene which contains a temperature-sensitive mutation may be activated or repressed by a 'permissive' or 'repressive' temperature, respectively.

Regulating the activity of genes *in vivo* often involves some *prior* engineering of the cell or organism. For example, a *lac* operator (see LAC OPERON) may be inserted in the regulatory region of the gene of interest – and the *lac* repressor protein, expressed from a plasmid vector, used to inhibit transcription of the gene. Removal of repression (activation of the gene), as and when required, can be achieved by using IPTG (which blocks the repressor). (See also entry CASCADE EXPRESSION SYSTEM.)

The successful activation of a gene *in vivo* can be signaled by the expression of a REPORTER GENE that has been fused, in-frame, with the gene of interest (see GENE FUSION) – and which is therefore co-expressed within the cell; the reporter gene may encode a fluorescent product (e.g. GREEN FLUORESCENT PROTEIN) which can be easily detected.

The *simultaneous* activation of *two* genes may be signaled e.g. by the SPLIT-CRE SYSTEM.

When using a site-specific recombination system to control the activation of gene(s), the *timing* of gene activation can be controlled e.g. by using a *ligand-activated* recombinase [see e.g. PLoS ONE (2009) 4(2):e4640].

The simultaneous conditional silencing of up to four genes in *Escherichia coli* was achieved by the use of vector-borne (IPTG-inducible) antisense RNAs [Nucleic Acids Res (2009) doi: 10.1093/nar/gkp498].

In a different strategy, the gene of interest is inactivated by *inverting* an essential sequence in the gene in such a way that the inversion can be reversed (thereby restoring the wild-type sequence) by using a SITE-SPECIFIC RECOMBINATION system.

The regulation of genes *in vivo* may involve blocking gene expression by various GENE SILENCING methods.

conditional knock-out Any KNOCK-OUT MUTATION which is targeted at a specific type of cell or tissue or is designed to be effected at a specific time, or during a specified time interval, or under specified conditions; that is, spatial, temporal and/or other restrictions are imposed on the required changes.

confocal microendoscope An instrument with which tissue can

be imaged *in vivo* at cell-level resolution. Commonly, tissues are made fluorescent by using exogenous fluorescent contrast agents such as fluorescein or acridine orange.

[Acridine orange and SYTO 16 for *in vivo* imaging of peritoneal tissues in mice: J Microsc (2009) 234(2):124–129.]

conjugation (*bacteriol*.) A (natural) process in which DNA is transferred from one cell (the *donor* or 'male') bacterium to another (*recipient* or 'female') bacterium while the cells are in physical contact. A bacterium that has received DNA in this process is a *transconjugant*. (See also RETROTRANSFER.)

In many cases the donor phenotype is conferred on a cell by the intracellular presence of a conjugative PLASMID; in other cases conjugation is mediated by a TRANSPOSON (rather than a plasmid) and may be considered as a form of (intercellular) transposition – see the entry CONJUGATIVE TRANSPOSITION.

This entry is concerned specifically with *plasmid*-mediated conjugation.

Plasmid-mediated DNA transfer can occur between Gram-negative species and it can also occur between Gram-positive species. The way in which the cells establish contact appears to differ mechanistically in Gram-negative and Gram-positive species, but there are similarities between Gram-negative and (certain) Gram-positive species in the mechanism(s) involved during mobilization and transfer of DNA.

Conjugation in Gram-negative bacteria

Conjugation in Gram-negative species involves a PILUS; pili are apparently essential for the establishment of conjugative cell–cell contact in these species.

Much of the published information on Gram-negative conjugation has been obtained from studies on the conjugative transfer of the F PLASMID, and of related plasmids, between strains of enterobacteria (such as *Escherichia coli*). However, the general features of the F plasmid–*E. coli* transfer system are not common to all forms of conjugative transfer in Gram-negative bacteria; for example, the type of pilus differs markedly in some other systems, and the *tra* (i.e. transfer) operon (which is constitutively de-repressed in the F plasmid – see FINOP SYSTEM) is often repressed in other types of plasmid.

The type of pilus expressed by a donor affects the physicochemical conditions under which conjugation can take place. For example, the F pilus (and similar types of pili) is able to mediate conjugation when the cells are freely suspended in a liquid medium, while other types of pili (e.g. those specified by IncN plasmids) mediate '*surface-obligatory*' conjugation – which occurs on an appropriate surface (for example, on an agar plate).

Early studies found that F-type conjugation in liquid media is greatly stimulated (in terms of the number of transconjugants per mating) by increased levels of electrolyte (200 mM Na^+ or K^+), and that IncN-mediated conjugation appears to need – or to be assisted by – surface tension.

In at least some cases, initial donor–recipient contact, via the pilus, is followed by pilus retraction and the subsequent establishment of wall-to-wall contact between the conjugants. Some authors have reported an (electron-dense) *conjugative junction* at the region of cell–cell contact. Effective cell–cell contact precedes the *mobilization* and transfer of DNA from the donor. In the F plasmid, a specific site (in the *T strand*) at the origin of transfer (*oriT*) is nicked, and a single strand of DNA (5′-end leading) is transferred to the recipient. Nicking is carried out by an endonuclease (*traI* gene product) which forms part of the so-called *relaxosome* complex. It is possible that the subsequent unwinding of the transferred strand may provide energy for the transfer process.

Within the recipient, a complementary strand is synthesized on the transferred strand, forming a circularized molecule of dsDNA – a process termed *repliconation*.

Within the donor, a complementary strand is synthesized on the non-transmitted strand, reconstituting the plasmid; this process, which may occur by a ROLLING CIRCLE mechanism, is referred to as *donor conjugal DNA synthesis* (DCDS).

In conjugation, *mobilization* can also refer to those cases in which DNA that is covalently continuous with a conjugative plasmid (e.g. a fused non-conjugative plasmid), or a bacterial genome which contains an integrated conjugative plasmid, is transferred to a recipient.

Conjugation in Gram-positive bacteria

Conjugation can occur in various Gram-positive bacteria, e.g. species of *Bacillus*, *Enterococcus* and *Staphylococcus* as well as in the mycelial organism *Streptomyces*.

Pheromone-mediated conjugation occurs in certain Gram-positive species. In this context, pheromones are small, linear peptides that are secreted by *potential recipients* of certain plasmids; a particular pheromone can act on a potential donor of the given plasmid – causing the donor to synthesize a cell-surface 'aggregation substance', promoting adhesion between potential conjugants. When a recipient receives the plasmid it stops secreting the particular pheromone – but may continue to secrete other pheromones corresponding to other plasmids.

Pheromone-mediated conjugation occurs e.g. within liquid cultures of certain strains of *Enterococcus faecalis*; plasmids which mediate this type of transfer frequently contain genes encoding resistance to antibiotic(s). In some cases a plasmid also encodes a peptide which antagonizes the corresponding pheromone; the function of this peptide may be to ensure that a mating response is not triggered in the donor cell unless the pheromone has reached a certain critical concentration – i.e. conditions under which there is a good chance of a random collision between potential conjugants.

Some strains of e.g. *Staphylococcus* and *Streptococcus* can carry out pheromone-independent conjugation on surfaces (as opposed to conjugation while freely suspended in liquids); the plasmids encoding this type of transfer are often >15 kb in size.

There appear to be two distinct conjugative processes in the Gram-positive bacteria. That operating in unicellular species is seen as functionally similar (in terms of DNA mobilization etc.) to conjugation in the Gram-negative species. However, a

different mechanism seems to operate in multicellular Gram-positive species; this process appears to be characterized by the transfer of *double*-stranded DNA.

conjugative pilus *Syn.* PILUS.

conjugative plasmid See the entry PLASMID.

conjugative transposition In bacteria: a plasmid-independent process – mediated by a *conjugative transposon* – in which DNA is transferred from a donor cell to a recipient cell while the cells are in physical contact. Conjugative transposition is known to occur between Gram-positive bacteria and between Gram-negative bacteria; it may also be able to mediate gene transfer between Gram-positive and Gram-negative bacteria.

Conjugative transposons resemble 'classic' transposons in that e.g. they can move from one DNA duplex to another and can occur in chromosomes and plasmids. Unlike most classic TRANSPOSABLE ELEMENTS they usually excise before transfer and they do not duplicate the target site when inserting in the recipient cell. The smallest of the conjugative transposons include Tn*916* (also referred to as CTn*916*; 18 kb), while the largest conjugative transposons are >50 kb.

Most or all conjugative transposons carry at least one gene that encodes antibiotic-resistance. Typically, the conjugative transposons are resistant to degradation by host cell restriction enzymes.

In one early model for conjugative transposition between Gram-positive bacteria, excision of the transposon is triggered by donor–recipient contact and it involves a transposon-encoded recombinase (*int* gene product). The excised transposon circularizes, and a single strand may be transferred to the recipient cell; synthesis of a complementary strand (in the recipient) is followed by insertion into the target site.

In the (Gram-negative) bacterium *Bacteroides* it has been reported that transfer of *part* of the conjugative transposon CTnERL can occur in an excision-deficient strain – transfer apparently involving an 'Hfr-type' mechanism [J Bacteriol (2006) 188(3):1169–1174].

Conjugative transposition of CTnDOT in *Bacteroides* was reported to involve homology-dependent and -independent steps [Nucleic Acids Res (2007) 35(17):5861–5873].

conjugative transposon See CONJUGATIVE TRANSPOSITION.

connector inversion probe (CIPer) technology A single-probe approach to the amplification of a given sequence of DNA [PLoS ONE (2007) 2(9):e915].

The (synthetic, linear) probe consists of (i) a 5′-end region and a 3′-end region (each about 20 bp long), the two regions being complementary to regions *flanking* the target sequence of DNA; (ii) a pair of primer-binding sites which occupy the internal region of the probe; and (iii) a central site – located between the primer-binding sites – which is the recognition sequence of a RESTRICTION ENDONUCLEASE.

Following strand separation in the (dsDNA) target molecule, the two terminal regions of the probe bind (anneal) to complementary regions on either side of the target sequence.

The 3′ end of the probe is extended by a DNA polymerase, thus copying the target sequence. Ligase action then results in the complete circularization of the probe.

Enrichment of circular DNA is then achieved by degrading any *linear* DNA (e.g. non-circularized probes) by using exonucleases I and III. These enzymes are inactivated before the next stage.

The circularized probes are linearized by cleavage with the restriction enzyme, and amplification (by PCR) is then carried out using the two primer-binding sites.

consensus sequence (in nucleic acids) In a particular genetic element (for example, a given type of promoter): an actual or theoretical sequence of nucleotides in which each nucleotide is the one found most commonly, at that location, in all the variant forms of the element that are found in different cells, organisms or strains.

constitutive gene *Syn.* HOUSEKEEPING GENE.

contig A word which has been used with various, incompatible meanings. When first used [see Nucleic Acids Res (1980) 8: 3673–3694] *contig* referred to '. . .a set of gel [electrophoresis] readings that are related to one another by overlap of their sequences.'. In the *Oxford Dictionary of Biochemistry and Molecular Biology* (revised edition, 2000), contig is defined as: 'one of a set of overlapping clones that represent a continuous region of DNA. Each contig is a genomic clone. . .. .'. Other uses of the term contig refer to a continuous sequence of DNA 'assembled' or 'constructed' from overlapping clones.

The current consensus view appears to be that a *contig* is a linear series of sequences that mutually overlap, throughout, and which, collectively, span a large, continuous sequence; the phrase *clone contig* has been used for a series of cloned DNA sequences with these properties.

contig mapping An approach used for the physical mapping of genomic DNA. Essentially, in this approach, the DNA is cut by restriction endonucleases into fragments which are cloned and sequenced. A (computer-based) search for overlapping regions then permits the fragments to be assembled into the correct overall sequence.

continuous-exchange cell-free protein synthesis See CELL-FREE PROTEIN SYNTHESIS.

contour-clamped homogeneous electric field electrophoresis (CHEF) See PFGE.

Cooley's anemia *Syn.* THALASSEMIA.

***copA* gene** See R1 PLASMID.

CopB See R1 PLASMID.

copia element In *Drosophila*: any of a category of transposable elements, each ~5 kb in length and present in a copy number of ~50; each element is bracketed by long terminal repeats (LTRs) of ~280 nt, and the coding region specifies products analogous to those of RETROVIRUSES.

copper-sulfate-inducible genes See e.g. entry DES, and vector pMT-DEST48 in the table in entry GATEWAY SITE-SPECIFIC RECOMBINATION SYSTEM.

copy DNA (cDNA) See CDNA.

copy mutant See COPY NUMBER.

copy number (1) (of bacterial plasmids) The number of copies of a given plasmid, per chromosome, in a bacterial cell. The

copy number is characteristic for a given plasmid in a host cell growing exponentially; the F PLASMID is an example of a low-copy-number plasmid, while pBR322 is an example of a multicopy plasmid.

Copy number is determined e.g. by the replication-control system; different control systems occur in different plasmids. Mutations can give rise to a *copy mutant* in which the copy number differs from that of the wild-type plasmid.

(2) The number of copies of a given gene, per cell, or e.g. the number of copies of a gene product per gene or per cell. (See also GENE AMPLIFICATION.)

(3) See COPY NUMBER VARIANT.

copy number variant (CNV) A phrase that refers to variation in the copy number of a given, repetitive sequence of DNA (of at least 1 kb), the number of copies in the genome varying from one individual to another. It has been suggested that the changes in contextual genomic DNA (brought about during the evolution of this type of variation) may, in themselves, be important in determining the characteristics of a given genotype.

Copy number variation is found in the human genome and e.g. in the mouse genome. In the latter case, the variation was reported to be more locally restricted than it is in the human genome, with copy number variation between strains of mice observed in large families of genes associated with particular functions such as immune response and pheromone signaling [Nature Genetics (2008) 40(7):909–914].

In humans, variation in CNV patterns among individuals has prompted 'association studies', i.e. attempts to correlate patterns of CNVs with particular phenotypic manifestations. Thus, various patterns of CNVs have been associated with conditions such as Alzheimer's disease, Crohn's disease and schizophrenia. (This is analogous to the association studies in relation to single-nucleotide polymorphisms – see the entry SNP GENOTYPING).

[A new wave in the genetics of psychiatric disorders – the copy number variant tsunami: J Psychiatry Neurosci (2009) 34(1):55–59.]

[Copy number variations and cancer: Genome Med (2009) 1(6):62.]

A CNV database was assembled from >2000 disease-free subjects [Genome Res (2009) doi: 10.1101/gr.083501.108].

The interest generated by the CNVs and, in particular, their predictive potential in medicine, has emphasized the need for rapid, accurate (and inexpensive) methods for sequencing the genome. As well as the DIDEOXY METHOD, a new generation of methods called MASSIVELY PARALLEL DNA SEQUENCING is under development.

cordycepin 3′-Deoxyadenosine: a nucleoside analog used e.g. to block RNA synthesis and polyadenylation.

[Use (0.6 mM) in a study of polyadenylated mitochondrial transcripts in *Arabidopsis thaliana*: PLoS ONE (2008) 3(8): e2889].

core promoter (promoter core) Regions of a PROMOTER that are essential for initiation of transcription of the given gene; they are involved in assembly of transcriptional machinery at the initiation site. For example, sequences found in at least some RNA polymerase II promoters include: a TATA box (*Goldberg–Hogness box*, *Hogness box*), an AT-rich sequence about 30 nt upstream of the start site; a TFIIB-recognition element; and a DOWNSTREAM PROMOTER ELEMENT (DPE). All of these sequences act as binding sites for transcription-related factors. Not all of these sequences may be present in the same promoter; for example, some core promoters have a DPE and lack a TATA box.

COS cells Cell lines, originating from monkey tissue, which encode the (large) T ANTIGEN of SV40 virus. COS cells are used e.g. for the replication of vectors that contain the SV40 origin of replication, and also in an EXON TRAPPING protocol.

***cos* site** In PHAGE LAMBDA: the 12-nt 5′ STICKY ENDS which form the terminal parts of the genome within the phage virion; when injected into a host cell the genome circularizes via the sticky ends.

In DNA technology, *cos* sites are exploited e.g. in various constructs (see e.g. COSMID).

cosmid A plasmid containing the (double-stranded) 12-bp *cos* site of PHAGE LAMBDA; cosmids have been widely used e.g. as cloning vectors for preparing genomic libraries.

(See also COSMID WALKING.)

Linearized cosmids (with a *non*-terminal *cos* site) can be mixed with DNA fragments of various lengths prepared with the same restriction enzyme as that used for linearizing the cosmid. Some fragments will bind a cosmid at both ends; if the *cos*-to-*cos* distance is about 40–50 kb (depending on the given fragment), the *cos–cos* sequence is of the correct size for packaging into the head of phage λ. *Cos* sites are cleaved enzymically, and packaging (*in vitro*) occurs in the presence of phage components (see diagram for the overall scheme).

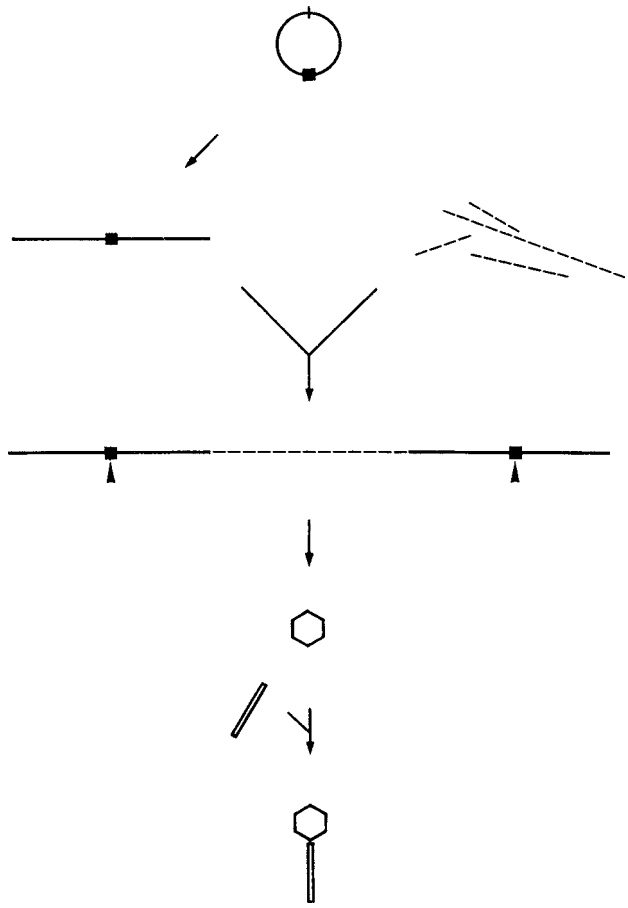
Products for packaging are available commercially: see e.g. GIGAPACK.

Once assembled, phage virions can inject their recombinant DNA into appropriate bacteria.

Several improvements can be made to this basic protocol. For example, the sample fragments (derived e.g. by digestion of genomic DNA with restriction endonucleases) can be size-fractionated prior to mixing with the cosmids; in this way the reaction can be enriched with fragments of appropriate length (i.e. fragments which, when flanked by cosmids, can be cut to provide inserts that are suitable, in size, for packaging).

One problem with the basic scheme is that many of the cosmids may bind to one another and form multimers. This can be prevented by cleaving the terminal 5′-phosphate from both strands of the cosmid (e.g. with ALKALINE PHOSPHATASE); in this case the terminal 5′-phosphate groups on the *inserts* will permit ligation to cosmids and subsequent packaging (see the entry SUPERCOS I VECTOR for an alternative approach).

cosmid walking A procedure, similar to CHROMOSOME WALKING (q.v.), in which a clone contig can be prepared from a COSMID library. Because the inserts in cosmids are quite short, it is possible to use an entire sequence as a probe –



COSMID: a cosmid cloning vector (diagrammatic). The cosmid (*top*) is a small plasmid into which has been inserted the *cos* site (*solid black square*) of phage λ; *cos* sites enable a construct, containing an insert, to be packaged (*in vitro*) in a phage λ head and subsequently to be injected into a bacterial cell for cloning.

Cosmids are cut at a single restriction site, linearizing the molecule and leaving the *cos* site in a non-terminal position. The linearized cosmid molecules are then mixed, in the presence of DNA ligase, with DNA fragments (*dashed lines*) which have been cut with the same restriction enzyme. Cosmids and fragments bind randomly, via their sticky ends, and in some cases a fragment will be flanked by two cosmids (*center*); if, in such a molecule, the distance *cos*-to-*cos* is about 40–50 kb long, then the *cos*–*cos* sequence is the right size for packaging in the phage λ head. Enzymic cleavage at the cos sites (*arrowheads*) creates sticky ends, and packaging of the recombinant molecule in the phage λ head occurs *in vitro* in the presence of phage components. After addition of the tail, the phage particle can inject its recombinant DNA into a suitable bacterium (just like an ordinary phage λ). The injected DNA circularizes, via the *cos* sites, and replicates as a plasmid; its presence in the cell can be detected e.g. by the expression of plasmid-encoded antibiotic-resistance genes.

Figure reproduced from *Bacteria in Biology, Biotechnology and Medicine*, 6th edition, Figure 8.15, page 231, Paul Singleton (2004) John Wiley & Sons Ltd, UK [ISBN 0-470-09027-8] with permission from the publisher.

thus avoiding the need to prepare end probes (as is necessary in chromosome walking).

co-suppression See PTGS.

cotrimoxazole See ANTIBIOTIC (synergism).

counterselection *Syn*. NEGATIVE SELECTION.

coupled transcription–translation (*in vitro*) See CELL-FREE PROTEIN SYNTHESIS.

coxsackieviruses See ENTEROVIRUSES.

cozymase See NAD.

CPD Cyclobutyl pyrimidine dimer (see THYMINE DIMER). (See also COCKAYNE SYNDROME and ULTRAVIOLET RADIATION.)

CPE Cytoplasmic polyadenylation element: a (U-rich) motif, in certain mRNA molecules, involved in CYTOPLASMIC POLYADENYLATION (q.v.).

CpG island One of a number of regions in genomic DNA that contains a high frequency of the dinucleotide 5′-CpG-3′. The DNA in a CpG island typically has a GC% greater than ∼60.

CpG islands (stretches ∼0.5–4 kb) are frequently associated with the 5′ end (i.e. promoter region) of genes, and in these locations the cytosines are often unmethylated; this is significant in that aberrant methylation of these cytosines can lead to gene silencing (which could result in tumor development if the inactivated gene is a tumor suppressor). However, it must not be assumed that methylation is always associated with gene silencing: see the entry METHYLATION. Nevertheless, of those studies carried out on CpG islands many are focused on some aspect of methylation.

Methylated cytosines are especially susceptible to mutation – spontaneous deamination of methylcytosine yields thymine, a normal base which is not a target for the cell's DNA repair mechanisms.

One novel technique for detecting methyl-CpGs is based on the construction of a multimeric form of a naturally occurring methyl-CpG-binding domain (MBD) that is found in a family of proteins associated with repression of transcription. These synthetic (poly-MBD) proteins were reported to be sensitive reagents for indicating DNA methylation levels in (isolated) DNA as well as for the cytological detection of CpG methylation in chromosomes [method: Nucleic Acids Res (2006) 34 (13):e96].

Detection of C5-methylated cytosine residues without using bisulfite treatment, or DNA amplification, was reported with monoclonal antibodies (mAbs) that bind specifically to C5-methylated cytosines in strands base-paired to oligoucleotide probes; mAbs bound to the target (C5-methylated) cytosines were detected by secondary antibodies (dye-conjugated anti-immunoglobulin antibodies) that are visualized via the dye's fluorescence. In this method, probe sequences were designed on the basis of analysis of predicted promoter sequences for possible methylation sites [DNA Res (2006) 13(1):37–42].

CpG methylation can also be monitored by METHYLATION-SPECIFIC PCR or by sequencing of BISULFITE-treated DNA.

(Other methods for monitoring CpG methylation: COBRA, METHYL-BINDING PCR, MULTIPLEX LIGATION-DEPENDENT PROBE AMPLIFICATION and RESTRICTION LANDMARK GENOMIC SCANNING.)

Observations on (human) lymphocytes have suggested that the composition of DNA in CpG islands (sequence, repeats, structure) has a significant role in predisposing CpG islands to methylation [PLoS Genet (2006) 2(3):e26].

A specific, 120-nt sequence from the CpG island associated with the gene encoding adenosine phosphoribosyltransferase has been linked to that gene's ability to resist methylation-directed silencing. This protective activity has been exploited by inserting a copy of the sequence into retroviral vectors in order to protect the (integrated) provirus from methylation-directed silencing [J Virol (2008) 82(16):7818–7827].

(Other relevant entries: BSSHII, ISLAND RESCUE PCR, KAISO PROTEIN, MBD PROTEINS, SSSI, TAGM.)

CpG island methylator phenotype (CIMP) In certain types of tumor cell: hypermethylation of numerous genes, a pattern of methylation greater than the level of abnormal DNA methylation usually seen in tumor cells.

An inverse relationship was reported between CIMP and the hypomethylation of long interspersed nucleotide element-1 (LINE-1) in colorectal cancer [Int J Cancer (2008) 122(12): 2767–2773].

CpGV *Cydia pomonella* GV: see GRANULOSIS VIRUSES.

CpNpG sites In the genomes of mammalian (and other) cells: sites in DNA at which, in at least some cases, methylation of cytosine has important physiologic effects.

The methylation pattern in the sequence CC(A/T)GG may be determined e.g. by carrying out comparative (differential) DNA cleavage with methylation-resistant and methylation-sensitive restriction endonucleases. The activity of restriction enzyme BstNI is not affected by methylation of the internal cytosine, and BstNI may be used as the methylation-resistant endonuclease.

EcoRII recognizes this site, and is sensitive to methylation; however, as it is a type IIE RESTRICTION ENDONUCLEASE, it requires *two* copies of the recognition site for activity. To be useful in comparative DNA cleavage (for detecting methylation at CpNpG sites) the methylation-sensitive enzyme must be able to act on *single* copies of the recognition sequence; this need has been met by an engineered version of EcoRII (designated EcoRII-C) which is able to cleave at a single site in a methylation-sensitive manner [BioTechniques (2005) 38: 855–856].

CR3 *Syn*. CD11b/CD18 (q.v.).

CR4 *Syn*. CD11c/CD18 (q.v.).

***cre* gene** See CRE–LOXP SYSTEM.

CRE The sequence 5′-TGACGTCA-3′, involved in regulating various genes (see also CREB PROTEIN).

Cre-ER^{T2} A chimeric form of the Cre recombinase: see entry CRE–LOXP SYSTEM.

Cre–*loxP* system A recombinase (product of gene *cre*) and its corresponding recognition site (*loxP*) derived from PHAGE P1. Cre–*loxP* is an example of a SITE-SPECIFIC RECOMBINATION system. The (34-bp) *loxP* site consists of a central 8-bp core flanked, on both sides, by a 13-bp PALINDROMIC SEQUENCE;

Cre binds to the palindromic sequences, and the core region contains the cleavage site.

(cf. FLP.)

A sequence of DNA flanked by two *loxP* sites (both in the same orientation) can be excised from a larger molecule, and circularized, by Cre.

Given an inducible *cre*, the excision of a *loxP*–DNA–*loxP* sequence can be achieved *in vivo*, i.e. within cells, when an appropriate technique is used for induction; *in vivo* excision can be manipulated e.g. by transfecting the cells with a Cre-expression vector: see e.g. CELL LINEAGE TRACKING.

Cre–*loxP* can also mediate the *exchange* of a sequence in a circular vector molecule with a sequence in genomic DNA; for this purpose, each of the sequences is flanked by a pair of *loxP* sites (in the same orientation). Dissimilar *loxP* sites (e.g. one wild-type and one modified) can be used to bracket each sequence; this promotes an exchange in which sequences are inserted in a known orientation within their respective molecules. (cf. RECOMBINASE-MEDIATED CASSETTE EXCHANGE.)

DNA flanked by a pair of *loxP* sites in *opposite* orientation can be inverted by Cre.

The Cre–*loxP* system can also mediate the insertion of one vector into another, e.g. the integration of two circular molecules to produce one circular molecule; for this purpose, each molecule contains a single *loxP* site.

Conditional, ligand-dependent expression of Cre – allowing *temporally*-controlled site-specific recombination within cells – was achieved by creating a chimeric recombinase by fusing the Cre gene with a nucleotide sequence encoding a mutated ligand-binding domain of the human estrogen receptor (ER); when transcribed and translated within cells, this chimeric enzyme (Cre-ER^{T2}) can be activated by the administration of the ligand tamoxifen, but it is not activated by estradiol. This approach was exploited for conducting temporally-controlled site-specific recombination in zebrafish [PLoS ONE (2009) 4(2):e4640].

The Cre recombinase (the protein) has been delivered direct to CNS (central nervous system) tissues by fusing it to certain cell-penetrating peptides, an alternative approach to the intra-cellular expression of Cre from transfected plasmids [BMC Biotechnol (2009) 9:40].

Fragments of the Cre recombinase were used in a technique (based on COMPLEMENTATION) which was reported to detect *simultaneous* activity of *two* different promoters in the same cell: see SPLIT-CRE.

Peptides which inhibit Cre have been reported from studies on the HOLLIDAY JUNCTION (q.v.). (See also DAM GENE.)

[Use of the Cre–*loxP* system for making deletions in the bacterium *Lactobacillus plantarum*: Appl Environ Microbiol (2007) 73(4):1126–1135.]

[*loxP* sites used for efficient generation of random chromosome deletions: BioTechniques (2007) 42(5):572–576.]

Cryptic loxP sites

Cryptic *loxP* sites have been reported in genomic DNA in the mouse [Nucleic Acids Res (2007) 35(5):1402–1410].

CREB protein Cyclic AMP (cAMP) response element binding protein: a protein which acts as a transcription factor when it is phosphorylated by an appropriate enzyme at the serine-133 residue.

When the CREB protein is phosphorylated at serine-133, it blinds to a sequence – known as the cAMP response element (CRE), 5′-TGACGTCA-3′ – which is involved in regulating various genes.

Phosphorylation of CREB can be carried out by the enzyme cAMP-dependent protein kinase (PKA).

(See also PATHDETECT SYSTEMS.)

CRF (of HIV-1) Circulating recombinant form.

(See also HIV-1.)

Crigler–Najjar syndrome type I An autosomal recessive disorder characterized by unconjugated hyperbilirubinemia due to deficiency of UDP-glucuronosyltransferase (UDPGT) activity; UDPGT is required for glucuronidation and the normal excretion of bilirubin.

The type I disorder, which apparently involves a complete absence of UDPGT activity, is more severe than the type II disorder – which involves a decrease in, but not total loss of, UDPGT activity; the type II disorder was reported earlier to be inherited in an autosomal dominant manner, but recently it has been reported to be an autosomal recessive condition.

In both forms of the disorder the site of the genetic lesion is reported to be 2q37.

CRM1 See KARYOPHERINS.

cRNA A designation used with various (different) meanings – and therefore able to cause confusion. The meanings ascribed to cRNA include:

(1) Antisense RNA transcribed from dsDNA templates which have been formed (a) in the EBERWINE TECHNIQUE (and in commercial methods based on this technique), and (b) in the NAIMA method (see the entry GMO). This antisense RNA is called aRNA in the MESSAGEAMP ARNA AMPLIFICATION KIT.

(2) Positive-sense, full-length RNA which is complementary to the negative-sense genomic RNA (vRNA) of certain types of virus and which is formed as an intermediate during viral replication (vRNA → cRNA → vRNA). [Examples of use: J Virol (2007) 81(5):2196–2204; J Virol (2008) 82(14):6902–6910; Virol J (2009) 6:22.]

(3) 'Antigenomic' RNA transcribed from dsDNA templates in the engineered generation of certain types of *negative-strand* RNA virus. This involves the intracellular production of full-length RNAs and the simultaneous expression of those viral proteins required for producing the actual ribonucleoprotein complex which is formed during the normal process of viral replication. [Use (e.g.): J Virol (2006) 80(12):5708–5715.]

(4) Positive-sense RNA transcribed from dsDNA during any wholly *in vitro* procedure.

cross-linking (in dsDNA targets) See e.g. TRIPLEX DNA.

cross-linking assay (photo-cross-linking assay) A probe-based assay that was used originally for detecting/quantitating DNA from the hepatitis B virus (HBV) in samples of serum.

Initially, the samples are incubated (65°C/30 minutes) with

a lysis reagent that includes proteinase K and sodium dodecyl sulfate in order to release viral DNA. The viral DNA is then denatured by heat at an alkaline pH. Following cooling and centrifugation, the supernatant is neutralized and an aliquot is used for the assay; probes are added and hybridization carried out at 45°C/20 minutes.

The probes are of two types, each type being targeted to its complementary sequence in the HBV DNA.

Each *capture probe* has a BIOTINylated tag.

Each *reporter probe* carries a number of fluorescein molecules.

Both types of probe incorporate a cross-linking agent (a derivative of 7-hydroxycoumarin) which can be activated by exposure to ultraviolet radiation; when this agent is activated, the bound probe becomes covalently linked to its target sequence.

Following the hybridization stage, the reaction mixture is exposed to ultraviolet radiation for 30 minutes. Subsequently, streptavidin-coated DYNABEADS are added, and these bind to the capture probes (via their biotinylated tags) – thereby also binding the target DNA and the (covalently linked) reporter probes. The beads are washed (while retained by a magnetic field) and then incubated with an antifluorescein antibody–alkaline phosphatase (AP) conjugate; this conjugate binds to reporter probes via their fluorescein tags. The washed preparation is then incubated with ATTOPHOS, at 37°C/60 minutes, and the fluorescent product detected by a microplate reader. For quantitation, the mean signal (from repeat tests) is interpreted against a standard curve.

The original assay [J Clin Microbiol (1999) 37:161–164] was improved by a modified arrangement that was reported to give a ∼10-fold increase in sensitivity when used to detect Factor V Leiden mutations in clinical samples [Clin Chem (2004) 50:296–305].

An advantage of a cross-linking assay is that manipulation/washing etc. can be carried out under conditions that would be likely to denature conventional probe–target complexes.

crossing over Recombination involving breakage of strands in each of two juxtaposed DNA duplexes, cross-wise exchange of ends, and ligation. Crossing over occurs e.g. between non-sister chromatids during meiosis.

(See also RECOMBINASE.)

crown gall A PLASMID-mediated plant neoplasm that is found typically at the crown (i.e. stem–root junction), primarily in a range of gymnosperms and dicotyledonous angiosperms.

Development of the neoplasm follows infection of a wound by a virulent (plasmid-containing) strain of certain species of the bacterium *Agrobacterium*, commonly *A. tumefaciens* (but crown gall in the grape is caused by *A. vitis*). Crown gall may cause stunting/death of affected plants; important economic losses have been caused (e.g. in stone-fruit trees and vines) in e.g. Europe, the USA and Australia.

In crown gall, bacterial virulence (the ability to promote tumorigenesis) is conferred by a large plasmid called the Ti (tumor-inducing) plasmid. The genes directly responsible for tumorigenesis are in a subregion of the plasmid referred to as T-DNA; during infection of the plant, a single-stranded form of T-DNA is transferred from the bacterium to the plant.

Some ∼25 of the Ti genes (in regions outside T-DNA) are needed for transfer of T-DNA to the plant. Certain of these genes encode elements of a TWO-COMPONENT REGULATORY SYSTEM – VirA (the sensor) and VirG (response regulator). VirA responds to certain phenolic compounds (e.g. ACETOSYRINGONE) that occur in plant wounds and to certain monosaccharides; variant forms of VirA may respond to sensor-specific plant signals.

When activated, VirG promotes transcription of a number of plasmid genes in the *vir* operon, some of which encode a site-specific endonuclease complex which nicks one strand of the Ti plasmid at unique sites in a pair of 24-bp direct repeats flanking T-DNA. The ssDNA copy of T-DNA, with a protein (VirD2) linked covalently to the 5′ end, is transferred from the bacterium to the plant cell nucleus.

In the plant cell, T-DNA becomes inserted into the nuclear DNA. T-DNA encodes certain phytohormones (e.g. AUXINS, CYTOKININS) considered to be responsible for tumorigenesis.

A detailed review [Microbiol Mol Biol Rev (2003) 67:16–37] covers DNA transfer from *Agrobacterium* to the plant.

Tumor cells (in the plant) produce certain T-DNA-encoded products (*opines*) that (according to the given strain of *Agrobacterium*) may be compounds of the *octopine* family or of the *nopaline* family. The opines produced by tumor cells can be used as the sole source of carbon and nitrogen by the particular strain of *Agrobacterium* which is infecting the plant; the Ti plasmid (in the bacterium) encodes enzymes which degrade the opines that are encoded by the T-DNA.

Transfer of the Ti plasmid from Ti^+ strains to Ti^- strains of agrobacteria (enlarging the pathogen population) is regulated e.g. by a QUORUM SENSING mechanism in which activation of transfer ability depends on the population density of *donor* cells. As in various other Gram-negative bacteria, the agents which mediate quorum sensing (*autoinducers*) are *N*-acyl-L-homoserine lactones (AHLs). In at least certain strains of *Agrobacterium*, this quorum sensing system is dependent on the presence of opines, formed by the tumor cells.

The ability of *A. tumefaciens* to promote transfer of DNA into plant cells has been exploited in a widely-used BINARY VECTOR SYSTEM for introducing viral DNA (or cDNA) into plants.

(See also AGROINFECTION.)

CRP cAMP-receptor protein: see CATABOLITE REPRESSION.

cruciform (1) (*adj*.) Cross-shaped.

(2) (*DNA technol*.) (*noun*) See e.g. PALINDROMIC SEQUENCE.

cryptic plasmid Any plasmid which has no known phenotypic effects.

cryptic splicing An aberrant form of SPLICING which occurs at a *cryptic splice site*. A cryptic splice site is a pre-existing but 'silent' splice site – either a donor site or an acceptor site – which may be activated (i.e. used as a splice site) if a normal, vicinal splice site is inactivated by mutation. Whether or not

a given cryptic site becomes active (following mutation in a vicinal splice site) may depend not only on the presence of an inactivated vicinal site but also on the local 'splicing environment' – including e.g. the presence of an appropriate exonic splicing enhancer (ESE).

Some draw a distinction between cryptic splice sites in the genome and splice sites formed *de novo* through mutation in an exon or an intron. Others do not recognize this distinction.

The involvement of a cryptic splice site can affect the biological activity of the product. Thus, studies on a splicing site mutation reported that the disruption of normal splicing was accompanied by activation of a downstream cryptic site; the product was full-length (with a short insert of 18 amino acid residues) but was functionally defective [J Mol Cell Cardiol (2008) 44(3):502–509].

cryptic viruses Virus-like particles seen in various (symptomless) plants. Some (e.g. beet cryptic viruses) contain dsRNA.

CspD In *Escherichia coli*: a protein which can be induced by nutrient deprivation; it also appears in the stationary phase of growth at 37°C. CspD is also included within the category of cold-shock proteins.

CSPD® A 1,2-dioxetane substrate that emits light (λ = 477 nm) when dephosphorylated by ALKALINE PHOSPHATASE. It is used e.g. to detect enzyme-labeled probes.

(See also CHEMILUMINESCENCE.)

CSRE Carbon source responsive element (in *Saccharomyces cerevisiae*): a consensus sequence in the promoters of certain genes (e.g. genes involved in gluconeogenesis); if cells are supplied with glucose, CSRE elements mediate repression of the corresponding genes.

(See also CATABOLITE REPRESSION.)

C_t (threshold cycle) See REAL-TIME PCR.

CTCF A protein which (in vertebrates) is reported to have roles in e.g. transcriptional activation, regulation of nuclear organization and blocking transcriptional enhancer activity; it is a major binding protein of CHROMATIN INSULATORS.

CTCF was reported to have various functions in oocyte and preimplantation embryo development [Development (2008) 135(16):2729–2738].

CTCF has also been implicated e.g. in X-INACTIVATION.

CTn*916* See CONJUGATIVE TRANSPOSITION.

CTnDOT See CONJUGATIVE TRANSPOSITION.

CTnERL See CONJUGATIVE TRANSPOSITION.

CTV Cancer terminator virus – see e.g. MELANOMA.

Culex A genus of mosquitoes that behave as vectors in certain diseases – e.g. Murray Valley fever, caused by a flavivirus.

Culicoides A genus of biting midges that behave as vectors e.g. in African horse sickness and bluetongue (both caused by a species of *Orbivirus*).

cullin-RING complexes A large group of ubiquitin ligases (see the entry PROTEASOME).

[Cullin-RING complexes: global regulation and activation cycles: Cell Div (2008) 3:7.]

Curie point pyrolysis PYROLYSIS involving a *ferromagnetic* wire which is heated inductively, i.e. within a high-frequency alternating magnetic field. Rapidly (e.g. in 0.1 second) the temperature of the wire reaches the Curie point: a temperature which is maintained by the alternating magnetic field – but which cannot be exceeded because, above a certain temperature, the wire is no longer susceptible to inductive heating. The Curie point varies with the chemical composition of the wire; iron–nickel wires often have a Curie point of 510°C.

curing (of bacterial plasmids) Loss of plasmid(s) from bacteria, which retain viability. Spontaneous curing may occur e.g. if the plasmid's partition system is defective. Experimentally, it can be achieved e.g. by the use of sublethal doses of certain INTERCALATING AGENTS which inhibit plasmid replication; increasing numbers of plasmid-free cells are produced when a growing population of bacteria is exposed to such agents. Thus, curing may be carried out e.g. with certain acridines or (in *Pseudomonas*) with mitomycin C.

Some (temperature-sensitive) plasmids can be cured simply by growing the host cells at a non-permissive temperature. For example, the derivative plasmid pWV01 is stable at 30°C but lost by growth at or above 37°C.

(See also SUICIDE VECTOR.)

cut-and-paste mechanism (in transposable elements) See the entry TRANSPOSABLE ELEMENT (figure and legend).

cutting DNA See DNA CUTTING AND ASSEMBLY.

CXCR4 (in HIV-1 infection) See the entry HIV-1.

***cya* gene** In *Escherichia coli*: the gene encoding ADENYLATE CYCLASE.

(See also BACTERIAL TWO-HYBRID SYSTEM.)

cyanobacteria ('blue-green algae') A (non-taxonomic) category of heterogeneous bacteria characterized by the ability to carry out oxygenic (oxygen-producing) photosynthesis, i.e. a mode of photosynthesis similar to that found in eukaryotes (such as algae) – and unlike the anoxygenic photosynthesis carried out (anaerobically) by some other bacteria. The organisms range from unicellular to filamentous and complex forms, with various types of pigmentation. The cell envelope is essentially of the Gram-negative type, although features of Gram-positive organization can be seen in some species. Flagella are absent, but some strains can move e.g. by a so-called *gliding motility*.

[Antisense RNAs in gene regulation: Mol Syst Biol (2009) doi: 10.1038/msb.2009.63; Genome database update: Nucleic Acids Res (2009) doi: 10.1093/nar/gkp915.]

cybrid ('cytoplasmic hybrid') A cell prepared by the transfer of mitochondria from a given type of cell to a RHO-ZERO CELL or to another type of cell; cybrids are used e.g. to investigate the effects of mutations in mitochondrial DNA (mtDNA) on functions such as oxidative phosphorylation and cell growth.

The mouse model is commonly used in studies on mtDNA – e.g. cytoplasts from a mouse cell line may be fused with ES cells (i.e. embryonic STEM CELLS) and the resulting cybrids injected into a BLASTOCYST for development; in this way the effects of specific mutations in mtDNA can be investigated in the progeny. Transmitochondrial mice have been discussed in the context of mouse models for studying defects in mtDNA [Biochim Biophys Acta (2009) 1793(1):171–180].

(See also HETEROPLASMY.)

CycLiC See CYCLIC LIGATION AND CLEAVAGE.

cyclic AMP (cAMP) Adenosine 3′,5′-monophosphate, in which phosphate links the 3′ and 5′ positions of the sugar residue. cAMP is synthesized from ATP by ADENYLATE CYCLASE.

cAMP is a key regulatory molecule in both eukaryotes and prokaryotes. It is involved e.g. in signal transduction pathways and (in prokaryotes) in CATABOLITE REPRESSION.

cAMP is converted to AMP by cAMP phosphodiesterase (EC 3.1.4.17).

cyclic AMP response element binding protein See the entry CREB PROTEIN.

cyclic ligation and cleavage (CycLiC) A DNA SEQUENCING method involving a cyclical sequence of reactions in which a fluorescently labeled (color-coded) oligonucleotide is ligated to the primer and is then cleaved to remove all except the (terminal) ligated nucleotide. This cycle is repeated for each of the unknown nucleotides in the strand being sequenced.

At each step, the primer is exposed to four populations of oligonucleotides; in a given population all the oligos have the same ligatable terminal nucleotide (i.e. A, T, C or G) and all have the same fluorescent label. Subsequent cleavage of the ligated oligo occurs at a *ribo*nucleotide adjacent to the ligated deoxyribonucleotide. At each step, in which a nucleotide is added to the primer, the identity of the paired nucleotide (in the unknown strand) is indicated by the fluorescent label of the corresponding oligo.

[Method: Nucleic Acids Res (2009) 37(1):e5.]

cyclin-dependent kinase 1 See NUCLEAR LAMINA.

cyclobutyl thymine dimer See e.g. THYMINE DIMER.

cycloheximide An antibiotic produced e.g. by certain strains of *Streptomyces griseus*: β-[2-(3,5-dimethyl-2-oxocyclohexyl)-2-hydroxyethyl]-glutarimide. Cycloheximide is active against various eukaryotes; it binds to the 60S ribosomal subunit and blocks translocation. This agent is used in TOEPRINTING [e.g. Mol Cell Biol (2008) 28(13):4227–4239].

cyclolysin See ADENYLATE CYCLASE.

cystic fibrosis A congenital disorder that typically involves a defective membrane protein (*cystic fibrosis transmembrane conductance regulator*, CFTR, ecoded by the *CFTR* gene at 7q31.2). [The 3905insT mutation: Eur J Hum Genet (2009) doi: 10.1038/ejhg.2009.140.] The CFTR is an ATP-energized 'ABC transporter' protein associated with the transmembrane transport of chloride (Cl^-). This disorder is characterized by abnormal secretion – most notably in the respiratory tract.

In those suffering from cystic fibrosis, respiratory-tract infections – e.g. those with mucoid strains of *Pseudomonas aeruginosa* – can be refractory to antibiotic treatment, constituting a major risk; the potential of GENE THERAPY is being actively investigated.

A caprine adeno-associated virus (AAV), which is resistant to human neutralizing antibodies, and which has a marked tropism for (murine) lung tissue, was suggested as a vector for gene therapy in cystic fibrosis.

cytidine A RIBONUCLEOSIDE. Deoxycytidine, the corresponding deoxyribonucleoside, is sometimes referred to as 'cytidine'.

cytidine deaminase Any enzyme which deaminates cytidine to uridine.

One example is APOBEC-1 – part of an RNA editing complex that acts e.g. on transcripts of apolioprotein B (see RNA EDITING).

Another example is AID (ACTIVATION-INDUCED CYTIDINE DEAMINASE), which may also be involved in RNA editing.

APOBEC-1 and AID were found to have distinct functional properties when examined in an experimental yeast system [Mol Immunol (2006) 43(4):295–307].

cytogenetics A branch of science in which cellular features, e.g. structure and function of the chromosomes, are studied in the context of inheritance.

cytokinins (phytokinins) Phytohormones (plant hormones) that stimulate metabolism and cell division; they are substituted adenines synthesized mainly at the root apex.

(See also AUXINS and CROWN GALL.)

cytoplasmic genes (extrachromosomal genes; also: extranuclear genes) Genes other than those which occur in a prokaryotic chromosome or in the eukaryotic nucleus. In prokaryotes, the term commonly refers to genes in an autonomously replicating PLASMID, while in eukaryotes it generally refers e.g. to genes in mitochondria and (when present) chloroplasts.

cytoplasmic incompatibility (CI) See, for example, *Wolbachia* in the entry GMMS.

cytoplasmic inheritance (extrachromosomal inheritance, non-Mendelian inheritance) Inheritance that involves factors such as CYTOPLASMIC GENES, i.e. inheritance unrelated to nuclear or chromosomal genes and not subject to Mendelian laws.

(See also MATERNAL INHERITANCE.)

cytoplasmic polyadenylation In the cytoplasm: elongation of the poly(A) tail on certain mRNA molecules in the oocytes and embryos of some species (including the mouse and frog). On initial synthesis, these mRNAs have a short poly(A) tail and are (as a consequence) translationally repressed. At the appropriate stage in development, the poly(A) tail is extended and the mRNAs become translationally active.

In one model for cytoplasmic polyadenylation, the mRNA contains a cytoplasmic polyadenylation element (CPE) which binds the CPEB1 protein. CPEB1 recruits various proteins – forming a large complex which includes the poly(A)-specific ribonuclease PARN and a poly(A) polymerase (Gld2); PARN is dominant over Gld2, keeping the tail short. Subsequently, on phosphorylation of CPEB1, PARN is released and Gld2 is free to elongate the tail.

Cytoplasmic polyadenylation was reported to be involved e.g. in development of the retinal axon in *Xenopus* [Neural Dev (2009) 4:8].

cytoplast A (eukaryotic) cell from which the nucleus has been removed (i.e. an *enucleated* cell).

(See also CYBRID and SCNT.)

cytosine deaminase (in siRNA selection) See RNA INTERFERENCE.

CytoTrap™ two-hybrid system A two-hybrid system (Strata-

gene, La Jolla CA) for detecting protein–protein interaction within the *cytoplasm* of the yeast *Saccharomyces cerevisiae*; this approach is therefore distinct from those systems (e.g. the YEAST TWO-HYBRID SYSTEM) that detect protein–protein interaction by assaying for *transcriptional* activation in the nucleus. (See also BACTERIAL TWO-HYBRID SYSTEM.)

As in other two-hybrid systems, the 'bait' protein forms part of a fusion protein encoded by a plasmid vector. The target protein (one of various proteins to be assayed) also forms part of a fusion protein encoded by a separate plasmid vector.

Both vectors are introduced into *S. cerevisiae*.

The pMyr vector includes the gene of the target protein that is fused to a sequence encoding the *myristylation membrane localization signal*; when expressed in the yeast host cell, the localization signal becomes anchored to the cell membrane – together with the target protein, which is therefore accessible in the cytoplasm.

The pSos vector contains the gene of the bait protein fused to a sequence from the gene of the human Sos protein, hSos. hSos is a *guanosine nucleotide exchange factor* which, when brought into close proximity to a (membrane-associated) Ras protein *by bait–target binding*, can activate Ras by promoting the exchange of bound GDP for GTP; when activated in this way, Ras initiates intracellular signaling, leading to growth of the cell.

The strain of *S. cerevisiae* used as host cell has a mutation in the *cdc25* gene (the homolog of the gene encoding hSos). This makes the host cell temperature-sensitive: growth occurs at 25°C but not at 37°C. However, when hSos is internalized, and activating Ras, it can complement the defect in *cdc25* – so that growth can occur at 37°C. Hence, growth of the host cell at 37°C is used as an indicator of bait–target interaction.

Examples of use of the CytoTrap™ two-hybrid system

The CytoTrap™ two-hybrid system has been used e.g. for: studies on the association of heat-shock transcription factor 4b with a kinase and a tyrosine phosphatase [Mol Cell Biol (2006) 26(8):3282–3294]; studying regulation of APOPTOSIS by the p8/prothymosin α complex [Proc Natl Acad Sci USA (2006) 103(8):2671–2676]; studying interaction of the Zfra protein with TRADD and e.g. its negative regulation of the activation of NF-κB [BMC Mol Biol (2007) 8:50]; studying calcium channels in *Cryptococcus neoformans* [Eukaryotic Cell (2008) 7(7):1118–1126]; and for studying the role of the Orc6 protein of *Drosophila* [Mol Biol Cell (2009) 20(1):270–281].

D

D (1) A specific indicator of ambiguity in the recognition site of a RESTRICTION ENDONUCLEASE, or in another nucleic acid sequence; for example, in the recognition site GDGCH↓C (enzyme SduI) the 'D' indicates A or G or T. (In RNA 'D' indicates A or G or U.) In the example, 'H' is A or C or T.
(2) L-Aspartic acid (an alternative to Asp).

D loop (displacement loop) A loop of ssDNA that is formed, for example, when a short ssDNA fragment hybridizes with a complementary sequence in one of the strands of dsDNA; the displaced, homologous strand is the D loop. (A D loop can be used e.g. for bisulfite-mediated site-directed mutagenesis.)

The formation of a D loop can be demonstrated *in vitro* in a negatively supercoiled ccc dsDNA molecule. The fragment is taken up, and the D loop forms spontaneously, as the creation of the D loop gives rise to a more relaxed state.

D loop formation is promoted e.g. by the RecA protein.

If a short *ssRNA* fragment hybridizes in this way, the resulting loop of ssDNA is called an *R loop*. If the ssRNA is a *mature* mRNA transcribed from a SPLIT GENE, then, when it binds to the gene, R loops will be formed at the exons; intron regions (which have no homologous sequences in the mature mRNA) remain as dsDNA.

DAB 3,3′-diaminobenzidine: a substrate for HORSERADISH PEROXIDASE.

dabcyl 4-((4-(dimethylamino)phenyl)azo) benzoic acid: a non-fluorescent quencher of fluorescence which is used e.g. as a FRET acceptor molecule; dabcyl appears to act as a universal quencher – i.e. it quenches the emission from a wide range of fluorophores regardless of their emission profile.

(See also PYRENE BINARY PROBE.)

[Example of use: PLoS ONE (2009) 4(1):e4183.]

(cf. DABSYL CHLORIDE.)

dabsyl chloride 4-dimethylaminobenzene-4′-sulfonyl chloride: a reagent used e.g. for labeling proteins. (cf. DABCYL.) [Use (e.g.): Appl Environ Microbiol (2007) 73:6450–6459.]

Dacogen® (decitabine) 5-AZA-2′-DEOXYCYTIDINE, marketed by MGI Pharma, Bloomington MN, USA.

DAF (direct amplification fingerprinting) See AP-PCR.

Dam-directed mismatch repair See MISMATCH REPAIR.

***dam* gene** In various prokaryotes, including *Escherichia coli*: a gene encoding the enzyme Dam methylase which methylates DNA at the N-6 position of adenine within the sequence 5′-GATC-3′ (a so-called 'Dam site'); during DNA replication the daughter strand is methylated soon after its synthesis. Inactivation of the *dam* gene blocks methylation at Dam sites. (See also MODIFICATION.)

In *E. coli*, Dam methylation is a factor in the MISMATCH REPAIR system and is involved in the regulation of certain genes that have Dam sites in their promoters. For example, in transposon Tn*10*, transposition is inhibited by Dam methylation within the promoter of the transposase gene; inhibition results from a reduction in transposase activity. Transposition may occur during DNA synthesis when the replication fork has passed the Dam site in the transposase promoter – but before Dam methylation has occurred at this site.

Other genes affected by Dam methylation include the *cre* gene of phage P1 (see CRE–LOXP SYSTEM) and the *trpR* gene of *E. coli*.

In *E. coli*, the chromosomal origin of replication, *oriC*, contains a number of Dam sites.

(See also DCM GENE.)

Dam methylase See DAM GENE.

Dam methylation See DAM GENE.

Dam site See DAM GENE.

damaged DNA See entries: CODIS, FORENSIC APPLICATIONS, LONG PCR, WHOLE-GENOME AMPLIFICATION and also Y FAMILY DNA polymerases.

(See also ANCIENT DNA.)

dangling ends (in a microarray) See MICROARRAY.

dansyl (DNS, Dns) The group [5-dimethylaminonaphthalene-1-sulfonyl-] which can be incorporated e.g. in a protein in order to provide a fluorescent derivative.

DAPI 4′,6-Diamidino-2-phenylindole dihydrochloride: a fluorescent stain for dsDNA which binds preferentially to AT-rich regions. DAPI binds to RNA with lower affinity; moreover, the fluorescence emission maximum of a DAPI–RNA complex (∼500 nm) differs from that of a DAPI–dsDNA complex (∼460 nm).

On binding to dsDNA the fluorescence of DAPI is reported to be enhanced by about 20-fold.

[Uses of DAPI (e.g.): PLoS Pathog (2009) 5(5):e1000417; Proc Natl Acad Sci USA (2009) 106(18):7479–7484.]

darbepoietin-α See BIOPHARMACEUTICAL (table).

dark quenching A FRET-type technique which has been used e.g. to study the mechanism of homologous recombination – specifically, separation of the two strands of the target duplex prior to hybridization of the presynaptic filament with one of the strands. The fluorophore FAM was used to label a strand of the target duplex, while the quencher DABCYL was used in the other strand, the fluorophore and quencher being in close proximity; hence, prior to separation of the two strands there was no signal. However, on strand separation the fluorophore becomes unquenched so that a fluorescent signal is produced. This procedure has the advantage of a signal-to-background ratio reported to be significantly higher than that obtained by the conventional FRET arrangement [BioTechniques (2006) 40(6):736–738].

Dark quenching was also used to study a strand-exchange event [Biochemistry (2007) 46(24):7163–7173].

DASH DYNAMIC ALLELE-SPECIFIC HYBRIDIZATION.

database In the present context: any organized and retrievable collection of data on particular sequences of nucleic acid or whole-genome sequences (see e.g. INTERNATIONAL NUCLEOTIDE SEQUENCE DATABASE COLLABORATION and TIGR MICROBIAL DATABASE); polypeptide sequences; gene expression profiles; genomic characteristics (see e.g. CODIS) etc. or

data on particular methods (see e.g. GMDD), types of enzyme (e.g. RESTRICTION ENDONUCLEASES: see REBASE), or plants (see e.g. the entry MAIZEGDB).

(See also ONLINE MENDELIAN INHERITANCE IN MAN, OMIA and the databases mentioned in entries ESCHERICHIA COLI, MITOCHONDRIAL DNA, P53, QUADRUPLEX DNA, RNA MODIFICATION PATHWAYS, SPOLIGOTYPING and STR.)

The phrase DNA DATABASE frequently refers to a database, maintained e.g. by a government or law-enforcement agency, which contains a record of the DNA sequences of specific individuals.

There are also *collections* of databases [e.g. The Molecular Biology Database Collection (2008 update): Nucleic Acids Res (2008) 36(Database issue):D2–D4].

Some databases have limited access; others are available to the public.

Reliability of databases

The reliability of some public databases has been questioned. For example, significant discrepancies have been noticed in the genome sequences of different strains of (the prokaryote) *Pyrococcus furiosus* [see J Bacteriol (2005) 187:7325–7332]. Among human SNP databases, it was suggested that some of the reported SNPs may be associated with *editing* sites rather than representing actual SNPs [Nucleic Acids Res (2005) 33 (14):4612–4617]. Moreover, the database sequences of many cancer-associated genes may reflect atypical splice forms of the type found in tumor cells [see Nucleic Acids Res (2005) 33(16):5026–5033].

dATPαS Deoxyadenosine α-thiotriphosphate (α-thiophosphoryl dATP).

This modified nucleotide is used in various types of method (see e.g. entries PYROSEQUENCING and SDA).

Dbh A DNA polymerase (of the Y FAMILY) which occurs in the archaean *Sulfolobus acidocaldarius*.

dbSNP The NCBI website on single-nucleotide polymorphisms (see the entry SNP), accessible at:
http://www.ncbi.nlm.nih.gov/projects/SNP/

DC-SIGN (CD209) DC-specific ICAM3-grabbing nonintegrin: a cell-surface receptor molecule (a C-type lectin) that occurs on dendritic cells (and also on other types of cell); it binds to the envelope glycoprotein gp120 of HIV-1 (human immunodeficiency virus), and is reported to act as a receptor for e.g. cytomegalovirus (CMV), the Ebola virus, and the bacterium *Mycobacterium tuberculosis*.

Expression of DC-SIGN on a subset of B lymphocytes (B cells) was increased, *in vitro*, on stimulation with interleukin 4 (IL-4) and the CD40 ligand. Activated B cells bound and internalized certain strains of HIV-1 and were able to mediate infection of T cells with the virus [PLoS Pathogens (2006) 2(7):e70].

It appears that DC-SIGN is involved in some of the cases in which dendritic cells promote the infection of $CD4^+$ T cells with HIV-1 [J Virol (2007) 81(5):2497–2507; J Virol (2007) 81(5):2519–2523].

DCDS Donor conjugal DNA synthesis: see CONJUGATION.

***dcm* gene** In *Escherichia coli*: a gene encoding an enzyme that methylates DNA at the C-5 position in the (second) cytosine residue within the sequence 5′-CC(A/T)GG-3′. Inactivation of the gene blocks methylation at these sites.

(See also DAM GENE.)

***dda* gene** (of phage T4) See HELICASE.

DDBJ DNA DataBank of Japan.

ddF DIDEOXY FINGERPRINTING.

DDMR Dam-directed mismatch repair: see MISMATCH REPAIR.

ddNTP 2′,3′-Dideoxyribonucleoside triphosphate: any synthesized nucleoside triphosphate used e.g. for chain termination in the DIDEOXY METHOD of DNA SEQUENCING (q.v.), and for SINGLE-BASE EXTENSION (q.v.).

DEAD-box proteins Various types of protein which have the motif: Asp-Glu-Ala-Asp ('DEAD' by the single-letter designations of the amino acids); they include, for example, eIF4A (a eukaryotic initiation factor in protein synthesis) and certain proteins involved in the SPLICING of pre-mRNA.

DEAE The diethylaminoethyl group which is bound to various types of matrix (e.g. cellulose) used for anion exchange.

decitabine *Syn.* 5-AZA-2′-DEOXYCYTIDINE.

decoyinine See PSICOFURANINE.

defective interfering particle (DI particle) A defective derivative of a normal, non-defective (*standard*) virus; the presence of the standard virus is needed for replication of the DI particle, although this lowers the yield of the standard virus. In cell cultures, and in experimental animals, at least some types of DI particle have been reported to attenuate the pathogenic effects of a virulent standard virus.

degeneracy (of the genetic code) See CODON.

degenerate primer One of a set of primers used to amplify a given region of DNA whose complete sequence is not known – but which has been predicted e.g. on the basis of the amino acid sequence of an encoded protein. Because, as a result of *degeneracy* in the genetic code, a given amino acid may be specified by more than one type of CODON, the nucleotide sequence of the coding region of the given protein could be any of various combinations of nucleotides; hence, a number of different primers (referred to as degenerate primers) may be used in an attempt to find one primer complementary to the *actual* sequence of nucleotides that encodes the protein.

(cf. DEGENERATE PROBE.)

degenerate probe One of a set of probes that are designed to be complementary, collectively, to most or all of the possible (predicted) variant forms of a given target sequence. A set of degenerate probes may be used e.g. when the target sequence is not completely known.

(cf. DEGENERATE PRIMER.)

degraded DNA See e.g. ANCIENT DNA and DAMAGED DNA.

(See also Y FAMILY (of DNA polymerases).)

degradosome (RNA degradosome) A multi-component complex (e.g. in *Escherichia coli*) that processes/degrades RNA molecules; it includes RNase E, polynucleotide phosphorylase and Rh1B RNA helicase, and has also been reported to include polynucleotide phosphate kinase, enolase, DnaK and

GroEL. The complex degrades certain types of mRNA.

degron See N-END RULE.

delavirdine A NON-NUCLEOSIDE REVERSE TRANSCRIPTASE INHIBITOR.

delayed enrichment method See AUXOTROPHIC MUTANT.

deletions (nested) See NESTED DELETIONS.

Δ*lon* (delta *lon*) See the entry ESCHERICHIA COLI (table).

ΔT_m°C See UNLABELED PROBE.

ΔVII-Ets-1 An alternatively spliced transcription factor which was used for activating latent HIV-1 (see entry AIDS, section *Quiescent HIV-1*).

demethylation (of DNA) Removal of the methyl group ($-CH_3$) from bases in DNA, or the inhibition of DNA methylation *in vitro* or *in vivo* by using certain inhibitors of DNA methyltransferases.

Methods involving methylation-inhibition are used e.g. to investigate METHYLATION in DNA and are also used in anti-cancer chemotherapy. For anti-cancer therapy, the object may be e.g. to de-methylate a tumor-suppressor gene which has been inactivated by methylation in the promoter region; such anti-cancer therapy may include both an inhibitor of DNA methyltransferases and an inhibitor of HDACs (see HDAC) – the latter in order to promote 'open' (accessible) chromatin in the region of the target gene(s). (See also CANCER THERAPY.)

A number of methods use agents such as 5-AZA-2′-DEOXYCYTIDINE which inhibit *in vivo* methylation. This nucleoside analog is phosphorylated and then incorporated into the DNA of actively dividing cells, after which it binds and inactivates DNA methyltransferase 1; this causes a generalized fall in the 'maintenance' methyltransferase activity in the cell, leading to global demethylation or under-methylation.

Laboratory studies on demethylation

One problem with drug-induced demethylation – for example with 5-aza-2′-deoxycytidine – is that, in a given culture, cells in different stages of the cell cycle may respond differently (by exhibiting different levels of demethylation) on exposure to the drug. Moreover, cells which are not actively dividing will not be affected by the treatment. Thus, a drug-treated cell culture can consist of a mixed population in which demethylation has occurred to varying degrees among the cells.

Drug-mediated DNA demethylation has been studied e.g. by the GFP REACTIVATION TECHNIQUE (q.v.).

(See also INTRACISTERNAL A-PARTICLE.)

Some of the techniques used for demethylation involve the direct inactivation of DNA methyltransferases e.g. by RNAi or by other means. Problems with methods that involve *direct* inactivation of Dnmt1 expression by RNAi include failure of the intervention in a proportion of the cells, again resulting in heterogeneity in demethylation among the population of cells in a culture.

denaturing-gradient gel electrophoresis See DGGE.

dendrichip A MICROARRAY prepared by the covalent binding of NH_2-modified DNA probes (via their 5′ termini) to dendrimer-activated glass slides (*dendrislides*); each *dendrimer* is a nanometric spherical structure, a polyfunctional polymer, that links the probe to the slide.

Dendrimers can be prepared with various types of reactive surface group – for example, aldehyde, thiol or epoxy. (See the entry below for further information on dendrimers – and thier uses in DNA technology.)

dendrimer An (engineered) macromolecular structure consisting of a central core from which arise many radially arranged dichotomously branching arms; the free ends of these multi-branched arms (i.e. at the surface of the dendrimer) may carry any of various functional groups, the choice depending on the purpose for which the dendrimer was designed.

In the nomenclature, the *core* is referred to as G0; the arms which are attached *directly* to it are the *first-generation* (G1) monomers. Each G1 monomer carries two second-generation (G2) monomers – and each of the G2 monomers carries two third-generation (G3) monomers, and so on.

Uses of dendrimers in DNA technology

Dendrimers have various uses in DNA technology. Thus, for example, DNA dendrimers (with arms consisting of branched ssDNA) were conjugated to HORSERADISH PEROXIDASE and used in a signal-amplification-based method for the detection of MICRORNAS in microtiter plates [Biotechniques (2006) 41 (4):420–424].

(See also DENDRICHIP.)

Dendrimers are also being considered for use as vectors for the delivery of nucleic acid into (e.g.) tumor cells. Given that dendrimers can bind nucleic acids, and given their potential for extensive surface-modification, there is clearly scope for combining targeting with gene-delivery functions in a single dendrimer. In this context, it has been reported that targeting of tumor cells can be promoted by surface pegylation of the dendrimers; moreover, ligand-based uptake of dendrimers by tumor cells has already been exploited for drug delivery to cells. [Nanoparticles and cancer (a review, with emphasis on dendrimers): Int J Nanomed (2009) 4:1–7.]

Dendrimers containing polyamidoamine (PAMAM) may be useful as vectors for the delivery of siRNA to cells. However, the primary amine groups at the surface of these dendrimers are associated with a certain amount of cytotoxicity. Studies have been carried out to determine the effect of acetylation of the amine groups on the ability of these dendrimers to deliver siRNA into cells; while the acetylation of amines was found to reduce cytotoxicity, a significant degree of acetylation was found to cause a reduction in the efficiency of gene silencing by the PAMAM vectors [BMC Biotechnol (2009) 9:38].

[Multiscale modeling of dendrimers, and their interactions with lipid bilayers and polyelectrolytes: Molecules (2009) 14 (1):423–438.]

dendrislide See DENDRICHIP.

deoxyadenosine α-thiotriphosphate (dATPαS) The modified, synthetic form of dATP in which sulfur replaces phosphorus at the α position.

dATPαS is used, for example, in PYROSEQUENCING and in

SDA.

deoxyribonuclease See DNASE.

deoxyribozyme (DNA enzyme or DNAzyme) Any molecule of DNA which has catalytic activity, typically for a nucleic acid [e.g. RNA: Nucleic Acids Res (2005) 33(19):6151–6163].

The sequence-specific, Mn^{2+}- and Zn^{2+}-dependent cleavage of DNA has also been reported [see Nature Chem Biol (2009) doi: 10.1038/nchembio.201].

Some Zn^{2+}-dependent deoxyribozymes are apparently able to ligate molecules of RNA.

(See also BINARY DEOXYRIBOZYME LIGASE.)

(cf. RIBOZYME.)

DEPC DIETHYLPYROCARBONATE.

derepressed Refers e.g. to a gene or operon which, previously inactive (i.e. repressed), is being expressed – or to a gene or operon which is not subject to the repression which is normal for related molecules (see e.g. F PLASMID).

DES® The *Drosophila* expression system (from Invitrogen, Carlsbad CA): a system designed for the expression of recombinant proteins from simple types of vector in *Drosophila* Schneider S2 cells. (Expression of heterologous proteins in insect cells can also be achieved in a baculovirus-mediated format: see BACULOVIRUS EXPRESSION SYSTEMS.)

The pMT/V5-His vector permits expression of recombinant proteins regulated by the *Drosophila* copper-sulfate-inducible metallothionein gene promoter. Purification of the expressed protein is facilitated by a SIX-HISTIDINE TAG. An alternative version of this vector includes the N-terminal signal sequence of the BiP gene (which promotes secretion of the expressed recombinant protein).

The pMT-DEST48 vector permits expression of a gene or sequence received from a Gateway® entry clone (see table in the entry GATEWAY SITE-SPECIFIC RECOMBINATION SYSTEM for details).

The linearized pMT/V5-His-TOPO® vector (3.6 kb) allows expression of a gene or sequence prepared as a *Taq*-created PCR product (see TOPOISOMERASE I CLONING).

destination vector In the GATEWAY SITE-SPECIFIC RECOMBINATION SYSTEM: a vector which can receive a gene or insert from an ENTRY CLONE or an ENTRY VECTOR. (See the table in entry GATEWAY SITE-SPECIFIC RECOMBINATION SYSTEM for various examples of destination vectors.)

desumoylation See SUMOYLATION.

detection of cancer See CANCER DETECTION.

DFRS plasmid Dual-fluorescence reporter/sensor plasmid: a plasmid which contains sequences that encode two different fluorophores, one of the fluorophores being used to signal the intracellular presence of the plasmid, the other being used to detect the presence/absence of a specific intracellular target sequence. DFRS plasmids were used e.g. for monitoring the dynamics of specific types of microRNA (see MICRORNAS) within individual (living) cells of the zebrafish (*Danio rerio*) and mouse (*Mus* sp) [BioTechniques (2006) 41(6):727–732].

In the above study, the GREEN FLUORESCENT PROTEIN was used as reporter and MONOMERIC RED FLUORESCENT PROTEIN was used as the sensor. Both genes were under the regulation of identical constitutive promoters. The 3′ untranslated region of the mRFP mRNA contained a tandem cassette which was complementary to the particular miRNA being studied. Given the *absence* of the target species of miRNA, the mRNA of mRFP is translated – and this is detected by red fluorescence. If the target miRNA is present, hybridization occurs between the miRNA and the complementary sequence in the mRNA of mRFP, inhibiting translation of mRFP; this is detected by an absence of red fluorescence. (The intracellular presence of the plasmid is confirmed by green fluorescence.)

DGCR8 A dsRNA-binding protein which is a component of the MICROPROCESSOR complex that is commonly involved in the biogenesis of MICRORNAS.

(See also MIRTRON.)

DGGE Denaturing-gradient gel electrophoresis: a method used for comparing *related*, *double-stranded*, PCR-generated fragments of DNA (or fragments of DNA generated by restriction endonucleases) by carrying out electrophoresis in a gel which contains a DNA-denaturing reagent in a known concentration gradient; in such a gradient, fragments whose nucleotide sequences differ are likely to have different electrophoretic mobilities owing to the differing effects of the denaturing agent on base-pairing/structure (see below).

(cf. SSCP ANALYSIS.)

The method may involve either one-dimensional or two-dimensional electrophoresis.

In the first (preliminary) phase of two-dimensional DGGE, the fragments are separated, by size, in a routine form of GEL ELECTROPHORESIS.

In the second phase, electrophoretic movement of the fragments occurs in a direction perpendicular (at right-angles) to the original path. During the second phase, fragments move into an increasing gradient of DNA-denaturing agents (such as urea + formamide). Hence, at given levels in the gradient, the dsDNA fragments will exhibit *sequence-dependent* melting (i.e. strand separation) – and this will affect their speed of movement through the gel; different fragments, with different sequences, are likely to have different melting characteristics under these conditions, and this can result in the separation of fragments through differential changes in their rates of movement in the gel. (Note that base-pairing is stronger in GC-rich regions of a fragment; strand separation in such regions will occur less readily than strand separation in AT-rich regions. Thus, for example, an AT→CG mutation in a given fragment is likely to affect the ease of strand separation in the fragment and, hence, to affect the electrophoretic mobility of that fragment.)

In some protocols, the PCR primers used to prepare sample fragments are modified with GC-rich tags at the 5′ end; this is useful e.g. in that it avoids the *complete* denaturation of fragments in the higher concentrations of denaturing agent – thus avoiding complex and unwanted patterns of movement in at least some of the fragments.

DGGE has been useful e.g. for studying *pncA* mutations in *Mycobacterium tuberculosis* associated with resistance to the drug pyrazinamide (used in anti-tuberculosis chemotherapy) [Antimicrob Agents Chemother (2005) 49:2210–2217].

DGGE has also been useful for TYPING various species of bacteria (for example, differentiating isolates of *Escherichia coli* by analysis of the 16S–23S intergenic spacers in the RRN OPERON).

A method related to DGGE uses an ongoing rise in temperature (instead of a chemical gradient) in the second phase of electrophoresis; this procedure is referred to as temporal temperature-gradient gel electrophoresis: see TTGE.

DGR Diversity-generating retroelement: any one of a family of genetic elements, occurring e.g. in bacteria, which encodes a REVERSE TRANSCRIPTASE that is used in mediating changes in DNA target sequences – thus modifying the characteristics of the encoded proteins.

An early example was the DGR reported in a phage (BPP-1) that infects a Gram-negative bacterium, *Bordetella*; in this case, the DGR is involved in making a (template-dependent) change in the gene encoding a phage tail-fiber protein – the modified protein allowing the phage to alter its host range by binding to different ligand(s) on the bacterial surface.

[DGRs: Curr Opin Microbiol (2007) 10(4):388–395.]

(See also RETRON.)

DHPA (*S*)-9-(2,3-dihydroxypropyl)adenine: an agent which has antiviral activity (against e.g. herpes simplex and varicella–zoster viruses).

DHPG 9-(1,3-dihydroxy-2-propoxymethyl)guanine – an agent with activity against e.g. herpes simplex virus. The mode of action resembles that of ACYCLOVIR.

DI particle See DEFECTIVE INTERFERING PARTICLE.

DIA DNA integrity assay: see CANCER DETECTION.

diabody Two *non-covalently* linked scFvs: see SINGLE-CHAIN VARIABLE FRAGMENT.

diagnostic tag See e.g. SAGE.

diamidines Compounds that have two amidine [$NH_2.C(=NH)–$] groups. Some aromatic diamidines can bind to KINETOPLAST DNA, and such binding may contribute to their trypanocidal properties.

6-diazo-5-oxo-L-norleucine See DON.

diazomethane See STREPTOZOTOCIN.

Dicer (DICER) An RNase III-like enzyme involved in various types of post-transcriptional gene silencing (see PTGS). The enzyme cleaves double-stranded RNA (dsRNA) into pieces of regular size, reported to be within the range 19–28 nucleotides in length in individual cases; in at least some cases these fragments have characteristic 2-nucleotide 3′ overhangs.

Dicer is widespread in eukaryotic organisms – where it has a range of essential roles in gene expression (see MICRORNAS and RNA INTERFERENCE).

The enzyme contains a number of domains which include a C-terminal dsRNA-binding domain, an (N-terminal) DEAD-box helicase domain, and two RNase III-like domains.

Bacterial RNASE III, and also recombinant forms of Dicer, are used in experimental RNAi.

(See also ENDO-PORTER.)

dichroic beamsplitter In fluorescence microscopy: a filter, interposed between the eyepiece and objective which (i) transmits light emitted from the fluorophore, but (ii) *reflects* the (shorter-wavelength) light from the excitation filter onto the sample.

didanosine See NUCLEOSIDE REVERSE TRANSCRIPTASE INHIBITORS.

dideoxy fingerprinting (ddF) An early method which has been used e.g. for detecting resistance to an anti-tuberculosis drug, rifampicin, in *Mycobacterium tuberculosis*.

In this method, resistance to rifampicin is detected by demonstrating the presence of certain mutations in the *rpoB* gene; *rpoB* encodes the β subunit of the enzyme RNA polymerase, the target of rifampicin.

Dideoxy fingerprinting involves an initial amplification, by PCR, of the target sequence (in *rpoB*). The resulting (dsDNA) amplicons are then used as templates in a modified form of the DIDEOXY METHOD of sequencing involving temperature cycling (as in PCR); temperature cycling permits the use of double-stranded DNA amplicons, avoiding e.g. the need for ssDNA templates (which are required, for example, in SSCP). In contrast to the standard procedure for dideoxy sequencing, only one type of dideoxyribonucleotide is used; this ddNTP is used for each strain being tested so that the results can be compared. The products of the sequencing reaction are examined by electrophoresis in a *non*-denaturing gel.

Mutations may be detected e.g. by changes in the mobility of full-length and/or chain-terminated products, such changes being due to alterations in intra-strand base-pairing resulting from mutation(s). Mutations can also be indicated by changes in the *number* of chain-terminated products in a given reaction; the number of products can change when mutation(s) increase or decrease the number of chain-terminating sites in the template strand, a chain-terminating site being a base that is complementary to the particular ddNTP being used in a given reaction.

Interpretation of the results from dideoxy fingerprinting is facilitated by a knowledge of common resistance-conferring mutations in *rpoB* (i.e. base change and location); these have been mapped.

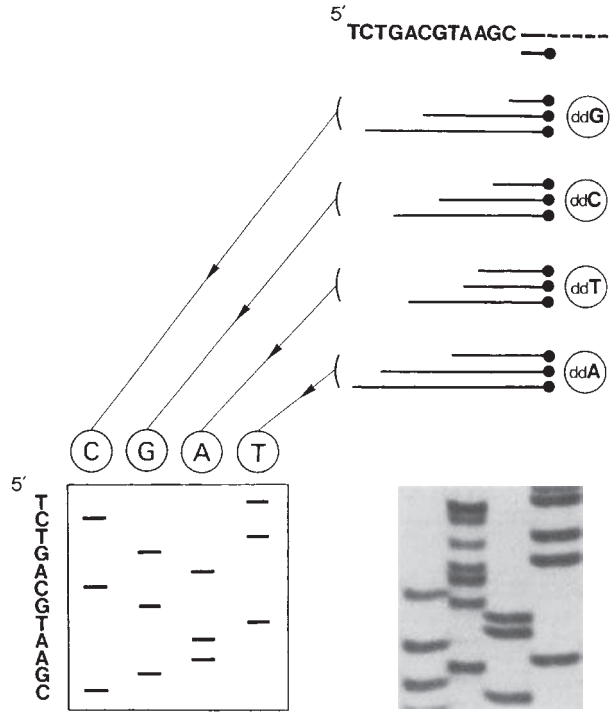
dideoxy method (DNA sequencing) (Sanger's method, or chain-termination method) A method of DNA SEQUENCING (q.v.) which is widely used for sequencing templates of up to ∼500 nucleotides; longer templates can be sequenced in stages.

The method is shown diagrammatically on page 84.

For so-called 'second generation' DNA sequencing methods see the entry MASSIVELY PARALLEL DNA SEQUENCING.

dideoxyribonucleoside triphosphate See the entry DDNTP.

Diels–Alder reaction A reaction that creates a six-membered ring structure as a result of interaction between a dienophile and a 1,3-diene. The Diels–Alder reaction has been exploited in the preparation of peptide–oligonucleotide conjugates; this has been achieved by interacting diene-modified oligonucleo-



DIDEOXY METHOD (DNA sequencing) The fragment to be sequenced is first obtained in single-stranded form. This can be achieved e.g. by cloning in a phage M13 vector or by asymmetric PCR. Either strand can be sequenced; for increased accuracy both strands are sequenced (in separate sequencing reactions).

In this example the unknown sequence is 5′-TCT. . ...AGC-3′ (*top*). Regardless of the method used for obtaining single-stranded templates, the unknown sequence will be flanked on its 3′ side by a known sequence of nucleotides; this permits the design of a primer which can bind next to the unknown sequence such that the first nucleotide to be added to the primer will pair with the first 3′ nucleotide of the unknown sequence. In the diagram, a primer (*short line*), carrying a label (*black disk*), has bound at a site flanking the 3′ end of the unknown sequence (*top*).

DNA synthesis *in vitro* is usually carried out with a reaction mixture that includes: (i) templates (in this case, single-stranded fragments that include the unknown sequence); (ii) primers; (iii) the four types of deoxyribonucleoside triphosphate (dNTP) – i.e. dATP, dCTP, dGTP and dTTP; and (iv) DNA polymerase. When base-paired to the template strand, the primer is extended (5′→3′) by the sequential addition of nucleotides, as dictated by the template.

For sequencing, there are four separate reactions. A given reaction determines each of the sites at which *one* of the four types of nucleotide (A, C, G or T) occurs in the unknown sequence. Each of the four reactions contains all of the constituents mentioned above (including many copies of the template, and of the primer). Additionally, each of the reaction mixtures contains *one* type of *di*deoxyribonucleoside triphosphate (ddNTP) – that is, ddT, ddG, ddC or ddA.

When a dideoxyribonucleotide is added to a growing strand of DNA it prevents the addition of the *next* nucleotide; this is because a dideoxyribonucleotide lacks the 3′-OH group which is necessary for making the next phosphodiester bond. Hence, extension of the primer will stop (= chain termination) at any position where a dideoxyribonucleotide has been incorporated. In a given reaction mixture, the concentration of the ddNTP is such that, in most growing strands, synthesis will be stopped – at some stage – by the incorporation of a dideoxyribonucleotide. Because a given type of ddNTP molecule can pair with its complementary base wherever it occurs in the template strand, chain termination can occur at different sites on different copies of the template strand – so that, in a given reaction, product strands of

(*continued*)

tides and maleimide-derivatized peptides [Nucleic Acids Res (2006) 34(3):e24].

This procedure may be used e.g. for conjugating antisense oligonucleotides to appropriate peptides in order to facilitate their uptake by target cells.

Nucleotides can be diene-modified to permit the tagging of DNA via the Diels–Alder reaction [Nucleic Acids Res (2009) 37(5):1477–1485].

diethylpyrocarbonate (DEPC) $C_2H_5.O.CO.O.CO.C_2H_5$ – an inactivator of RNases used e.g. for treating glassware etc. in order to avoid loss of sample RNA through exogenous RNase activity; autoclaving, after treatment, is used to remove any remaining DEPC.

DEPC can also modify the exposed N-7 position of an unpaired adenine in RNA or DNA.

(See also RNALATER.)

differential allelic expression See the entry ALLELE-SPECIFIC GENE EXPRESSION.

differential DNA cleavage See e.g. CPNPG SITES.

differential display A method which is used e.g. for detecting those genes which are expressed, in a given type of cell, only under certain conditions.

Essentially, the mRNAs are isolated from each of two (or more) populations of the given cell that have grown under, or been exposed to, different conditions – and which are therefore expected to have different patterns of gene expression. *Selected* mRNAs from each of the populations being studied are first converted to double-stranded cDNAs; the selection is achieved by carrying out the reverse transcription stage with primers whose 3′ terminal base pairs with the complementary base in only some of the mRNAs in a given population. (The selection permits an easily manageable number of mRNAs to be converted to cDNAs; a different subgroup of mRNAs can be studied later simply by employing primers with a different 3′ terminal base.) The cDNAs are amplified by PCR, using the original 'selective' primer and a primer of arbitrary sequence; the PCR products therefore represent the 3′ ends of selected mRNA molecules.

The amplicons from each population are then subjected to GEL ELECTROPHORESIS in separate lanes, and the fingerprints of the different populations are compared. A band of interest (e.g. one present in a given fingerprint but absent in others)

DIDEOXY METHOD (DNA sequencing) (*continued*) different lengths will be formed. For example, with ddG (see diagram) the three products are of different lengths because, during extension of primers, ddG has paired with a cytosine residue at three different locations in the template; note that, in this case, *the length of a given product strand is related to the location of a particular cytosine residue in the unknown sequence*. Analogous comments apply to reaction mixtures containing the other types of ddNTP.

At the end of the reaction, new product strands are separated from templates by formamide. Each of the four reaction mixtures is then subjected to electrophoresis in a separate lane of a polyacrylamide gel. During electrophoresis, small products move further than larger ones, in a given time; products that differ in length by only one nucleotide can be distinguished in this way – the shorter product moving just a little further.

The gel used for electrophoresis contains a denaturing agent which inhibits intra-strand base-pairing. This is essential: in order to be able to deduce the locations of a given base in the template strand it is necessary to compare the *lengths* of all the product strands (by comparing their positions within the gel), and this requires proportionality between strand length and electrophoretic mobility, i.e. between the length of a given strand and the position of its band in the gel. Were intra-strand base-pairing to occur, there would be no fixed relationship between a product's length and the position of its band in the gel.

After electrophoresis, bands in the gel may be revealed by autoradiography (if the primers had been labeled e.g. with ^{32}P) or by exposure to ultraviolet radiation (if the primers had been labeled with a fluorophore).

An alternative approach to detecting the bands of products exploits chemiluminescence (see entry) and uses e.g. biotin-labeled primers; after electrophoresis, bands in the gel are transferred to a membrane – which is exposed to a streptavidin–alkaline phosphatase conjugate and a substrate (such as CSPD®) which generates a light signal.

The locations of the bands in the gel (*bottom, left*) indicate the relative lengths of the product strands – the shorter products having moved further along the gel (from top to bottom in the diagram). Note that the first unknown 3′ nucleotide (C) is identified by (i) the shortest product strand (which has moved the furthest), and (ii) the fact that this product came from the ddG reaction mixture, indicating a base that pairs with G, i.e. C. Similarly, the next unknown nucleotide (G) is indicated by the next shortest product strand – which came from the ddC reaction mixture, thus indicating G in the template. The whole of the unknown sequence can be deduced in this way.

Bottom, right. Part of an autoradiograph of a sequencing gel (courtesy of Joop Gaken, Molecular Medicine Unit, King's College, London).

Figure reproduced from *Bacteria in Biology, Biotechnology and Medicine*, 6th edition, Figure 8.23, pages 260–261, Paul Singleton (2004) John Wiley & Sons Ltd, UK [ISBN 0-470-09027-8] with permission from the publisher.

can be extracted from the gel and amplified (using the same primers) prior to e.g. sequencing/further investigation.

Differential display was used e.g. to study gene expression during the development of different zones of the neural tube in the chick embryo [BMC Dev Biol (2006) 6:9]. It has been used for studying the changed pattern of gene expression in cells infected with respiratory syncytial virus [Virol J (2008) 5:114], and also for studying genes differentially expressed in human ovarian cancer [Int J Med Sci (2008) 5(3):133–142].

differential splicing *Syn*. ALTERNATIVE SPLICING.

Digene Hybrid Capture® 2 assay (for human papillomavirus) An FDA-approved assay (Digene/QIAGEN) used for detecting the DNA of any of a range of cancer-associated human PAPILLOMAVIRUSES, including the high-risk types 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59 and 68.

Essentially, exfoliated cells from (gynecological) swabs are proteolytically lysed and the DNA denatured. Target DNA is then exposed to a cocktail of type-specific RNA probes that are designed to bind to HPV DNA of the types being assayed. The DNA:RNA hybrids, which are captured on a solid phase, are detected with fluorophore-tagged antibodies; fluorescence under appropriate excitation is measured by a luminometer.

digital PCR An approach which has been used e.g. to assess the number of PCR-amplifiable fragments of DNA in a given sample; this has been applied to the calibration of sequencing libraries in MASSIVELY PARALLEL DNA SEQUENCING (q.v.), as outlined below.

Essentially, a (small) aliquot of the sequencing sample (less than a femtogram) is extensively diluted, and aliquots of the *diluted* sample are then examined by a large number of individual PCR reactions; because the original sample was highly diluted, the majority of the PCR reaction mixtures will have no amplifiable fragment – and so will give a negative result. However, the number of *positive* PCR reactions can be used to indicate the number of amplifiable fragments present in the original sequencing sample.

As each of the fragments in a sequencing library is flanked by ligated adaptors, at both ends, the primers used in the PCR reactions, mentioned above, are complementary to sequences in these adaptors.

To obtain fluorescent signals from the amplified fragments (in positive reactions), the procedure incorporates a *universal template* (UT) approach. In this approach, the PCR forward primer has a 5′ tag to which is bound a TaqMan®-style probe (see the entry TAQMAN PROBE). Thus, when the PCR reverse primer is extended on its template strands during the reaction, the DNA polymerase cleaves the (bound) probe and produces a fluorescent signal. Notice that, in this UT assay, production of a fluorescent signal is independent of the actual amplified target sequence; this approach is used because the fragments in a sequencing library, being random, do not permit the use of a TaqMan probe in its usual target-sequence-specific role.

In calibrating a sequencing library by digital PCR, each of the positive PCR reactions is taken to represent a single fragment in the (original) sequencing library – regardless of the amount of product formed in a given positive reaction; that is, the results obtained by this approach are not dependent on the efficiency of amplification. Moreover, use of digital PCR for this purpose is also advantageous in that the assessment it produces is based on the actual number of fragments which are amplifiable.

[Digital PCR provides sensitive and absolute calibration for high throughput sequencing: BMC Genomics (2009) 10:116.]

digoxigenin (DIG) A steroid (from the plant *Digitalis*) used e.g. for labeling PROBES. DIG-labeled RNA probes can be made e.g. by *in vitro* transcription on a DNA target sequence using labeled nucleotides.

DIG-dUTP can be incorporated into DNA probes during synthesis by various types of DNA polymerase – including *Taq* polymerase, phage T4 polymerase, Klenow fragment and at least some types of reverse transcriptase.

Detection of DIG-labeled probes can be achieved with anti-digoxigenin antibodies covalently linked to a fluorophore, or linked to an enzyme such as alkaline phosphatase that gives a detectable reaction with substrates such as the 1,2-dioxetane CSPD®.

(See also EMEA.)

dihydroorotase (control of synthesis in *Escherichia coli*) See TRANSLATIONAL ATTENUATION (sense 1).

7,8-dihydro-8-oxoguanine See 8-OXOG.

dikaryon (1) A cell with two genetically dissimilar (haploid) nuclei, or a (fungal) mycelium comprising such cells.
(2) A pair of nuclei that may be genetically dissimilar.

***dinB* gene** See DNA POLYMERASE IV.

dioecism See HETEROTHALLISM.

1,2-dioxetanes A category of agents which include commercial products used for the generation of CHEMILUMINESCENCE – see e.g. AMPPD and CSPD.

diphosphate See PYROPHOSPHATE.

diphosphopyridine nucleotide See NAD.

diploid See PLOIDY.

direct amplification fingerprinting (DAF) See AP-PCR.

direct repeat (DR) One of two (or more) identical or closely similar sequences, in a given molecule of nucleic acid, which have the same polarity and orientation. Repeats may or may not be contiguous. An example of a pair of direct repeats:

5′ . . . . . GCCTA . . . . . GCCTA . . . . . 3′
3′ . . . . . CGGAT . . . . . CGGAT . . . . . 5′

(cf. INVERTED REPEAT.)

Direct repeats are formed during INSERTION–DUPLICATION RECOMBINATION, and also during transposition (see diagram in the entry TRANSPOSABLE ELEMENT).

Direct repeats in the genome of *Mycobacterium tuberculosis* and related bacteria have been used in a form of TYPING (see SPOLIGOTYPING).

(See also ITERON and TANDEM REPEAT.)

direct repeat (DR) locus (of *Mycobacterium tuberculosis*) A distinct region of the chromosome, found in members of the

M. tuberculosis complex, consisting of a number of highly conserved 36-bp direct repeats (DRs) which are interspersed with spacers of 34–41 bp. The number of repeats, as well as the presence/length of the spacers, are strain-dependent characteristics; such variation may have arisen through intramolecular homologous recombination.

Variability in the DR locus among different strains of these organisms has been exploited in SPOLIGOTYPING.

direct selection *Syn*. POSITIVE SELECTION.

directed evolution Any technique whose object is, effectively, to enact a rapid 'evolutionary' process in the laboratory with the object of selecting molecules with certain desired characteristics; this approach can be used e.g. with nucleic acids and proteins.

[Laboratory-directed protein evolution: Microbiol Mol Biol Rev (2005) 69(3):373–392; the directed evolution of homing endonucleases: Nucleic Acids Res (2005) 33(18):e154.]

Directed evolution has been used to improve the efficacy of viral vectors in gene therapy by designing viral capsids that remain functional in the presence of selective pressure [Gene Therapy (2009) 16:311–319].

(See also DNA SHUFFLING.)

directional cloning *Syn*. FORCED CLONING.

directional TOPO® pENTR™ vectors Vectors (Invitrogen, Carlsbad CA, USA) used for cloning an insert, in a specific orientation, using the TOPOISOMERASE I CLONING approach.

Directional insertion of a fragment is achieved by using an asymmetric arrangement in which one end of the (linearized) vector is blunt-ended and the other end has a four-nucleotide overhang; as in other forms of topo cloning the enzyme topoisomerase I is bound to phosphate at each end of the vector – on opposite strands – and carries out the same function.

These vectors include *att*L sequences, permitting recombination with any of a range of Gateway® destination vectors – see GATEWAY SITE-SPECIFIC RECOMBINATION SYSTEM; they also include sites suitable for sequencing.

disjunction Separation of chromosomes during the process of nuclear division.

dispersed repeat A non-specific phrase which has been used to refer to any of various, relatively small sequences of nucleotides, copies of which may be present at more than one site in a genome.

(cf. SEGMENTAL DUPLICATION.)

displacement loop (D loop) See D LOOP.

dissymmetry ratio *Syn*. BASE RATIO.

distal box See RNASE III.

ditag See SAGE.

dithiothreitol (DDT, Cleland's reagent) A reagent used e.g. to reduce disulfides to thiols and to maintain thiol groups in the reduced state. High concentrations of dithiothreitol facilitate denaturation of proteins by chaotropes and detergents.

In DNA technology, the reagent has also been used e.g. as a mucolytic agent for the treatment of mucoid specimens of sputum prior to examination by nucleic-acid-based tests for the detection of *Mycobacterium tuberculosis*.

divergent transcription Transcription from two closely adjacent promoters, in opposite directions.

diversity-generating retroelement See the entry DGR.

***DMD* gene** The gene encoding the muscle protein dystrophin – see e.g. DUCHENNE MUSCULAR DYSTROPHY.

(See also HUMAN ARTIFICIAL CHROMOSOME.)

DNA Deoxyribonucleic acid: typically, a double-stranded (i.e. 'duplex') molecule or a single-stranded molecule – referred to as dsDNA and ssDNA, respectively – that consists of linearly polymerized deoxyribonucleotides, with adjacent nucleotides linked by a phosphodiester bond.

DNA encodes the genetic information in cells, and in some types of virus. In cells, this information is incorporated in the GENETIC CODE – and e.g. in the pattern of methylation in the DNA molecule (see e.g. GENETIC IMPRINTING).

The genomic DNA in eukaryotes may contain an additional code that signals the position of nucleosomes [Nucleic Acids Res (2009) doi: 10.1093/nar/gkp689].

In the prokaryotes, DNA is found in the nucleoid (see also PLASMID); in eukaryotes it occurs primarily in the NUCLEUS, mitochondria and (in photosynthetic species) in chloroplasts. (See also KINETOPLAST and MITOCHONDRIAL DNA.)

Major bases found in the nucleotides of DNA are adenine, guanine, cytosine and thymine. In RNA (i.e. ribonucleic acid) the major bases are adenine, guanine, cytosine and uracil.

In double-stranded DNA the two strands are wound around each other to form a helical structure (the so-called 'double helix') (compare with the SIDE-BY-SIDE MODEL). (See also B-DNA and Z-DNA.)

Individual molecules of DNA can be studied by appropriate methods: see entry SINGLE-MOLECULE DNA MANIPULATION/VISUALIZATION.)

(See also TRIPLEX DNA and QUADRUPLEX DNA.)

(Compare DNA with GNA, PNA and TNA.)

In linear molecules of DNA, each strand has *polarity*: a 5′ end and a 3′ end. At one end of a strand, 5′ refers to the 5′ carbon atom in the terminal ribose residue; at the other end of the strand, the 3′ refers to the 3′ carbon atom in the terminal ribose residue.

In dsDNA molecules, the two strands are arranged in an *antiparallel* mode: a 5′-to-3′ strand is bound to a 3′-to-5′ strand. The hybridization (binding) between the two strands involves hybrogen bonding between the bases in opposite strands. Hydrogen bonding occurs between specific pairs of bases: adenine, in one strand, base-pairs with thymine in the other strand, while cytosine base-pairs with guanine; adenine and thymine are called *complementary* bases, as are cytosine and guanine. Moreover, one strand of the duplex is said to be complementary to the other (when all the bases are correctly paired).

In 'circular' molecules of dsDNA the polymer is continuous – i.e. there are no free 5′ and 3′ ends. Such molecules may be in either a relaxed state or a supercoiled state. (See the entry TOPOISOMERASE.)

ssDNA can also occur in a linear or circular form. (See also

DNA NANOCIRCLE.)

DNA technology involves a wide range of methods – such as those concerned with DNA AMPLIFICATION, DNA CUTTING AND ASSEMBLY, DNA ISOLATION and DNA SEQUENCING.

(See also DNA STAINING and ZIP NUCLEIC ACID.)

DNA amplification (*in vitro*) Any of a range of procedures in which, typically, a defined sequence of nucleotides in DNA is copied ('amplified') – usually with the formation of a large number of copies. (cf. RNA AMPLIFICATION.) See the entry NUCLEIC ACID AMPLIFICATION (table) for specific examples of methods.

(cf. WHOLE-GENOME AMPLIFICATION.)

(See also DNA REPLICATION.)

DNA amplification is widely used in biomedical sciences; it also has FORENSIC APPLICATIONS.

Isothermal amplification (amplification at a fixed temperature) can be carried out by methods such as strand displacement amplification (see entry SDA). For some other examples of isothermal amplification of DNA see the table in the entry NUCLEIC ACID AMPLIFICATION. For isothermal amplification of RNA see e.g. NASBA.

Other methods, such as the LIGASE CHAIN REACTION and polymerase chain reaction (see entry PCR), involve repetitive temperature cycling (e.g. for ~25 to 40 cycles). Temperatures used in these processes include the initial high temperature (e.g. 94°C) that is designed to separate the strands of dsDNA in order to expose target sequence(s). These methods require the use of thermostable enzymes (e.g. a thermostable ligase in LCR and a thermostable polymerase in PCR).

A single-probe approach to DNA amplification is involved in CONNECTOR INVERSION PROBE TECHNOLOGY.

Without adequate care, any DNA amplification method can yield false-positive results – aberrant amplification occurring when the given sequence of nucleotides is not present in the test sample; this can occur if the sample is contaminated with extraneous DNA, and such contamination must be rigorously excluded (see e.g. AMPLICON CONTAINMENT and AMPLICON INACTIVATION). False-negative results, i.e. failure to amplify a sequence which *is* in the test sample, can be due to any of a range of problems with the protocol and/or failure adequately to extract DNA from the source organisms or cells.

Most of the methods mentioned above are well established and have been available for some years. More recent methods include e.g. HELICASE-DEPENDENT AMPLIFICATION and the MOLECULAR ZIPPER.

Recombinase polymerase amplification is a method which combines recombinase-mediated binding of primers with the strand-displacement synthesis of DNA: see RPA.

(See also *in situ* solid-phase DNA amplification in SOLID-PHASE PCR.)

DNA assembly See the entry DNA CUTTING AND ASSEMBLY.

DNA barcoding An approach that aims to use DNA-based data for identifying all currently known species, and for providing criteria for the recognition of new species.

Note. This topic is not related to the *barcoding* mentioned in MOLECULAR INVERSION PROBE GENOTYPING.

Note. Some authors have described *DNA barcoding* as a misnomer, preferring instead the name 'Barcoding Life' [Philos Trans R Soc Lond B Biol Sci (2005) 360(1462):1905–1916]; this suggestion has not been widely taken up.

Brief overview

Since its inception, DNA barcoding has (perhaps inevitably) fragmented into at least two sectors – animal and plant – that differ not only in the type of barcoding marker(s) but also in the source of the marker(s).

In animals, the (mitochondrial) gene encoding cytochrome oxidase *c* subunit I (*COI*) has been a useful barcoding marker across a wide range of taxa. While it has not been universally successful as a barcoding marker, it seems to have been used for the majority of animal studies. Nevertheless, other genes have been suggested as useful markers. In some cases NUMTS have raised difficulties in the interpretation of data; thus, for example, studies on grasshoppers and crayfish (both of which exhibit *COI*-related numts that may differ significantly from orthologous mtDNA) reported that the presence of numts can result in overestimates in the number of species [Proc Natl Acad Sci USA (2008) 105(36):13486–13491].

In plants, the relatively slow evolution of the mitochondrial DNA offers too little variation on which to base the species differentiation. Instead, studies have been focused mainly on plastid DNA [e.g. chloroplast *trn*L intron: Nucleic Acids Res (2007) 35(3):e14]. Moreover, there appears to be a majority view that, for plants, one barcoding region is not enough – and that chloroplast DNA will have to be the source of the required barcoding markers. [How many loci does it take to DNA barcode a crocus? PLoS ONE (2009) 4(2):e4598.]

DNA-based identification See FORENSIC APPLICATIONS.

(See also IDENTIFIER DNA.)

DNA-binding proteins (in assays) See CHROMATIN IMMUNOPRECIPITATION, EMSA, FOOTPRINTING and SPA (scintillation proximity assay).

(See also DNA–PROTEIN INTERACTIONS.)

DNA caging See CAGED DNA.

DNA chip See CHIP; see also MICROARRAY.

DNA clean-up resin See e.g. STRATACLEAN RESIN.

DNA cloning See CLONING.

DNA combing (*syn*. molecular combing) A technique, used for SINGLE-MOLECULE DNA MANIPULATION/VISUALIZATION (q.v.), in which one end of a DNA molecule is fixed to a solid surface and the molecule itself is extended away from the point of attachment. Extension of the DNA molecule (stretching) was first achieved by using the forces (surface tension) found in a receding meniscus; however, the surface tension present at an air–water interface is so high that it results in *over*-stretching of the DNA molecule in a manner which may affect potential DNA–protein interaction. In a modified approach, the surface tension was lowered by adding a monolayer of 1-dodecanol (a fatty alcohol) which apparently has no significant affect on the properties of the sample DNA.

One advantage of the DNA combing approach is that many

molecules of DNA can be studied simultaneously on a given solid surface.

DNA attached to surfaces

During the combing procedure, the extended DNA molecule adheres to the surface at a number of points along its length – and it remains attached. The number of points of attachment should be minimized (by suitable modification of the surface) – and over-stretching of the molecule should be avoided – if the study aims to investigate DNA–protein interactions.

Making the DNA visible

To make the DNA visible it can be stained with certain types of INTERCALATING AGENT, although changes in the physical characteristics of the molecule brought about by intercalation might make the system unsuitable for studying DNA–protein interactions.

Among other approaches, the DNA under investigation can be prepared as PCR amplicons, using a reaction mixture that includes both fluorophore-labeled and unlabeled nucleotides – thus allowing some labeled nucleotides to be inserted along the length of the DNA molecule.

End-labeling of the DNA molecule (e.g. with a QUANTUM DOT) is a further option.

Visualization of the combed DNA may be carried out by means of conventional epifluorescence microscopy. However, as the combed DNA is closely associated with the (glass) surface, it is possible to use total internal reflection fluorescence microscopy (see TIRF MICROSCOPY).

Uses of DNA combing

Uses of DNA combing include e.g. genomic mapping with fluorescent probes. With a suitable level of stretching of the DNA molecule, combing has also been used e.g. to study the movement of an immunolabeled RNA polymerase along the DNA during both free translocation and transcription.

Disadvantages of DNA combing

A major weakness in this approach is the lack of control over the force used to 'stretch' the DNA molecule. Moreover, the multiple, non-specific attachment points formed between the DNA molecule and the surface may preclude some investigations which require unhindered access to the DNA.

One alternative to combing is SURFACE TETHERING.

[DNA combing in studies on DNA replication: PLoS ONE (2009) 4(2):e4516.]

DNA curtain See SURFACE TETHERING.

DNA cutting and assembly Cutting and assembly of DNA – *in vitro* or *in vivo* – is fundamental to DNA technology and is an integral part of a wide range of procedures.

Cutting DNA

In some cases, the requirement is simply to break DNA into small pieces; for example, in the new generation sequencing methods (see MASSIVELY PARALLEL DNA SEQUENCING), the genomic DNA is first prepared as a large number of small fragments. In such cases, DNA may be broken up by shearing forces; alternatively, the fragmentation may be carried out by random endonuclease activity.

Typically, however, DNA needs to be cut at *precise* sites – i.e. between two *particular* nucleotides in one strand and two *particular* nucleotides in the other strand. In order to do this, the usual approach is to use an enzyme called a RESTRICTION ENDONUCLEASE (q.v.) – often a type IIP restriction enzyme; the type IIP enzymes characteristically cut at a fixed site in a symmetric (palindromic) recognition site. The wide range of type IIP enzymes available permits the researcher to cleave DNA in such a manner that the two cut ends have the desired characteristics (see e.g. STICKY ENDS, and cf. BLUNT-ENDED DNA); for example, sticky ends having a particular nucleotide sequence may be needed for ligation to another molecule that has compatible sticky ends.

Even when cutting produces blunt-ended fragments, it may be required to convert the terminal regions of (selected) fragments to sticky ends. This may be achieved e.g. by covalently attaching a LINKER, to both ends of the fragments, by BLUNT-END LIGATION; then, if the linker contains the recognition site of an appropriate enzyme, cleavage at the recognition site can form the required sticky ends.

In some cases, when fragments have the same sticky end at both ends, it may be desirable to modify their sticky ends to prevent self-ligation, i.e. circularization of the fragments; this may be achieved e.g. by a *partial fill-in* procedure – see the entry LAMBDA FIXII VECTOR for a description of a partial fill-in procedure.

Self-ligation can also be avoided by de-phosphorylating the 5′ termini of the fragments.

The various constructs used in DNA technology (such as CLONING VECTORS, EXPRESSION VECTORS) typically include a POLYLINKER in order to offer a choice of cutting sites.

When cutting fragments with a given restriction enzyme it may be necessary to cut at a *particular* recognition site, even when other copies of that site are present elsewhere in the fragment. This may be achieved e.g. with an approach such as the ACHILLES' HEEL TECHNIQUE.

Sometimes, the requirement is for a restriction enzyme that cuts *outside* its recognition sequence – as, for example, in the SAGE method, and in procedures used to prepare fragments for subsequent assembly into longer sequences. Type IIS restriction enzymes are often used in this context. One problem with type IIS enzymes is that their recognition sites are quite common – limiting their use in the preparation of fragments for assembly. One solution to this limitation was to construct a fusion consisting of the catalytic domain of FokI (a type IIS enzyme) fused with a (non-cleaving) mutant of the homing endonuclease I-SceI; this type IIS construct has the 18-bp site specificity of I-SceI together with the ability to cut DNA at positions 2 nt and 6 nt from its recognition site – creating 5′ four-nucleotide overhangs [Nucleic Acids Res (2009) 37(9): 3061–3073].

Targeted IIS-mediated cutting *in vivo* can be achieved with a ZINC-FINGER NUCLEASE.

Although classic restriction endonucleases are widely used for cutting DNA, there are alternative modes of cutting – e.g. cutting mediated by some types of HOMING ENDONUCLEASE.

Additionally, cutting may be carried out by certain types of SINGLE-STRAND-SPECIFIC NUCLEASE – for example, cutting (nicking) one strand in a heteroduplex (as in ECOTILLING).

Cutting DNA is also an integral part of the various forms of recombination mediated (*in vitro* and *in vivo*) by the various phage-encoded systems (see e.g. CRE-LOXP SYSTEM, FLP and RECOMBINEERING).

Assembly of DNA

DNA can be assembled in various contexts and with a variety of methods. In a simple scenario, fragments with compatible sticky ends are first hybridized and then covalently joined by a suitable ligase. Blunt-ended DNA fragments may be joined by blunt-end ligation (using e.g. the T4 DNA ligase).

Blunt-ended PCR amplicons can be inserted, directionally, into specially designed vector molecules e.g. by the technique known as STICKY RICE (q.v.).

Ligase-independent covalent joining can be conducted e.g. with a topoisomerase (see TOPOISOMERASE I CLONING).

The assembly of two sequences of DNA (on different molecules, or on the same molecule) may be carried out by SPLICE OVERLAP EXTENSION PCR.

Pieces of dsDNA can be joined together, or a segment of dsDNA can be inserted into a linearized vector, with the PCR-based, ligase- and restriction-enzyme-independent IN-FUSION protocol which involves the creation of sticky ends by $3' \rightarrow 5'$ exonuclease activity.

Assembly of the exons of a split gene (to form a continuous coding sequence) can be achieved *in vitro* by a method called SPLICE (q.v.).

HOMOLOGOUS RECOMBINATION can be used to mediate the insertion of DNA *in vivo*: see e.g. INSERTION–DUPLICATION RECOMBINATION.

A form of facilitated homologous recombination is used for the insertion (and deletion) of DNA sequences *in vivo* by the LAMBDA (λ) RED RECOMBINATION system.

Insertion/exchange of DNA sequences can be mediated e.g. via *att* sites in the GATEWAY SITE-SPECIFIC RECOMBINATION SYSTEM, while various forms of insertion, inversion etc. can be achieved with the CRE–LOXP SYSTEM and the FLP system. Site-specific recombination systems such as these can usually mediate insertions according to the required orientation in the receiving molecule.

Insertion of a fragment of DNA into a circular vector can be carried out by a (commercial) process involving PCR and a type IIS restriction enzyme (see SEAMLESS CLONING KIT).

Insertions using the group II intron homing principle can be achieved by the TARGETRON.

Some methods can insert a linear fragment or gene into a construct – e.g. for cloning – by a technique in which all the component parts are prepared by PCR with uracil-containing primers. One method permits the assembly of a complete gene, ready for *in vitro* transcription and translation, in less than 2 hours (see the entry USER CLONING).

Yeast (*Saccharomyces cerevisiae)* can be used to assemble overlapping DNA sequences internalized by transformation [see e.g. Nucleic Res (2009) doi: 10.1093/nar/gkp687].

DNA database A phrase that frequently refers to a DATABASE containing a record of the DNA sequences from a number of individuals; such databases are often maintained e.g. by law-enforcement agencies.

(See also CODIS and FORENSIC APPLICATIONS.)

DNA delay mutant (of phage T4) A strain with mutation(s) in topoisomerase-encoding gene(s) which exhibits delayed synthesis of DNA and a smaller burst size. For replication, a T4 DNA delay mutant depends on a functional gyrase in the host cell.

DNA demethylation See DEMETHYLATION.

DNA dendrimer See DENDRIMER.

DNA-dependent DNA polymerase ('DNA polymerase') Any enzyme which polymerizes dNTPs (deoxyribonucleoside triphosphates) in the $5'$-to-$3'$ direction on a DNA template, with elimination of $P_2O_7^{4-}$. (See also STRAND DISPLACEMENT.)

Escherichia coli encodes two main DNA polymerases that are designated I and III. (See also DNA POLYMERASE II, DNA POLYMERASE IV and DNA POLYMERASE V.)

Eukaryotes typically have a number of different DNA polymerases, including separate enzymes for replicating the mitochondrial and the nuclear DNA. The nuclear polymerases include DNA polymerases α and δ (see also OKAZAKI FRAGMENT).

Many types of DNA polymerase are encoded by the human genome. These include high-fidelity polymerases involved in routine replication of the genomic DNA, but also encoded are polymerases that permit ongoing synthesis at locations where damage to the DNA would block the activity of high-fidelity enzymes. [The human polymerases (function and regulation): Nature Reviews Genetics (2008) 9:594–604.]

DNA polymerases comprise at least four major families – families being distinguished on the basis of e.g. composition and activity: see A FAMILY, B FAMILY, X FAMILY, Y FAMILY. Polymerases of families A and B are generally high-fidelity enzymes involved in processes such as DNA replication and the correction of errors arising during replication. X family polymerases include DNA repair enzymes, and those of the Y family are characteristically involved in *translesion* synthesis of damaged DNA (see the entry TRANSLESION SYNTHESIS) which may help to prevent stalling during replication.

There is an extensive range of (commercial) DNA polymerases – many of which are thermostable enzymes intended for use in PCR (see e.g. ACCUPRIME GC-RICH DNA POLYMERASE, AMPLITAQ GOLD DNA POLYMERASE, PFUTURBO, PLATINUM TAQ DNA POLYMERASE and RTTH DNA POLYMERASE XL).

Various forms of the well-known TAQ DNA POLYMERASE (q.v.), including recombinant forms, are available.

The THERMINATOR DNA POLYMERASE is useful for polymerizing *atypical* nucleotides, including acyclonucleotides.

Some DNA polymerases (e.g. r*Tth* DNA polymerase XL) have PROOF READING ability – that is, $3'$-to-$5'$ exonuclease activity; others (including *Taq* DNA polymerase) do not.

The enzyme *PfuUltra*™ DNA polymerase (Stratagene, La

Jolla CA, USA) was reported to avoid the problem of DUTP POISONING.

Two DNA polymerases are used, simultaneously, in LONG PCR.

An *error-prone* DNA polymerase (Mutazyme™) is used for generating random mutations during PCR (in the GENEMORPH PCR MUTAGENESIS KIT).

(See also TWO-METAL-ION CATALYSIS.)

DNA-dependent RNA polymerase See RNA POLYMERASE.

DNA-dependent TNA polymerase See TNA.

DNA DIRECT™ See DYNABEADS.

DNA enzyme See DEOXYRIBOZYME.

DNA extraction See DNA ISOLATION.

DNA fingerprinting (restriction enzyme analysis) (in *bacterial* TYPING) In the original method, genomic DNA from a test strain is cleaved by certain restriction enzymes and the resulting fragments separated by gel electrophoresis. The bands of fragments are stained *in situ*, or are denatured in the gel and then blotted onto a membrane before staining. The pattern of bands in the gel, or on the membrane, is the *fingerprint* of the test strain; strains are compared and classified on the basis of similarities and differences in their fingerprints.

A disadvantage of this approach is that it may generate too many fragments and thus make a complex fingerprint which is difficult to interpret. Fewer, larger fragments are generated if a so-called 'rare-cutting' restriction enzyme is used for the initial restriction. These large fragments can be separated by PFGE, although this procedure requires 2–3 days and may be susceptible to endogenous nucleases.

The problem of too many fragments can also be addressed by using a labeled probe which binds only to those fragments that contain the probe's target sequence; in this case, only fragments which bind the probe are made visible – so that the fingerprint consists of only a limited number of bands. This principle is exploited in RIBOTYPING.

DNA fragmentation Fragmentation of genomic DNA occurs, naturally, during APOPTOSIS in eukaryotic cells, and also e.g. in (prokaryotic) cells of *Escherichia coli* during infection by the lytic bacteriophage T7; in the latter case, fragmentation of the bacterial DNA is brought about by phage-encoded endo- and exonucleases whose activity provides nucleotides for the synthesis of phage DNA.

(See also TUNEL ASSAY.)

In DNA technology, fragmentation of genomic DNA is an initial stage in various procedures such as the preparation of a genomic LIBRARY and certain methods for sequencing – see MASSIVELY PARALLEL DNA SEQUENCING; it is also used in CHROMATIN IMMUNOPRECIPITATION. Fragmentation can be achieved e.g. by mechanical shearing or by the use of certain types of RESTRICTION ENDONUCLEASE.

DNA-free cell (*bacteriol.*) A cell, modified *in vitro*, in which chromosomal DNA has undergone cleavage and degradation. This can be achieved in at least two ways. First, treatment of certain types of cell with ULTRAVIOLET RADIATION (see e.g. the entry MAXICELL). Second, treatment of *dam* mutants of *Escherichia coli* with 2-aminopurine, resulting in mismatch repair-related degradation of the chromosome [see J Bacteriol (2006) 188(1):339–342].

DNA glycosylase Any enzyme which cleaves an aberrant base from a nucleotide in DNA; DNA glycosylases are active e.g. in the BASE EXCISION REPAIR pathway.

(See e.g. URACIL-N-GLYCOSYLASE; see also 8-OXOG.)

DNA glycosylase I (in *Escherichia coli*) *Syn*. Tag protein (see DNA REPAIR).

DNA glycosylase II (in *Escherichia coli*) *Syn*. AlkA (see DNA REPAIR).

DNA gyrase See TOPOISOMERASE.

DNA helicase See HELICASE.

DNA identification See FORENSIC APPLICATIONS.

DNA identification tag See IDENTIFIER DNA.

DNA immunization *Syn*. GENETIC IMMUNIZATION (q.v.).

DNA integrity assay (in the detection of cancer) See CANCER DETECTION.

DNA isolation (DNA extraction) Any procedure for removing and concentrating DNA from cells or tissues, or from specific environments (see e.g. biofilm, below).

The methods used for isolating DNA from cells or tissues are influenced by various factors – such as the type of DNA sought (genomic, plasmid DNA etc.), the quality of the DNA available, the amount of sample available, and also the nature of the sample (e.g. fungal, bacterial, mammalian cells).

[Optimization of procedures for isolation of *ancient* DNA: BioTechniques (2007) 42(3):343–352; see also SPEX.]

[Isolation of nucleic acids from fixed, paraffin-embedded tissues (methods compared): PLoS ONE (2007) 2(6):e537.]

(See also NUCLEIC ACID ISOLATION.)

Protocols include a wide range of *ad hoc* in-house methods and commercial systems.

Isolation of DNA from skin

For 'difficult' tissues – such as skin – the method may include an initial stage of mechanical disaggregation of the tissue that is followed by digestion with a proteinase; DNA may then be recovered by phenol–chloroform–isoamyl alcohol extraction and precipitated by ethanol (or e.g. by isopropanol).

Isolation of DNA from brain or liver

DNA can be extracted from some tissues – e.g. brain, liver – without organic solvents. The RecoverEase™ DNA isolation procedure (Stratagene, La Jolla CA) employs an initial stage of physical disaggregation and (coarse) filtration. The nuclei are separated by centrifugation. Incubation with proteinase K is continued overnight within a dialysis cup, yielding purified high-molecular-weight DNA.

Isolation of DNA from whole blood

DNA from whole blood can be obtained rapidly e.g. with the QIAamp® blood kit (QIAGEN, Hilden, Germany). After the buffer-mediated lysis, the lysate is loaded into a spin column. On centrifugation, lysate passes through a specialized membrane to which DNA is adsorbed. Following spin washes, the DNA can be eluted from the membrane. This system can be used with blood containing common anticoagulants such as

DNA ISOLATION: some common problems and their possible causes

Problem	Possible cause(s)
Low yield	Failure to ensure adequate lysis of cells, preventing release of cellular genomic DNA
Low yield	In spin columns: use of buffers which have incorrect pH or incorrect electrolyte concentration
Low yield	Pellet over-dried, causing incomplete re-suspension in buffer
Low yield of supercoiled plasmid DNA	Prolonged exposure to alkaline lysis buffer, resulting in denaturation and anomalous migration in gel electrophoresis
No yield (in gels)	Bands of *small* fragments migrated to full length of gel – avoided by use of lower voltage or shorter run
Quality (plasmid DNA) poor	Effect of endogenous nucleases. High-level activity of endonucleases is common in some bacteria
Quality (plasmid DNA) poor	Over-vigorous pipeting or vortexing, resulting in shearing
Contamination of plasmid DNA with chromosomal DNA	In controlled lysis methods: over-vigorous agitation of the lysate, with escape of both genomic and plasmid DNA – instead of only plasmid DNA
Contamination with RNA	A routine finding in some procedures; the RNA can be eliminated with RNase
Smeared bands in gel electrophoresis	Assuming correct conditions for electrophoresis: possible degradation of DNA via contamination with nucleases, or loading of gel with excess DNA; electrolyte levels in the DNA may be incorrect

citrate and heparin. (Other types of QIAamp kit are available for different sources of DNA.)

Isolation of plasmid DNA

Plasmid DNA may be recovered from bacteria e.g. with the QIAGEN® plasmid mini kit. The starting point is a pellet of cells of (e.g.) *Escherichia coli*. The pellet is resuspended in a buffer which disrupts the bacterial outer membrane. The use of (controlled) alkaline lysis with sodium hydroxide and the detergent sodium dodecyl sulfate (SDS) then releases soluble proteins, RNA and plasmids from the cells. RNA is degraded by RNase A. Conditions are optimized for maximum release of plasmid DNA without release of the chromosomal DNA; the protocol also avoids irreversible denaturation of plasmid DNA. Neutralization of the reaction mixture by a high-salt buffer (pH 5.5) precipitates the SDS; proteins are entrapped within the salt–SDS complexes – but plasmids renature and remain in solution. Centrifugation removes the proteins etc., leaving plasmids in the supernatant. The supernatant is then passed through a QIAGEN anion-exchange resin; given the pH and electrolyte concentration of the supernatant, plasmid DNA is retained by the resin. After washing, plasmid DNA is eluted in buffer containing 1.25 M NaCl (pH 8.5), desalted, and concentrated by precipitation with isopropanol.

Some of the problems (and possible causes) experienced in DNA isolation from cells/tissues are listed in the table.

Isolation of extracellular DNA from biofilms

Isolation of the *extra*cellular DNA that is present in biofilms is associated with certain difficulties. These include e.g. the problem of avoiding contamination with DNA that is released by the lysis of biofilm organisms as a result of the extraction process itself. Different methods for extracting the extracellular DNA from biofilms (with avoidance of contamination) have been evaluated [Appl Environ Microbiol (2009) doi: 10.1128/ AEM.00400.09].

DNA ladder See GEL ELECTROPHORESIS.

DNA ligase Any LIGASE which can catalyze a phosphodiester bond between the 3′-OH terminal of a strand of DNA and the correctly juxtaposed 5′-phosphate terminal of the same strand or of a different strand.

A thermostable ligase, Ampligase® – from a thermophilic prokaryote – has uses which include e.g. the LIGASE CHAIN REACTION and REPEAT-EXPANSION DETECTION; however, it is not used for BLUNT-END LIGATION.

The (ATP-dependent) phage T4 DNA ligase (EC 6.5.1.1) and the (NAD^+-dependent) *Escherichia coli* DNA ligase (EC 6.5.1.2) are used e.g for closing nicks.

DNA looping See the entry LOOPING.

DNA manipulation/visualization (at the molecular level) See SINGLE-MOLECULE DNA MANIPULATION/VISUALIZATION.

DNA melting See MELTING.

DNA methylation See METHYLATION (of DNA).

DNA-methylation maps See METHYLATION (of DNA).

DNA methyltransferase (MTase) Any enzyme that catalyzes the METHYLATION of specific bases in DNA; these enzymes are sometimes referred to as 'methylases'.

DNA methyltransferases in prokaryotes

One example of a prokaryotic methyltransferase is M.SssI, an enzyme which carries out SYMMETRIC METHYLATION at the 5-position in cytosine residues in the sequence 5′-CG-3′.

In general, the prokaryotic methyltransferases are involved in the (routine) process of MODIFICATION; methyltransferases involved in modification include many types of RESTRICTION

ENDONUCLEASE that carry out specific methylation as well as restriction.

The primary role of methylation in prokaryotes is generally understood to be protection of genomic DNA from the cell's own restriction enzymes (enzymes which degrade DNA, e.g. 'foreign' DNA that lacks the particular pattern of methylation characteristic of the cell's own DNA).

See PHAGE MU for an example of a DNA methyltransferase encoded by a bacteriophage.

Specific methyltransferases have been used in a method for enhancing the efficiency of *in vitro* bacterial transformation (see the entry TRANSFORMATION).

[Use of prokaryotic DNA methyltransferases as analytical and experimental tools in biology: Anal Biochem (2005) 338 (1):1–11.]

DNA methyltransferases in mammals

Mammals have a number of DNA methyltransferases which are essential for normal function or survival. These enzymes are required, throughout development, to maintain an overall pattern of methylation consistent with the normal functioning of cells. (This includes GENETIC IMPRINTING.) Dysregulation of methylation – including aberrant demethylation or aberrant hyper-methylation – can give rise to disorders or death; thus, for example, mutation in *DNMT3B*, the gene encoding DNA methyltransferase 3B (Dnmt3B), results in the disorder called ICF SYNDROME. Moreover, methylation of genome sequences that are normally unmethylated can cause disease by repressing gene function – as may occur, for example, when a tumor suppressor gene is repressed. (See METHYLATION for further information.)

Methyltransferases Dnmt3A and Dnmt3B appear to be required for *de novo* methylation. On the other hand, Dnmt1 is a 'maintenance' enzyme which e.g. methylates hemimethylated CpG sites shortly after DNA replication; for this task it works in association with an accessory factor (UHRF1) that is able to recognize hemimethylated DNA. Dnmt1 also appears to have a methylation-independent role in that it binds to some transcription inhibitors, contributing to gene repression. The stability of Dnmt1 is reported to be regulated through SET7-mediated lysine methylation [Proc Natl Acad Sci USA (2009) 106(13):5076–5081].

Studies on (and manipulation of) DNA methyltransferases

Studies on DNA methyltransferase activity have been carried out e.g. with *inhibitors* of DNA methyltransferases – such as 5-AZA-2′-DEOXYCYTIDINE (q.v.).

(See also DEMETHYLATION and TAGM.)

DNA modification (*bacteriol.*) See MODIFICATION.

DNA nanocircle A small, synthetic circular ssDNA molecule (up to ∼120 nt) which can be transcribed *in vitro* (or in bacteria after insertion by heat shock); RNA synthesis occurs by the ROLLING CIRCLE mechanism. RNA polymerase may be able to initiate transcription with GTP at any cytosine residue in the circle, although the conformation of the template can strongly inhibit or promote transcription.

A nanocircle containing an optimized promoter sequence – and encoding a self-cleaving monomeric hammerhead RIBOZYME – was able to express the active ribozyme in cells of *Escherichia coli* [Proc Natl Acad Sci USA (2002) 99(1):54–59].

[DNA nanocircle used for synthesizing the human telomere repeat: Nucleic Acids Res (2004) 32(19):e152.]

DNA notation See e.g. DNA SKYLINE.

DNA-only transposon Any TRANSPOSABLE ELEMENT that can transpose without the involvement of an RNA intermediate.

(cf. RETROTRANSPOSON.)

DNA photolyase See PHOTOLYASE.

DNA polymerase See DNA-DEPENDENT DNA POLYMERASE for a definition and for some commercial examples.

See subsequent entries for some prokaryotic and eukaryotic examples.

DNA polymerase I (Kornberg enzyme) In *Escherichia coli*, the DNA-DEPENDENT DNA POLYMERASE which is involved e.g. in EXCISION REPAIR and in the degradation of RNA primers during DNA replication; it has proof-reading (3′-to-5′ exonuclease) and 5′-to-3′ exonuclease activities. It can also act as a reverse transcriptase, although with low efficiency.

Proteolytic cleavage of the polymerase (e.g. with subtilisin) gives rise to the (75-kDa) *Klenow fragment*, an enzyme with polymerase and proof-reading activities – but lacking the 5′ exonuclease activity. Klenow fragment can be used e.g. in the Sanger (chain-termination) method of DNA sequencing and also for 3′-end-labeling of dsDNA, for second-strand synthesis of cDNA, and for filling in 5′ overhangs. The enzyme also has strand-displacement activity.

(See also EXO KLENOW FRAGMENT.)

DNA polymerase II (in *Escherichia coli*) A DNA-DEPENDENT DNA POLYMERASE that is induced following damage to DNA or inhibition of replication (in the so-called SOS SYSTEM).

(See also DNA POLYMERASE IV.)

DNA polymerase III (in *Escherichia coli*) The (major) DNA-DEPENDENT DNA POLYMERASE which is used e.g. in chromosomal replication; it has both 3′ and 5′ exonuclease activities.

DNA polymerase IV (in *Escherichia coli*) A DNA polymerase (product of the *dinB* gene) which is involved in mutagenesis when induced in the SOS SYSTEM.

(See also DNA POLYMERASE II and DNA POLYMERASE V.)

In *recA730* strains of *Escherichia coli*, constitutive expression of the SOS response gives rise to a mutator phenotype in which DNA polymerase IV is reported to contribute significantly [J Bacteriol (2006) 188(22):7977–7980].

(See also Y FAMILY (of DNA polymerases).)

DNA polymerase V (in *Escherichia coli*) A DNA polymerase that consists of two UmuD′ proteins and one UmuC protein (products of the *umuD* and *umuC* genes respectively), which is involved in the error-prone repair of DNA (TRANSLESION SYNTHESIS) when induced in the SOS SYSTEM.

(See also DNA POLYMERASE IV, TOL PLASMID and Y FAMILY (of DNA polymerases).)

DNA polymerase α See B FAMILY and APHIDICOLIN.

DNA polymerase β See X FAMILY.

DNA polymerase δ See B FAMILY.

DNA polymerase ε See B FAMILY.

DNA polymerase ζ (pol ζ; pol zeta) A DNA-DEPENDENT DNA POLYMERASE (of the B FAMILY) that is found e.g. in the yeast *Saccharomyces cerevisiae* (a homolog is found in humans); it is an 'error-prone' enzyme involved in DNA-damage-induced mutagenesis. The yeast enzyme is a heterodimer; the catalytic subunit is encoded by gene *REV3*, and the accessory subunit is encoded by gene *REV7*.

DNA polymerase η See Y FAMILY.

DNA polymerase θ (pol *θ*; pol theta) See A FAMILY.

DNA polymerase ι See Y FAMILY.

DNA polymerase κ See Y FAMILY and 8-OXOG.

DNA polymerase λ See X FAMILY.

DNA polymerase μ See X FAMILY.

DNA polymerase Q (POLQ, pol Q, Pol Q etc.) See A FAMILY.

DNA Polymorphism Discovery Resource A collection of 450 samples of DNA which were obtained from residents of the USA with ancestry from all major regions of the world; these samples are intended for use in studies to facilitate discovery of human genetic variation [Genome Res (1998) 8(12):1229–1231; erratum: Genome Res (1999) 9(2):210].

[Examples of use of resource (discovery of rare nucleotide polymorphisms): Nucleic Acids Res (2006) 34(13):e99; and studies on the *SLC17A8* gene: Am J Hum Genet (2008) 83 (2):278–292.]

DNA profile Data specific to the DNA of a given individual. For example, a profile may consist of details of the number of short tandem repeats at given genomic loci (see e.g. CODIS).

DNA–protein interactions Proteins interact with DNA, sometimes more or less transiently, in a variety of situations which include replication, transcription, recombination and repair.

Methods for studying such interactions (e.g.): EMEA, EMSA, CHROMATIN IMMUNOPRECIPITATION, the scintillation proximity assay (see SPA), the BIOTIN INTERFERENCE ASSAY and FRET-based systems. At the *molecular* level, such interactions may be studied by the techniques used in SINGLE-MOLECULE DNA MANIPULATION/VISUALIZATION.

A method for studying intracellular DNA–protein binding in *Escherichia coli* K-12 uses a plasmid containing a cassette that includes the target DNA and repeated copies of the LacI binding site (see LAC OPERON) – the cassette being bracketed by cutting sites for the I-SceI homing endonuclease; the test strain (*E. coli* K-12) synthesizes a FLAG-tagged LacI protein. Target DNA, with any bound protein, is isolated by excising the cassette with (induced) I-SceI nuclease, and the complex is released from the bacteria and then purified by the use of bead-bound anti-FLAG antibodies [Nucleic Acids Res (2009) 37(5):e37].

DNA quadruplex See QUADRUPLEX DNA.

DNA repair Any of various (normal) physiologic processes which recognize and repair damaged/abnormal DNA.

Some repair processes remove the damaged (or abnormal) part of a strand and replace it by synthesis (see e.g. EXCISION REPAIR).

By contrast, direct reversal of damage caused by a methylating agent in *Escherichia coli* can be carried out by the Ada protein; this protein is a bifunctional methyltransferase which can transfer a methyl group from e.g. O^6-methylguanine to a cysteine residue (on the Ada protein itself). Methylation of a phosphodiester bond can also be reversed by Ada: the methyl group is transferred to a different cysteine residue on the Ada protein.

Some repair systems are constitutive (i.e. usually present in the cell). By contrast, the Ada protein is part of an inducible repair system (the so-called 'adaptive response') produced in *E. coli* in response to the presence of low levels of certain alkylating agents (e.g. *N*-methyl-*N′*-nitro-*N*-nitrosoguanidine (MNNG) and *N*-methyl-*N*-nitrosourea (MNU)). Another enzyme produced in the adaptive response is 3-methyladenine DNA glycosylase II (the AlkA protein) which can cleave the methylated bases from N^3-methyladenine, N^3-methylguanine and N^7-methylguanine as well as from O^2-methylcytosine.

The Tag (TAG) protein (*tag* gene product) is a constitutive glycosylase that specifically cleaves N^3-methyladenine. [Activity of Tag: EMBO J (2007) 26:2411–2420.]

The DNA repair enzyme O^6-alkylguanine-DNA alkyltransferase (AGT) has been used as a partner in fusion proteins (see GENE FUSION (uses)).

(See also BASE EXCISION REPAIR and NON-HOMOLOGOUS DNA END-JOINING.)

DNA replication *In vivo*: the synthesis of a (dsDNA) molecule in a *semi-conservative* mode, a process in which each strand is used as a template for synthesis of a new, complementary strand – resulting in the formation of two dsDNA molecules, each consisting of one original (parental) strand and a newly synthesized strand.

Typically, initiation of replication *in vivo* (on each strand of duplex DNA) requires prior synthesis of a PRIMER (see also OKAZAKI FRAGMENT). However, in certain cases initiation of replication depends on so-called protein primimg (see e.g. the entry PHAGE φ29).

A distinct form of replication is used to maintain chromosomal length: TELOMERASE extends a (shortened) strand on the RNA template, the extended strand then being copied.

(See also DNA SYNTHESIS IN VIVO (monitoring).)

[DNAReplication (a database of information and resources on (eukaryotic) DNA replication): Nucleic Acids Res (2009) 37(Database issue):D837–D839.]

DNA synthesis in vitro

For *in vitro* processes of DNA synthesis see the entries DNA AMPLIFICATION and NUCLEIC ACID AMPLIFICATION.

DNA sampling See, for example, BUCCAL CELL SAMPLING and FORENSIC APPLICATIONS.

DNA sequencing Determining the identity and position of each of the constituent nucleotides in a molecule of DNA (e.g. a genome or fragment); this provides definitive information on a given molecule – information needed e.g. for identification, characterization or taxonomic studies. (Rapid, accurate (and

also inexpensive) methods for sequencing genomic DNA are required e.g. for FORENSIC APPLICATIONS and for scientific/medical studies.)

Once determined, the sequence of nucleotides in a given sample of DNA may or may not give a direct indication of the composition of any protein(s) encoded by the DNA; this is because the actual sequence of amino acid residues in a protein can be influenced by factors such as INTRONS and by events such as FRAMESHIFTING during translation – i.e. the sequence of nucleotides does not necessarily indicate a series of codons (see CODON).

As the strands in duplex DNA are normally complementary (so that the sequence of one strand can be deduced from the sequence of the other), it might be supposed that sequencing of only one strand would be adequate; however, sequencing of both strands increases the level of accuracy.

Some of the techniques for sequencing require the sample to be available as a single-stranded template (prepared e.g. by asymmetric PCR, or cloning in a phage such as M13); however, such templates are not necessary in the 'new generation' methods for DNA sequencing.

Methods used for sequencing DNA

1. Recent methods

'2nd generation' or 'new generation' methods are described in the entry MASSIVELY PARALLEL DNA SEQUENCING (q.v.).

Third-generation sequencing methods have been discussed [Nature (2009) doi: 10.1038/news.2009.86].

A nanopore (with adaptor) identified unlabeled nucleotides with average accuracy of 99.8% [Nature Nanotechnol (2009) doi: 10.1038/nnano.2009.12].

The guanine bases in single-molecule DNA were reported to have been identified with a scanning tunneling microscope [Nature Nanotechnol (2009) 4:518–522].

2. Established methods

One well-known procedure – the DIDEOXY METHOD (q.v.) – is used for sequencing templates of up to ~500 nt in length. Longer templates may be sequenced by using a solid-phase approach (see PRIMER WALKING). The dideoxy method can

DNA SEQUENCING (summary of methods)[a]

Method	Features
Dideoxy (chain-termination) method (Sanger's method)	Labeled primers are extended on template strands in a reaction mixture which includes a proportion of dideoxyribonucleoside triphosphates (of one type), resulting in chain-terminated products of various lengths (corresponding to locations of the complementary nucleotide within the template). A similar reaction is carried out for each of the four types of nucleotide. Products from each of the four reactions are examined (in separate lanes) by gel electrophoresis, the location of each band in the gel indicating the identity of the nucleotide at a particular location within the sequence. (See entry DIDEOXY METHOD.)
Hybridization-based sequencing	Sample strands are hybridized to an array of oligonucleotides which, collectively, represent all possible combinations of nucleotides; the particular oligonucleotides to which the sample strands bind are analysed by computer and these data are used to indicate the sequence of the sample DNA. (See entry HYBRIDIZATION-BASED SEQUENCING.)
Massively parallel sequencing	'New generation' sequencing: see entry MASSIVELY PARALLEL DNA SEQUENCING for details
Maxam–Gilbert (chemical cleavage) method	End-labeled strands of sample DNA are cleaved (ideally at only one location in each strand) by a base-specific chemical agent, resulting in products of various lengths (as different strands are cut at different locations of the given base within the strand). A similar reaction is carried out for each of the four types of nucleotide, using a different base-specific agent in each reaction. Products from each of the four reactions are examined by gel electrophoresis to identify the nucleotides at specific locations. (See entry MAXAM–GILBERT METHOD.)
Nanopore-based sequencing	A possible future method based on translocation of DNA molecules through a nanopore with the identification of each nucleotide in transit
Pyrosequencing™	A primer, bound to the template strand, is extended by the sequential addition of each of the four types of dNTP (dATPαS being used in place of dATP – see entry). The addition of a nucleotide to the primer releases pyrophosphate which is converted, enzymatically, to ATP. The ATP generates a light signal, via the luciferase system, thus indicating the identity of the nucleotide incorporated at a given location. (See entry PYROSEQUENCING.)
Solid-phase sequencing	End-bound sample strands are sequenced by a series of primers, each primer being extended over a different region of the strand. (See entry PRIMER WALKING.)

[a]See separate entries for further details.

be automated (see AUTOMATED SEQUENCING).

Pyrosequencing™ (see PYROSEQUENCING) can be used for relatively short templates (e.g. ~40 nt) (and is used e.g. in the massively parallel method GS FLX).

The MAXAM–GILBERT METHOD (= the chemical cleavage method) is a mechanistically distinct approach involving the base-specific cleavage of end-labeled fragments and analysis of the resulting products.

Visualization of sequencing products in gels (which originally involved the use of radioactive labels) can be achieved with e.g. fluorescent labels; alternatively, use can be made of CHEMILUMINESCENCE. Fast, capillary-based electrophoresis of the products can be used in place of the gel slab technique.

HYBRIDIZATION-BASED SEQUENCING is a further option.

DNA sequencing has been reported with a MALDI–TOF-based approach [Nucleic Acids Res (2007) 35(8):e62].

Sequencing of ancient DNA

See e.g. the entry SPEX.

DNA sex determination assay A (DNA-based) assay which is used for determining the gender of the individual from which a given sample of DNA has been obtained.

One widely used assay (e.g. in archeological and forensic work) involves amplification of a specific sequence in each of the two genes encoding AMELOGENIN (q.v.); these genes occur on the X and Y chromosomes. In each of these genes, a sequence in the region of the first intron is targeted by a pair of PCR primers; these targets are chosen as they yield easily distinguishable products of different sizes and sequences. In the assay, the formation of products of one type only (from the X chromosome) generally indicates a female source of the sample DNA – while products of two types (from the X and Y chromosomes) indicates a male source.

Various commercial assay kits are available, and different sets of primers are used, but the same principle applies in all of the assays.

Failures in accuracy of the amelogenin assay can occur as a result of mutation. For example, mutation in the Y allele, blocking amplification of the specific sequence from the Y chromosome, could lead to an incorrect interpretation of the result: products from the X chromosome (only) indicating a female source.

Failures in the amelogenin assay may be affected by ethnic group or geographic location. Thus, in one study on an Indian population, relevant deletions were observed in the Y copy of the amelogenin gene in 10 of 4257 males [BMC Med Genet (2006) 7:37].

The SRY-BASED ASSAY is a gender-determination assay that involves a different principle.

(See also BARR BODY, FORENSIC APPLICATIONS, and SEX SORTING.)

DNA shuffling (gene shuffling; sometimes referred to as 'sexual PCR') A technique in which the properties of a given protein can be modified by extensive *in vitro* recombination between the encoding gene and a number of allelic or related forms of that gene.

The isolated genes are initially cleaved by DNase I to form fragments of ~50 bp. The fragments from all of these genes are then used in a primer-less form of PCR in which *the fragments prime each other*; fragment–fragment binding occurs owing to high-level homology among the genes. During this process there is frequent switching of templates (successive priming of different fragments), giving rise to a wide range of chimeric forms. These chimeric forms are then amplified by conventional PCR, and the resulting library of recombinants are screened for products with the desired characteristics.

One or more further rounds of shuffling can be carried out if required.

(See also PSEUDORECOMBINATION.)

DNA Skyline A form of notation in which a given sequence of nucleotides is shown as a series of symbols, rather than as a series of the letters A, C, G and T; in this system, the bases adenine, cytosine, guanine and thymine are represented by four symbols of different height, and these symbols are used instead of the usual letters (see figure). Skyline was designed to facilitate *visual* comparison of strand complementarity and conserved motifs etc. [BioTechniques (2006) 40:740].

(See also OB GENE.)

DNA staining A procedure in which a dye, often a fluorophore, is bound to DNA. Such binding may be non-covalent (e.g. by INTERCALATING AGENTS such as ETHIDIUM BROMIDE, or by minor-groove-binding agents like DAPI and HOECHST 33342) – or covalent (e.g. following amine modification: see the entry PROBE LABELING). (See also CLICK-IT EDU.)

DNA may be stained e.g. to detect bands of amplification products in gel electrophoresis strips; to study preparations of chromosomes, or *in situ* genomes; to demonstrate chromosome banding; to facilitate flow cytometry (see, for example, SEX SORTING); and to distinguish between living and dead cells; the latter role depends on the capacity of some dyes to penetrate the cell membrane in living cells (while other dyes are excluded) – see e.g. LIVE/DEAD BACLIGHT BACTERIAL VIABILITY KIT.

The long list of DNA-staining dyes include e.g. ACRIDINES, BENA435, BOXTO, DAPI, ETHIDIUM BROMIDE and ETHIDIUM MONOAZIDE, HEXIDIUM IODIDE, HOECHST 33342, PROPIDIUM IODIDE, QUINACRINE, SYBR GREEN I and various SYTO DYES, and YOYO-1.

Some (fluorescent) dyes can indicate the local conditions at their binding sites by modifying their emission spectra: see SOLVATOCHROMIC FLUORESCENT DYE.

ssDNA may be quantitated e.g. with OLIGREEN.

(See also ULTRAVIOLET ABSORBANCE.)

DNA stretching See e.g. DNA COMBING.

DNA synthesis See the entries DNA AMPLIFICATION (*in vitro* synthesis) and DNA REPLICATION (*in vivo* synthesis).

DNA synthesis *in vivo* (monitoring) Monitoring the occurrence of DNA synthesis *in vivo* is often achieved by allowing cells/organisms to incorporate a labeled nucleotide or a nucleotide analog in their DNA during growth; after growth, the label or analog is detected/quantitated. Incorporation *per se* indicates that synthesis occurred, while the amount of label, or analog,

DNA Skyline: a form of notation which facilitates visual inspection/comparison of nucleic acid sequences. The design of the Skyline GATC font is shown in the top panel. The center panel shows a sequence of nucleotides in the obesity (OB) gene from different species of mammal. In the lower panel the nucleotides shown in the center panel are displayed in the GATC version of the Skyline font.

Figure reproduced, with permission, from 'DNA Skyline: fonts to facilitate visual inspection of nucleic acid sequences' by Jonas Jarvius and Ulf Landegren, *BioTechniques* (2006) 40:740.

incorporated indicates the extent of synthesis in the period of growth.

^{3}H-labeled thymidine has been widely used in this context; the label can be detected e.g. by autoradiography.

The analog BRDU can be detected (in DNA) by denaturing the DNA and then using labeled anti-BrdU antibodies (as in the procedure for a CELL PROLIFERATION ASSAY).

A simpler and more rapid and convenient approach uses the analog 5-ethynyl-2′-deoxyuridine (EdU); when incorporated in DNA, EdU can be detected by certain fluorescent azides which bind covalently to the EdU. [EdU-based detection of DNA synthesis *in vivo*: Proc Natl Acad Sci USA (2008) 105 (7):2415–2420.]

In some cases it is required to detect DNA synthesis which is occurring *specifically* to repair *damaged* DNA, and in such cases synthesis associated with (S-phase) genome replication must be excluded. One example of this is found in a type of test used for diagnosis of XERODERMA PIGMENTOSUM (q.v.).

DNA tagging (by diene modification) See entry DIELS–ALDER REACTION (reference).

DNA thermal cycler *Syn*. THERMOCYCLER.

DNA toroid A highly compacted form of DNA which is generally believed to be the form of DNA in the nucleoid of the bacterium *Deinococcus radiodurans* and to occur in bacterial endospores. It has been suggested that the toroidal structure may contribute to the resistance of *D. radiodurans* (and of endospores) to DNA-damaging agents such as certain types of radiation – a toroidal structure facilitating repair of double-stranded breaks in DNA in a template-independent, RecA-independent way [J Bacteriol (2004) 186:5973–5977]. (DNA in human sperm cells was also reported to occur in nucleoprotamine toroids [J Cell Sci (2005) 118(19):4541–4550].)

Some authors question the toroidal model of DNA in the *D. radiodurans* nucleoid [J Bacteriol (2006) 188:6053–6058].

DNA unwinding protein See HELICASE.

DNA uptake site See TRANSFORMATION.

DNA vaccination *Syn*. GENETIC IMMUNIZATION (q.v.).

DNA vaccine Any parenterally administered vaccine containing DNA encoding specific antigen(s) or other agents that are *synthesized within the vaccinated individual*; these vaccines may be designed e.g. to induce humoral (i.e. antibody) and/or cell-mediated immune responses to given antigen(s).

(See also the destination vector pVAX^{TM}200-DEST in the table of vectors in GATEWAY SITE-SPECIFIC RECOMBINATION SYSTEM.)

A DNA vaccine may be administered intradermally using a biolistic approach (see also GENE GUN). Administration may also be carried out by intramuscular injection, by aerosol, or by intravenous delivery.

The efficacy/immunogenicity of a vaccine can be improved e.g. by optimization of codon usage [Infect Immun (2005) 73 (9):5666–5674] and by using a DNA vector encoding certain proteins *in addition to* target sequence(s), e.g. CALRETICULIN [J Virol (2004) 78(16):8468–8476] or VP22 protein of bovine herpesvirus 1 [J Virol (2005) 79(3):1948–1953].

Effective induction of specific antibodies within laboratory animals has been achieved by the intravenous delivery of a plasmid encoding the corresponding antigen [BioTechniques (2006) 40(2):199–208].

Bacteriophages have been used as vehicles for delivering DNA vaccines in a mammalian setting [Infect Immun (2006) 74(1):167–174].

A DNA vaccine encoding antigen E7 of HPV-16 (human papillomavirus type 16) and calreticulin – in association with BORTEZOMIB treatment – was reported to generate a more potent E7-specific CD8^{+} T cell immune response, and better therapeutic effects against tumors in mice, as compared with monotherapy [J Mol Med (2008) 86(8):899–908].

Survival in mice with sepsis was reported to be improved by a DNA vaccine against macrophage migration inhibitory factor (MIF) [Gene Therapy (2008) 15:1513–1522].

[DNA vaccination of horses: Acta Vet Scand (2008) 50(1): 44.]

A DNA vaccine which expresses a mutant, non-toxic form of the Shiga-like toxin Stx2 gave partial protection against Stx2 challenge in mice; such a vaccine, either alone or as part of a composite treatment, may be useful in the prophylaxis or therapeutic intervention against enterohemorrhagic strains of *Escherichia coli* (EHEC) [Clin Vaccine Immunol (2009) 16 (5):712–718].

(See also GENE THERAPY.)

***dnaA* gene** (*Escherichia coli*) A gene whose product (DnaA) is required for DNA replication. During initiation of replication, copies of DnaA bind to so-called 'DnaA boxes' located in the origin of replication (*oriC*); this is a prelude to strand separation ('melting'), an essential feature of the replication process.

Some early reports indicated that DnaA may be needed for the replication of certain plasmids (including pSC101), but it now appears that this is not the case.

In eukaryotic cells, six proteins, Orc1–6, bind to the origin of replication; they constitute the *origin recognition complex* (ORC).

In archaeans, the heterodimer Cdc6/Orc1 appears to carry out a similar function.

dnaB* gene (Escherichia coli*) See HELICASE.

***dnaF* gene** (*Escherichia coli*) *Syn*. NRDA GENE.

***dnaG* gene** (*Escherichia coli*) See RNA POLYMERASE.

DNAReplication A database of information and resources for researchers on (eukaryotic) DNA replication [Nucleic Acids Res (2009) 37(Database issue):D837–D839].

DNase (deoxyribonuclease) Any enzyme that cleaves phosphodiester bonds in DNA. An *exo*deoxyribonuclease is a DNase that cleaves terminal bonds, while an *endo*deoxyribonuclease is a DNase that cleaves internal bonds. Examples of DNases: DNASE I, DNASE II, ENDONUCLEASE S1, EXONUCLEASE III and EXONUCLEASE IV.

(See also caspase-activated DNase in APOPTOSIS.)

DNases are produced by various microorganisms and e.g. by the human pancreas.

Microbial DNases include e.g. the staphylococcal 'thermo-

nuclease' – a Ca^{2+}-dependent thermostable enzyme with both exonuclease and endonuclease activity.

(See also STREPTODORNASE.)

To detect DNase production by bacteria, the test strain is plated on an agar medium that contains DNA and a calcium salt. Following development of colonies, the plate is flooded with hydrochloric acid (1 N HCl) to precipitate non-degraded DNA; any DNase-producing colonies are encircled by a clear halo in which the DNA has been degraded.

DNase I A DNASE whose products have 5′-phosphate terminal groups (compare DNASE II). This enzyme is EC 3.1.21.1.

(See also STREPTODORNASE.)

DNase I footprinting See FOOTPRINTING.

DNase II A DNASE whose products have 3′-phosphate terminal groups (compare DNASE I). This enzyme is EC 3.1.22.1.

DNAzyme *Syn*. DEOXYRIBOZYME.

Dnmt1 (DNMT1) DNA METHYLTRANSFERASE 1.

(See also DEMETHYLATION.)

DNMT3B The gene encoding (mammalian) DNA methyltransferase 3B.

dodecahedron base (adenoviral) See the entry ADENOVIRUS.

doi Digital object identifier: a unique reference which identifies a specific item.

Dolly the sheep See SCNT.

dominant templates Those templates, present in great excess, which may, as a result of their dominance, interfere e.g. with the amplification of rare or low-copy-number templates.

For approaches to dealing with dominant templates see e.g. entries PCR CLAMPING and SUICIDE POLYMERASE ENDONUCLEASE RESTRICTION for DNA templates, and COMPETIMER for RNA templates.

DON (6-diazo-5-oxo-L-norleucine) A compound which acts as an analog of L-glutamine and interferes with the biosynthesis of purines (and, hence, nucleotides); DON is produced by the Gram-positive bacterium *Streptomyces*.

DON has antimicrobial and antitumor activity.

(cf. AZASERINE; see also HADACIDIN.)

donor (conjugative) See CONJUGATION and F PLASMID.

donor conjugal DNA synthesis (DCDS) See CONJUGATION.

donor splice site (donor splice junction) In a molecule of pre-mRNA: the splice site (consensus sequence GU) at the 5′ end of an intron.

(cf. ACCEPTOR SPLICE SITE.)

dot blot A simple PROBE-based method for detecting and/or quantitating a given sequence of nucleotides in a sample. A drop of the sample, on a membrane, is treated so that all the nucleic acid molecules are released and bound to the membrane in single-stranded form. Target-specific probes are then used to detect the given sequence; the probes are usually labeled with e.g. DIGOXIGENIN or BIOTIN. For quantitation of the target, the sample can be diluted serially and each dilution probed in the way described; quantity is estimated by comparing the strength of label from one or more dilutions of the sample with the strength of label from control(s) of known concentration.

Slot blot is a similar technique involving elongated (rather than dot-like) inoculations of sample on the membrane.

double helix See B-DNA.

double inhomogeneous field electrophoresis See PFGE.

double-strand break (in DNA) See DSB.

down mutation (down-promoter mutation) Any mutation, in a promoter sequence, resulting in decreased transcription from that promoter.

Down's syndrome (detection) Down's syndrome is commonly associated with trisomy (in this case, three copies of chromosome 21). This and other forms of aneuploidy can be detected e.g. by COMPARATIVE GENOMIC HYBRIDIZATION; FISH (that is, hybridization of fluorophore-labeled probes to preparations of metaphase chromosomes); and also MULTIPLEX LIGATION-DEPENDENT PROBE AMPLIFICATION.

A non-invasive prenatal diagnosis of fetal chromosomal aneuploidy – from fetal DNA present in maternal plasma – has been successfully carried out by adapting a new-generation DNA sequencing technique. From 28 (1st and 2nd trimester) maternal plasmas, all 14 'trisomy 21' fetuses were correctly identified [Proc Natl Acad Sci USA (2008) 105(51):20458–20463]. (cf. EPIGENETIC ALLELIC RATIO ASSAY.)

(See also entry PGD.)

(cf. EDWARDS' SYNDROME.)

downstream Refers (e.g.) to the direction in which a strand of nucleic acid is synthesized, i.e. the 5′ → 3′ direction.

The converse is *upstream*.

downstream box In some prokaryotic and phage genes: a DNA sequence, downstream of the transcription start site, which is associated with enhancement of translation; it was suggested that this sequence may bind to certain bases in the 30S ribosomal subunit (the *anti-downstream box*) but this notion was questioned [J Bacteriol (2001) 183(11):3499–3505].

downstream promoter element (DPE) In certain eukaryotic genes: a region of the CORE PROMOTER located about 30 nt downstream of the START SITE. It binds an initiation factor and is apparently functionally analogous to the TATA box.

The DPE mediates efficient initiation of transcription in at least some TATA-less promoters.

DPE DOWNSTREAM PROMOTER ELEMENT.

DPN Diphosphopyridine nucleotide: an earlier name for NAD.

DpnI A methylation-dependent restriction endonuclease – see entry RESTRICTION ENDONUCLEASE (table). DpnI is used e.g. in the QUIKCHANGE SITE-DIRECTED MUTAGENESIS KIT and in the EXSITE SITE-DIRECTED MUTAGENESIS KIT.

Dpo4 A DNA polymerase (of the Y FAMILY) which occurs in the archaean *Sulfolobus solfataricus*.

(See also BYPASS DNA POLYMERASE.)

DR DIRECT REPEAT.

DR locus In members of the *Mycobacterium tuberculosis* complex: a distinct region of the genome consisting of a number of highly conserved 36-bp direct repeats that are interspersed with spacers of 34–41 bp. The number of direct repeats varies between strains, and the presence/absence/length of spacers also varies; such variation is the basis of SPOLIGOTYPING.

***drd* plasmid** A conjugative plasmid whose transfer operon is permanently DEREPRESSED.

Driselase A product (Sigma-Aldrich) used e.g. for degradation/permeabilization of cell walls in fungi and plants. Although it is usually effective with the fungus *Aspergillus fumigatus*, the *rasA* mutant of *A. fumigatus* was reported to be resistant to it [Eukaryotic Cell (2008) 7(9):1530–1539].

Uses (examples): maize [Plant Cell (2009) 21(1):234–247]; *Aspergillus nidulans* [Mol Biol Cell (2009) 20(2):673–684]; *Arabidopsis thaliana* [Plant Physiol (2009) 149(1):424–433].

(See also ZYMOLYASE.)

drive system (gene drive system) See e.g. GMMS.

Drosha An RNase III-like enzyme commonly involved in the biogenesis of MICRORNAS (cf. MIRTRON).

***Drosophila* expression system** A system (Invitrogen, Carlsbad CA, USA) used for the synthesis of recombinant proteins in insect cells with simple types of vector: see the entry DES.

Drosophila melanogaster A species of fruitfly much studied in classical genetics and also used e.g. for studying POLYTENE CHROMOSOMES. The P ELEMENT is used for mutagenesis in this species.

A useful source of information: www.fruitfly.org

[Genome sequence of *Drosophila melanogaster*: Science (2000) 287:2185–2195.]

Drosophila Schneider S2 cells are used e.g. in a system for expressing recombinant proteins: see DES.

(Relevant entries: AUBERGINE, COPIA ELEMENT, MARINER FAMILY, POLYCOMB-GROUP GENES, TELOMERASE.)

DSB (1) Double-strand break: refers to cleavage of both strands at a given site in a duplex molecule of nucleic acid.

(See also NON-HOMOLOGOUS DNA END-JOINING.)

(2) An abbreviation for 3-*O*-(3′,3′-dimethylsuccinyl)betulinic acid (see BEVIRIMAT).

dsDNA Double-stranded DNA.

DsrA A bacterial sRNA: see the entry SRNAS. As is typical of bacterial sRNAs, DsrA binds to the 5′-UTR of the mRNA (in this case *rpoS*); however, binding of DsrA brings about the opening of a stem–loop structure which, when extant, acts to repress transcription. Sequences upstream of this stem–loop apparently increase the sensitivity of *rpoS* mRNA to the Hfq chaperone and DsrA [RNA (2008) 14(9):1907–1917].

DsRed A red fluorescent protein from the reef coral *Discosoma* which forms tetramers. (cf. MONOMERIC RED FLUORESCENT PROTEIN; see also entries for GREEN FLUORESCENT PROTEIN and YELLOW FLUORESCENT PROTEIN.)

[Examples of use: BioTechniques (2007) 42(3):285–288; Neoplasia (2009) 11(7):683–691 and PLoS ONE (2009) 4(7): e6193.]

DTT DITHIOTHREITOL.

dual-fluorescence reporter/sensor plasmid DFRS PLASMID.

dual-function codons See GENETIC CODE.

dual-tropic strains (of HIV-1) See HIV-1.

Duchenne muscular dystrophy A genetic disorder characterized by the absence (or abnormality) of *dystrophin*, a protein found in normal muscle. (See also table in the entry GENETIC DISEASE.)

(cf. NONSENSE-ASSOCIATED DISEASE.)

The (2.4 Mb) gene for dystrophin has been incorporated in a HUMAN ARTIFICIAL CHROMOSOME (q.v.).

Antisense oligomers were used *in vitro* (in human skeletal muscle cells) to induce exon skipping in the *DMD* gene – this giving rise to an internally truncated but functional form of dystrophin [Mol Therapy (2009) doi: 10.1038/mt.2008.287].

(See also ALTERNATIVE SPLICING.)

dUTP poisoning In PCR: an inhibitory effect on certain types of DNA polymerase which results from deamination of dCTP to dUTP [Proc Natl Acad Sci USA (2002) 99:596–601].

Some commercial enzymes, such as the *PfuUltra*™ DNA polymerase (Stratagene, La Jolla CA, USA), are designed to overcome dUTP poisoning.

(See also ARCHAEMAXX.)

Dyggve–Melchior–Clausen syndrome A syndrome, involving e.g. skeletal abnormalities; it is associated with mutation in the *DYM* gene (location 18q12–q21.1).

Dynabeads® Microscopic spheres of precise size that contain a mixture of iron oxides (δFe_2O_3 and Fe_3O_4) encased within a thin polymer shell, giving the spheres SUPERPARAMAGNETIC properties. Dynabeads of 2.8 μm diameter are referred to as Dynabeads M-280, those of 4.5 μm diameter are referred to as Dynabeads M-450. The surface of the polymer shell can be modified so as to exhibit any of a wide range of ligands, the choice of ligand being determined by the particular use for which the beads are required.

In use, the required entity (e.g. macromolecule) is allowed to bind to Dynabeads (exhibiting a particular ligand), and the Dynabeads are retained, by a magnetic field, while unwanted macromolecules (etc.) are removed by washing; the required entity can then be eluted from the Dynabeads.

Dynabeads® (Dynal) have at least three main types of use:

- separation of a specific component from a mixture of similar components;
- isolation of a given component from a crude preparation;
- provision of a solid-phase support.

Separation (= *biomagnetic separation*) may involve various types of target, including macromolecules (e.g. nucleic acid species and fragments), viruses and specific types or subtypes of cell. For example, mRNAs can be separated (and isolated) from a sample of total RNA by using Dynabeads Oligo $(dT)_{25}$ – that is, Dynabeads coated with (covalently attached) 25-nucleotide-long chains of deoxythymidines. These dT ligands bind the oligo (dA) tails of mRNA molecules; retention of the bead–mRNA complexes – by means of a magnetic field – permits washing etc. and the removal of unwanted components. If required, subsequently, the mRNA molecules can be eluted from the beads by incubation in a suitable buffer solution.

Separation of a particular type of blood cell from other cells (in e.g. whole blood) can be achieved by means of a BIOTIN-linked antibody (specific to the target cell) bound to streptavidin-coated Dynabeads. This procedure is called *immuno-*

magnetic separation. If required (after washing), the bound, magnetically retained cells can be lysed (while still attached to the Dynabeads), and mRNAs can be recovered from the supernatant with Dynabeads Oligo $d(T)_{25}$.

The isolation of PCR-ready high-molecular-weight genomic DNA from a crude preparation, or e.g. from whole blood, can be achieved e.g. by Dynabeads DNA DIRECT™. The DNA, released by cell lysis, is adsorbed to Dynabeads coated with an appropriate ligand; then, magnetic separation of the beads is followed by a series of washings. One advantage of this procedure is that the washing stages may assist in eliminating inhibitors of PCR.

The solid-phase characteristics of Dynabeads have been exploited e.g. for DNA sequencing by primer walking; for gene assembly; and for hybridization reactions.

The advantages of a solid-phase support include the ability to re-use a DNA template (e.g. for sequencing) several times.

Dynabeads*: *examples of use

Uses include e.g. purification of mRNA from total RNA with Dynabeads Oligo $d(T)_{25}$ [Plant Cell (2006) 18:1750–1765]; separation of T lymphocytes (T cells) from peripheral blood mononuclear cells of (cancer) patients [J Clin Invest (2006) 116(7):1946–1954]; selection of specific immunoglobulins from sera by Dynabeads coated with PROTEIN A and Dynabeads coated with PROTEIN G [Infect Immun (2005) 73(9): 5685–5696]; mRNA isolation from bovine peripheral blood mononuclear cells [BMC Immunol (2006) 7:10]; isolation of a streptavidin–tRNA fusion with BIOTIN-coated Dynabeads [Nucleic Acids Res (2006) 34(5):e44]; separation of B cells and T cells by using Pan B Dynabeads and Pan T Dynabeads, respectively [Proc Natl Acad Sci USA (2006) 103(13):4819–4824]; isolation of biotinylated transcripts with streptavidin-coated Dynabeads in studies on DNA methylation [Nucleic Acids Res (2009) 37(8):2658–2671]; elimination of B cells by incubation with sheep anti-mouse IgG-coated Dynabeads in cell cycle studies [PLoS ONE (2009) 4(4):e5245]; retrieval of biotinylated fragments with streptavidin-coated Dynabeads [BMC Plant Biol (2009) 9:40].

dynamic allele-specific hybridization (DASH) A method used for scoring SNPs (and other short polymorphisms) in single-stranded samples of the DNA under test. The single-stranded samples are usually prepared by PCR amplification using one BIOTINylated primer; the strand formed by extension of this primer is then immobilized on a (streptavidin-coated) surface to form the target sequence.

Essentially, a probe – with known complementarity to one possible version of the target sequence – is hybridized to the target sequence, and the probe–target duplex is monitored for denaturation with increasing temperature, the object being to determine the temperature at which denaturation occurs at a maximum rate. The different values of temperature which are found with different samples indicate differences in relative stability; these data are used to reveal the nature of the target sequences being examined.

[Uses (e.g.): a study on bidirectional transcription [PLoS Genet (2008) 4(11):e1000258]; detection of gene expression in a particular type of cell in a mixture of cells [PLoS ONE (2009) 4(2):e4427]; identification of peptide mimics of the functional epitope of IgG1-Fc [FASEB J (2009) 23(2):575–585]; studies on activated T cells [FEBS Lett (2009) 583(1): 61–69].

E

E L-Glutamic acid (alternative to Glu).

E. coli See the entry ESCHERICHIA COLI.

E1A protein (adenoviral) See ADENOVIRUS and MELANOMA.

EAF plasmid See PATHOGENICITY ISLAND.

EBER Epstein–Barr virus-(EBV)-associated RNA: either of the two small (∼170-nucleotide) virus-encoded RNA molecules (EBER-1 and EBER-2) which are expressed at high levels in cells latently infected with the EBV; these molecules provide target sequences for the detection of EBV infection.

In latently infected B lymphocytes (B cells), EBERs inhibit α-interferon-induced APOPTOSIS, and thus appear to promote viral persistence in these cells. It had been thought that such inhibition of apoptosis is due to the influence of EBERs on the role of dsRNA-dependent protein kinase PKR; however, direct inhibition of PKR by EBERs was considered unlikely [J Virol (2005) 79(23):14562–14569] and a later study [J Cell Biol (2006) 173(3):319–325] confirmed that EBERs are confined to the cell's nucleus (whereas PKR is cytoplasmic).

(See also EPSTEIN–BARR VIRUS.)

Eberwine technique A technique used for amplifying RNA in a mixed population of mRNA molecules. Essentially, all the mRNAs are reverse transcribed with a poly(T) primer whose 5′ end carries a tag incorporating a promoter sequence for the phage T7 RNA polymerase. RNA–DNA duplexes are treated with RNase H to cause *partial* degradation of the RNA strand in each duplex; fragments of RNA that remain are then used for priming synthesis of DNA and, hence, completion of the second strand. The formation of the second strand provides a functional T7 promoter that is used for transcribing antisense copies of RNA.

[Uses: BMC Genomics (2008) 9:635; Mol Vision (2009) 15:296–311.]

EBNA-1 See EPSTEIN–BARR VIRUS.

Ebola virus A causal agent of viral hemorrhagic fever in man; other animals can be infected experimentally – such infection often being fatal. Ebola virus can be grown in some types of cell culture. The virus is found in certain parts of Africa.

The genome of the Ebola virus consists of one molecule of negative-sense ssRNA. The virion is a helical nucleocapsid enclosed within an envelope that bears surface spikes.

(See also the note on Ebola virus in entry SUMOYLATION.)

EBV EPSTEIN–BARR VIRUS.

EBV-associated RNA See the entry EBER.

ecdysone A steroidal compound which is produced by certain insect cells. It has been used e.g. to induce the expression of mammalian genes whose control system has been engineered to respond to this agent.

Ecl18kI A type IIP RESTRICTION ENDONUCLEASE (recognition sequence: ↓CCNGG).

[Specificity (mechanism): EMBO J (2006) 25:2219–2229.]

EcoCyc database See the entry ESCHERICHIA COLI.

EcoKI A type I RESTRICTION ENDONUCLEASE which cleaves at random sites. The recognition sequence AA^{m}C(6N)GTGC (in which adenine is methylated at N-6) is not cleaved by EcoKI.

Viewed by atomic force microscopy, EcoKI – when bound to DNA – showed dimerization and looping of DNA [Nucleic Acids Res (2009) 37(6):2053–2063].

(See also HSD GENES.)

EcoKMcrA *Syn*. McrA (see the entry MCRA GENE).

EcoKMcrBC *Syn*. McrBC (see the entry MCRBC).

EcoRII-C See CPNPG SITES.

EcoR124I A type IC RESTRICTION ENDONUCLEASE which has random cleavage sites. [Mechanism of translocation: EMBO J (2006) 25:2230–2239.]

Ecotilling A method used for detecting SNPs (and other polymorphisms) (cf. TILLING). The principle of the method is outlined below.

Genomic DNA under test is initially mixed with a sample of the corresponding wild-type/reference DNA. With labeled, gene-specific PCR primers, a target sequence of 1–1.5 kb is then amplified from the mixture of DNA. Denaturation of the products by heat is followed by cooling and then annealing of the single-stranded products.

If the target sequence in *genomic* DNA contains an SNP, or a point mutation, then – during the annealing stage – some single-stranded amplicons, copied from genomic DNA, will hybridize with single-stranded wild-type amplicons (copied from the wild-type/reference DNA). The heteroduplex DNA amplicons (each containing a single-base-pair mismatch) are then exposed to a SINGLE-STRAND-SPECIFIC NUCLEASE which nicks at the site of the mismatch. The two parts of the nicked strand, on either side of the nick, are then examined by gel electrophoresis to determine the site of the nick and, hence, the location of the SNP (or point mutation).

[Example of use (discovery of rare human nucleotide polymorphisms): Nucleic Acids Res (2006) 34(13):e99.]

[An optimized procedure for the design and evaluation of Ecotilling assays: BMC Genomics (2008) 9:510.]

Ecotilling was used to identify base substitutions and indels in *Magnaporthe oryzae* [Plant Cell (2009) 21(5):1573–1591].

EDC 1-ethyl-3-(dimethylaminopropyl) carbodiimide: a fixative used e.g. to prevent the loss of microRNAs in *in situ* hybridization studies on mouse tissue [Nature Methods (2009) doi: 10.1038/nmeth.1294].

editing (1) See PROOF READING.

(2) See RNA EDITING.

editosome A complex involved in RNA EDITING. In trypanosomatids, the editosome may include proteins with functions such as: endonuclease, helicase, terminal uridylyl transferase, exouridylylase and ligase.

EdU 5-ethynyl-2′-deoxyuridine: a thymidine analog used in the CLICK-IT EDU procedure.

EdU has been used e.g. in a diagnostic test for XERODERMA PIGMENTOSUM (q.v.).

Edwards' syndrome (trisomy 18) A condition, associated with

the presence of three copies of chromosome 18, characterized by severe mental impairment, skeletal abnormalities, and an increased incidence of cardiac disease. Most infants with this condition die within the first year.

Edwards' syndrome has been detected pre-natally by means of an EPIGENETIC ALLELIC RATIO ASSAY.

(cf. DOWN'S SYNDROME.)

efavirenz A NON-NUCLEOSIDE REVERSE TRANSCRIPTASE INHIBITOR.

EFC ENZYME FRAGMENT COMPLEMENTATION.

eGFP *Syn*. EGFP – see the next entry.

EGFP (or eGFP) Enhanced GREEN FLUORESCENT PROTEIN: a mutant form of GFP which produces a higher level of fluorescence compared to the wild-type GFP.

[Uses (e.g.): Brain Res (2008) 1190:15–22; J Comp Neurol (2009) 512(5):702–716; PLoS Pathog (2009) 5(2):e1000311; J Biomed Biotechnol (2009) doi:10.1155/2009/149254; Virol J (2009) 6:16.]

EGFR family See EPIDERMAL GROWTH FACTOR RECEPTOR FAMILY.

EGS (1) External guide sequence: see SNRNAS.

(2) Ethyleneglycol-bis(succinimidyl succinate): see the entry CHROMATIN IMMUNOPRECIPITATION.

EGTA Ethyleneglycol-bis(β-aminoethyl-ether)-*N*,*N*,*N*′,*N*′-tetra-acetic acid, a chelating agent.

EGTA complexes various types of cation – including Ca^{2+}, Mg^{2+} and Zn^{2+}. Ca^{2+} is complexed much more strongly than Mg^{2+}.

[EGTA (in a Ca^{2+}-dependence assay): PLoS Pathog (2009) 5(2):e1000301.]

EHEC Enterohemorrhagic *Escherichia coli*: strains of typically food-borne pathogenic *E. coli* which produce disease ranging from mild diarrhea to severe, sometimes fatal bloody diarrhea (*hemorrhagic colitis*). These strains produce one (or both) of two phage-encoded SHIGA-LIKE TOXINS (SLT-I, SLT-II, also referred to as Stx1, Stx2, respectively, or as 'verocytotoxins' VT1, VT2, respectively); these toxins have been associated with vascular damage.

Of the various strains of EHEC, the (serologically defined) strain O157:H7 has been particularly prominent and has been well studied [genome: Nature (2001) 409:529–533 – erratum 410:240]. Note the use of an upper-case 'O' – *not a zero* (0) – in the designation of this strain.

A DNA vaccine, encoding a mutant non-toxic form of Stx2, was reported to give partial protection against Stx2 challenge in mice [Clin Vaccine Immunol (2009) 16(5):712–718].

eIF4E Part of the eukaryotic translation initiation complex that recognizes the CAP region of mRNA.

EK-Away™ resin An agarose-based product from Invitrogen (Carlsbad CA) which is designed for removing enterokinase (such as EKMax™) from reactions involving the cleavage of fusion proteins by this enzyme; the product contains soybean trypsin inhibitor, which has a high affinity for enterokinase.

[Use (e.g.): Proc Natl Acad Sci USA (2008) 105:641–645; Chem Biol Drug Des (2008) 72(1):34–39.]

EKMax™ An ENTEROKINASE (from Invitrogen, Carlsbad CA) which is reported to cleave proteins after the lysine residue in the sequence Asp-Asp-Asp-Asp-Lys.

[Uses (e.g.): Infect Immun (2008) 76(8):3569–3576.]

(See also EK-AWAY RESIN.)

electroblotting A form of BLOTTING, faster than the method based on capillary action (SOUTHERN BLOTTING), in which nucleic acid fragments or proteins are transferred from gel to matrix by means of an electrical potential difference ('electric field') between gel and matrix.

electrocompetent cells See ELECTROPORATION.

electropermeabilization See ELECTROPORATION.

electrophoresis Any method which is used e.g. for separating charged fragments, particles or molecules, in a heterogeneous population, by exploiting their different rates of movement produced by an electrical potential gradient in a liquid-based medium.

(cf. ISOTACHOPHORESIS.)

Electrophoresis is often used for identification/quantitation of nucleic acid fragments, proteins etc., determination of the molecular weight of proteins, preparation of fingerprints (e.g. in typing procedures), and confirmation of specific amplicons in nucleic-acid-amplification techniques such as NASBA and PCR.

GEL ELECTROPHORESIS is widely used in nucleic acid technology.

electrophoretic mobility-shift assay See EMSA.

electroporation A method for inserting nucleic acid into cells by exposing a mixture of cells and nucleic acid to an electric field for a fraction of a second (or for several seconds). Field strengths of up to ~16 kV/cm have been used (although field strengths are often lower than this). Electroporation has been used e.g. for bacterial, yeast, plant and mammalian cells.

(See also NUCLEOFECTION.)

[Example of protocol: PLoS Biology (2006) 4(10):e309.]

Specialized 'electrocompetent' cells, i.e. cells selected to be particularly suitable for electroporation, as indicated by their successful uptake and expression of DNA, are available from various commercial sources – e.g. ElectroTen-Blue™ from Stratagene (La Jolla, CA), and Electrocomp™ GeneHogs® and ElectroMax™ from Invitrogen (Carlsbad, CA).

A procedure with >20% transfection efficiency in mammalian neurons was reported in a 96-well electroporation format [BioTechniques (2006) 41(5):619–624].

Electroporation of single cells

Electroporation of *single* cells can be carried out with the use of a micromanipulator, a microscope and microelectrodes; an advantage (when dealing with adherent cells) is that contact with the electrode can create an increased tension in the cell membrane – an effect which can decrease the level of voltage required for electroporation. Difficulties associated with e.g. manual control of electrodes are addressed in an automated form of single-cell electroporation [BioTechniques (2006) 41 (4):399–402].

in vivo electroporation
Although 'electroporation' is generally understood to refer to the electrically assisted insertion of nucleic acid into *cultured* cells, a similar approach is used to facilitate uptake of nucleic acid into the cells (especially skin) of *living* animals. Such a procedure (designated *in vivo* electroporation, electrotransfer, or electropermeabilization) has been reported e.g. to enhance the immune response to plasmid-encoded antigens (i.e. DNA vaccines).

[Studies on *in vivo* electroporation (e.g.): J Control Release (2007) 124:81–87; Genet Vaccine Ther (2008) 6:16; PLoS ONE (2008) 3(6):e2517.]

electrospray ionization See ESI

ElectroTenBlue® See COMPETENT CELLS.

electrotransfer See ELECTROPORATION (under: *in vivo* electroporation).

elementary body (EB) The infectious form of the organism in the development cycle of CHLAMYDIA (q.v.).

ELISA Enzyme-linked immunosorbent assay: a highly sensitive method for detecting specific antigens or antibodies. For example, to detect a given antigen, use is made of antibodies (specific to that antigen) which are conjugated to a reporter system (such as the enzyme ALKALINE PHOSPHATASE). If the given antigen is present, and immobilized, in the test sample, it should bind the antibodies and thus become labeled with the reporter system. Any unbound antibodies are removed by washing. If the reporter is alkaline phosphatase, addition of a color-generating (or e.g. chemiluminescent) substrate gives rise to an amplified signal which can be used for assaying the antigen.

(See also ELISPOT, EMEA and IMMUNO-PCR.)

ELISPOT Enzyme-linked immunospot: an assay, based on the ELISA principle, that is used e.g. for the detection of specific molecules or cells.

In one example, a sample of blood is assayed specifically to quantitate those T lymphocytes (T cells) which respond to the tuberculosis-associated antigen ESAT-6 (q.v.) by secreting the cytokine γ-interferon (gamma interferon). In this assay, cells in a given sample are first allowed to settle on the floor of the well. Then, on addition of ESAT-6, ESAT-6-specific cells in the sample react by secreting the cytokine – which is in high concentration in the immediate vicinity of each secreting cell. The cytokine is detected by adding a conjugate that consists of: (i) a molecule that binds to γ-interferon, linked to (ii) an enzyme capable of a colorigenic reaction with an appropriate (added) substrate. After adding the substrate, and incubating, the presence of ESAT-6-specific T cells is indicated by a visible spot of color at the location of each secreting T cell.

The ELISPOT approach has been used in various ways, e.g. to screen dominant peptides recognized by $CD8^+$ T cells in mice and in humans infected with *Trypanosoma cruzi* [PLoS Pathogens (2006) 2(8):e77]; for the characterization of T cell responses to immediate early antigens in humans with genital herpes [Virol J (2006) 3:54]; for detecting T cell responses in HIV infection [PLoS Pathog (2009) 5(5):e1000414]; and for assaying IFN-γ and IL-2 (interleukin-2) secretion by lymphocytes in a study of immune responses to Ebola virus [PLoS ONE (2009) 4(5):e5547].

Elk-1 A protein which behaves as a transcription factor when activated appropriately (see also PATHDETECT SYSTEMS).

This protein also occurs in extranuclear locations and was reported to be associated with the mitochondrial permeability transition pore complex [Proc Natl Acad Sci USA (2006) 103 (13):5155–5160].

Elongase® amplification system A product (from Invitrogen, Carlsbad CA, USA) which is used for LONG PCR; it includes *Taq* DNA polymerase and a proofreading polymerase (*Pyrococcus* species *GB-D* polymerase).

elongation arrest See PCR CLAMPING.

Elongator complex A complex of proteins that is associated with elongation during transcription mediated by RNA polymerase II.

elvitegravir An inhibitor of the HIV-1 integrase which blocks integration of the HIV-1 cDNA into the host cell's chromosome by inhibiting DNA strand transfer.

[Antiretroviral action and resistance profile of elvitegravir: J Virol (2008) 82(2):764–774.]

(See also INTEGRASE INHIBITOR.)

EMBL European Molecular Biology Laboratory. A source of nucleotide sequence data (EMBL-Bank) is available at: www.ebi.ac.uk/embl.html

EMBL-Bank See EMBL.

embryonic stem cells (ES cells) See STEM CELL.

EMCV Encephalomyocarditis virus.

(See also IRES.)

EMEA Exonuclease-mediated ELISA-like assay: a probe-based method used for detecting the binding of transcription factors to DNA.

The method involves dsDNA probes, immobilized at one end, which include a distal binding site for a specific transcription factor and a proximal (internal) site labeled (in one strand) with DIGOXIGENIN. The plate with the immobilized probes is incubated with a sample of nuclear extract in order to permit the binding of the given transcription factor, if it is present. The whole is then exposed to the exonucleolytic activity of EXONUCLEASE III.

If the given transcription factor is present, and has bound to a probe, then exonuclease III will not be able to degrade that strand of the (dsDNA) probe which contains the digoxigenin label – because its activity will be blocked at the site of the bound transcription factor. As a consequence, the (internal) digoxigenin label will remain intact, and this label can be detected (as in the basic ELISA approach) by using an anti-digoxigenin antibody that is conjugated to an enzyme such as ALKALINE PHOSPHATASE; adding the enzyme's (colorigenic) substrate then permits a *quantitative* estimation of the given transcription factor in the sample – as the signal's intensity is proportional to the quantity of transcription factor present.

If there are no copies of the given transcription factor in the sample then the dsDNA probes will not be protected from the

activity of exonuclease III; in this case the digoxigenin label will be lost during strand degradation – leading to an absence of signal in the ELISA phase.

This method was used to assay the nuclear factor NF-κB [BioTechniques (2006) 41(1):79–90].

emission filter In fluorescence microscopy: a *bandpass* filter (see EXCITATION FILTER) with two primary functions: (i) to transmit radiation corresponding to the emission spectra of the fluorophore(s) in the sample, and (ii) to block radiation of other wavelengths.

emPCR Emulsion PCR: see e.g. the entry GS FLX.

EMRSA Epidemic MRSA.

EMSA Electrophoretic mobility-shift assay: a technique which has been used to investigate the interactions between proteins (e.g. those found in cell lysates) and particular sequences of nucleotides in dsDNA. In one basic approach, differentially labeled fragments of DNA (containing different target sequences) are incubated with a lysate in order to permit DNA–protein binding. On subsequent electrophoresis, any protein-bound fragment of DNA will exhibit a lower electrophoretic mobility (i.e. it will move more slowly through the gel than unbound fragments) and can be identified, isolated and then examined.

A two-dimensional EMSA procedure was employed for mapping target sites of DNA-binding proteins along a 563-kb region of the human genome [BioTechniques (2006) 41(1): 91–96].

EMSA has been used in a study on the effects of different reagents on the DNA-binding activity of the RAD51 protein [Nucleic Acids Res (2009) doi: 10.1093/nar/gkp200].

Other uses of EMSA (e.g.): studies on monoclonal antibody binding [PLoS ONE (2009) 4(2):e4614]; and investigations on transcription regulation factors [Nucleic Acids Res (2009) 37(3):778–792].

(See also DNA–PROTEIN INTERACTIONS.)

emulsion PCR (emPCR) See e.g. the entry GS FLX.

Enbrel® See BIOPHARMACEUTICAL (table).

End-It™ DNA end-repair kit A product (Epicentre) which includes enzymes T4 DNA polymerase and T4 polynucleotide kinase that convert PCR-derived amplicons with 3′ adenosine overhangs (see the entry EXTENDASE ACTIVITY) to blunt-ended fragments; in this reaction, 3′ overhangs are cleaved, and the 5′ termini are phosphorylated – leaving the amplicons ready for blunt-ended ligation to a suitable type of vector.

[Uses: Genetics (2006) 173(2):1007–1021; Infect Immun (2006) 74(1):449–460; Nucleic Acids Res (2006) 34(12):e84; Proc Natl Acad Sci USA (2008) 105:17273–17277; Genetics (2009) 181(3):1129–1145; Infect Immun (2009) 77(1): 152–161.]

(See also POLISHING.)

end labeling Addition of a label to the 3′ or 5′ end of a strand of nucleic acid.

3′ labeling may be achieved e.g. by adding labeled nucleotide(s) to the 3′-OH group with the enzyme terminal deoxynucleotidyl transferase.

5′ labeling may be achieved e.g. by cleaving 5′-phosphate (with alkaline phosphatase) and replacing it with a labeled phosphate using (ATP-dependent) polynucleotide kinase.

end probe See CHROMOSOME WALKING.

***endA* gene** See ESCHERICHIA COLI (table).

endogenote See MEROZYGOTE.

endogenous retroviral sequence A sequence of DNA, within a eukaryotic genome, which corresponds to, or resembles, the DNA copy of part(s) of a retroviral genome.

endogenous retrovirus Any complete, or defective, retroviral *provirus*, i.e. DNA copy of the retroviral genome, integrated in the host cell's genome. Endogenous retroviruses are widely distributed in the genomes of vertebrates, including humans. Some of the endogenous retroviruses can be activated (e.g. by exposure to radiation, or to chemicals such as 5-bromo-2-deoxyuridine) and may then form infectious or non-infectious virions.

endonuclease Any enzyme which can cleave *internal* phosphodiester bonds in a molecule of nucleic acid – either in a single strand or in both strands of a duplex molecule, according to enzyme. For examples, see: RESTRICTION ENDONUCLEASE, STRUCTURE-SPECIFIC ENDONUCLEASE and also SINGLE-STRAND-SPECIFIC NUCLEASE.

(See also HOMING ENDONUCLEASE.)

endonuclease CEL I A SINGLE-STRAND-SPECIFIC NUCLEASE. It is used e.g. in the CEL I NUCLEASE MISMATCH ASSAY.

endonuclease P_1 A SINGLE-STRAND-SPECIFIC NUCLEASE.

endonuclease S_1 (endonuclease S1, nuclease S1, S1 nuclease) An endonuclease (first obtained from the fungus *Aspergillus oryzae*) which specifically degrades ssDNA and ssRNA. The enzyme is one of a range of nucleases in the SINGLE-STRAND-SPECIFIC NUCLEASE group. Endonuclease S_1 is used e.g. to remove single-stranded overhangs during the preparation of NESTED DELETIONS.

Endonuclease S_1 differs from e.g. the mung bean nuclease in having the ability to cleave the DNA strand opposite a nick in a DNA duplex.

(See also RIBONUCLEASE PROTECTION ASSAY.)

Endo-Porter A peptide reagent (Gene-Tools, Philomath OR, USA) which facilitates the delivery of materials into cells by enhancing the permeability of endocytic vesicles, thereby releasing vesicle contents into the cytoplasm.

This reagent was used e.g. for the delivery of MORPHOLINO ANTISENSE OLIGOMERS into cells of a kidney organ culture for the knockdown of an apoptotic regulator [BioTechniques (2008) 44(4):547–549]. It has also been used to knockdown the DICER function in PC-3 cells by facilitating transfection of morpholino antisense oligomers [Proc Natl Acad Sci USA (2008) 105(5):1608–1613], and for the transfection of splice-switching oligonucleotides [Oligonucleotides (2009) 19(1): 1–14].

(See also MELITTIN.)

endoreduplication See RE-REPLICATION.

endoribonuclease See RIBONUCLEASE.

endotoxin The term commonly used, specifically, to refer to a

macromolecular complex (a PYROGEN) present in the normal cell envelope of Gram-negative bacteria (such as *Escherichia coli*): see LIPOPOLYSACCHARIDE.

enfuvirtide A FUSION INHIBITOR (q.v.).

engineered transposable element A RECONSTRUCTED TRANSPOSABLE ELEMENT.

engineered underdominance constructs See GMMS.

enhanced green fluorescent protein (eGFP) See EGFP.

enhancer (in chemiluminescence systems) See CHEMILUMINESCENCE ENHANCER.

enhancer (transcriptional) A *cis*-acting element that promotes the transcription of certain genes; enhancers occur upstream or downstream of their promoters, within and outside genes, and may function independently of orientation or of distance from the promoter (cf. UAS). Some enhancers can function in a range of different cells, while others are cell-specific. In at least some cases, the activity of an enhancer seems to depend on *trans*-acting proteins.

In yeast, DNA LOOPING was reported between an enhancer and a promoter – suggesting that the apposition of these two elements is involved in transcriptional enhancement, at least in this case [Nucleic Acids Res (2005) 33(12):3743–3750].

(See also entries CHROMATIN INSULATOR and ENHANCER TRAPPING.)

enhancer (translational) Any sequence which is associated with the enhancement of translation (see e.g. DOWNSTREAM BOX). The adenine-rich sequences that are found downstream of the start codon in many *Escherichia coli* mRNAs appear to influence the level of gene expression. One study, using site-directed mutagenesis, assessed the contribution made by the adenine-rich motifs to *in vivo* gene expression and to *in vitro* binding of mRNAs to ribosomes; the results suggested that these adenine-rich motifs may assist gene expression e.g. by facilitating mRNA–ribosome binding [J Bacteriol (2007) 189 (2):501–510].

(See also IRES.)

enhancer trapping A procedure used to detect (transcriptional) ENHANCER sequences in a given genome. TRANSPOSONS that are carrying a REPORTER GENE which has a *weak* PROMOTER are first allowed to insert randomly into the given genome in a population of cells; conditions are chosen so that each cell will receive, on average, only one transposon. Occasionally, a transposon inserts at a location where transcriptional activity of the reporter gene is enhanced by a nearby transcriptional enhancer. When this occurs, DNA of the relevant clone can be isolated and investigated by sequencing the transposon's *flanking* sequences; the flanking sequences may be amplified e.g. by INVERSE PCR to provide the templates for sequencing.

[Transposons as tools for enhancer trap screens: Genome Biol (2007) 8(suppl 1):S8.]

enoxacin See QUINOLONE ANTIBIOTICS.

enterobacteria A general (non-taxonomic) term that may refer to bacteria associated with the gut (i.e. intestine) but which usually refers specifically to bacteria belonging to the family Enterobacteriaceae. All the bacteria in this family are Gram-negative, rod-shaped, non-spore-forming organisms that are facultatively anaerobic; they may adopt a fermentative or an oxidative metabolism according to conditions. The organisms usually grow well on simple (basal) media such as NUTRIENT AGAR. Enterobacteriaceae includes genera such as *Erwinia*, *Escherichia*, *Klebsiella*, *Proteus*, *Salmonella*, *Serratia*, *Shigella*, *Sodalis*, *Wigglesworthia* and *Yersinia*.

(See also ESCHERICHIA COLI and SALMONELLA.)

enterohemorrhagic *Escherichia coli* See EHEC.

enterokinase (*syn*. enteropeptidase) A serine proteinase which is used e.g. for cleaving the components of fusion proteins; in at least some cases it is reported to cleave proteins/peptides between the residues of lysine and isoleucine. EKMax™ (an enzyme from Invitrogen (Carlsbad, CA)) is reported to cleave after the lysine residue in the sequence Asp-Asp-Asp-Asp-Lys.

(cf. THROMBIN; see also AFFINITY PROTEIN EXPRESSION AND PURIFICATION.)

enteropathogenic *Escherichia coli* See EPEC.

enteropeptidase *Syn*. ENTEROKINASE.

enterotoxigenic *Escherichia coli* See ETEC.

enteroviruses A category of PICORNAVIRUSES, many of which cause disease in humans and animals – e.g. polioviruses (the causal agents of poliomyelitis) and coxsackieviruses (causal agents of aseptic meningitis, Bornholm disease, myocarditis and herpangina). Porcine enteroviruses can cause e.g. swine vesicular disease and Teschen disease.

entry clone See entry GATEWAY SITE-SPECIFIC RECOMBINATION SYSTEM.

(cf. ENTRY VECTOR.)

entry vector In the GATEWAY SITE-SPECIFIC RECOMBINATION SYSTEM: a vector containing a polylinker between a pair of *att*L sites; a sequence of interest can be inserted into the polylinker by using restriction enzymes and ligase. After cloning in the entry vector, the gene or sequence of interest can be transferred to a *destination vector* (see entry GATEWAY SITE-SPECIFIC RECOMBINATION SYSTEM (table) for examples of destination vectors). Transfer of the gene/sequence of interest from the entry vector to the destination vector involves site-specific recombination between the *att*L sites (in the entry vector) and *att*R sites (in the destination vector); such transfer is mediated by LR Clonase.

enucleated cell *Syn*. CYTOPLAST.

(See also SCNT.)

env See RETROVIRUSES.

EnzChek™ pyrophosphate assay A system for the assay of inorganic PYROPHOSPHATE (PP_i) (Invitrogen, Carlsbad CA, USA).

The assay involves two enzymes. Initially, pyrophosphatase cleaves the pyrophosphate to phosphate (P_i). The second enzyme, purine nucleoside phosphorylase, uses the phosphate to mediate a reaction in which 2-amino-6-mercapto-7-methylpurine ribonucleoside (MESG) is cleaved to ribose-1-phosphate and 2-amino-6-mercapto-7-methylpurine; in this

second reaction, the consumption of phosphate by MESG is accompanied by a proportional increase in absorbance at 360 nm, which is monitored spectrophotometrically.

The EnzChek approach is also used to monitor ATPase activity [Proc Natl Acad Sci USA (2009) 106(13):5111–5116; J Biol Chem (2009) 284(16):10296–10300].

(See also PIPER PYROPHOSPHATE ASSAY KIT.)

enzyme I (PTS) See PTS.

enzyme II (PTS) See PTS.

enzyme fragment complementation (EFC) A rapid technique for detecting specific recombinant proteins, either in purified form or in crude whole-cell lysates; unlike immunoblotting (a Western blot probed with labeled antibodies), this procedure does not require the use of antibodies.

The principle of EFC is related to α-complementation (see α-PEPTIDE). Essentially, a protein of interest is expressed as a fusion protein which includes the α-peptide as fusion partner. Such a fusion protein (present e.g. on a nitrocellulose filter) can complement the defective form of β-galactosidase, yielding a positive signal with e.g. a chemiluminescent substrate that is activated by the enzyme [BioTechniques (2006) 40(3): 381–383].

(See also COMPLEMENTATION.)

enzyme-linked immunosorbent assay See ELISA.

enzyme-linked immunospot See ELISPOT.

EPEC Enteropathogenic *Escherichia coli*: strains of food- and/or water-borne pathogenic *E. coli* that are responsible e.g. for diarrhea-related infant mortality in developing countries. The characteristic lesions on the intestinal epithelium result from products encoded by the LEE PATHOGENICITY ISLAND.

[Molecular evolution of EPEC (an analysis by multilocus sequence typing): J Bacteriol (2007) 189(2):342–350.]

eperezolid See OXAZOLIDINONE ANTIBIOTICS.

epidermal growth factor receptor family (the EGFR family or ERBB receptors) A family of membrane-spanning proteins characterized by (i) the (cell-surface) ligand-binding domain, (ii) a transmembrane section, and (iii) a cytoplasmic (intracellular) domain associated with TYROSINE KINASE activity.

The family includes:

EGFR (gene: c-*erbB1*, also called *ERBB1*)
HER2 (gene: c-*erbB2*, also called *ERBB2*)
HER3 (gene: c-*erbB3*, also called *ERBB3*)
HER4 (gene: c-*erbB4*, also called *ERBB4*)

(EGFR is sometimes referred to as HER1.)

Of these receptors, EGFR (170 kDa) and HER2 (185 kDa) have known roles in certain cancers. The ligand (activator) of EGFR is the epidermal growth factor (EGF). HER2 appears to have no known natural ligand. In cancer cells, these receptors may show GENE AMPLIFICATION and/or overexpression. Thus, e.g. overexpression of HER2 occurs in various tumors, including up to ∼30% of breast carcinomas; overexpression is associated with a poor clinical outcome. [The amplification of HER2/*neu* (tested with multiplex ligation-dependent probe amplification): BMC Cancer (2009) 9:4.]

Types of anticancer therapy include e.g. agents that inhibit the ligand-binding domain, or the tyrosine kinase activity, of EGFR; additionally, there are monoclonal antibodies (mAbs) against HER2. Herceptin® (= trastuzumab) is a (humanized) mAb that binds to HER2; it has been used therapeutically (in combination therapy) against HER2-overexpressing cancers.

(See also ONCOLYTIC VIRUS.)

epigenetic allelic ratio assay An assay designed for the non-invasive, pre-natal diagnosis of fetal chromosome *aneuploidy* (see ANEUPLOID); it depends on the presence of cell-free fetal DNA in maternal plasma.

One approach [Clin Chem (2006) 52:2194–2202], used for detecting trisomy 18 (EDWARDS' SYNDROME), monitors the target gene *SERPINB5* – encoding MASPIN – which occurs on chromosome 18 (at 18q21.33). In the *fetal* DNA the promoter of this gene is *un*methylated – whereas it is methylated in the maternal DNA; thus, after BISULFITE treatment, the fetal gene can be amplified by PCR with high specificity, even against a background of the maternal (methylated) gene in the plasma sample.

In this assay for trisomy 18, the target gene must be heterozygous for a given single-nucleotide polymorphism (SNP) – so that if a *third* copy of the chromosome is present (trisomy 18) it will either include or lack the given SNP. Hence, the primer extension products (derived from PCR) should consist of a mixture in which *one* type of primer extension product, *either* the one with *or* the one without the SNP, is numerically dominant; the existence of a third copy of the chromosome can therefore be deduced by determining the *ratio* of these two types of primer extension product.

(See also ALLELE-SPECIFIC GENE EXPRESSION.)

A different approach to non-invasive pre-natal diagnosis of fetal chromosome aneuploidy (which is independent of polymorphisms) employs a new-generation sequencing technique: see the entry DOWN'S SYNDROME.

epigenetic therapy Chemotherapy that uses e.g. inhibitors of certain DNA methyltransferases (for example, 5-AZA-2′-DEOXYCYTIDINE), and/or inhibitors of histone deacetylases (HDACs), in order e.g. to reverse the inactivation of particular gene(s).

epigenetics The study of certain (heritable) factors – other than nucleotide sequence – which influence gene expression. One factor is the pattern of methylation of bases in DNA. Methylation can influence gene expression e.g. by affecting transcription; for example, the methylation of specific bases may block access by the transcription initiation factors in a simple, mechanical way, and/or may promote recruitment of specific repressor protein(s). The inappropriate methylation of certain genes can give rise to specific diseases, such as cancer.

(See also GENETIC IMPRINTING.)

The Human Epigenome Project looked at methylation over the entire human genome. Quantitative analysis of data from this project revealed complex patterns of variation in methylation between individuals. Regions of CpG-rich DNA exhibit low but relatively similar levels of methylation in the overall

population, but random variation occurs between individuals in the order with which the methylated CpGs occur in these regions; the CpG-poor regions were reported to show higher levels of inter-individual variation in methylation – but they showed a significant level of conservation, between unrelated subjects, of specific methylation patterns [Nucleic Acids Res (2008) 36(10):e55].

(See also POLYCOMB-GROUP GENES.)

epigenotyping A term that has been used, generally, to refer to any method for assaying the methylation status of DNA – in particular the methylation of specific CPG ISLANDS in a given genome.

Some authors appear to extend the meaning to include e.g. any method for assaying the acetylation status of histones.

epimutation Events such as DNA hypermethylation that affect gene expression without a change in nucleotide sequence.

episome (1) Any PLASMID which can exist, and replicate, autonomously, i.e. extrachromosomally, in host cells – but which, in some cases, may be able to integrate into the chromosome of the host cell.

(2) *Syn*. PLASMID.

epitope tagging Addition of a specific peptide tag to a protein – e.g. by fusing the gene encoding the protein to a sequence encoding the peptide; the tag may be added to the C-terminal or the N-terminal of the protein.

Tagging may be carried out e.g. to facilitate the detection or recovery of an expressed protein, or as part of a procedure for labeling the protein (see e.g. PROTEIN LABELING). Detection/recovery of the given protein may be achieved e.g. by using monoclonal antibodies specific for the tag; one advantage of this approach is that the same tag can be re-used in different experiments with different proteins – thereby avoiding the need to generate a specific antibody for each of the proteins.

Various tags (both *ad hoc* and commercial) are used. One example is the FLAG® sequence (DYKDDDDK) – see the entry FLAG.

(See also e.g. PESC VECTORS and SIX-HISTIDINE TAG.)

Detection of certain tagged proteins from an *in vivo* source may be problematic owing to low-level expression of such proteins. One approach to improving the signal-to-noise ratio has been to insert a sequence encoding *tandem* copies of the epitope into the genome to form a fusion that specifies both the target protein and the tandem epitope; this was carried out by using LAMBDA (λ) RED RECOMBINATION (for insertion of the sequence), and the resulting tagged proteins (transcription factors of *Escherichia coli*) were detected by Western blot analysis and CHROMATIN IMMUNOPRECIPITATION (the latter method to check that tagging had not negated the proteins' DNA-binding function) [BioTechniques (2006) 40:67–72].

[Epitope tagging (mini-review): BioTechniques (2008) 44: 693–695.]

Epstein–Barr virus (EBV) A (ubiquitous) human herpesvirus associated e.g. with various types of malignancy – including nasopharyngeal carcinoma [comprehensive profiling of EBV miRNAs in nasopharyngeal carcinoma: J Virol (2009) 83(5): 2357–2367]; it is also found in some breast cancers, in which it is reported to confer resistance to the drug paclitaxel, and to promote overexpression of a gene encoding multidrug resistance [J Virol (2006) 80(2):845–853].

The EBV genome is linear dsDNA. Structurally, the virion resembles other herpesviruses. EBV infects various types of human cell, including B lymphocytes (B cells) and epithelial cells; the B cell membrane molecule CD21 is a receptor for EBV. In proliferating B cells, the EBV genome occurs intranuclearly as a circular extrachromosomal molecule with an origin of replication, *oriP*. A virus-encoded protein, EBNA-1, binds to *oriP* and may be necessary for persistence of the viral genome within the cell's nucleus; EBNA-1 is detectable within several hours of infection by EBV.

(See also MINICHROMOSOME (sense 1).)

EBV infection is commonly latent and is associated with a life-long carrier state. (See also BART, EBER, MICRORNAS.) [Control of latency: J Virol (2007) 81(12):6389–6401.] (See also BORTEZOMIB.)

Activation of viral replication in latently infected cells can be induced by certain types of chemical (see entry ZEBRA).

erb (*ERB*) An ONCOGENE initially detected in an avian erythroblastosis virus. v-*erb* comprises two contiguous genes: v-*erb*-A and v-*erb*B; the cellular homologs (c-*erb*A, c-*erb*B) are not contiguous e.g. in the chicken genome.

The c-*erb*A protein lacks tyrosine kinase activity.

The c-*erb*B protein has TYROSINE KINASE activity.

The cell-transforming (i.e. oncogenic) agent of the virus is apparently a fusion protein containing the v-*erb*B product.

ERBB receptors See EPIDERMAL GROWTH FACTOR RECEPTOR FAMILY.

ERIC sequence (enterobacterial repetitive intergenic consensus sequence) A highly conserved, repetitive element found in the genomes of e.g. enteric bacteria such as *Escherichia coli*; each contains a PALINDROMIC SEQUENCE. ERIC sequences have been exploited in TYPING (see REP-PCR).

[Use in typing *Escherichia coli* and *Klebsiella pneumoniae*: BMC Infect Dis (2008) 8:68.]

erlotinib An anticancer agent which inhibits oncogenic (EGF receptor) tyrosine kinase activity.

(cf. IMATINIB.)

error-prone DNA polymerases See e.g. Y FAMILY (of DNA polymerases).

error-prone repair (in bacteria) DNA repair associated with TRANSLESION SYNTHESIS.

(See also SOS SYSTEM.)

ERV sequence ENDOGENOUS RETROVIRAL SEQUENCE.

ES cells Embryonic stem cells: see STEM CELL.

ESAT-6 A 6-kDa protein secreted by the bacterial pathogen *Mycobacterium tuberculosis* and by the species *M. bovis* and *M. africanum*; ESAT is an acronym for: *e*arly *s*ecreted *a*ntigenic *t*arget. The operon that encodes ESAT-6 was reported to encode a low-molecular-weight protein, CFP-10.

The specificity of ESAT-6 for members of the so-called *M. tuberculosis* complex has been used for tracing contacts in

epidemiological studies of human tuberculosis. When used in a skin test for tuberculosis in bovines, ESAT-6 was found to be more specific than PPD (purified protein derivative) [Vet Rec (2000) 146:659–665].

[Use of ESAT-6 in studies on latent tuberculosis in health-care workers: PLoS Med (2007) 4(2):e55.]

(See also ELISPOT.)

Escherichia coli A much-studied, well-characterized species of Gram-negative, rod-shaped bacteria (of the family Enterobacteriaceae) that is found e.g. in the mammalian gut. *E. coli* is often used in DNA-based technology; it grows rapidly at ~37°C and can be cultured (grown) on/in inexpensive media such as NUTRIENT AGAR. (See also e.g. MACCONKEY'S AGAR and TERRIFIC BROTH.)

Certain strains of *E. coli* are well-characterized pathogens: see e.g. EHEC, EPEC and ETEC.

The genome of *E. coli* is ~5 × 10^6 base pairs. [Genome of *E. coli* (sequence): Science (1997) 277:1453–1474; genome of *E. coli* strain O157:H7: Nature (2001) 409:529–533, erratum: Nature 410:240.]

The GC% of the genomic DNA is ~50.5.

[Extensive annotation of the genome of *E. coli* strain K-12

ESCHERICHIA COLI: genetic markers associated with some of the frequently used and commercial strains of *Escherichia coli*[a]

Genetic marker	Notes
Amp^r	Resistant to the antibiotic ampicillin
Cam^r	Resistant to the antibiotic chloramphenicol
dcm	See entry DCM GENE
endA	*endA* strains (inactivated endonuclease I) have been used e.g. to improve the quality/yield of cloned plasmid DNA
$endA^+$	$endA^+$ strains have been used e.g. to reduce the level of contamination by double-stranded DNA when producing ssDNA
F′	Indicates the presence of an F PLASMID – which permits infection of an *Escherichia coli* host cell by phage M13 and related phages; these phages adsorb to specific plasmid-encoded pili (see PILUS) and do not infect cells which lack such pili. Infection by M13 and related phages is used e.g. for the production of ssDNA copies of a given sequence
galK	Inability to use galactose
gyrA	*gyrA* encodes a subunit of gyrase (a type II topoisomerase); mutation in *gyrA* can result in resistance to the antibiotic nalidixic acid
hsd	See entry HSD GENES
hsdR17(r_K^-, m_K^+)	Restriction-negative and modification-positive for this specific restriction–modification system
$lacI^q$	High-level production of the *lac* operon repressor protein (LacI) which regulates activity of the *lac* promoter. This permits tight control of expression of proteins when transcription is regulated via the *lac* operator system
*lacZ*Δ*M15*	Deletion of part of the *lacZ* gene (encoding the enzyme β-galactosidase); this deletion is requird e.g. in strains used for 'blue–white screening' (see entry PBLUESCRIPT)
leuB	Requires leucine for growth in minimal media
Δ*lon*	The Lon protease degrades abnormal proteins in the heat-shock response (and is also functional in the SOS system in *E. coli*). Engineered (i.e. 'abnormal') fusion proteins and heterologous proteins are less susceptible to degradation in Δ*lon* host cells, and Δ*lon* strains are used e.g. in studies on fusion products and the overproduction of heterologous proteins (see also *rpoH*, below)
mcrA	See entry MCRA GENE. Some commercially available strains contain an inactivated *mrcA* gene. Such strains are used e.g. when cloning methylated DNA from eukaryotic sources
mcrBC	See entry MCRBC. Comments are similar to those for *mcrA* (above)
metB	Requires methionine for growth in minimal media
mrr	See entry MRR GENE. Strains with an inactivated *mrr* gene are used e.g. when cloning DNA containing methyladenine residues

(*continued*)

ESCHERICHIA COLI (*continued*)

Genetic marker	Notes
ompT	See entry OMPT GENE. Cells lacking a functional OmpT protease are used e.g. for the display of heterologous (i.e. 'foreign') proteins at the cell surface; the heterologous gene, fused to one part of an autotransporter, is expressed as a cell-surface fusion protein in an *ompT*$^-$ strain
proAB	Requires proline for growth in minimal media
recA	*recA* mutants are used to inhibit homologous recombination (e.g. to prevent recombination between copies of a given plasmid present in the same cell). If required, a temporary *recA* facility can be provided on a SUICIDE VECTOR (q.v.)
relA	The *relA* gene encodes the ribosome-associated enzyme pyrophosphotransferase (*stringent factor*). When activated by an uncharged tRNA at the A site, this enzyme catalyzes the synthesis of ppGpp; this results in inhibition of protein and RNA synthesis (*stringent response*). A null mutation in *relA* produces a so-called *relaxed* phenotype in which e.g. RNA synthesis can continue under conditions that would otherwise trigger the stringent response
rne131	See entry BL21 STAR.
rpoH	The *rpoH* gene encodes a sigma factor (σ^{32}) involved in activation of the heat-shock regulon – which includes the synthesis of Lon protease (see Δ*lon*, above). Intracellular accumulation of abnormal/ heterologous proteins triggers the heat-shock response; *rpoH* mutants, deficient in the Lon protease, have been found to give increased yields during overproduction studies
rpsL	*rpsL* mutants used in recombinant studies are resistant to the antibiotic streptomycin (see RPSL GENE)
slyD	A *slyD* mutant is used to prepare a lysate for use in a commercial CELL-FREE PROTEIN SYNTHESIS system that employs a LUMIO tag in the fusion product; the *absence* of the SlyD protein in the lysate improves the signal-to-background ratio
supE/*supF*	Genes involved in suppression of amber mutations
Tet^r	Resistance to the antibiotic tetracycline
thi-1	Strains with this mutation need vitamin B_1 (thiamine) for growth in minimal medium
Tn*5*	A transposon that encodes e.g. resistance to kanamycin
Tn*10*	A transposon that encodes e.g. resistance to tetracyclines
tonA	Mutation in the *tonA* gene prevents infection by certain phages (e.g. phages T1, T5 and φ80) – for which the TonA protein acts as a cell-surface receptor
tonB	*tonB* mutants are resistant to various colicins
traD	In F$^+$ cells: *traD* mutations can inhibit conjugative transfer of the F plasmid

[a]An inactivated, mutant form of a gene may be written as e.g. *lon*, *lon*$^-$ or Δ*lon*.

in the EcoCyc database: Nucleic Acids Res (2007) 35(22): 7577–7590.]

Strains of *E. coli* frequently contain plasmids (see e.g. the F PLASMID).

(See also the *E. coli* strain JM109.)

A strain of the so-called REDUCED-GENOME ESCHERICHIA COLI has been useful e.g. for cloning lentiviral vectors (which are usually unstable in normal *E. coli* hosts).

Some of the genetic markers found in a number of strains of *E. coli* are listed in the table.

Commercial strains of *E. coli* are widely used: see e.g. BL21 STAR.

***E. coli* in recombinant protein production**

E.coli is often used for OVEREXPRESSION (q.v.).

The use of *E. coli* for producing heterologous proteins has a number of advantages, such as:

- Well-characterized genetics and metabolism.
- Easily cultured on inexpensive media.
- High-level expression of recombinant proteins is achievable with appropriately designed systems. In some cases, the recombinant protein can form >25% of the cell's total protein.
- The cells are readily transfected by electroporation and by other procedures.

However, despite the above advantages, this organism is also

associated with certain disadvantages which, in some cases, preclude its use for the synthesis of recombinant proteins. For example, there are the following factors:

• Lack of ability to carry out post-translational modification (such as glycosylation) which is necessary for the biological activity of many recombinant proteins. (Despite this, *E. coli* has been used for the production of some (normally glycosylated) proteins: see e.g. Neupogen® in BIOPHARMACEUTICAL (table).)

• The recombinant protein may accumulate *intra*cellularly; this complicates the downstream processing as it is necessary then to isolate the recombinant protein from the producing cells and to separate the required protein from the other (cell-associated) proteins.

• Part of the normal structure of the *E. coli* cell (and of other Gram-negative bacteria) is the so-called *outer membrane* – a permeability barrier that includes *lipopolysaccharide* (LPS). LPS, also called endotoxin, is a complex macromolecule that is harmful to humans (e.g. it behaves as a pyrogen) and its removal from any product is essential. Removal of LPS from recombinant products (to make them safe for therapeutic use) may require e.g. chromatographic fractionation. The presence of LPS in a product can be detected e.g. by the LAL TEST.

ESE Exon(ic) splicing enhancer: see SPLICING.

ESI Electrospray ionization: a technique used in mass spectrometry e.g. for studying reaction kinetics by continuous monitoring.

Unlike the procedure in MALDI (q.v.), analytes are initially present in the *liquid* phase; the analyte-containing liquid is sprayed from an electrically charged capillary – forming fine droplets which quickly evaporate. As in MALDI, there is a detector system such as the time-of-flight (TOF) analyzer.

EST EXPRESSED SEQUENCE TAG.

***estA* gene** The gene encoding a heat-stable toxin (Sta) produced by strains of ETEC (q.v.).

ET recombination A type of facilitated homologous recombination, in cells of *Escherichia coli* or other bacteria, mediated by the RecET recombinase of PHAGE RAC (genes *recE* and *recT*); this system permits *linear* fragments of DNA (whose terminal sequences are homologous to the target sequences) to be inserted into cells (e.g. by ELECTROPORATION or transformation) and then undergo recombination at targeted sites in the chromosome (or plasmid etc.). The terminal homologous sequences on a fragment can be as short as (e.g.) ~50 bp; the fragments may be prepared by PCR using primers which carry the terminal (homologous) sequences as 5′ tags.

The RecE product is a 5′-to-3′ exonuclease which produces 3′ single-stranded overhangs in fragments of linear dsDNA.

The RecT product is a *single-strand-annealing protein* that pairs RecE-produced overhangs to ssDNA target sequences.

Like LAMBDA (λ) RED RECOMBINATION, ET recombination is used in RECOMBINEERING (q.v.).

etanercept See BIOPHARMACEUTICAL (Enbrel® in the table).

EtBr ETHIDIUM BROMIDE.

ETEC Enterotoxigenic strains of ESCHERICHIA COLI: strains of food-borne or water-borne *E. coli* which are common causes of diarrhea in developing countries and in travelers. The cells bind to gut epithelium via plasmid-encoded fimbriae, forming plasmid-encoded heat-stable toxins (Sta and Stb, also referred to as ST-I and ST-II) and heat-labile toxins (LT-I and LT-II); the gene for Sta (*estA*) occurs on a plasmid-borne transposon (Tn*1681*).

ethidium bromide (EtBr) 5-ethyl-3,8-diamino-6-phenylphenanthridinium bromide: an INTERCALATING AGENT used e.g. as a fluorophore for staining nucleic acids in electrophoresis strips; when bound to dsDNA, fluorescence occurs at ~605 nm (red) on appropriate excitation. Ethidium bromide is generally excluded from living cells, being unable to penetrate the cell membrane.

(See also DNA STAINING and PROPIDIUM IODIDE.)

Ethidium bromide is also used for separating ccc dsDNA from linear or nicked circular molecules of dsDNA by isopyknic ultracentrifugation. This depends on the ability of the linear or nicked circular molecules to bind more molecules of ethidium bromide when compared to ccc dsDNA molecules; because binding ethidium bromide makes a molecule more buoyant, molecules that bind more ethidium bromide separate in an upper (i.e. less dense) band during ultracentrifugation.

ethidium monoazide A fluorescent phenanthridinium intercalating dye, reported to be excluded by the cell membrane in living cells.

In dead cells, the dye can be covalently photo-cross-linked to DNA.

1-ethyl-3-(dimethylaminopropyl) carbodiimide See EDC.

ethyleneglycol-bis(β-aminoethylether)-*N*,*N*,*N*′,*N*′-tetraacetic acid See EGTA.

ethyleneglycol-bis(succinimidyl succinate) See CHROMATIN IMMUNOPRECIPITATION.

etravirine A NON-NUCLEOSIDE REVERSE TRANSCRIPTASE INHIBITOR (q.v.).

euchromatin See CHROMATIN.

eukaryote Any animal, plant or microorganism in which the cellular structure includes a membrane-bounded nucleus and in which mitosis and/or meiosis precedes cell division.

(cf. PROKARYOTE.)

euploid (1) Refers to a cell or organism in which the number of chromosomes is the same as that characteristic of the species, i.e. a cell/organism containing the *normal* number of chromosomes.

(cf. HETEROPLOID.)

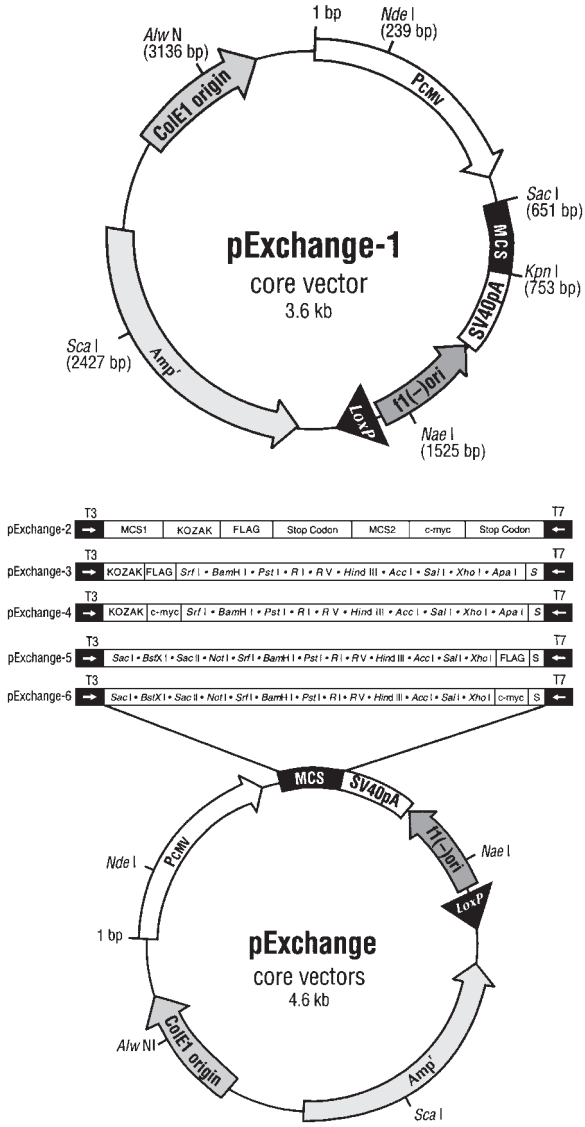
(2) Having a complement of chromosomes in which all the chromosomes are present in the same number of copies, e.g. one copy of each chromosome or two copies of each chromosome etc.

evanescent wave (in TIRF microscopy) See TIRF MICROSCOPY.

***ex vivo* mode** (in gene therapy) See GENE THERAPY.

exBiFC *Extended* BIMOLECULAR FLUORESCENCE COMPLEMENTATION (q.v.).

Exchanger™ system A range of expression vectors (Strata-



Exchanger™ SYSTEM VECTORS These vectors permit the *in vitro* insertion of a given antibiotic-resistance module by site-specific recombination at the *loxP* site. The vector below shows the expanded MCS (polylinker) in each of a range of vectors. See the entry for details of the system. (See also separate entries for FLAG and KOZAK SEQUENCE.)

Courtesy of Stratagene, La Jolla CA, USA.

gene, La Jolla CA, USA) that can be used e.g. for cloning in *Escherichia coli* (by using the ColE1 origin of replication and selection on an ampicillin-containing medium), and protein expression in eukaryotic (including mammalian) cells (using the CMV promoter sequence and the SV40 polyadenylation signal). In this system, the required selectable marker can be introduced (*in vitro*) by site-specific recombination in which a linearized drug-resistance module is added by Cre-mediated recombination between *loxP* sites (CRE–LOXP SYSTEM); each module encodes resistance to hygromycin, neomycin or puromycin. In this way, the gene of interest can be located in each of several different vectors, with different selectable markers, without the need to move the gene itself.

In each vector the multiple cloning site (MCS, polylinker) is bracketed by T3 and T7 promoters, in opposite orientation. The vectors contain a FLAG or c-*myc* sequence that provides an EPITOPE TAGGING facility.

A phage f1 origin of replication in the vectors permits the recovery of ssDNA, if required.

excimer See e.g. BINARY PROBE.

excipient Any component of a BIOPHARMACEUTICAL having the primary function of *stabilizing* the product, i.e. helping to prevent any change in the chemical/molecular structure of the active constituent(s).

Excipients, which are often used in (protein-based) biopharmaceuticals, include e.g. serum albumin, certain sugars (glucose, maltose etc.), amino acids (e.g. glycine, threonine), glycerol, sorbitol and polyethylene glycol (PEG).

excision repair Any of various DNA REPAIR processes in which the damaged/abnormal region of a strand of DNA is removed and replaced by the activity of a DNA polymerase; see e.g. BASE EXCISION REPAIR, MISMATCH REPAIR and UVRABC-MEDIATED REPAIR.

excitation filter (fluorescence microscopy) A bandpass filter (i.e. one which transmits radiation only within a given band of wavelengths) which has two main functions: (i) to transmit radiation corresponding to the excitation wavelength(s) of the *specific* fluorophore(s) in the sample, thus minimizing non-specific fluorescence, and (ii) to block radiation of any other wavelengths.

(cf. EMISSION FILTER; see also DICHROIC BEAMSPLITTER.)

Exo$^-$ Klenow Fragment A recombinant form of Klenow fragment (see DNA POLYMERASE I), marketed by Ambion (Austin TX, USA), which lacks the 3′-to′-5′ exonuclease activity, i.e. it has only DNA polymerase activity.

exogenote See MEROZYGOTE.

exon See the entries SPLIT GENE and SPLICING.

The entry SPLICE describes an *in vitro* (PCR-based) method for the production of complete coding sequences from intron-containing genomic DNA.

(See also ALU SEQUENCE and EXONIZATION, EXON ARRAY, EXON SKIPPING, EXON TRAPPING.)

exon amplification PCR-based amplification using, as template, cDNA copies of the (spliced) RNA produced during EXON TRAPPING. This approach may be used e.g. to recover exon(s) derived from the sample DNA.

exon array A MICROARRAY used e.g. for the analysis of gene expression in terms of individual exons within a given transcript, including investigation of the quantitative changes that may occur in ALTERNATIVE SPLICING – information that can be useful for studying splicing patterns in certain diseases.

Exon arrays include GeneChip® Exon 1.0ST (Affymetrix), which uses several probes to interrogate each exon. Target molecules (DNA) are generated from mRNAs by the use of random hexamer primers which carry a T7 promoter on a 5′ tag (instead of the more usual 5′ oligo-d(T) primers directed at poly(A) tails of mRNAs); hence, this exon array can detect *non*-polyadenylated (as well as polyadenylated) transcripts – a feature that was reported to be a causal factor of differences in performance between this exon array and the THREE-PRIME ARRAYS [BMC Genomics (2008) 9:529].

Other types of exon array include those in which the probes span exon–exon junctions; these arrays are able to interrogate particular splicing events.

[Dissecting an alternative splicing analysis workflow for GeneChip® Exon 1.0ST: BMC Genomics (2008) 9:571.]

[Use of an exon array for detection of copy number variants in mice: PLoS ONE (2009) 4(3):e4649.]

An exon array was also used for analysis of antisense transcripts (see TRANSCRIPTOME).

exon-prediction program Computer software that is designed to detect exons within sequences of nucleotides from cloned DNA.

exon skipping (1) (*in vivo*) The exclusion of one or more exons during the expression of a given gene (see also ALTERNATIVE SPLICING).

(2) (*in vitro*) The *engineered* exclusion of exon(s) during the expression of a given gene – see e.g. DUCHENNE MUSCULAR DYSTROPHY.

exon(ic) splicing enhancer (ESE) See SPLICING.

exon trapping A method used e.g. for demonstrating/detecting the presence of exon(s) – with functional splice sites – within a given fragment of sample DNA. Essentially, the fragment is inserted into a construct (a *minigene*) which already contains known exon and intron sequences; when the minigene (which is part of a plasmid) is transcribed *in vivo*, exons in the minigene may be spliced to any exon(s) that may be present in the fragment. If cDNA copies of the spliced transcript are then amplified by PCR, the length and composition of the amplicons can indicate whether the transcript contains only minigene exons or includes exon(s) from the sample fragment. If required, specific amplified exon(s) can then be recovered by PCR.

In one scheme for exon trapping, the minigene forms part of a 6-kb plasmid shuttle vector which includes an origin of replication and an ampicillin-resistance gene for replication (and selection) in *Escherichia coli*. For replication and RNA splicing in suitable mammalian cells, one terminal region of the minigene consists of an origin of replication and promoter region from simian virus 40 (SV40), while the other terminal

region consists of the SV40 polyadenylation signal (encoding the 3′ end of the transcript). The sequence between these two regions consists of two exons that are separated by an intron; the intron includes a multiple cloning site (i.e. POLYLINKER) into which the sample DNA fragment is inserted. Following transcription of the minigene in COS cells (recombinant cells prepared from a monkey cell line), several types of splicing pattern may be detected by amplifying cDNA copies of the transcripts. In the absence of (functional) exon(s) in the DNA sample fragment, the two exons in the minigene are spliced together (forming a sequence of known length). However, a functional exon present in the fragment may be spliced to the minigene's exons; this can be detected from the composition of the transcripts.

This general approach – using minigenes in an expression system – has also been valuable for investigating the effect on splicing of mutations or SNPs and for studying the overall regulation of pre-mRNA splicing – including ALTERNATIVE SPLICING (which includes a number of tissue-specific forms of splicing). A (proposed) minigene plasmid vector may be useful e.g. for (i) assessing the effect(s) of disease-associated mutations on the splicing of relevant gene(s); (ii) identifying *cis*-acting elements that are involved in normal regulation of splicing; and (iii) examining the effects of specific regulatory factors/regions on alternative splicing [BioTechniques (2006) 41(2):177–181].

exonic splicing enhancer See SPLICING.

exonic splicing silencer See SPLICING.

exonization An *in vivo* phenomenon in which a new exon is created by e.g. (i) the introduction of a sequence of DNA into a gene and/or (ii) development of mutations in a pre-existing non-exonic sequence.

(See also ALU SEQUENCE.)

exonuclease Any enzyme which cleaves terminal nucleotides from a nucleic acid molecule.

The lambda exonuclease acts preferentially 5′-to-3′ on a 5′-phosphorylated strand of blunt-ended dsDNA. This enzyme is therefore useful e.g. for preparing single-stranded products (e.g. for SSCP ANALYSIS) from PCR amplicons. To prepare the ssDNA products, PCR is conducted with a reaction mixture in which only *one* of the primers has a 5′-terminal phosphate – so that only *one* strand of a (double-stranded) amplicon will be subsequently digested by the lambda exonuclease.

The *Pfu* DNA polymerase has 3′ exonuclease activity and is used for POLISHING (q.v.).

An exonuclease is also used e.g. in the IN-FUSION technique (for joining molecules of DNA) – and for preparing NESTED DELETIONS.

Some exonucleases can also act as a STRUCTURE-SPECIFIC ENDONUCLEASE.

Some examples of exonucleases are given in the following entries.

exonuclease III An enzyme (product of the *Escherichia coli* gene *xthA*) which degrades blunt-ended dsDNA (or dsDNA with 5′ overhangs) from the 3′ end of each strand, leaving the 5′ ends intact and producing 5′-phosphomononucleotides; the enzyme – which does not degrade 3′ ssDNA overhangs – has been used, for example, for preparing NESTED DELETIONS in cloned DNA.

The enzyme also has activity similar to that of RNASE H and it can function as an AP endonuclease in the BASE EXCISION REPAIR process.

(See also EMEA and LONG PCR.)

exonuclease VII An enzyme (product of the *Escherichia coli* gene *xseA*) which degrades the ssDNA of 3′ and 5′ overhangs but does not degrade dsDNA.

exonuclease-mediated ELISA-like assay See EMEA.

5′-exonuclease PCR *Syn*. 5′-NUCLEASE PCR.

exoribonuclease See RIBONUCLEASE.

exosome (1) In the cells of eukaryotes: a multiprotein complex which is associated with 3′-to-5′ exoribonuclease activity and which has functions that include degradation of fragmentary and aberrant forms of RNA, and the processing of rRNA and snRNA; part(s) of the exosome are apparently functional in both the nucleus and the cytoplasm. Cells which are defective in nuclear exosomal activity may accumulate RNA molecules produced as a result of aberrant splicing or incomplete 3′-end processing.

An analogous complex has been reported in at least some archaeans.

(2) The term has also been used to refer to a DNA fragment that may replicate, and be expressed, when taken up by a cell but which may not be integrated in the host's DNA.

expanded gold standard See GOLD STANDARD.

exponential silencing Downregulation of promoters of specific bacterial gene(s) associated with rapid growth in rich media.

exportins See KARYOPHERINS.

expressed sequence tag (EST) A recorded sequence of nucleotides (often ∼200–400 nt), derived e.g. from a cDNA library, which is kept in a database as a record of the expression of an unknown gene from a particular type of cell under given conditions.

A given EST, with a sequence of interest, may be located in a database by computer sequence-matching programs.

A pair of primers, specific for a given EST, can be designed and used in a PCR assay of a large, cloned random fragment of genomic DNA; if the assay yields products that correspond to the given EST (i.e. if the fragment contains a homologous sequence), then the relevant sequence in the random fragment of DNA is referred to as a *sequence-tagged site* (STS).

Given an assortment of large, cloned random fragments of DNA which, collectively, cover a major part of an unknown genome, each of the fragments can be assayed for STSs that correspond to a range of different ESTs. If various STSs are located in each random fragment, this could permit alignment of fragments with overlapping sequences; thus, for example, if fragment A has STSs X and Y, and fragment B has STSs Y and Z, then fragments A and B may overlap at Y – 'may'

because the Y STS may be a sequence that occurs more than once (or even many times) in the genome.

Analysis of 10000 ESTs, derived from a cDNA library, has been used to increase understanding of the immune system in the cynomolgus monkey (*Macaca fascicularis*). Bearing in mind the similarities between the human genome and that of *M. fascicularis*, it was suggested that this might be useful for revising annotations of relevant genes in the human genome [BMC Genomics (2006) 7:82].

[Assembly of approximately 185000 ESTs from the cotton plant (*Gossypium*): Genome Res (2006) 16(3):441–450.]

expression array A MICROARRAY designed for investigating one or more aspects of gene expression – see, for example, EXON ARRAY and THREE-PRIME ARRAY.

expression bacmid See BAC-TO-BAC.

expression platform See e.g. RIBOSWITCH.

expression site (in trypanosomes) See VSG.

expression vector A VECTOR which encodes functions for the transcription and/or translation of a gene (or other sequence) carried by the vector, or subsequently inserted into the vector.

The expression vectors used in bacteria are generally extrachromosomal elements – i.e. they replicate independently of the chromosome; they may include: (i) an appropriate origin of replication; (ii) a marker gene (e.g. an antibiotic-resistance gene); (iii) one or more promoters (allowing e.g. transcription of the insert in either direction); (iv) a segment corresponding to the Shine–Dalgarno sequence in the transcript (for translation in bacteria); (v) a POLYLINKER (allowing flexibility when inserting target DNA); (vi) a transcription terminator (to avoid unwanted readthrough); (vii) an appropriate control system for initiating transcription of the insert (see also entry CONDITIONAL GENE ACTIVATION/REGULATION).

Expression vectors used in yeasts include the YCp (YEAST CENTROMERE PLASMID) vectors, and those (such as the YIp (yeast integrating plasmids) which integrate in the host cell's DNA. (See also YEAST ARTIFICIAL CHROMOSOME.)

Expression vectors used in mammalian cells include those used for generating replication-deficient viral vectors – e.g. in certain commercial gene-delivery systems (see AAV HELPER-FREE SYSTEM, VIRAPORT RETROVIRAL GENE EXPRESSION SYSTEM and VIRAPOWER LENTIVIRAL EXPRESSION SYSTEM). They also include the vectors used for producing certain biopharmaceutical products (e.g. Aranesp® and Benefix®) in mammalian cell lines.

Viral expression vectors such as the adeno-associated virus vectors (see AAVS), and the lentiviral vectors, may integrate into the host's DNA.

(See also BACULOVIRUS EXPRESSION SYSTEMS.)

Target DNA may be initially cloned (by replication) in the expression vector. Transcription may be initiated, as needed, e.g. via a *lac* operator (see LAC OPERON) located upstream of the coding sequence and controlled by LacI repressor protein; in this case, repression can be lifted (transcription initiated) by adding IPTG.

Cloning and expression can be carried out as separate processes, in different organisms, in a SHUTTLE VECTOR.

Expression of a gene carried by an expression vector may be regulated by a chromosomally located control system (see e.g. CASCADE EXPRESSION SYSTEM) – or by a control system located on the vector itself.

Some expression vectors include a sequence that encodes a fusion partner for the gene of interest (see GENE FUSION and GENE FUSION VECTOR) in order e.g. to increase the solubility of the target protein – the fusion product being more soluble than the target protein itself: see e.g. CHAMPION PET SUMO VECTOR.

The use of an EPITOPE TAGGING system may facilitate the detection/isolation of expressed proteins.

(See also PESC VECTORS.)

The EXCHANGER SYSTEM of expression vectors is useful e.g. for epitope tagging of proteins and also for changing the selectable marker in a vector (by site-specific recombination) without the need to re-locate the insert/gene of interest.

Engineering gene expression in trypanosomes – a class of eukaryotic microorganisms – is associated with an additional problem. Trypanosomes generally do not regulate expression of their protein-encoding genes by controlling the initiation of transcription. (Gene expression may be regulated e.g. by (i) controlling the *stability* of mRNAs, and/or (ii) regulating the rate of translation of mRNAs into protein.) Consequently, for studies on gene expression with an inducible transcription system in these organisms, there is a need to construct a system from exogenous sources. For the species *Trypanosoma cruzi* (causal agent of Chagas' disease in Central and South America), a stable (tetracycline-inducible) expression vector (pTcINDEX), containing a multiple cloning site (MCS), was designed to integrate into the genome at a specific transcriptionally silent location, thus avoiding some of the problems of integration into a routinely transcribed site [BMC Biotechnol (2006) 6:32].

(See also the table of destination vectors in entry GATEWAY SITE-SPECIFIC RECOMBINATION SYSTEM.)

Expressway™ Plus expression system A system (Invitrogen, Carlsbad CA, USA) for *in vitro* protein synthesis (see CELL-FREE PROTEIN SYNTHESIS).

ExSite™ site-directed mutagenesis kit A kit (Stratagene, La Jolla CA) used for PCR-based SITE-DIRECTED MUTAGENESIS in which deletions, insertions or point mutations can be introduced into almost any circular dsDNA plasmid.

Essentially, the type of mutation produced depends on the design of the primers. A point mutation can be introduced by using primers with a single-base substitution. A deletion can be produced by primers which bind at sites separated by the sequence to be deleted. An insertion can be introduced by the use of primers with a 5′ extension.

Following PCR, the reaction mixture includes the following molecules: original template DNA (which had been methylated *in vivo*); linear, newly synthesized (unmethylated) DNA; and linear, hybrid DNA which consists of parental and newly synthesized strands. The reaction mixture is then subjected to

restriction enzyme DpnI (see table in the entry RESTRICTION ENDONUCLEASE for details) which eliminates both parental and hybrid DNA. Blunt-ended ligation of newly synthesized DNA is then carried out with T4 DNA ligase – this providing circular, mutagenized molecules that are suitable for insertion into *Escherichia coli* by transformation.

ExSite™ has been used e.g. to replace a GTC codon with a GAG codon [J Bacteriol (2006) 188(7):2604–2613]; to introduce mutations into the *hmuR* gene (which encodes a hemin receptor) in a periodontitis-associated Gram-negative pathogen, *Porphyromonas gingivalis* [Infect Immun (2006) 74(2): 1222–1232; and to study the regulation of translation termination [RNA (2008) 14(2):240–245]. [Other uses of ExSite™: Mol Biol Cell (2009) 20(8):2196–2206; PLoS Pathog (2009) 5(3):e1000332.]

extein See INTEIN.

extendase activity (template-independent polymerase activity) During PCR: the addition of an extra residue of an adenosine nucleotide to the 3′ end of the product strand (i.e. beyond the final, template-determined 3′ nucleotide) – when synthesis is mediated e.g. by TAQ DNA POLYMERASE. Single-nucleotide (adenosine) 3′ overhangs can be exploited e.g. in facilitating insertion of the amplicons into certain types of vector which have terminal single-nucleotide thymidine overhangs – see entries TOPOISOMERASE I CLONING and CHAMPION PET SUMO VECTOR. (See also TOPO TA CLONING KIT.)

The expression of extendase activity is strongest when the last (template-determined) residue at the 3′ end of the product strand is a pyrimidine; if required, extendase activity can be promoted by including a purine (rather than a pyrimidine) in the relevant position in the template. If the last residue at the 3′ end of the product strand is a purine, extendase activity is less efficient.

If blunt-ended products are required from PCR, amplicons with the 3′ overhang can be subjected to exonuclease action: see POLISHING.

(See also END-IT DNA END-REPAIR KIT.)

extended BiFC See BIMOLECULAR FLUORESCENCE COMPLEMENTATION.

external guide sequence See SNRNAS.

extrachromosomal genes *Syn*. CYTOPLASMIC GENES.

extrachromosomal inheritance *Syn*. CYTOPLASMIC INHERITANCE.

extraction of DNA See DNA ISOLATION.

extranuclear genes *Syn*. CYTOPLASMIC GENES.

extrinsic fluorescence Fluorescence due to the excitation of a small molecule that has been bound, covalently or otherwise, to a larger one in order to provide a fluorescent derivative.

EYFP *Enhanced* YELLOW FLUORESCENT PROTEIN (q.v.).

F

F L-Phenylalanine (alternative to Phe).

F^+ donor, F^- recipient See F PLASMID.

F factor See F PLASMID.

F-like pili See PILUS.

F plasmid (formerly F factor) A specific, IncFI, conjugative, low-copy-number plasmid (~95 kb) found e.g. in *Escherichia coli*; the F plasmid can exist either as an extrachromosomal replicon or integrated in the host chromosome – and is able to mediate CONJUGATION in either location.

Cells containing the free, circular plasmid are conjugative *donor* cells (F^+ cells); potential *recipient* cells (F^- cells) lack the plasmid. F^+ cells express certain (plasmid-encoded) donor functions (e.g. a PILUS) and are *derepressed* for conjugation – meaning that all, or nearly all, cells in an F^+ population can function as donors (see FINOP SYSTEM); hence, in a mixed population of F^+ and F^- cells, most or all F^- cells will receive a copy of the F plasmid (through conjugation) and will therefore be converted to F^+ cells. (During conjugation, F^+ donor cells retain the plasmid – a *copy* of the plasmid is transmitted to the recipient.)

Chromosomal integration of the F plasmid gives rise to an *Hfr donor* (Hfr means high frequency of recombination); Hfr donors transfer chromosomal (as well as plasmid) DNA to F^- recipients, the amount of DNA transferred increasing with increased length of uninterrupted mating. (The F plasmid in an Hfr cell is said to *mobilize* the chromosome for transfer.) (See also INTERRUPTED MATING.)

An F plasmid may excise from the chromosome aberrantly to form an F′ plasmid (see PRIME PLASMID).

Certain *par* loci are involved in plasmid *partition* (that is, the appropriate segregation of plasmids to daughter cells on cell division); the plasmid-encoded SopA protein is important in this context (see PAR LOCI). The stable maintenance of the F plasmid in a bacterial population is promoted by the CCD MECHANISM.

Transcription from the P_Y promoter of the F plasmid's *tra* (transfer) operon was reported to be repressed by the host's H-NS protein [J Bacteriol (2006) 188(2):507–514].

(See also FEMALE-SPECIFIC PHAGE and PIF.)

F′ plasmid See F PLASMID and PRIME PLASMID.

f1 phage See PHAGE F1.

Fab fragment (of an antibody) See ZENON ANTIBODY-LABELING REAGENTS.

Fabry disease An X-linked disease involving deficiency in the activity of the enzyme α-galactosidase A. Various manifestations are reported, including e.g. vascular-type skin lesions and cardiac symptoms; the cause of death is usually reported to be renal failure.

A recombinant enzyme, Fabrazyme®, is used for treatment (see entry BIOPHARMACEUTICAL (table)).

facile (of procedures) Able to be carried out easily, i.e. without specific or unspecified difficulties.

facilitated homologous recombination See the entry RECOMBINEERING.

FACS FLUORESCENCE-ACTIVATED CELL SORTER.

factor VIIa A factor involved in the blood-clotting mechanism: see NovoSeven® in the entry BIOPHARMACEUTICAL (table).

factor VIII A factor involved in the blood-clotting mechanism: see 'hemophilia' in the table in entry GENETIC DISEASE, and see also KOGENATE.

factor IX A factor involved in the blood-clotting mechanism: see CHRISTMAS FACTOR, and see also hemophilia in the entry GENETIC DISEASE (table).

factor essential for methicillin resistance (*fem*) See MRSA.

FAM 6-Carboxyfluorescein: a fluorescent reporter dye with an emission maximum at 525 nm and an excitation maximum at 493 nm.

farnesol A sesquiterpene alcohol reported to act as a QUORUM SENSING molecule in *Candida albicans*.

[Autoregulation and quorum sensing in fungi: Eukaryotic Cell (2006) 5(4):613–619].

Fas (CD95, APO-1) The cell-surface receptor for the Fas ligand (see APOPTOSIS).

Fas ligand (Fas L, FasL) See APOPTOSIS.

FBS Fetal bovine serum.

Fc portion (of an antibody) See ZENON ANTIBODY-LABELING REAGENTS.

fd phage See PHAGE FD.

feature A given type of probe at a specific location in a MICROARRAY. In a GeneChip® (Affymetrix) microarray a feature is ~5 μm in size.

[Use of term (e.g.): BioTechniques (2007) 43(6):791–797.]

fecal DNA test (colorectal cancer) See CANCER DETECTION.

feeder cells Cells which may be included in cultures of STEM CELLS in order to promote growth of the latter.

For mouse ES cells in culture, the engineered expression of Bcl-2 was reported to override the need for feeder cells, but leukemia inhibitory factor was still a requirement for the ES cells [Proc Natl Acad Sci USA (2005) 102(9):3312–3317].

(See also *Nanog* in the entry STEM CELL.)

feline immunodeficiency virus See LENTIVIRINAE.

feline syncytial virus See SPUMAVIRINAE.

***fem* genes** See MRSA.

female-specific phage Any of certain phages – e.g. T7, W31 – whose replication is inhibited in those 'male' cells that contain particular plasmids; for example, production of intermediate and late phage proteins of T7 is inhibited in cells of *Escherichia coli* containing the F PLASMID.

Note that cells containing conjugative plasmids are said to be 'male' cells; cells lacking these plasmids are 'female' cells.

In *Pseudomonas aeruginosa*, the DNA of phages F116 and G101 is degraded by RESTRICTION ENDONUCLEASE activity in cells containing plasmid pMG7.

(See also PIF.)

femto- A prefix meaning 10^{-15}. For a list of prefixes, see the 'Ready reference' section at the front of the dictionary.

FEN1 (FEN-1) See FLAP ENDONUCLEASE 1.

Fenton reaction The reaction in which ferrous iron is oxidized to ferric (III) iron by hydrogen peroxide with the production of hydroxyl radical (·OH):

$$Fe(II) + H_2O_2 \rightarrow Fe(III) + \cdot OH + OH^-$$

A Fenton reaction at the endoplasmic reticulum was reported to influence the expression of hypoxia-inducible genes [Proc Natl Acad Sci USA (2004) 101(12):4302–4307].

The highly reactive hydroxyl radical, produced in a Fenton reaction, can be employed to obtain fine resolution in FOOTPRINTING studies [Nucleic Acids Res (2006) 34(6):e48].

In vivo footprinting (in frozen cells) has been achieved with a brief exposure to a synchrotron X-ray beam that produces hydroxyl radical footprints similar to those obtained *in vitro* using the Fenton reaction [Nucleic Acids Res (2006) 34(8): e64].

fermentation (1) In an industrial/commercial context: any process involving the large-scale culture (i.e. growth) of microorganisms, regardless of whether metabolism is fermentative or respiratory (oxidative). This use of the term 'fermentation' is in sharp distinction to that given in meaning (2), below.

(2) (*energy metab.*) A specific energy-converting process in which a substrate is metabolized without the involvement of an external electron acceptor, oxidation and reduction within the process being balanced. ATP is characteristically produced by substrate-level phosphorylation.

(cf. RESPIRATION.)

fes (*FES*) An ONCOGENE in strains of feline sarcoma virus. The v-*fes* product has TYROSINE KINASE activity.

fetal stem cell See STEM CELL.

FFPE (of tissue samples) Formalin-fixed, paraffin-embedded.

FHA The filamentous hemagglutinin produced by the (Gram-negative) bacterium *Bordetella pertussis*.

(See also INTEGRINS.)

FhuA protein See TONA GENE.

Fidelity Index A proposed measure of STAR ACTIVITY (q.v.); it was defined as the ratio of the maximum amount of enzyme that shows no star activity to the minimum amount of enzyme needed for complete digestion at the cognate recognition site [Nucleic Acids Res (2008) 36(9):e50].

field inversion electrophoresis See PFGE.

filament pyrolyser See PYROLYSIS.

filamentous hemagglutinin (FHA) In *Bordetella pertussis*: a cell-surface ligand which binds the $\alpha_M\beta_2$ INTEGRIN.

filamentous phages A category of non-lytic BACTERIOPHAGES whose genome is ssDNA; these phages are characterized by their morphology – each phage being within the approximate range 800–2000 × 6 nm (according to phage).

The filamentous phages include e.g. f1, fd and PHAGE M13, the latter being used in DNA technology e.g. for preparation of single-stranded target DNA.

On a LAWN PLATE of sensitive bacteria, filamentous phages typically form a turbid PLAQUE (rather than a clear plaque) because they merely reduce the growth rate of infected cells (rather than lyse them).

filgrastim See Neupogen® in BIOPHARMACEUTICAL (table).

fimbria (*bacteriol.*) A term that is often regarded, illogically, as a synonym of PILUS (q.v.). Fimbriae are thin, proteinaceous filaments that project from the surface of certain types of cell and which have roles e.g. in cell–cell adhesion, cell–substrate adhesion, and (in some cases) in a type of motility referred to as 'twitching motility'.

Fimbriae are found on many Gram-negative bacteria and on some types of Gram-positive bacteria (as well as e.g. on some fungi, including *Saccharomyces cerevisiae*).

The fimbriae of at least some Gram-positive bacteria differ from those of the Gram-negative species e.g. in their mode of assembly and in the location of the genes which encode them.

The enzyme SORTASE plays a role in the development of at least some of the fimbriae in Gram-positive species.

The fimbriae in the Gram-positive bacterium *Streptococcus agalactiae* were reported to have a role in biofilm formation and also in adherence to epithelial cells [PLoS Pathog (2009) 5(5):e1000422].

FimS The sensor kinase in a two-component regulatory system in *Pseudomonas aeruginosa*: see the entry ALGR.

fingerprint (DNA) A pattern formed by a number of bands of stained (or labeled) fragments of DNA, of different sizes, that are present in a gel medium (following gel electrophoresis) – or on a nitrocellulose (or other) membrane following transfer in a BLOTTING procedure. The nature of any given fingerprint will depend on the particular sample of DNA which is being examined and on the method used to produce the fingerprint.

The term *fingerprint* is also used to refer to other forms of sequence-specific 'signature' used e.g. for the characterization of genomic DNA or particular fragments of nucleic acid.

Fingerprints are generated, for example, by various TYPING procedures (which may produce strain-specific fingerprints).

FinOP system A regulatory system, involving plasmid genes *finO* and *finP*, which controls conjugative transfer in many F-like plasmids. The products of *finO* and *finP* (jointly) inhibit expression of *traJ*, a gene in the 'transfer operon' which has a key role in the initiation of conjugative transfer.

The product of *finP* is a molecule of ANTISENSE RNA which binds to the *traJ* transcript, while the product of *finO*, a polypeptide referred to as an 'RNA chaperone', stabilizes the activity of the antisense RNA. With a fully active FinOP system in operation, the transfer system remains repressed and conjugation remains inhibited.

The F PLASMID is constitutively derepressed for transfer – i.e. it permanently confers on the host cell the ability to act as a conjugative donor. Earlier, it was thought that this was due to the absence of a *finO* locus in the F plasmid. Later it was shown that *finO* is present but is inactivated by an insertion sequence, IS*3*, which disrupts the *finO* coding sequence. As a consequence, FinO is not available to act jointly with the *finP* product so that the *tra* operon is not inhibited.

first strand (of cDNA) The strand of DNA synthesized on an mRNA molecule by reverse transcriptase. If only a few types

of mRNA are being targeted then gene-specific primers may be used. If many, or all, mRNAs are being targeted (e.g. for a microarray experiment) then oligo(dT) primers may be used; these primers bind to the 3′ poly(A) tails that are commonly present on eukaryotic mRNAs. Another possibility is the use of random primers (e.g. random hexamers); in one study it was found that the use of random pentadecamer (i.e. 15-mer) primers resulted in an improved range and yield of cDNAs compared with the use of random hexamers [BioTechniques (2006) 40(5):649–657].

mRNA is removed from the RNA/DNA hybrid – e.g. by RNASE H or by alkaline hydrolysis – and the *second strand* of DNA is synthesized, making ds cDNA.

In the Gubler–Hoffmann procedure, the hybrid RNA/DNA product is treated by *controlled* enzymic action, leaving short sequences of RNA; DNA polymerase I (which has 5′-to-3′ exonuclease activity) is then used to synthesize the second strand. Hence, the second strand is produced without the use of an exogenous primer.

[Use of Gubler–Hoffmann method: BMC Genomics (2007) 8:363; Cell (2008) 134(6):1019–1029; and PLoS ONE (2009) 4(2):e4404.]

(See also entry CDNA.)

FISH Fluorescence *in situ* hybridization: a particular form of *in situ* hybridization (ISH) in which the label is a fluorophore. In one approach, BIOTINylated probes, bound to their target, are detected by a STREPTAVIDIN-conjugated fluorophore.

In a different approach the probe incorporates a fluorophore (for example, ChromaTide™: see entry PROBE LABELING).

Increased sensitivity may be achieved by using TYRAMIDE SIGNAL AMPLIFICATION.

fitness cost In a transgenic organism: reduction in the overall biological fitness that may result from carrying foreign DNA.

FLAG Fluorescent amplicon generation: a technique used for signal generation in REAL-TIME PCR (cf. the next entry: FLAG®). Essentially, the primers carry a 5′ tag with a quenched fluorophore; during temperature cycling, and production of double-stranded amplicons, a thermostable restriction endonuclease (PspGI) cleaves that part of the amplicon containing the quencher and, hence, allows the fluorophore to fluoresce with suitable excitation [Nucleic Acids Res (2007) 35(19):e131].

FLAG® A peptide (sequence: DYKDDDDK) used for tagging proteins by creating protein–peptide fusions (see entry GENE FUSION); a protein tagged with this sequence can be detected/recovered by monoclonal antibodies specific for the tag.

(See also EPITOPE TAGGING.)

FLAG® was used in various studies on prion proteins [e.g. Mol Cell Biol (2006) 26(7):2697–2715; Mol Biol Cell (2006) 17(8):3356–3368].

[Other uses of FLAG® tag (e.g.): Nucleic Acids Res (2008) 36(19):e127; BMC Genetics (2009) 10:4; PLoS ONE (2009) 4(2):e4466; BMC Microbiol (2009) 9:73; Proc Natl Acad Sci USA (2009) 106(11):4189–4194; and Biochem J (2009) 419 (3):577–584.]

flap (1) (in lagging strand synthesis) See OKAZAKI FRAGMENT. (2) (*DNA technol*.) See INVADER.

flap endonuclease 1 (FEN1; FEN-1) A structure-specific endonuclease involved e.g. in replication of DNA (see OKAZAKI FRAGMENT). FEN1 has also been reported to exhibit a gap endonuclease function which may have a role in resolution of stalled replication forks [FEN1 activity with gap substrates: Nucleic Acids Res (2006) 34(6):1772–1784].

(See also STRUCTURE-SPECIFIC ENDONUCLEASE.)

FLASH® chemiluminescent gene mapping kit A kit (Stratagene) used for high-resolution restriction mapping.

Essentially, the gene (or fragment) – flanked by T3 and T7 promoter sites – is first cloned in a vector. The gene is then excised from the vector, still flanked by the T3 and T7 sites. (Excision occurs at two NotI recognition sites; the NotI sites are used as they are the recognition sites of a 'rare-cutting' enzyme, i.e. an enzyme which is unlikely to interfere with the activity of the particular enzyme whose target sites are being mapped.)

The excised gene is then partially (i.e. incompletely) digested with the particular RESTRICTION ENDONUCLEASE whose target sites are being mapped.

The fragments produced by the restriction enzyme are subjected to electrophoresis in an agarose gel, and the bands of fragments are transferred to a hybridization membrane.

Bands of fragments on the membrane are probed with a T3 sequence that is conjugated to ALKALINE PHOSPHATASE. The hybridization membrane is then incubated with a chemiluminescent substrate (see CHEMILUMINESCENCE), and the light signal from the target–probe complex is detected on film.

The position of each band in the resulting ladder reflects the distance between one end of the gene (i.e. the T3 site) and *one* of the cutting sites of the given restriction enzyme. The bands in the ladder therefore indicate the various cutting sites of the given enzyme within the gene.

By replicating the procedure with an alkaline phosphatase-conjugated T7 probe, a *complementary* set of bands are seen – each band reflecting the distance between one end of the gene (i.e. the T7 end in this case) and one of the cutting sites of the restriction enzyme. Hence, the second determination (with the T7 probe) serves to confirm the results obtained in the first determination.

flavivirus A member of a family of enveloped viruses in which the monopartite genome is linear positive-sense ssRNA. The viruses can infect invertebrate and vertebrate hosts, and many cause mosquito-borne or tick-borne disease in humans and in animals; they include e.g. West Nile virus, yellow fever virus and the causal agents of dengue and Murray Valley fever.

flipping (nucleotide flipping) See NUCLEOTIDE FLIPPING.

Flock House virus (FHV) A virus of the genus *Alphanodavirus* (see NODAVIRUSES). FHV encodes a protein, B2, which can suppress RNA INTERFERENCE; B2 is essential for FHV infection of *Drosophila melanogaster*. A recombinant B2-encoding Sindbis virus was used for studying the role of RNAi in the (arbovirus) infection of mosquitoes [BMC Microbiol (2009)

9:49].

flood plate A PLATE whose surface has been flooded with a (concentrated) liquid suspension of a given organism (usually a suspension of bacteria) and excess liquid removed e.g. with a sterile Pasteur pipet.

A ('dried') flood plate (in which excess surface moisture has been eliminated) is used e.g. in PHAGE TYPING.

(See also LAWN PLATE.)

floral dip method A method that was first used for the transfection of *Arabidopsis thaliana* by *Agrobacterium tumefaciens*; it involves dipping developing floral tissues (including floral buds) into a solution containing 5% sucrose, 500 μL/L of the surfactant Silwet L-77 and cells of *A. tumefaciens*. [See also Methods Mol Biol (2005) 286:91–102, (2005) 286:103–110.] It appears that, during this procedure, the bacterial cells are delivered to the inside of the developing gynoecium, there carrying out transformation of female gametes.

(See also the entry AGROBACTERIUM.)

[Use of the floral dip method for transcriptional profiling of the *Arabidopsis* embryo: Plant Physiol (2007) 143:924–940. Other uses of the method: Plant Physiol (2009) 149(2):938–948; Plant Cell Physiol (2009) 50(2):413–422; BMC Plant Biol (2009) 9:5; PLoS Genet (2009) 5(1):e1000355.]

[Evidence for stable transformation of wheat by floral dip in *Agrobacterium tumefaciens*: Plant Cell Rep (2009) 28(6): 903–913.]

[Streamlined sub-protocols for the floral-dip transformation and selection of transformants in *Arabidopsis thaliana*: Plant Methods (2009) 5:3.]

flow cytometer An instrument used in FLOW CYTOMETRY. In the instrument, the cells of a given sample (in suspension) are made to move – singly, and at high speed – down a fine tube; within this tube each cell passes through a beam from a laser or other source, generating a signal. Thus, for example, a cell passing through a beam from a 488-nm argon laser scatters the beam and gives rise to both diffracted and reflected rays – which are picked up by separate scatter detectors. The signals are sent to a computer and provide information regarding the cell's size and internal complexity.

The flow cytometer can also be used e.g. to determine the number of cells (if any) in a given population which express a specific cell-surface molecule. For example, a population of cells can be exposed to fluorophore-labeled antibodies which bind (specifically) to the given cell-surface molecule. When each cell passes through the detector system it is exposed to an excitatory beam that causes the label to fluoresce, and this signal is transmitted to the computer. More than one type of cell-surface molecule can be investigated in a given assay by using different populations of antibodies that are labeled with different fluorophores (with distinct emission spectra).

By incorporating a *cell sorter* in the flow cytometer, it is possible to separate the cells of a population into subsets, in different collecting vessels, on the basis of specified criteria and the signals received by the detector: see FLUORESCENCE-ACTIVATED CELL SORTER and SEX SORTING.

Online information on flow cytometers

Online information on flow cytometers may be obtained via several commercial websites, e.g.:

www.beckman.com

www.bdbiosciences.com/home/

flow cytometry The use of a specialized instrument (a FLOW CYTOMETER) for counting cells (or cell components etc.), or for analyzing/measuring specific feature(s) of those cells or components. If the flow cytometer incorporates a *cell sorter* it can be used to separate the cells of a given population into subsets according to specified properties/criteria – see e.g. FLUORESCENCE-ACTIVATED CELL SORTER.

***flox*ed DNA** A segment of DNA flanked, on both sides, by a *loxP* site (see the entry CRE–LOXP SYSTEM).

[Examples of use of term: Proc Natl Acad Sci USA (2008) 105(7):2521–2526, and J Biomed Biotechnol (2009) 985140 (doi: 10.1155/2009/985140).]

(See also FLRTED.)

Flp (FLP) A site-specific RECOMBINASE, encoded by the TWO-MICRON PLASMID, which was initially isolated from the yeast *Saccharomyces*; Flp is a member of the integrase family of recombinases, and a tyrosine residue occurs at the active site of the enzyme.

(See also SITE-SPECIFIC RECOMBINATION.)

Within *S. cerevisiae* Flp catalyzes the inversion of a section of the plasmid flanked by a pair of (inverted) recognition sequences.

In recombinant DNA technology, Flp is used e.g. for the excision of specific sequences *in vivo*. The enzyme is active in both eukaryotes (e.g. *Drosophila*) and bacteria (including *Escherichia coli*). The enzyme can excise a sequence that is flanked, at both ends, with a so-called *Flp recognition target* (FRT) sequence – both FRTs being in the same orientation; the excised fragment is in the form of a circular molecule.

A plasmid containing one FRT site can be inserted, by the action of the Flp recombinase, into another molecule which also contains one copy of the FRT site.

[Modular and excisable molecular switch for the induction of gene expression by the Flp recombinase: BioTechniques (2006) 41(6):711–713.]

The Flp/FRT system has been used e.g. in a gene targeting vector for engineering conditional knock-outs in mice [PLoS ONE (2009) 4(2):e4341].

(See also the vector pEF5/FRT/V5-DEST™ in the table for entry GATEWAY SITE-SPECIFIC RECOMBINATION SYSTEM.)

Certain peptides which inhibit Flp have been reported from studies on the HOLLIDAY JUNCTION (q.v.).

(See also CRE–LOXP SYSTEM.)

Flp-In™ cell lines Cell lines from Invitrogen (Carlsbad CA) in which the target site (FRT) of the Flp recombinase (see FLP) has been inserted stably into a transcriptionally active region of the genome. These cells are designed to be co-transfected with an FRT-containing Flp-In™ expression vector (carrying the gene of interest) and a plasmid (pOG44) encoding the Flp

recombinase; (targeted) integration of the expression vector occurs at the same location in each of the transfected cells – promoting uniform levels of expression of the given gene of interest.

Flp-In™ cells have been prepared from various types of parental cell, including BHK, CHO, Jurkat and 293. In some of these cells a CMV promoter works well, but in others (e.g. BHK) expression from a CMV promoter was reported to be downregulated; in the latter cells, Flp-In™ vectors with the EF-1α promoter are recommended by the manufacturer.

[Uses (e.g.): Nucleic Acids Res (2008) 36(12):3926–3938; PLoS ONE (2009) 4(1):e4200; Mol Syst Biol (2009) 5:235; PLoS ONE (2009) 4(5):e5710; Nucleic Acids Res (2009) 37 (11):3635–3644; Proc Natl Acad Sci USA (2009) 106 (13): 5053–5058.]

(See also vector pEF5/FRT/V5-DEST™ in GATEWAY SITE-SPECIFIC RECOMBINATION SYSTEM (table).)

flrted Flanked by FRT sites (see the entry FLP).

[Example of use of term: PLoS ONE (2009) 4(2):e4341.]

(See also FLOXED DNA.)

fluctuation test A classic, early (1943) test, devised by Luria and Delbrück, for examining the way in which populations of bacteria respond to changes in the environment.

Two hypotheses co-existed at the time: (i) genetic change is induced by environmental influences (i.e. an adaptive mechanism), and (ii) genetic change occurs spontaneously (that is, independently of environmental influences) – cells in which such changes have occurred being selected (and proliferating) if conditions become suitable.

Essentially the (statistics-based) test measured the variance between numbers of phage-resistant (mutant) cells in each of a set of individual, separate liquid cultures of phage-sensitive bacteria; this was compared with the variance in numbers of phage-resistant mutants in each of a number of samples taken from a single (bulk) liquid culture. (The numbers of phage-resistant mutants were ascertained by plating aliquots from the cultures on separate plates and then exposing the plates to virulent phage.)

The high variance (wide fluctuation) in numbers of mutants from the separate cultures indicated that the (phage-resistant) mutant cells had arisen spontaneously – at different times in different cultures – i.e. that the appearance of mutants was unrelated to exposure of the bacteria to phage.

(See also NEWCOMBE EXPERIMENT.)

fluorescence (*in vivo*) See IN VIVO FLUORESCENCE.

fluorescence-activated cell sorter (FACS) A type of FLOW CYTOMETER (q.v.) which incorporates a cell sorter and which is used for separating the cells of a given population into subsets on the basis of fluorescent signals. Thus, in order to be sorted by FACS, the cells must carry a fluorophore – either at the surface or intracellularly; specific cell-surface molecules can bind (fluorophore-labeled) antibodies, or the cells can be stained with a fluorophore (such as HOECHST 33342) which is able to bind to intracellular DNA.

In the FACS, each cell passing through the detector tube (see FLOW CYTOMETER) is blown out within a micro-droplet. As it falls, the micro-droplet is given a charge (either positive or negative) according to computer guidance (based on the signals received by the detector system); thus, e.g. a cell that gives a particular fluorescence signal may be given a positive charge while one lacking that signal may be given a negative charge. All cells then pass between a pair of electrostatically charged plates, one positive, the other negative; the positively charged cells will therefore be deflected towards the negative plate, and vice versa, so that each cell will be directed to an appropriate collecting vessel below.

The FACS principle is used e.g. for SEX SORTING (q.v.).

fluorescence *in situ* hybridization See FISH.

fluorescence resonance energy transfer See FRET.

fluorescent amplicon generation See FLAG.

5-fluorocytosine (in siRNA selection) See RNA INTERFERENCE (section: *Selection of potent siRNAs from an siRNA library*).

fluoromycobacteriophage A mycobacteriophage that carries a fluorescent reporter gene; such phages were constructed for detecting antibiotic resistance in *Mycobacterium tuberculosis* [PLoS ONE (2009) 4(3):e4870].

5-fluoro-orotic acid See entry YEAST TWO-HYBRID SYSTEM.

fluorophore (fluorophor) Any molecule/substance that exhibits fluorescence on stimulation by radiant energy (of appropriate wavelength).

(See DNA STAINING and SOLVATOCHROMIC FLUORESCENT DYE.)

fluoroquinolones See QUINOLONE ANTIBIOTICS.

FluoroTrans® PVDF membranes See PVDF.

5-fluorouracil (in siRNA selection) See RNA INTERFERENCE (section: *Selection of potent siRNAs from an siRNA library*).

FMRP Fragile X mental retardation protein: the protein whose absence is responsible for FRAGILE X DISEASE (q.v.).

FMRP binds, with apparent selectivity, to mRNAs, and has been implicated in regulating the efficiency of translation. It is an important factor e.g. in the development and function of neurones, and the absence of FMRP-regulated translation in response to synaptic activity has been suggested as a possible causal mechanism of fragile X disease.

FMRP was reported to bind to mRNA encoding superoxide dismutase and to enhance translation of that enzyme; it was suggested that a lack of FMRP (in fragile X disease) might lead to reduced levels of superoxide dismutase and increased oxidative stress in the brain, the associated features of which typically affect individuals with fragile X disease [PLoS Biol (2009) 7(1):e1000016].

5-FOA 5-fluoro-orotic acid: see YEAST TWO-HYBRID SYSTEM.

foamy viruses Viruses of the subfamily SPUMAVIRINAE (q.v.).

folding modulators See OVEREXPRESSION (*Inclusion bodies*).

footprinting A technique that is used to determine the location (i.e. sequence) at which a given ligand (e.g. a DNA-binding protein) binds to DNA.

Note. An *entirely unrelated* technique is known as GENETIC FOOTPRINTING (q.v.) '*Genomic footprinting*' has been used to refer to footprinting conducted with *living* cells: see the entry

IN VIVO FOOTPRINTING.

This entry refers to footprinting *in vitro*.

Initially, two identical populations of end-labeled sample fragments of DNA are prepared, and the ligand of interest is allowed to bind to fragments in *one* of the populations. Both populations are then subjected to either (i) enzymatic *or* (ii) chemical cleavage under conditions in which each fragment is (ideally) cut at only one of a large number of possible sites. In each population of fragments, the result is a collection of subfragments, present in a range of different lengths, that are produced by cleavage at different sites.

In the population of ligand-bound fragments, cleavage will not have occurred at those cleavage sites which had been shielded from the enzyme (or chemical agent) by the bound ligand; this population will therefore lack a certain range of subfragments that are present in the other population. Hence, the *absence* of certain subfragments in the ligand-bound population will be seen as a *gap* when the two populations of subfragments are compared by gel electrophoresis; this gap is the ligand's *footprint*, and its position in the gel (in relation to the bands of subfragments) can help to indicate the ligand's binding site on the DNA.

In one approach, cleavage of fragments is carried out by the enzyme DNase I. This produces a large and clear footprint because DNase I is a large protein that will not cut the DNA at sites close to a large DNA-bound ligand (such as a DNA-binding protein).

Resolution can be improved (i.e. the footprint can be better defined) by using a small chemical agent to cleave the DNA. One such agent is hydroxyl radical (·OH) which brings about sequence-independent cleavage of DNA: any phosphodiester bond which is unprotected by the bound ligand is susceptible to cleavage. Footprinting with the hydroxyl radical can be achieved by means of the FENTON REACTION [Nucleic Acids Res (2006) 34(6):e48]. Footprinting with hydroxyl radical in frozed cells (i.e. *in vivo*) can be achieved by using a brief pulse of X-radiation from a synchrotron – which can generate hydroxyl radical footprints similar to those generated *in vitro* by the Fenton reaction [Nucleic Acids Res (2006) 34(8):e64].

forced cloning (directional cloning) Any procedure designed to ensure that target DNA is inserted into a vector molecule in a definite, known orientation; insertion in this way is required e.g. when the target DNA is subsequently to be transcribed from a promoter sequence within the vector.

Methods used for directional cloning include preparation of target DNA (and vector) with asymmetric termini – see e.g. DIRECTIONAL TOPO PENTR VECTORS.

forensic applications (of DNA-based technology) DNA-based technology can offer a range of possibilities for assisting with investigations. Moreover, given that DNA may survive for long periods of time, particularly in the right kind of tissue, under appropriate conditions, current methods can sometimes resolve questions that have remained unanswered for many years. (One recent investigation reported identification of two missing Romanov children (from the last Russian monarchy) by DNA analysis [PLoS ONE (2009) 4(3):e4838].)

DNA-based identification; ***sample quantity and condition***

DNA is naturally important in the context of identification – including the identification of those killed in natural disasters such as earthquakes, floods, tsunamis etc. as well as victims (or perpetrators) of crime; moreover, this technology also has a role e.g. in cases involving paternity issues.

Each investigation starts with one or more samples that are sources of DNA. The quality and quantity of DNA available from a given sample can vary enormously according to (i) the type of sample, (ii) the conditions under which that sample had existed prior to collection, and (iii) the time for which it was exposed to those conditions.

Positive, unambiguous analysis of DNA requires a sample of at least a certain minimum quantity – that is, of the order of nanograms/picograms – and of a quality which is suitable for use. Fortunately, with modern methods, minute quantities of DNA can be characterized, and some commercial methods (e.g. CHARGESWITCH TECHNOLOGY) can efficiently recover the DNA from various types of crude sample.

PCR (q.v.) is widely used for amplifying the sample DNA in a forensic context; this approach can handle samples of DNA of the order of picograms. (See also LCN DNA TEST.)

PCR amplification of DNA from contaminated specimens, – even from e.g. crude soil samples – has been reported with a mutant form of *Taq* DNA polymerase [Nucleic Acids Res (2009) 37(5):e40].

Even poor quality/damaged DNA may still be of use if it is susceptible to PCR amplification by the Y FAMILY (q.v.) of DNA polymerases.

Degraded DNA in small amounts is reported to be susceptible to isothermal whole-genome amplification with primers optimized, bioinformatically, for coverage of the human genome. Degraded DNA, pre-amplified by this procedure, was reported to be satisfactory for STR-based genotyping [DNA Res (2006) 13(2):77–88].

Degraded DNA in small amounts has been assayed with an STR-based PCR involving short amplicons (<280 bp) and a low-copy-number protocol (33 PCR cycles, reaction mixture 12.5 μL) [J Forensic Sci (2009) 54(4):862–865].

Whole-genome amplification (WGA) may be useful when sample DNA is degraded and/or available in only insufficient quantities for STR-based genotyping. Use of such low levels of DNA may result in an incomplete DNA profile which has insufficient discriminatory power. The applicability of WGA (multiple displacement amplification) for generating a usable quantity of sample DNA (from a minimal source) – for use in forensic SNP genotyping – was assessed by a comparative analysis of genomic versus MDA-amplified DNA samples that involved a 26-SNP TaqMan panel; the results suggested that MDA can generate large quantities of genome-equivalent DNA by using small initial samples [BMC Genomics (2009) 10:159].

A system for *quantifying* the level of damage in a sample of DNA has been devised in a *zoological* context and may be

useful for forensic investigations involving domestic animals or wildlife [Frontiers Zool (2006) 3:11].

Nuclear versus mitochondrial DNA

Both nuclear DNA and MITOCHONDRIAL DNA are of value for identification purposes. In general, analysis of nuclear (as opposed to mitochondrial) DNA is frequently the preferred option because nuclear DNA has a potentially greater power to discriminate between individuals. Samples of mtDNA may be unsatisfactory in certain cases – e.g. in that they can fail to distinguish between individuals from the same maternal line. Nevertheless, in the USA there is an extensive database of mtDNA – known as the mtDNA Popstats Population Database – containing over 5000 mtDNA profiles; this database is intended primarily for law-based and academic use. A given mtDNA profile in the database is described in relation to the (revised) standard 'Cambridge Reference Sequence'. Ongoing scrutiny continues to reduce errors in this database, which is considered to be a reliable resource for forensic case analysis.

Sequencing of 'whole-genome' mtDNA by a procedure that involves two overlapping PCR fragments (each of ~8500 nt) was reported to produce high-quality data [BMC Genomics (2009) 10:139].

Investigations based on mtDNA have been facilitated with the T5000™ BIOSENSOR SYSTEM.

A potential problem with the use of mtDNA targets is the existence of so-called 'mitochondrial pseudogenes': mtDNA-like sequences which occur in nuclear (chromosomal) DNA; these sequences are called *numts* (nuclear mitochondria-like sequences). Without appropriate precautions, these sequences could be co-amplified by PCR when attempting *specifically* to amplify particular mtDNA targets. The importance of this would be underestimated if it is assumed that the high copy number of the mtDNA genes (compared with the nuclear pseudogenes) may mask the effect of these nuclear targets: some of the pseudogenes are represented at *multiple* sites in nuclear DNA. Some 46 fragments of nuclear DNA – covering all of the human mitochondrial DNA – have been sequenced [BMC Genomics (2006) 7:185].

There are indications that mtDNA can survive better than nuclear DNA in ancient samples taken from the permafrost regions [Nucleic Acids Res (2009) doi: 10.1093/nar/gkp159].

Obtaining samples of DNA

Samples of material containing DNA can be obtained e.g. by BUCCAL CELL SAMPLING (from both human and non-human subjects) or by taking small amounts of blood etc.

A proportion of DNA samples obtained from saline mouth-washes were reported to be degraded [see: PLoS ONE (2009) 4(7):e6165].

Other sources of DNA include blood, teeth, hair, and cells from the skin.

(See also DNA ISOLATION.)

Blood as a source of DNA

While blood is a useful source of DNA, uncertainty has been expressed concerning oxidative damage to the DNA in *dried* specimens. Such damage was thought possible by the release of Fe^{2+} from hemoglobin and the involvement of these ions in a FENTON REACTION that produced DNA-damaging hydroxyl radicals (·OH). Fe^{2+} ions are released from dried hemoglobin, and this presents a risk of oxidative damage to DNA in the vicinity; it appears that any damage to DNA is likely to have occurred before the specimen reaches the analyst [PLoS ONE (2009) 4(4):e5110].

Teeth as a source of DNA

Given adverse conditions, the remains of dead individuals (including isolated body parts) may be poor sources of DNA. However, the more-durable hard tissues, such as teeth, can be a valuable source; moreover, compared with other types of sample, hard-tissue samples may be easier to free from contamination by extraneous DNA.

Hair as a source of DNA

Hair can be useful as a source of DNA, and is particularly valuable when the origin of a given sample is a major factor in an investigation. Prior to DNA analysis, it is important to carry out a thorough macroscopic and microscopic examination of the hair sample (particularly with a very small sample) because the portion used for DNA analysis will no longer be available for physical characterization.

A hair in the active phase of growth is likely to provide a good source of both nuclear and mitochondrial DNA from cells in the root and sheath regions. However, hairs at the end of the growth cycle – lacking follicular components, and shed from the body – are likely to be poor/inadequate sources of nuclear DNA, although they may be adequate for the analysis of mitochondrial DNA.

[Guidelines for the forensic examination of human hair (including a glossary of descriptive terms): Forensic Sci Comm (2005) 7(2).]

Types of investigation used for identification purposes

Given adequate samples, human and non-human DNA can be readily distinguished (if necessary).

The *gender* of a subject may be ascertained e.g. by a PCR-based assay of sequences from alleles of the amelogenin gene on the X and Y chromosomes: see DNA SEX DETERMINATION ASSAY. (See also BARR BODY.) If this assay is not applicable (e.g. through mutation in the amelogenin gene), an alternative approach may be an SRY-BASED ASSAY.

Identification can involve e.g. the analysis of genomic *short tandem repeats* (see STR) – an approach used in CODIS (q.v.); this is a widely used approach.

Male-specific STRs (Y-STRs – on the Y chromosome) may be useful in sex-based and other cases. [Interpretation guidelines for Y-STRs: Forensic Sci Comm (2009) 11(1).]

[A 26plex autosomal STR assay to aid human identity testing: J Forensic Sci (2009) 54(5):1008–1015.]

Although the use of STRs in CODIS is a firmly established procedure, *single-nucleotide polymorphisms* (see entry SNP) could be useful e.g. (i) when only short fragments of sample DNA (60–80 bp) are available; (ii) when the sample DNA is degraded to the point at which normal procedures cannot be used; (iii) if an STR-based match cannot be found in CODIS.

Four classes of (forensically relevant) SNPs were described [BioTechniques (2008) 44(5):603–610]. Assuming an insufficient source of sample DNA, then whole-genome amplification may be of use for preparing quantities of DNA which are suitable for SNP genotyping (see above).

Investigations may also include an analysis of mtDNA (see comments above).

In criminal investigations, results that exclude a suspect can be as valuable as those which incriminate.

DNA derived from *dogs* may be useful e.g. to link a suspect to a crime scene, or may be presented as evidence in cases of dog attacks. Canine DNA may be examined by a method that involves amplification, by PCR, of 18 STRs [developmental validation of an STR kit for forensic DNA profiling of canine biological materials: Croat Med J (2009) 50(3):268–285].

Ethical considerations

DNA-based sampling and identification has important ethical aspects which have raised concerns. [Balancing crime detection, human rights and privacy (UK National DNA Database): EMBO Rep (2006) 7 (Special Issue):S26–S30. Race, genes, and justice: a call to reform the presentation of forensic DNA evidence in criminal trials: Brooklyn Law Rev (2008) 74(2): 325–375.]

Disasters with mass fatalities

In exceptional events, e.g. major earthquakes and tsunamis, the problem of identification may be exacerbated by an overwhelming number of victims and, in many cases, by the lack of adequate technical facilities. Moreover – in any situation – dental, fingerprint and DNA-based data can be useful only if comparative data (e.g. in a database) are also available.

In the South Asian tsunami of December, 2004, an international collaborative effort assisted with the identification of victims. At the TTVI (Thai Tsunami Victim Identification) center, some 2010 victims had been identified 7 months post-tsunami. Of these victims, 1.3% had been identified on the basis of DNA analyses – as compared with 61% by dental records and 19% from fingerprints (some of the victims were identified by a combination of methods). In this scenario it has been considered that DNA-based identification should be reserved for those cases in which attempts at identification by other methods have been unsuccessful (instead of being used as a first-line approach).

[Management of mass fatality following the South Asian tsunami of 2004: PLoS Medicine (2006) 3(6):e195.]

Microbial forensics

Microbial forensics is a multi-disciplinary field of endeavor created in response to the ongoing threat of attack, on populations and/or individuals, by terrorists/criminals using pathogenic microorganisms or their toxins. A high-profile example of such attacks was the letter containing spores of *Bacillus anthracis* (causal agent of anthrax) sent to a member of the US government in 2001. A specialized laboratory facility, the National Bioforensics Analysis Center (NBFAC), was set up to examine and analyze relevant material.

[Sample collection, handling and preservation for an effective program of microbial forensics: Appl Environ Microbiol (2006) 72(10):6431–6438.]

As, in this context, the results of microbiological tests may have far-reaching consequences (in e.g. legal/political terms), it is clearly important that both the quality and reliability of the methods used should conform to highly stringent criteria. [Criteria for validation of the methods in microbial forensics: Appl Environ Microbiol (2008) 74(18):5599–5607.]

Guidance is available for those physicians who believe that a given patient is a victim of bioterrorism (or other type(s) of biocrime). [Biocrimes, microbial forensics and the physician: PLoS Medicine (2005) 2(12):e337.]

The role of DNA-based technology in this field could be seen as central to tasks such as the identification of pathogens and determination of their source(s). Ideally, methods should be available to characterize a given isolate of a pathogen *and to link it indisputably to a specific origin*. However, this was not possible in the case of letter-borne spores of *B. anthracis*; in that case, TYPING based on the analysis of tandem repeats in DNA pointed to a laboratory source – but could not offer information that was more specific regarding the origin of the strain. A paper summarizing the requirements of the nascent field of microbial forensics [Appl Environ Microbiol (2005) 71(5):2209–2213] referred to a more informative and highly sophisticated (but not yet realized) typing system that could exploit our expanding knowledge of genome plasticity and bioinformatics. A future system for comparing genomic sequence data (suitable for detecting differences among isolates of a pathogen) was seen in terms of an approach that could accomodate genomic variation due to recombinational events and also possible genome-wide patterns of mutation; such a system was subsequently to be brought closer with the use of mass spectrometry – see later.

Reliance on selected sequences of nucleotides for identification purposes can be problematic – as was shown e.g. by the finding of sequences of *B. anthracis* virulence plasmid pXO2 in plasmids from other species of *Bacillus* [J Clin Microbiol (2006) 44(7):2367–2377]. Precision-based identification and comparison of isolates by whole-genome sequencing would require the availability of techniques that are faster and less expensive than those in current use; moreover, even with the development of such techniques, maximum use of these sequencing data is possible only if (accessible) databases on the pathogens, and related species/strains, are in place.

Typing of *B. anthracis* on the basis of its single-nucleotide repeats (see SNR) has been reported to permit differentiation of isolates with extremely low levels of genetic diversity – i.e. very closely related strains – even when such isolates could not be distinguished by other methods of typing [J Clin Microbiol (2006) 44(3):777–783].

Any typing procedure involving copying specific sequences of genomic DNA is likely to be facilitated by techniques that accelerate amplification – such as an ultra-fast form of PCR [Nucleic Acids Res (2006) 34:e77].

It was suggested [Appl Environ Microbiol (2005) 71(5): 2209–2213] that nucleic-acid-based techniques are unlikely to be able to pinpoint the unique source of a given isolate of a pathogen (a requirement e.g. for tracing perpetrators of bioterrorism). This pointed to the need for chemical and physical analyses to supplement nucleic-acid-based methods in order to supply additional information; in particular, it was thought that the (chemical) composition of microbial cells might be able to reveal more information regarding their origin if more information were available on the way in which the chemical nature of miroorganisms reflects the type of environment in which they were grown.

Later, however, DNA-based technology in the form of the T5000^{TM} Biosensor System (Ibis Biosciences) demonstrated that strains of microorganisms could be distinguished from one another by the use of mass spectrometry to detect the fine differences in base composition of DNA in related strains.

B. anthracis is, of course, only one of a number of possible microorganisms that can be considered in this context. Other, more common, types of pathogen, including those associated with natural outbreaks of disease, are also candidates. Less dramatic, covert events caused by this kind of pathogen may be correspondingly more difficult to detect against the background of routine, natural outbreaks.

Plant pathogen forensics/plant forensics

Bioterrorism directed at agricultural targets has the potential for creating widespread nutritional and economic damage as well as political instability. The diffuse nature of agriculture, and the (relatively) lower level of sophistication needed to target this area, make this a particularly difficult problem to tackle. Moreover, the majority of plant pathogens are species of fungi which are, in general, technically less amenable than bacteria to laboratory investigation. Additionally, plant pathogens are generally harmless to humans, so that any would-be bioterrorist in the agricultural sphere would be exposed to minimal risk. The total prevention of an attack on agricultural targets was assessed as impossible by the Banbury Microbial Forensics Group – so that vigilence and an effective response to any incident remain the optimal approach.

Information, technology and resources available at present in the agricultural sphere have been assessed for potential use in plant pathogen forensics [Microbiol Mol Biol Rev (2006) 70(2):450–471].

Forensic monitoring of individual plant species, as opposed to the detection of 'agricultural bioterrorism' in general, is an ongoing study which, as in human forensics, can involve e.g. STR markers [developmental validation of a *Cannabis sativa* STR multiplex system: J Forensic Sci (2008) 53:1061–1067].

formamide ($H.C{=}O.NH_2$) A denaturing agent which lowers T_m in the THERMAL MELTING PROFILE of dsDNA.

forward genetics The classical approach to genetics in which, in general, the object is to identify gene(s) whose function(s) relate to particular aspect(s) of an organism's PHENOTYPE.

(cf. REVERSE GENETICS.)

forward mutation Any mutation which gives rise to a mutant phenotype, commonly by inactivating or modifying a gene.

forward primer Any PRIMER which is designed to bind to the *non-coding* strand of a gene (as shown below).

The fragment of dsDNA shown below identifies a potential forward primer as a sequence of nucleotides – in heavy type – within the coding strand. As an *isolated* oligonucleotide, this 8-mer sequence (5′-GCTATGCT-3′) would bind to the underlined sequence in the non-coding strand.

5′-**GCTATGCT**CTGCC/. . . . . /GGGACTCGGCACGA-3′
(*coding strand*)

3′-CGATACGAGACGG/. . . . . /CCCTGA**GCCGTGCT**-5′
(*non-coding strand*)

Consequently, when extended by the DNA polymerase, this (forward) primer would give rise to a sequence of nucleotides identical to that in the coding strand.

A *reverse primer* binds to the coding strand. In the fragment shown, a potential reverse primer is identified (in heavy type) as a sequence of nucleotides in the non-coding strand. As an *isolated* oligonuleotide, 5′-TCGTGCCG-3′ would bind to the underlined sequence in the coding strand.

fosmid Any COSMID based on the F PLASMID; intracellularly, it occurs as a low-copy-number vector.

[Some uses of fosmids: BMC Dev Biol (2008) 8:119; PLoS ONE (2009) 4 (1):e4192; J Hematol Oncol (2009) 2:3.]

[Recombineering of fosmids for making reporter constructs in the nematode *Caenorhabditis elegans*: PLoS ONE (2009) 4(3):e4625.]

454 sequencing (four-five-four sequencing) See GS FLX.

FOXC1 See AXENFELD–RIEGER SYNDROME.

FPET Formalin-fixed paraffin-embedded tissue.

Fpg protein *Syn*. MutM: see the entry 8-OXOG.

FPRL1 A potential cell-surface co-receptor for HIV-1.

fps (*FPS*) An ONCOGENE in strains of avian sarcoma virus. The v-*fps* product has TYROSINE KINASE activity.

fragile X disease A hereditary disease characterized by mental retardation; the defect, on the X chromosome, arises through instability of MICROSATELLITE DNA (see also the table in the entry GENETIC DISEASE).

Fragile X disease is associated with absence of the protein FMRP (q.v.).

fragmentation (of DNA) See DNA FRAGMENTATION.

frame-shift mutation (phase-shift mutation) Any mutation in which the addition or deletion of nucleotide(s) results in an out-of-phase message (altered reading frame) in the sequence downstream of the mutation. A frame-shift mutation may be induced e.g. by INTERCALATING AGENTS.

(cf. FRAMESHIFTING.)

frameshifting (ribosomal frameshifting; programed ribosomal frameshifting; translational frameshifting) A mechanism in which the precise product of translation is determined by an induced (programed) movement of the ribosome at a specific site in the mRNA – either a movement of one nucleotide up-

stream (i.e. in the 5′ direction) or one nucleotide downstream (in the 3′ direction), translation then continuing in the new –1 or +1 reading frame, respectively.

The ribosome's movement is induced by signals which are incorporated within the nucleotide sequence of the mRNA. In general, two kinds of signal have been reported. First, a so-called *slippery sequence*, downstream of the initiator codon, which contains groups of repeated nucleotides. Second, a so-called *stimulator sequence* – a short distance downstream of the slippery sequence – which appears to be often in the form of an RNA pseudoknot. It appears that, in at least some cases, frameshifting can occur through the influence of the slippery sequence alone, i.e. without assistance from a stimulator.

Frameshifting has been reported in relation to genes from viruses, bacteria, insects and animals. One of the functions of frameshifting is apparently to confer on a given coding sequence the ability to yield more than one type of product.

Studies on the translation of the 3a protein from the SARS (severe acute respiratory syndrome) virus have indicated that heptauridine and octouridine stretches in the mRNA are the sole signals needed for frameshifting in this instance [Nucleic Acids Res (2006) 34(4):1250–1260]. [Predicting ribosomal frameshifting efficiency: Phys Biol (2008) 5(1):16002.]

framycetin See the entry AMINOGLYCOSIDE ANTIBIOTIC.

FRET (fluorescence resonance energy transfer) A phenomenon in which energy (that would otherwise be dissipated as fluorescence) is transferred from one molecule or moiety (donor) to another, *closely adjacent* molecule or moiety (acceptor) by a mechanism that does not involve emission of a photon; the energy received by the acceptor molecule or moiety is typically converted to fluorescence, but in some cases is absorbed (*quenched*) without the production of fluorescence.

(Compare BRET; see also IFRET.)

FRET can be exhibited between two regions on the same molecule, one region acting as the donor and the other as the acceptor. One example is the CCF2 substrate (which includes both coumarin and fluorescein moieties): see pcDNA^{TM}6.2/GeneBLAzer^{TM}-DEST vector in the table in GATEWAY SITE-SPECIFIC RECOMBINATION SYSTEM.

In general, FRET requires that donor and acceptor molecules be separated by no more than ∼100 Å; the separation in experimental protocols is often e.g. 30–70 Å. The donor and acceptor dipoles must also be orientated correctly for effective transmission of energy.

In a TAQMAN PROBE (q.v.), energy received by the donor is quenched by the acceptor molecule that is bound adjacently on the oligonucleotide. When the probe is degraded (by the exonuclease activity of the DNA polymerase), the donor and acceptor molecules become separated and the donor molecule (i.e. reporter dye) can therefore exhibit fluorescence because its energy is no longer quenched.

The probes used in the LightCycler^{TM} instrument involve a different approach (see entry REAL-TIME PCR). Fluorescence from the reporter dye (at a wavelength of 640 nm) is detected only if both halves of the probe bind correctly to the target sequence. When this happens, FRET-based excitation of the reporter dye, LightCycler^{TM} Red 640 (present on one half of the probe), occurs by transfer of energy from the fluorescein molecule (on the other half of the probe) because, under these conditions, the donor and acceptor molecules are within the correct distance of one another for effective energy transfer.

In some donor–acceptor pairs the acceptor molecule is a non-fluorescent quencher such as DABCYL or one of the QSY dyes. The QSY dyes are derivatives of diarylrhodamine; their absorption spectra are largely insensitive to a wide range of pH values. The absorption maxima of QSY dyes include e.g. ∼660 nm (QSY 21) and ∼472 nm (QSY 35). Non-fluorescent acceptors are particularly useful in that they avoid the background fluorescence produced by non-specific excitation of fluorescent acceptor molecules.

The FRET principle has been exploited in a wide variety of studies involving e.g. detection of hybridization of nucleic acid molecules, the intracellular localization and distribution of specific molecules, the conformation of proteins, enzyme–substrate interactions etc. Note that, in all of these studies, the object is to determine whether or not the donor and acceptor molecules are located sufficiently close to one another in the experimental material to permit FRET-based interaction.

Uses of FRET

Examples of FRET-based studies include: determination of cell-cycle-dependent localization of proteins in the bacterium *Bacillus subtilis* [EMBO J (2002) 21:3108–3118]; studies on interactions between enzyme and substrate [Proc Natl Acad Sci USA (2002) 99:6603–6606]; studies on quadruplex and Watson–Crick DNA equilibrium [Nucleic Acids Res (2005) 33(21):6723–6732]; the detection of SNPs [BioTechniques (2006) 40(3):323–329]; studies on DNA bending [Proc Natl Acad Sci USA (2006) 103(49):18515–18520], and on homodimer formation of the enzyme signal peptide peptidase [Mol Neurodegener (2006) 1:16; studies on binding between DNA and the transcription factor NF-κB [BioTechniques (2007) 43 (1):93–98; protein interactions in the *Escherichia coli* chemotaxis pathway [Mol Syst Biol (2009) 5:238; and work on the RNA exit channel during transcription with RNA polymerase II [Proc Natl Acad Sci USA (2009) 106(1):127–132].

(See also DARK QUENCHING.)

Frog Prince See RECONSTRUCTED TRANSPOSABLE ELEMENT.

FRT The recognition site of the FLP recombinase (q.v.).

***ftsZ* gene** In *Escherichia coli*: a gene encoding the FtsZ protein, molecules of which form the Z RING (a mid-cell structure that develops during cell division).

fucosidosis An autosomal recessive disease involving a defective lysosomal enzyme, α-fucosidase, and the accumulation of glycoproteins.

(See also GENETIC DISEASE (table).)

furocoumarins See PSORALENS.

fusidic acid A steroidal antibiotic, synthesized by various fungi (e.g. some species of *Acremonium* and *Cylindrocarpon*), that is active against a range of Gram-positive bacteria, including staphylococci; it inhibits protein synthesis by preventing the

dissociation of the elongation factor EF-G from the ribosome following translocation.

Resistance to fusidic acid (in e.g. *Staphylococcus aureus*) can arise through mutations that affect EF-G. Such mutations in *S. aureus* are linked to a marked decrease in the biologic fitness of the cells – e.g. reduced growth in comparison with wild-type strains; this effect was reported to be compensated for by an additional mutation (S416F) in the gene for EF-G [Antimicrob Agents Chemother (2005) 49:1426–1431].

fusion inhibitors (anti-HIV-1) ANTIRETROVIRAL AGENTS that bind to the gp41 site on the HIV-1 virion and, hence, block viral infection of healthy cells. Fusion inhibitors include e.g. enfuvirtide, which is administered parenterally.

These agents have been used to suppress viral replication. (See also HIV-1 and RETROVIRUSES.)

fusion product Any product prepared by the GENE FUSION protocol, e.g. a FUSION PROTEIN (in which the expression vector encodes the protein of interest fused to a partner protein), or a tagged protein (in which the expression vector encodes the protein of interest fused to a *tag* – see EPITOPE TAGGING).

fusion protein (1) In some viruses (e.g. paramyxoviruses such as the Sendai virus): a protein that promotes fusion between the viral envelope and the membrane of target cells; fusion proteins can also promote fusion between host cells. (Another example is the GP64 protein of baculoviruses: see BACULOVIRIDAE.)

For an example of a fusion protein see the entry VOYAGER VECTORS.

(2) In recombinant DNA technology/genetic engineering: a protein consisting of amino acid residues specified by coding sequences of two different proteins – that is, a hybrid protein. A fusion protein can be prepared by the techniques of GENE FUSION (q.v.).

(See also ENZYME FRAGMENT COMPLEMENTATION.)

(cf. FUSION PRODUCT.)

fusion vector Any vector which is used for the synthesis of a fusion product, e.g. a FUSION PROTEIN (sense 2) – see GENE FUSION.

(Example of a fusion vector (using a FRET-based reporter system): pcDNA^{TM}6.2/GeneBLAzer^{TM}-DEST – see the table in GATEWAY SITE-SPECIFIC RECOMBINATION SYSTEM.)

fusogen See CELL FUSION.

G

G Glycine (alternative to Gly).

G-less cassette A sequence of DNA which, when transcribed, yields RNA containing no guanosine residues; such a G-less cassette is often used in *in vitro* transcription assays, and is inserted into DNA which is to be transcribed. In such assays, portions of the transcripts containing guanosine residues may be eliminated by treating the reaction mixture with RNASE T1 (q.v.).

A G-less cassette was used for studying transcriptional start sites in the yeast *Saccharomyces cerevisiae*. In this study, a 140-bp G-less cassette was fused immediately downstream of the *ADH1* –38 start site, and the transcripts (after treatment with RNase T1) were examined by electrophoresis to assess transcription from the –38 start site [Mol Cell Biol (2008) 28(11):3757–3766].

In a study on mutant strains with impaired *lacZ* expression, the efficiency of transcription *elongation* was determined by inserting a G-less cassette at each end of the DNA target sequence, the two cassettes being of different lengths; following transcription, the transcripts were examined – after RNase T1 treatment – to find the *ratio* of the downstream G-less RNA fragments (376-bp) to the upstream G-less RNA fragments (84-bp) [Mol Biol Cell (2008) 19(10):4310–4318].

A pair of G-less cassettes was also used in a genome-wide study of factors affecting transcription elongation and DNA repair [PLoS Genet (2009) 5(2):e1000364].

G loop A loop of unpaired ssDNA in the 'G region' of PHAGE MU DNA that can be observed when DNA from an induced lysogenic population of the phage is first denatured and then re-annealed; a G loop reflects a local region of non-homology due to inversion of the G region in a proportion of the cells.

G protein In the (ssRNA) vesiculostomatitis virus: an envelope protein which is used e.g. for PSEUDOTYPING other viruses.

G-quadruplex See QUADRUPLEX DNA.

G-quartet A square, planar group of four guanine molecules held together by hydrogen bonds. A *stack* of G-quartets, with their planes parallel, forms the *core* of QUADRUPLEX DNA – in which each corner of the stack is formed by a strand – or part of a strand – of DNA that contributes a row of guanine residues.

Note that some authors [Nucleic Acids Res (2006) 34(Database issue):D119–D124] have used G-quartet as synonymous with G-quadruplex (quadruplex DNA); these authors refer to a G-quartet as a G-tetrad.

G-tetrad See G-QUARTET (second paragraph).

G418 sulfate An antibiotic, related to gentamicin (see AMINOGLYCOSIDE ANTIBIOTIC), which is active against prokaryotic, yeast, plant and mammalian cells. Cells expressing the gene for an aminoglycoside phosphotransferase – e.g. a neomycin-resistance gene – exhibit resistance to G418 sulfate.

[Uses of G418 (examples): Genome Res (2008) 18(4):597–609; Genome Res (2008) 18(10):1670–1679; (at 400 μg/mL) Neoplasia (2009) 11(1):57–65; (400 μg/mL) J Immune Based Ther Vaccines (2009) 7:1.]

(See also HYGROMYCIN B.)

GadY A bacterial sRNA: see entry SRNAS. GadY is a regulator of acid-response genes in *Escherichia coli*; it is an atypical bacterial sRNA in that (i) it binds to the 3′-UTR (rather than the 5′-UTR), and (ii) it promotes (rather than inhibits) gene expression.

(See also GCVB.)

gag See RETROVIRUSES.

GAL1 (or Gal1) In *Saccharomyces cerevisiae*: the enzyme galactokinase. The *GAL1* gene promoter is widely used e.g. in yeast-based gene expression vectors.

GAL4 A eukaryotic transcription activator. In yeast (*Saccharomyces cerevisiae*), GAL4 is involved in the utilization of galactose; in this role it binds to the corresponding upstream activator sequence (UAS).

(See also entries CHROMATIN IMMUNOPRECIPITATION and PATHDETECT SYSTEMS.)

Galacton™ See CHEMILUMINESCENCE.

β-galactosidase An enzyme (EC 3.2.1.23), encoded by the *lacZ* gene in *Escherichia coli*, which cleaves lactose to glucose and galactose.

(See also LAC OPERON and ONPG.)

The gene for β-galactosidase is widely used as a REPORTER GENE. It is used in various indicator systems such as BLACK–WHITE SCREENING, blue–white screening (see PBLUESCRIPT) and ENZYME FRAGMENT COMPLEMENTATION.

(See also CHEMILUMINESCENCE, α-PEPTIDE, Q-TAG (sense 1) and X-GAL.)

gall A plant CANCER. (See also CROWN GALL.)

gamete A type of cell involved in sexual reproduction – e.g. an egg or a sperm. (cf. ZYGOTE.)

gamete intrafallopian transfer (GIFT) See the entry IVF.

γδ (gamma delta) *Syn.* IS1000.

ganciclovir A nucleoside analog which inhibits DNA synthesis after it has been phosphorylated intracellularly.

[Example of use (selecting for loss of the thymidine kinase (TK) marker): PLoS ONE (2009) 4(2):e4341.]

gap (1) (in dsDNA) A single-stranded region produced by loss of one or more nucleotides. (cf. NICK; see also AP SITE.)

(2) Following SHOTGUN SEQUENCING of a genome (or other large molecule of DNA): a region for which no cloned fragments are available.

gapped vector A linear VECTOR prepared e.g. by introducing a double-strand break in a ccc DNA vector.

Gateway® site-specific recombination system Various types of product (Invitrogen, Carlsbad CA, USA) that can be used e.g. to move a given sequence into, or between, vector molecules for expression or for directional cloning; the system is based on a SITE-SPECIFIC RECOMBINATION SYSTEM from the λ phage (see PHAGE LAMBDA) that involves ATT SITES.

The essential features of the system are as follows.

A sequence flanked by *att*B sites can replace a sequence

Gateway® DESTINATION VECTORS: examples of some of the wide range of vectors to which a gene or fragment of interest can be transferred from a Gateway® entry clone. Every destination vector contains *att*R sites which can be used for recombination with any sequence that is flanked by *att*L sites[a]

Vector	Size	Properties/notes
Cell-free protein expression systems		
pEXP3-DEST	4.6 kb	T7 promoter and terminator for high-level expression. N-terminal SIX-HISTIDINE TAG and LUMIO tag to assist purification and detection; cleavage site for efficient removal of tags. (See also entry CELL-FREE PROTEIN SYNTHESIS.)
pEXP4-DEST	4.4 kb	T7 promoter and terminator for high-level expression. C-terminal SIX-HISTIDINE TAG and LUMIO tag for purification and detection
Expression of proteins in Escherichia coli		
Champion™ pET104-DEST	7.6 kb	T7 promoter and terminator for strong expression of recombinant proteins. Transcription regulation via a LacO operator. N-terminal BIOEASE tag permitting IN VIVO BIOTINYLATION of fusion proteins
pBAD-DEST49	6.2 kb	*araBAD* promoter providing tight regulation of expression of toxic proteins. Improved solubility of protein product provided by N-terminal thioredoxin fusion partner. SIX-HISTIDINE TAG to facilitate detection/purification of product
Expression of proteins in Saccharomyces cerevisiae		
pYES-DEST52	7.8 kb	High-level expression of protein product provided by the *GAL1* promoter. Detection and purification of the product facilitated by a C-terminal SIX-HISTIDINE TAG
Expression of proteins in insect cells		
BaculoDirect™		Insertion of gene/fragment of interest generates a recombinant baculovirus genome. Any non-recombinant vectors are negatively selected by expression of the thymidine kinase gene in the presence of ganciclovir. The *lacZ* gene (excised during insertion of the gene/fragment) permits checking of viral stock purity. C-terminal or N-terminal SIX-HISTIDINE TAG for purification. One version of the vector includes a MELITTIN secretion signal to promote secretion of the expressed protein
pMT-DEST48	5.3 kb	Designed for expression in *Drosophila* S2 cells. Transcription from a metallothionein promoter can be induced e.g. with copper sulfate or with cadmium chloride. SIX-HISTIDINE TAG for purification
Expression of proteins (and shRNAs) in mammalian cells		
pAd/CMV/V5-DEST™	36.7 kb	The plasmid includes a replication-deficient part of the adenovirus genome (lacking sequences in the E1 and E3 regions). A population of the recombinant molecules, consisting of the gene of interest integrated in the plasmid, is linearized (by restriction enzyme PacI), leaving the ITRs (inverted terminal repeats) of the viral genome as the flanking sequences. This construct is used to transfect specialized cells (cell line 293A) within which replication-incompetent (but infectious) virus particles are produced; these particles can be used to infect *target* cells and, hence, to deliver the gene of interest. The CMV promoter provides a high level of expression
pBLOCK-iT™3-DEST	5.9 kb	A vector designed for long-term expression of shRNAs (see entry SHORT HAIRPIN RNA). Expression may be made either constitutive or inducible, depending on the type of entry vector used. Geneticin®-selectable (neomycin-resistance gene)
pBLOCK-iT™6-DEST	5.5 kb	Similar to pBLOCK-iT™3-DEST but smaller, and blasticidin-selectable
pcDNA™6/BioEase™-DEST	7.0 kb	High-level expression from the CMV promoter. Blasticidin-selectable. The BIOEASE tag facilitates purification of *in vivo* biotinylated recombinant proteins; the fusion tag can be excised via an adjacent enterokinase cleavage site

(*continued*)

Gateway® DESTINATION VECTORS (*continued*)

Vector	Size	Properties/notes
pcDNA^{TM}6.2/GeneBLAzer^{TM}-DEST		A fusion vector in which the gene of interest is fused (in N- or C-terminal orientation) to a mammalian-optimized β-lactamase gene (*bla*) which acts as a reporter gene. Reporter activity depends on the FRET (q.v.) principle. Cells are permeated with a FRET substrate (CCF2) – the molecule of which incorporates two fluorophores: coumarin and fluorescein. Cells in which *bla* is *not* expressed emit green light (520 nm) on excitation (409 nm) due to FRET-based interaction between coumarin and fluorescein. If *bla* is expressed, CCF2 is cleaved; excitation (at 409 nm) then causes simple (non-FRET-based) fluorescence from the coumarin moiety: blue light at 447 nm. The CMV promoter provides high-level expression of the fusion protein. Blasticidin-selectable. A related vector includes the option of topo cloning (see TOPOISOMERASE I CLONING) as well as the facility of a Gateway® destination vector; it also contains a CMV promoter and is blasticidin-selectable
pcDNA^{TM}6.2/nLumio^{TM}-DEST^{TM}	6.8 kb	High-level constitutive expression provided by the CMV promoter. The LUMIO tag facilitates detection and/or localization of the protein of interest either *in vivo* (in living cells) or *in vitro*
pEF5/FRT/V5-DEST^{TM}	7.5 kb	For inserting a gene of interest into cells of an FLP-IN CELL LINE. EF-1α promoter for use in those parental cells (e.g. BHK) in which the CMV promoter is not optimal. Bovine growth hormone (BGH) polyadenylation signal and transcription termination to promote stability of transcript. Hygromycin-resistance gene for selection of recombinant vectors
pLenti6/V5-DEST^{TM}	8.7 kb	A lentiviral expression vector incorporating the 5′-LTR and 3′-LTR sequences for insertion into the target cell's genome. The vector, carrying the gene of interest, is co-transfected into a producer cell line (293FT) together with plasmids encoding (i) the lentiviral packaging proteins, and (ii) the VSV-G envelope protein (the latter protein confers on a viral particle the ability to infect a broad range of types of mammalian cell). The recombinant (replication-incompetent) lentiviral particles, produced in the 293FT cells, can be used to infect *target* cells, within which the gene of interest is expressed. The CMV promoter normally provides high-level expression. The pLenti6/UbC/V5-DEST^{TM} vector contains the (human) ubiquitin C (UbC) promoter (instead of CMV) for efficient expression when CMV is inappropriate. A blasticidin-resistance gene is incorporated for selection of stable cell lines
pT-REx^{TM}-DEST31	7.5 kb	Strong, tetracycline-regulated CMV promoter, two copies of the $TetO_2$ operator being downstream of the promoter. Geneticin®-selectable. SIX-HISTIDINE TAG to facilitate purification
pVAX^{TM}200-DEST		A small vector designed specifically for use in the development of DNA VACCINES. In order to minimize the possibility of integration into the chromosome, the plasmid includes only the minimal amount of eukaryotic DNA necessary for expression. The polylinker permits insertion of DNA from PCR-generated fragments etc. as well as from Gateway® entry clones

[a]Descriptions of products are derived from data provided by courtesy of Invitrogen Corporation, Carlsbad CA, USA.

flanked by *att*P sites in a reaction mediated by BP CLONASE. When this occurs, the inserted sequence becomes flanked by hybrid (*att*B–*att*P) sites referred to as *att*L sites; the displaced sequence is also flanked by hybrid *att*B–*att*P sites, referred to as *att*R sites.

A sequence flanked by *att*R sites can be replaced with one flanked by *att*L sites; this results in reconstitution of the *att*B and *att*P sites – *att*B sites then flanking the inserted sequence and *att*P sites flanking the replaced sequence. This reaction is mediated by another recombinase: LR CLONASE.

Entry clones, entry vectors and destination vectors

Entry into the Gateway® system (for cloning or expression) can be achieved e.g. via PCR. Thus, the sequence of interest can be amplified with primers whose 5′ end incorporates an

*att*B site. The resulting amplicons, which are flanked on both sides by *att*B sites, can now be inserted *in vitro* into e.g. a pDONR™ vector. This vector has a pair of *att*P sites flanking the *ccdB* gene (see CCD MECHANISM). By inserting into the pDONR™ vector, the amplicon forms an *entry clone*; within the vector, the amplicon is flanked by *att*L sites. (During this reaction (carried out by BP Clonase) the *ccdB* gene is excised – so that *recombinant* vector molecules (i.e. those containing the sequence of interest) can be replicated in bacteria.) The vectors (which contain the amplicon/sequence of interest) can then be inserted, e.g. by transformation, into *Escherichia coli*; following growth, those cells containing recombinant vectors are selected on a medium containing kanamycin or Zeocin™ (pDONR™ vectors contain either a kanamycin-resistance or a ZEOCIN-resistance gene). Recombinant vectors can then be recovered.

The sequence of interest within the entry clone can now be transferred to any of a range of *destination vectors* (see the accompanying table) in order to carry out a given task; each of the destination vectors contains a pair of *att*R sites, so that the sequence of interest (flanked by *att*L sites) can be inserted into any desired destination vector. This reaction is mediated by LR Clonase.

Also available are *entry vectors* that contain a pair of *att*L sites; between these sites is either a POLYLINKER (which is used for insertion of a sequence by restriction and ligation) – or a site at which a sequence can be inserted by topo cloning (see TOPOISOMERASE I CLONING). A gene/fragment inserted into an entry vector can subsequently be transferred to one of the destination vectors, using the *att*L sites.

(***Note***. An *entry clone* differs from an *entry vector* in that, in the entry clone, the *att*L sites are *created by* the insertion of the gene/fragment into the pDONR™ vector; by contrast, the *att*L sites are already present in an entry vector, even before the gene or fragment is inserted.)

Examples of the use of the Gateway® system: production of synthetic human chemokine receptors [PLoS ONE (2009) 4(2):e4509] and a study on protein localization in asymmetric bacteria [Proc Natl Acad Sci USA (2009) 106:7858–7863].

(See also MULTISITE GATEWAY TECHNOLOGY and see TOPOISOMERASE I CLONING.)

gatifloxacin See QUINOLONE ANTIBIOTICS.

GC-profile See ISOCHORE.

GC type See BASE RATIO.

GC% (GC value; %GC; mol% G + C etc.) Of *bacterial* DNA: the ratio (G + C)/(A + T + G + C) in which A, T, G and C are the relative molar amounts of adenine, thymine, guanine and cytosine residues, respectively, in the DNA of a given organism; the ratio is expressed as a percentage. (cf. BASE RATIO.)

The value of GC% depends on genus and species. It is one of the criteria used in taxonomy, although similarity in GC%, in itself, is not an indicator of taxonomic affinity. In bacteria, GC% values range from ∼25 to ∼75; for example, the GC% of *Staphylococcus aureus* is ∼33, and that of *Mycobacterium tuberculosis* is ∼66.

GC% is useful as a descriptor e.g. for chromosomal DNA, for specific sequences of nucleotides in chromosomal DNA, and for plasmid DNA etc. In some pathogenic bacteria the genome includes certain sequences of nucleotides (so-called *pathogenicity islands*) in which the GC% differs from that of the remainder of the chromosomal DNA; this has been taken to indicate that such sequences of nucleotides were imported into the genome from an extracellular source.

A knowledge of the GC% of an organism's DNA may be useful for assessing the susceptibility of the DNA to a given RESTRICTION ENDONUCLEASE. For example, the restriction enzyme NotI (recognition site GC↓GGCCGC) lacks cutting sites in the genomic DNA of *Staphylococcus aureus* but has more than 250 cutting sites in *Mycobacterium tuberculosis*. This can be seen in one of the accompanying charts, chart B, which shows the cutting frequency of NotI in the genomic DNA of a range of bacteria and archaeans; this information can be found on the Internet (see the entry REBASE for details of access). The charts A and C show the cutting frequencies of the two restriction enzymes MseI and XmnI, respectively. The choice of restriction enzyme is important, for example, in RFLP ANALYSIS (q.v.).

An (indirect) indication of the GC% of a given sample of DNA may be obtained from its 'melting point' (see THERMAL MELTING PROFILE) owing to the linear relationship between T_m and GC% (a high T_m correlating with a high GC%). It is also possible to assess GC% by ultracentrifugation in a CsCl

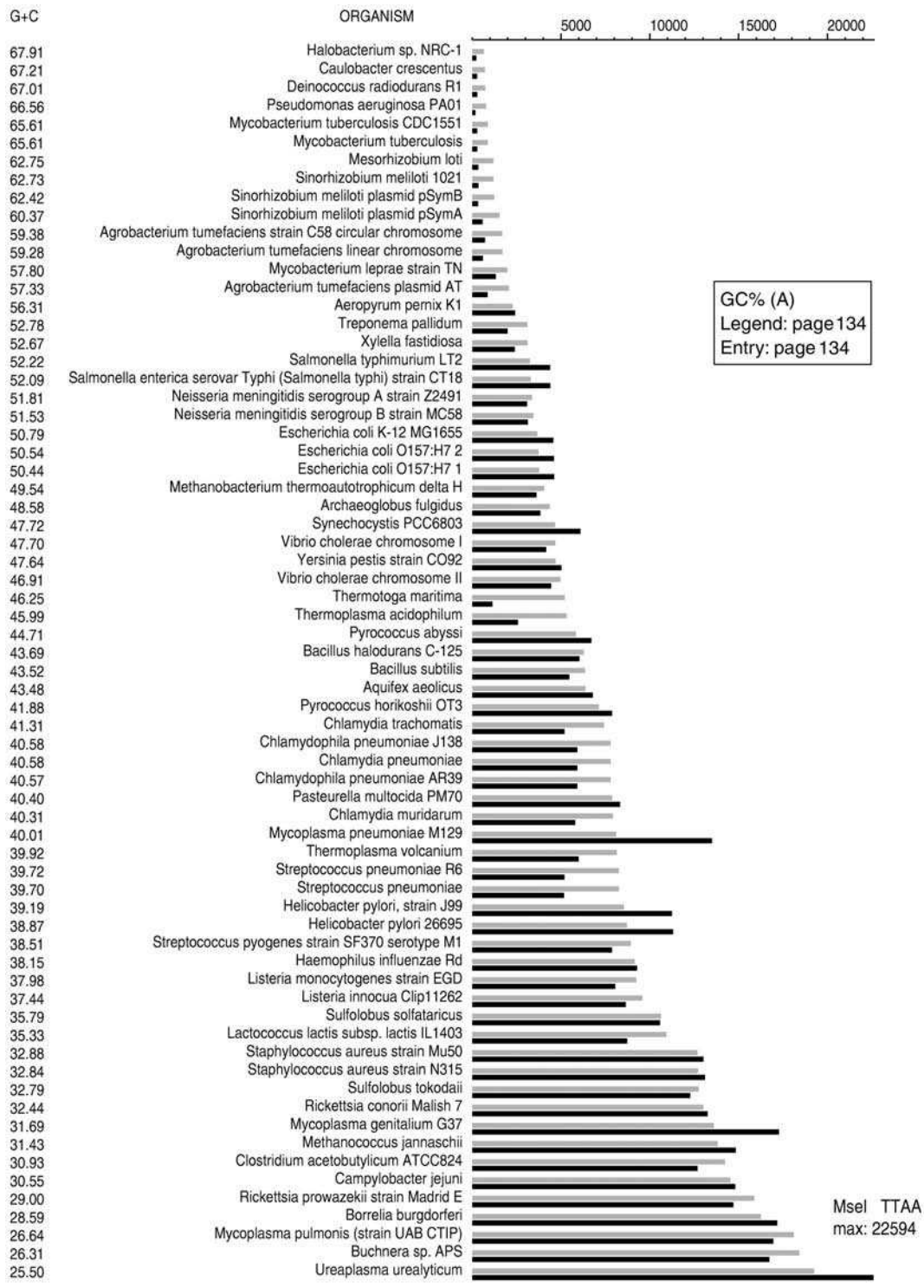
GC% (*opposite*) The effect of GC% on the cutting frequencies of three restriction endonucleases in the genomic DNA of various bacteria and archaeans.

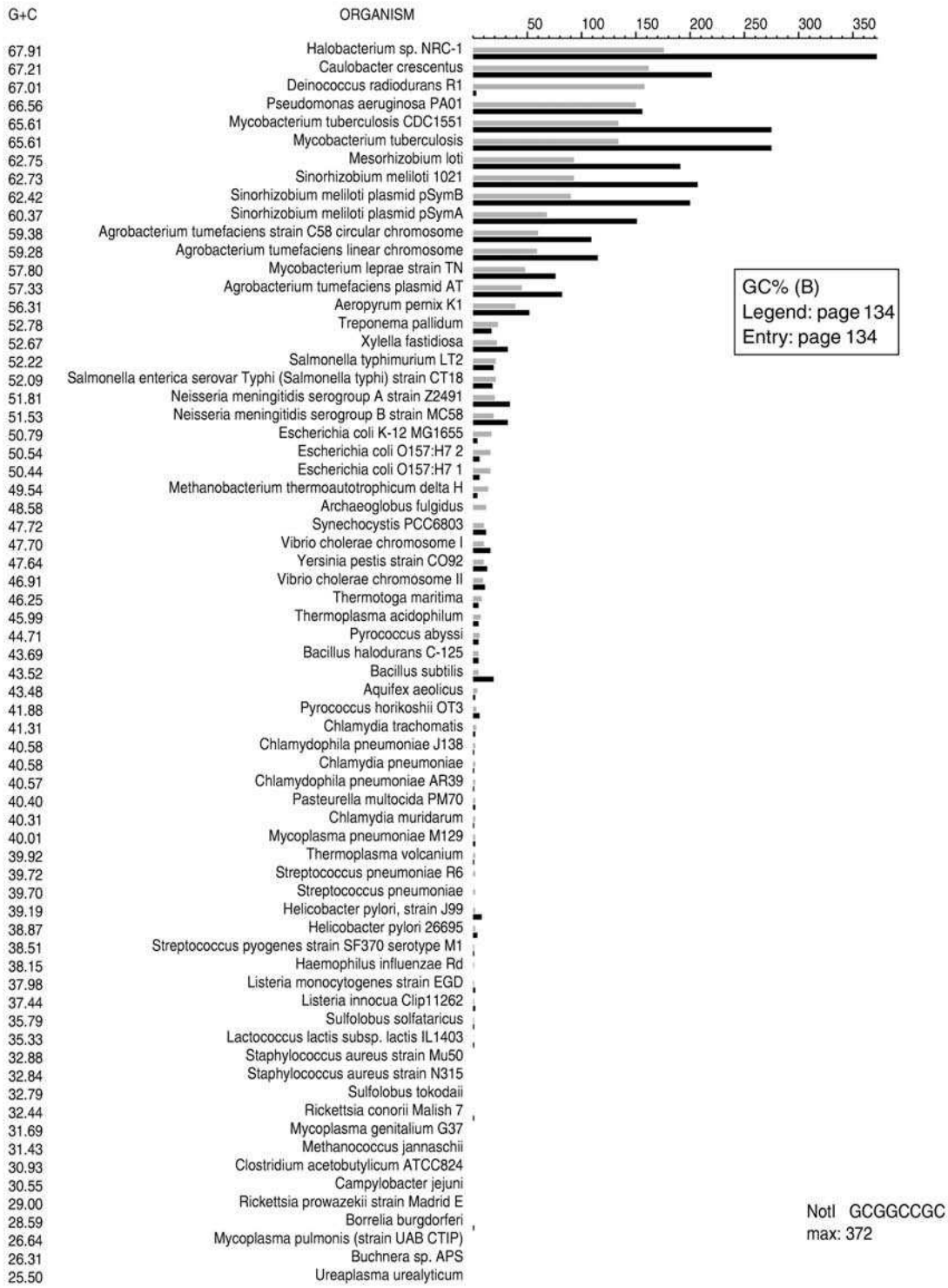
The organisms are arranged in order of the GC% of their genomic DNA – the highest value (67.91) at the top, the lowest (25.5) at the bottom.

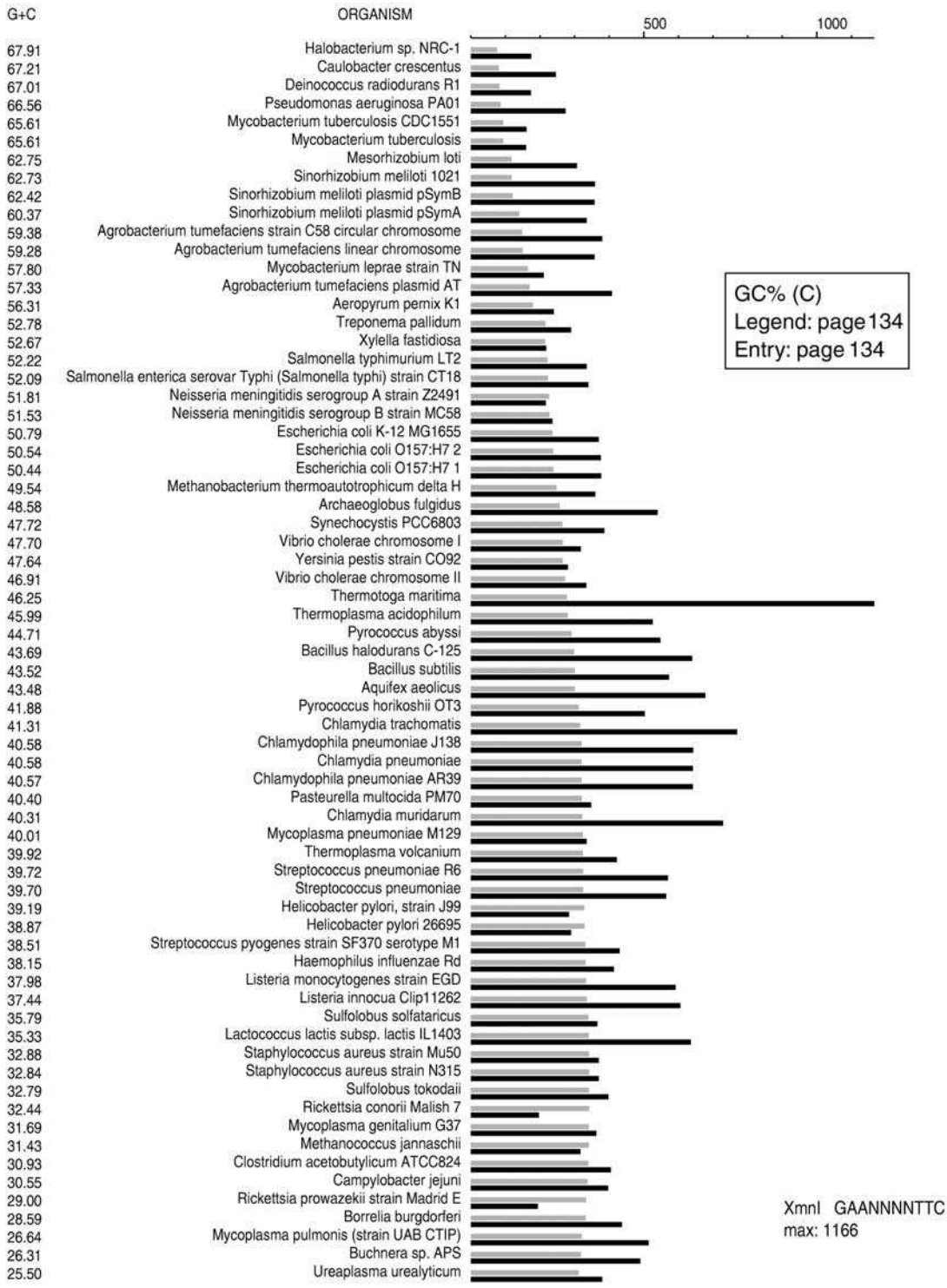
For a given organism, the solid black bar represents the observed number of recognition sites (see scale at top of chart) in the genomic DNA of that organism. The adjacent gray bar represents the expected number of cutting sites.

(A) Restriction endonuclease MseI. Cutting site: T↓TAA
(B) Restriction endonuclease NotI. Cutting site: GC↓GGCCGC
(C) Restriction endonuclease XmnI. Cutting site: GAANN↓NNTTC

Charts supplied by, and reproduced by courtesy of, Dr Janos Posfai and Tamas Vincze, New England Biolabs, Ipswich MA, USA.


G+C
ORGANISM
5000
10000
15000
20000
67.91 Halobacterium sp. NRC-1
67.21 Caulobacter crescentus
67.01 Deinococcus radiodurans R1
66.56 Pseudomonas aeruginosa PA01
65.61 Mycobacterium tuberculosis CDC1551
65.61 Mycobacterium tuberculosis
62.75 Mesorhizobium loti
62.73 Sinorhizobium meliloti 1021
62.42 Sinorhizobium meliloti plasmid pSymB
60.37 Sinorhizobium meliloti plasmid pSymA
59.38 Agrobacterium tumefaciens strain C58 circular chromosome
59.28 Agrobacterium tumefaciens linear chromosome
57.80 Mycobacterium leprae strain TN
57.33 Agrobacterium tumefaciens plasmid AT
56.31 Aeropyrum pernix K1
52.78 Treponema pallidum
52.67 Xylella fastidiosa
52.22 Salmonella typhimurium LT2
52.09 Salmonella enterica serovar Typhi (Salmonella typhi) strain CT18
51.81 Neisseria meningitidis serogroup A strain Z2491
51.53 Neisseria meningitidis serogroup B strain MC58
50.79 Escherichia coli K-12 MG1655
50.54 Escherichia coli O157:H7 2
50.44 Escherichia coli O157:H7 1
49.54 Methanobacterium thermoautotrophicum delta H
48.58 Archaeoglobus fulgidus
47.72 Synechocystis PCC6803
47.70 Vibrio cholerae chromosome I
47.64 Yersinia pestis strain CO92
46.91 Vibrio cholerae chromosome II
46.25 Thermotoga maritima
45.99 Thermoplasma acidophilum
44.71 Pyrococcus abyssi
43.69 Bacillus halodurans C-125
43.52 Bacillus subtilis
43.48 Aquifex aeolicus
41.88 Pyrococcus horikoshii OT3
41.31 Chlamydia trachomatis
40.58 Chlamydophila pneumoniae J138
40.58 Chlamydia pneumoniae
40.57 Chlamydophila pneumoniae AR39
40.40 Pasteurella multocida PM70
40.31 Chlamydia muridarum
40.01 Mycoplasma pneumoniae M129
39.92 Thermoplasma volcanium
39.72 Streptococcus pneumoniae R6
39.70 Streptococcus pneumoniae
39.19 Helicobacter pylori, strain J99
38.87 Helicobacter pylori 26695
38.51 Streptococcus pyogenes strain SF370 serotype M1
38.15 Haemophilus influenzae Rd
37.98 Listeria monocytogenes strain EGD
37.44 Listeria innocua Clip11262
35.79 Sulfolobus solfataricus
35.33 Lactococcus lactis subsp. lactis IL1403
32.88 Staphylococcus aureus strain Mu50
32.84 Staphylococcus aureus strain N315
32.79 Sulfolobus tokodaii
32.44 Rickettsia conorii Malish 7
31.69 Mycoplasma genitalium G37
31.43 Methanococcus jannaschii
30.93 Clostridium acetobutylicum ATCC824
30.55 Campylobacter jejuni
29.00 Rickettsia prowazekii strain Madrid E
28.59 Borrelia burgdorferi
26.64 Mycoplasma pulmonis (strain UAB CTIP)
26.31 Buchnera sp. APS
25.50 Ureaplasma urealyticum
GC% (A)
Legend: page 134
Entry: page 134
MseI TTAA
max: 22594


G+C
ORGANISM
50
100
150
200
250
300
350
67.91 Halobacterium sp. NRC-1
67.21 Caulobacter crescentus
67.01 Deinococcus radiodurans R1
66.56 Pseudomonas aeruginosa PA01
65.61 Mycobacterium tuberculosis CDC1551
65.61 Mycobacterium tuberculosis
62.75 Mesorhizobium loti
62.73 Sinorhizobium meliloti 1021
62.42 Sinorhizobium meliloti plasmid pSymB
60.37 Sinorhizobium meliloti plasmid pSymA
59.38 Agrobacterium tumefaciens strain C58 circular chromosome
59.28 Agrobacterium tumefaciens linear chromosome
57.80 Mycobacterium leprae strain TN
57.33 Agrobacterium tumefaciens plasmid AT
56.31 Aeropyrum pernix K1
52.78 Treponema pallidum
52.67 Xylella fastidiosa
52.22 Salmonella typhimurium LT2
52.09 Salmonella enterica serovar Typhi (Salmonella typhi) strain CT18
51.81 Neisseria meningitidis serogroup A strain Z2491
51.53 Neisseria meningitidis serogroup B strain MC58
50.79 Escherichia coli K-12 MG1655
50.54 Escherichia coli O157:H7 2
50.44 Escherichia coli O157:H7 1
49.54 Methanobacterium thermoautotrophicum delta H
48.58 Archaeoglobus fulgidus
47.72 Synechocystis PCC6803
47.70 Vibrio cholerae chromosome I
47.64 Yersinia pestis strain CO92
46.91 Vibrio cholerae chromosome II
46.25 Thermotoga maritima
45.99 Thermoplasma acidophilum
44.71 Pyrococcus abyssi
43.69 Bacillus halodurans C-125
43.52 Bacillus subtilis
43.48 Aquifex aeolicus
41.88 Pyrococcus horikoshii OT3
41.31 Chlamydia trachomatis
40.58 Chlamydophila pneumoniae J138
40.58 Chlamydia pneumoniae
40.57 Chlamydophila pneumoniae AR39
40.40 Pasteurella multocida PM70
40.31 Chlamydia muridarum
40.01 Mycoplasma pneumoniae M129
39.92 Thermoplasma volcanium
39.72 Streptococcus pneumoniae R6
39.70 Streptococcus pneumoniae
39.19 Helicobacter pylori, strain J99
38.87 Helicobacter pylori 26695
38.51 Streptococcus pyogenes strain SF370 serotype M1
38.15 Haemophilus influenzae Rd
37.98 Listeria monocytogenes strain EGD
37.44 Listeria innocua Clip11262
35.79 Sulfolobus solfataricus
35.33 Lactococcus lactis subsp. lactis IL1403
32.88 Staphylococcus aureus strain Mu50
32.84 Staphylococcus aureus strain N315
32.79 Sulfolobus tokodaii
32.44 Rickettsia conorii Malish 7
31.69 Mycoplasma genitalium G37
31.43 Methanococcus jannaschii
30.93 Clostridium acetobutylicum ATCC824
30.55 Campylobacter jejuni
29.00 Rickettsia prowazekii strain Madrid E
28.59 Borrelia burgdorferi
26.64 Mycoplasma pulmonis (strain UAB CTIP)
26.31 Buchnera sp. APS
25.50 Ureaplasma urealyticum
GC% (B)
Legend: page 134
Entry: page 134
NotI GCGGCCGC
max: 372


G+C
ORGANISM
500
1000
67.91 Halobacterium sp. NRC-1
67.21 Caulobacter crescentus
67.01 Deinococcus radiodurans R1
66.56 Pseudomonas aeruginosa PA01
65.61 Mycobacterium tuberculosis CDC1551
65.61 Mycobacterium tuberculosis
62.75 Mesorhizobium loti
62.73 Sinorhizobium meliloti 1021
62.42 Sinorhizobium meliloti plasmid pSymB
60.37 Sinorhizobium meliloti plasmid pSymA
59.38 Agrobacterium tumefaciens strain C58 circular chromosome
59.28 Agrobacterium tumefaciens linear chromosome
57.80 Mycobacterium leprae strain TN
57.33 Agrobacterium tumefaciens plasmid AT
56.31 Aeropyrum pernix K1
52.78 Treponema pallidum
52.67 Xylella fastidiosa
52.22 Salmonella typhimurium LT2
52.09 Salmonella enterica serovar Typhi (Salmonella typhi) strain CT18
51.81 Neisseria meningitidis serogroup A strain Z2491
51.53 Neisseria meningitidis serogroup B strain MC58
50.79 Escherichia coli K-12 MG1655
50.54 Escherichia coli O157:H7 2
50.44 Escherichia coli O157:H7 1
49.54 Methanobacterium thermoautotrophicum delta H
48.58 Archaeoglobus fulgidus
47.72 Synechocystis PCC6803
47.70 Vibrio cholerae chromosome I
47.64 Yersinia pestis strain CO92
46.91 Vibrio cholerae chromosome II
46.25 Thermotoga maritima
45.99 Thermoplasma acidophilum
44.71 Pyrococcus abyssi
43.69 Bacillus halodurans C-125
43.52 Bacillus subtilis
43.48 Aquifex aeolicus
41.88 Pyrococcus horikoshii OT3
41.31 Chlamydia trachomatis
40.58 Chlamydophila pneumoniae J138
40.58 Chlamydia pneumoniae
40.57 Chlamydophila pneumoniae AR39
40.40 Pasteurella multocida PM70
40.31 Chlamydia muridarum
40.01 Mycoplasma pneumoniae M129
39.92 Thermoplasma volcanium
39.72 Streptococcus pneumoniae R6
39.70 Streptococcus pneumoniae
39.19 Helicobacter pylori, strain J99
38.87 Helicobacter pylori 26695
38.51 Streptococcus pyogenes strain SF370 serotype M1
38.15 Haemophilus influenzae Rd
37.98 Listeria monocytogenes strain EGD
37.44 Listeria innocua Clip11262
35.79 Sulfolobus solfataricus
35.33 Lactococcus lactis subsp. lactis IL1403
32.88 Staphylococcus aureus strain Mu50
32.84 Staphylococcus aureus strain N315
32.79 Sulfolobus tokodaii
32.44 Rickettsia conorii Malish 7
31.69 Mycoplasma genitalium G37
31.43 Methanococcus jannaschii
30.93 Clostridium acetobutylicum ATCC824
30.55 Campylobacter jejuni
29.00 Rickettsia prowazekii strain Madrid E
28.59 Borrelia burgdorferi
26.64 Mycoplasma pulmonis (strain UAB CTIP)
26.31 Buchnera sp. APS
25.50 Ureaplasma urealyticum
GC% (C)
Legend: page 134
Entry: page 134
XmnI GAANNNNTTC
max: 1166

density gradient.

In any determination of genomic GC%, the presence of a plasmid (especially a high-copy-number plasmid) may affect the observed value because the GC% of the plasmid may not be similar to that of the chromosomal DNA. When isolating chromosomal DNA, to determine its GC%, it is necessary to discriminate against any extrachromosomal DNA.

GC% in eukaryotic/mammalian DNA

G + C content is also used to describe regions in eukaryotic/mammalian DNA. For example, human genomic DNA can be divided into >3000 regions called *isochores*, each isochore characterized by a uniform, or near-uniform, G + C content; the isochores in human genomic DNA fall into five groups on the basis of their G + C values. [Isochore mapping of the human genome: Genome Res (2006) 16(4):536–541.]

The DNA of CPG ISLANDS is typically characterized by a value higher than ~60%. This high value of G + C content is associated with susceptibility to certain types of 'rare-cutting' RESTRICTION ENDONUCLEASE; for example, more than 85% of the recognition sites of the restriction endonuclease NotI in the human genome are found in CpG islands.

G-CSF Granulocyte colony-stimulating factor. (A recombinant form is produced by DNA technology: see Neupogen® in the entry BIOPHARMACEUTICAL (table).)

GcvB A bacterial sRNA (see entry SRNAS) which was reported to be involved as a regulator of acid resistance in *Escherichia coli* [BMC Genomics (2009) 10:165].

(See also GADY.)

gDNA A designation sometimes used for *genomic* DNA.

gefitinib An anticancer agent which inhibits oncogenic (EGF receptor) TYROSINE KINASE activity.

(cf. IMATINIB.)

gel electrophoresis A form of ELECTROPHORESIS used e.g. for separating fragments of nucleic acid, of different sizes and/or conformations, in a gel medium. (cf. ISOTACHOPHORESIS.)

Following electrophoresis, the separated fragments may be stained *in situ* (i.e. within the gel); extracted from the gel (for ongoing study); or transferred to another medium (see e.g. SOUTHERN BLOTTING). In some cases, fragments are labeled (e.g. with a fluorophore) before electrophoresis; in such cases the label can be read from the bands of fragments that form in the gel during electrophoresis.

In methods such as NASBA and PCR, gel electrophoresis is used e.g. to check that amplicons of the *expected* length are produced in the reaction. To facilitate measurement of length, use can be made of a 'DNA ladder'; this consists of a set of DNA fragments, in a range of known sizes, which undergo electrophoresis in a lane parallel to the lane(s) in which test sample(s) are being examined. Following electrophoresis, the size of the fragment in a given 'experimental' band is assessed by comparison with those bands in the DNA ladder that have migrated at a similar rate.

For nucleic acids, the matrix used in gel electrophoresis is commonly AGAROSE or POLYACRYLAMIDE; agarose is used for separating large fragments, while polyacrylamide is used for the smaller fragments (see table).

(See also NOVEX GELS.)

GEL ELECTROPHORESIS: examples of concentrations of agarose and acrylamide in gels used for separating nucleic acid fragments of different sizes (from various sources)

Nucleic acid fragment (linear dsDNA, in bp)	Agarose gel (% agarose)	Polyacrylamide gel (% acrylamide)
20000–100000	0.3	
1000–30000	0.5	
800–12000	0.7	
500–10000	1.0	
400–7000	1.2	
200–3000	1.5	
100–1000		3.5
50–2000	2.0	
60–400		8.0
50–200		12.0
10–500	4.0	
5–100		20.0

gel retardation assay An alternative name for gel shift assay or electrophoretic mobility-shift assay: see the entry EMSA.

gel shift assay *Syn*. EMSA.

geldanamycin A benzoquinoid compound, produced by strains of *Streptomyces*, that inhibits heat-shock protein 90 (Hsp90) and which was reported to inhibit the expression of c-*myc* in murine lymphoblastoma cells.

gellan gum A bacterial product – a linear polymer containing residues of glucose, rhamnose and glucuronic acid – that has been recommended as a gelling agent (in place of AGAR) for bacterial growth media used in the context of PCR (because agar inhibits PCR) [J Med Microbiol (2001) 50:108–109].

A purified gellan gum is available commercially (Gelrite®; Kelco/Merck, San Diego CA, USA).

[Use (e.g.) for the growth of *Thermus thermophilus* (7 g/L): Genetics (2008) 180(1):17–25.]

Gellan gum is also used in media for growing plants, e.g. *Arabidopsis thaliana* [J Exp Bot (2008) 59(11):3069–3076; Plant Cell Physiol (2009) 50(1):58–66], and in a medium for the nematode *Caenorhabditis elegans* (1.5 g/L) [BMC Ecol (2009) 9:14].

Gelrite® See GELLAN GUM.

gemifloxacin See QUINOLONE ANTIBIOTICS.

GenBank® A major DATABASE forming part of the INTERNATIONAL NUCLEOTIDE SEQUENCE DATABASE COLLABORATION which includes nucleotide sequences for >300000 organisms.

[Organization of the database, retrieving data etc.: Nucleic Acids Res (2009) 37(Database issue):D26–D31.]

(See also gene ACCESSION NUMBER.)

genBlastA See BLAST.

gender determination assay An assay that is employed to

determine whether a given sample of DNA is from a female source or a male source: see e.g. DNA SEX DETERMINATION ASSAY and SRY-BASED ASSAY.

gene A specific sequence of nucleotides in DNA (in cells and in DNA viruses), or in RNA (in RNA viruses), encoding a given polypeptide – or encoding a molecule of RNA which has a biosynthetic or a control function.

(See also ALLELE.)

gene activation/regulation (*DNA technol.*) See CONDITIONAL GENE ACTIVATION/REGULATION.

gene amplification (*in vivo*) An increase in the COPY NUMBER of a given gene, or genes, in specific cells or tissues; it can be detected e.g. by COMPARATIVE GENOMIC HYBRIDIZATION or by MICROARRAY-based approaches.

Gene amplification is an important feature in the initiation and progression of various types of tumor, including breast cancer. For example, amplification and overexpression of the *HER2* (= *ERBB2*) gene (at location 17q12; see EPIDERMAL GROWTH FACTOR RECEPTOR FAMILY) is found in up to ~25–30% of human breast cancers and is associated with a poor prognosis. Other genes in the 17q12 region, including *GRB7* and *MLN64*, were reported to be co-amplified with *HER2*.

Amplification of *HER2* was reported to be associated with global genomic instability [BMC Cancer (2008) 8:297].

[Amplification in bacterial antibiotic resistance: Nature Rev Microbiol (2009) 7:578–588; and in fuel strains of *Saccharomyces*: Genome Res (2009) doi: 10.1101/gr.094276.109.]

gene cassette See CASSETTE.

gene cloning See CLONING.

gene conversion Conversion of one allele into another. Gene conversion can occur e.g. following hybridization between a strand from each of a pair of non-identical alleles; during the repair of this heteroduplex – in which *one* strand is used as template – the sequence of one strand is converted to a sequence which is complementary to the sequence of the other strand.

gene-delivery system Any system used for inserting gene(s), or other pieces of nucleic acid, into cells – often with the object of expression, mutagenesis, or gene silencing within the cells.

(See also VECTOR.)

Vehicles for gene-delivery include both circular molecules of nucleic acid (e.g. various PLASMIDS) and linear molecules (e.g. some viral genomes) into which the gene (or a fragment of DNA) is inserted prior to TRANSFECTION. (See also USER CLONING and BIGEASY LINEAR CLONING SYSTEM.)

Plasmids – and other molecules of naked nucleic acid – are often inserted into cells by TRANSFORMATION or by ELECTROPORATION. (See also DNA VACCINE and NUCLEOFECTION.)

Delivery of plasmid DNA to a strain of the Gram-negative bacterium *Pseudomonas putida* was reported with the use of 40 kHz ultrasound [Nucleic Acids Res (2007) 35(19):e129].

The intradermal transfection of plasmid DNA in mice was reported to be enhanced by a short (femtosecond) pulse with a low-power, near-infrared laser [J Biomed Sci (2009) 16(1): 36].

[Plasmid vector design/production: Gene Therapy (2009) 16:165–171.]

Viral vectors, in which the insert DNA is incorporated in a recombinant or modified viral genome, internalize this DNA when they infect cells.

Viral vectors include certain lentiviruses (a class of RETROVIRUSES), ADENOVIRUSES and adeno-associated viruses (see AAVS); these are viruses which can infect both dividing and non-dividing cells.

[Imaging *in vivo* of targeted radioactive (lentiviral) vectors: Gene Therapy (2009) 16:894–904.]

Commercial virus-based gene-delivery systems include the AAV HELPER-FREE SYSTEM, the ADEASY XL ADENOVIRAL VECTOR SYSTEM and the VIRAPORT RETROVIRAL GENE EXPRESSION SYSTEM.

Some viral vectors are modified so that they are targeted to *specific* type(s) of cell: see PSEUDOTYPING.

Bacteria are sometimes used as gene-delivery vehicles: see e.g. GENE THERAPY and INSERTIONAL MUTAGENESIS.

Other gene-delivery systems include e.g. the GENE GUN, LIPOFECTION, MAGNETOFECTION, MICROINJECTION and also SPINOCULATION.

(See also DENDRIMER, ENDO-PORTER, MELITTIN and PEI.)

Important constructs used as gene delivery systems include e.g. BACTERIAL ARTIFICIAL CHROMOSOMES, HUMAN ARTIFICIAL CHROMOSOMES and YEAST ARTIFICIAL CHROMOSOMES.

Nanoparticle-based gene delivery

Delivery of DNA to human lung fibroblasts (in culture) was achieved with nano-scale colloidal silica carrying collagen-bound viral vectors [BioTechniques (2007) 43(2):213–221].

An anti-metastasis gene, *NM23-H1*, was delivered on iron oxide nanoparticles by intravenous injection; this resulted in a significantly extended survival of mice in an experimental pulmonary metastasis model [Cancer Gene Therapy (2008) doi: 10.1038/cgt.2008.97].

Plasmid DNA has been used in polyethyleneimine nanogels [Cancer Gene Therapy (2009) doi: 10.1038/cgt.2009.11].

Cationic polybutyl cyanoacrylate nanoparticles are reported to be useful for delivering plasmid DNA into human hepatocellular carcinoma cells; these (positively charged) particles might protect DNA from degradation by nucleases [J Biomed Biotechnol (2009) doi: 10.1155/2009/149254].

(See also MAGNETOFECTION and NANOPARTICLES.)

Gene delivery to cells of the central nervous system (brain and spinal cord)

Gene delivery to cells of the brain and spinal cord has been carried out by methods such as cranial inoculation [e.g. Proc Natl Acad Sci USA (2006) 103(27):10373–10378]. However, it was found later that genes can be delivered to these cells, in mice, by the use of AAV (i.e. adeno-associated virus) vectors injected *intravenously* – a simpler approach which suggested the possibility of GENE THERAPY for some neurodegenerative diseases in humans [Nature Biotechnol (2008) 27:59–65].

Gene delivery in the context of gene therapy

Major requirements for effective GENE THERAPY include e.g.

a suitable vector, the means for efficient and specific delivery of that vector to appropriate type(s) of cell, and persistence of the therapeutic effect for an adequate period; also important is the need to avoid, or minimize, the risk of adverse effects of the therapy (cf. INSERTIONAL ONCOGENESIS).

In *cancer* gene therapy, one of the challenges is to ensure that an adequate level of therapy is directed – persistently – at a sufficiently large area of the tumor. In one approach, oncolytic adenoviruses, adsorbed onto silica implants, were used in an *extended release* scheme for treating advanced orthotopic gastric and pancreatic tumors. In mice with pancreatic tumors, this treatment doubled the survival, and silica-based delivery was also reported to retard the development of anti-adenovirus antibodies [Gene Therapy (2009) 16:103–110].

Efforts are being made to increase the activity against solid tumors by using *replicating* retroviral vectors (see the section on RCR vectors in the entry RETROVIRUSES).

In gene therapy there is a need to allow for the possibility of adverse effects. This may be achieved by a mechanism for *in situ* reversal of a gene's activity; for example, engineered *transcriptional switches* might be incorporated in the gene-delivery systems [Mol Cancer Ther (2008) 7(3):439–448].

gene drive system (drive system) See e.g. GMMS.

gene expression Conversion of the information encoded in a *gene* (i.e. a sequence of nucleotides) to a product or function. Expression, and the extent of expression, depend on various factors which include e.g. the operational ability and strength of the gene's promoter (and the presence of accessory factors required for expression) and permissive epigenetic conditions (such as the appropriate methylation/acetylation status of the nucleotide sequence/histones etc.) – as well as basic physical factors (such as temperature). Moreover, the expression of a given gene may depend on the expression, or silence, of other gene(s).

Maximization of gene expression is sometimes required in DNA technology: see e.g. GENE EXPRESSION (optimization).

(See also ALLELE-SPECIFIC GENE EXPRESSION.)

gene expression (optimization) Optimization of gene expression *in vitro* may be achieved in various ways, some of which are examined in the entries CODON HARMONIZATION, CODON OPTIMIZATION and OVEREXPRESSION.

gene expression profile The pattern of expression (i.e. active, non-active, level of activity) of gene(s) in a given tissue, cell or organism at a particular time under given conditions.

gene-for-gene concept (*plant pathol.*) An early concept – relating to interactions between plants and plant pathogens – in which e.g. the product(s) of specific *avirulence gene*(*s*) of a given pathogen interact with the product(s) of corresponding *resistance gene*(*s*) in one or more strains of the host plant – the outcome of infection (development of disease or resistance to disease) being determined by the interaction. The host plant may respond to an infection with a defense mechanism such as: (i) a *hypersensitivity reaction* (in which host cells die rapidly at the local site of infection); (ii) the production of PATHOGENESIS-RELATED PROTEINS; or (iii) the production of PHYTOALEXINS.

Interaction between the products of avirulence and resistance genes was reported e.g. in rice plants challenged with the pathogen causing rice blast disease, *Magnaporthe grisea* (*Pyricularia oryzae*) [EMBO J (2000) 19:4004–4014]. However, *Magnaporthe oryzae* (*sic*) appears to infect *Arabidopsis* by a distinct mechanism in which three fungal genes essential for disease-formation in rice had only limited roles in causing disease in *Arabidopsis* [Plant Physiol (2009) 149:474–486].

gene fusion In recombinant DNA technology/genetic engineering: any process in which coding sequences from two distinct genes are joined together, with their reading frames in phase, so that transcription from a single (upstream) promoter region produces an mRNA transcript that encodes a hybrid *fusion protein*. Some of the reasons for carrying out gene fusion are given in *Uses of gene fusion technology* (see later).

While the above refers to the fusion of two *genes*, a similar procedure involves the fusion of a gene with a sequence of nucleotides encoding a peptide *tag* (see EPITOPE TAGGING). The principle is the same: two sequences are fused together, with their reading frames in phase, and the FUSION PRODUCT (protein+tag) is transcribed from a single promoter as a single transcript. The tag may be fused either to the N-terminal or to the C-terminal of the protein.

Gene fusion can be carried out in various ways. Thus, e.g. it is possible to delete a sequence of nucleotides between two contiguous genes so as to provide a single, continuous coding sequence controlled by the first (i.e. upstream) gene, but this would be exceptional and is not a common approach.

More generally, gene fusion is achieved by engineering the join between the two sequences.

To carry out the fusion *in vitro*, one approach is to use an EXPRESSION VECTOR that incorporates the partner gene or the tag. In such cases, the junction between the partner gene (or tag) and the inserted gene has already been considered during the design of the expression vector. In some cases a gene may be inserted into an expression vector by the hybridization of sticky ends, and ligation; alternatively, it may be inserted by the use of a topoisomerase (see TOPOISOMERASE I CLONING).

In vivo (within living cells), fusion (as well as other tasks) can be carried out with techniques such as LAMBDA (λ) RED RECOMBINATION (see also RECOMBINEERING), and by using SITE-SPECIFIC RECOMBINATION systems (e.g. the CRE–LOXP SYSTEM).

Translational and transcriptional fusions

Fusions of the type described above are sometimes referred to as *translational fusions* because they give rise to a fusion protein whose synthesis requires both transcription and translation signals from the upstream fusion partner. By contrast, a *transcriptional fusion* is one in which, through engineering, the genes of interest are subjected to a common mode of regulation; in this case the genes themselves are not fused, and no fusion protein is formed.

'Translational fusion' is also used to refer to those instances

in which a single gene is inserted into a vector that supplies translation signals, as well as a promoter; the product of the expression of such a fusion is also called a 'fusion protein' – even though only one gene is involved.

'Transcriptional fusion' is also used to refer to instances in which a single gene is inserted into a vector and fuses with a promoter, promoter and gene forming a viable transcriptional unit.

Variation in the meanings given to the terms *translational fusion* and *transcriptional fusion*, by different authors, does suggest that these terms are best avoided in order to avoid the risk of confusion.

Uses of gene fusion technology

• To facilitate detection/isolation of a specific gene product. If a gene product is difficult to detect (or to assay), the gene may be fused with a partner gene (= *carrier gene*) encoding a readily detectable product. This partner gene might encode, for example, glutathione *S*-transferase (GST) – in which case the fusion protein (with a GST *affinity tail*) can be isolated by using AFFINITY CHROMATOGRAPHY (with glutathione in the stationary phase). A target protein from a thermophilic organism might be isolated e.g. by fusion with a temperature-stable lectin-based affinity tail [BioTechniques (2006) 41(3):327–332].

One method involves fusing the protein of interest to a *tag* (rather than a partner gene) which undergoes BIOTINylation *in vivo*; after cell lysis the biotinylated proteins can be readily detected/separated by a STREPTAVIDIN-based system. (See IN VIVO BIOTINYLATION.)

(See also FLAG and PESC VECTORS.)

• To determine the intracellular location of a given protein using a reporter system such as GFP (GREEN FLUORESCENT PROTEIN) as the fusion partner. Fusion with GFP permits the (intracellular) target protein to be detected by fluorescence microscopy. In one early study, the gene for FtsZ protein (see Z RING) was fused with that of GFP, enabling a remarkable photographic image of the Z ring of *Escherichia coli* to be obtained [Proc Natl Acad Sci USA (1996) 93:12998–13003].

The use of GFP (or other fluorescent protein partners such as MONOMERIC RED FLUORESCENT PROTEIN) places limitations on this technique in terms of excitation/emission wavelengths. In an alternative approach, the fusion protein consists of the protein of interest fused with O^6-alkylguanine-DNA alkyltransferase (AGT). The advantage here is that AGT can be labeled *in vivo*, i.e. within living cells, by certain types of fluorophore, derivatives of O^6-benzylguanine, of which many are cell-permeable; thus, an AGT-containing fusion protein that is expressed intracellularly can be labeled by incubating the cells in a medium that contains one of the cell-permeable fluorophores. (O^6-benzylguanine derivatives that are not cell-permeable may be internalized e.g. by microinjection.) Labeling of AGT fusion proteins with an O^6-benzylguanine derivative involves a covalent bond. The AGT approach permits flexibility because O^6-benzylguanine derivatives are available with a wide range of emission spectra (from 472 nm to 673 nm) [BioTechniques (2006) 41(2):167–175].

AGT is a normal (eukaryotic) DNA repair enzyme – which transfers alkyl groups from a guanine residue to a cysteine residue on the enzyme (thereby inactivating the enzyme). The (unwanted) labeling of *endogenous* molecules of AGT by the benzylguanine derivatives within the cells under test has been regarded as a disadvantage of the AGT technique – although such labeling appears not to occur to any significant degree in at least some mammalian cells. Moreover, use can be made of a mutant form of AGT (^{M}AGT); ^{M}AGT can be labeled in the presence of N^9-cyclopentyl-O^6-(4-bromothenyl)guanine, a compound that blocks labeling of wild-type AGT. Another advantage of ^{M}AGT is that its affinity for alkylated DNA is lower than that of the normal wild-type protein.

• To investigate protein–protein interactions *in vivo* (within living cells); thus, gene fusion is used to prepare the hybrid proteins involved in techniques such as the BACTERIAL TWO-HYBRID SYSTEM, the YEAST TWO-HYBRID SYSTEM, and the CYTOTRAP system.

• To prepare a protein for labeling (see PROTEIN LABELING).

• To increase the solubility of a target protein: see, for example, CHAMPION PET SUMO VECTOR.

• To obtain a gene product which is normally not highly expressed. The relevant (target) gene may be fused *downstream* of a highly expressed gene in an expression vector; given an appropriate choice, the partner gene may contribute increased solubility/stability to the fusion product.

• To modify the mobility of the target protein. For example, the fusion partner VP22 confers on a target protein the ability to translocate from a given mammalian cell to an adjacent cell in a cell culture (see VOYAGER VECTORS).

• To engineer the secretion of a given protein that normally remains intracellular. Fusion with a gene whose product is normally secreted may enable secretion of the target protein as part of the fusion product; this can be advantageous e.g. in a production process (in which secreted proteins are easier to isolate and purify).

• To study the regulation of gene expression. For example, if there are problems in monitoring the expression of a given gene (perhaps due to difficulty in assaying the gene product), it may be possible to monitor activity of the gene's *promoter* (rather than the whole gene). For this purpose, a promoter-less *reporter gene* (encoding a readily detectable product) can be fused downstream of the target gene's promoter; hence, in this way, promoter activity can be monitored via the product of the reporter gene.

Some examples of reporter genes used for studying gene expression include those encoding GFP, galactokinase and chloramphenicol acetyltransferase. Expression of gene *lacZ* (encoding β-galactosidase), in a *lacZ* fusion product, can be monitored on media containing X-GAL.

The use of reporter genes in this way generally works well. However, an early report [J Bacteriol (1994) 176:2128–2132]

indicated that certain reporter genes are able to influence the activity of certain promoters; this serves as a useful reminder that unpredictable interactions between components may take place in recombinant DNA systems.

• To facilitate ligation of a pair of polypeptides or proteins – see 'ligation tags' (relevant to *split intein*) in the entry INTEIN.

• To detect/identify specific organisms. In a novel approach, GFP was fused to the coat protein of a strain-specific phage and used for the rapid detection of the pathogenic bacterium *Escherichia coli* strain O157:H7.

• See also ENZYME FRAGMENT COMPLEMENTATION.

Cleavage of fusion proteins

It may be possible to split the product of a fusion gene into two separate portions corresponding to the two genes. Given an appropriate and accessible sequence of amino acids at the junction region, cleavage may be achieved by a site-specific protease (see e.g. GST in GENE FUSION VECTOR).

(See also CHAMPION PET SUMO VECTOR and TEV protease in the entry TEV.)

Cleavage may also be carried out with chemical agents, e.g. hydroxylamine (which cleaves between adjacent residues of glycine and asparagine), or cyanogen bromide (which cleaves at methionine residues); the problems with chemical cleavage include unwanted modification (by the agent) and cleavage at sites *within* the target protein.

gene fusion vector Any EXPRESSION VECTOR which includes a sequence encoding a *partner gene* – such as glutathione *S*-transferase (GST); the sequence from a target gene is inserted in the correct reading frame, adjacent to the partner gene, and both genes are jointly transcribed and translated as a single product (a so-called *fusion protein*). The fusion protein may be cleavable to separate products; for example, a GST partner has a recognition site for a site-specific protease close to the fusion boundary.

gene gun A device used e.g. for the intraepidermal inoculation of a DNA VACCINE. DNA may be coated onto μm-scale gold particles which are internalized via a pressure-based system. The device has also been used for delivering DNA that is not coated onto gold particles.

[Examples of use of gene gun: Genet Vaccines Ther (2009) 7:2; BMC Cell Biol (2009) 10:51; J Biomed Sci (2009) 16 (1):49; and (plasmid DNA coated on tungsten particles) Plant Methods (2009) 5:9.]

(See also BIOLISTIC METHOD.)

gene knock-in *Syn*. KNOCK-IN MUTATION.

gene knock-out *Syn*. KNOCK-OUT MUTATION.

gene knockdown See the entry KNOCKDOWN.

gene product (gp) The (final) product formed by the expression of a given gene: either a protein (formed via the mRNA intermediate) or an RNA molecule which is not translated – e.g. a molecule of rRNA or tRNA.

Specific gene products, i.e. products formed by the expression of specific genes, may be designated e.g. as gp32, gp120 etc., the numbers reflecting particular genes in a given organism.

gene shuffling See DNA SHUFFLING.

gene silencing Any (engineered or natural) process that blocks gene *expression* – either *in vitro* or *in vivo* – and which may involve *post*-transcriptional events (see (e.g.) MICRORNAS, MORPHOLINO ANTISENSE OLIGOMER, PTGS, RIBOSWITCH, RNA INTERFERENCE, SHORT HAIRPIN RNA, SRNAS and also the U1 ADAPTOR) or *pre*-transcriptional events (see e.g. under heading *Transcriptional gene silencing* in the entry SIRNA).

(cf. GENE TARGETING and KNOCKDOWN.)

In vivo, METHYLATION (q.v.) of DNA is involved in gene silencing in contexts such as GENETIC IMPRINTING and some cases of tumor development (in which methylation blocks the transcription of tumor-suppressor genes). The methylation of DNA sequences may promote recruitment of specific binding proteins which – either alone or as part of a complex – recruit other agents, such as histone deacetylases, leading to a local region of repressed (inactive) chromatin.

gene subtraction A phrase that has been employed e.g. to refer to the use of antisense RNA – synthesized *in vivo* (i.e. within cells) – to inhibit the expression of specific gene(s).

The phrase is also used to refer to SUBTRACTIVE HYBRIDIZATION.

gene targeting Any procedure which is used to make change(s) in the nucleotide sequence of specific gene(s), or to inactivate those gene(s), or to modify the *expression* of those gene(s) – *in vivo* (i.e. within living cells).

Note. Specific gene(s) may be targeted simply to block their *expression* – rather than to cause any permanent change in a gene: see GENE SILENCING.

Some of the available methods can be used either for gene modification or for the complete inactivation of a gene.

To modify or inactivate a gene – e.g. by replacing part(s) of the coding sequence and/or a regulatory sequence – methods based on HOMOLOGOUS RECOMBINATION may be used. This has typically involved transfecting cells with a vector, such as a plasmid, containing an insert (and selectable marker) that is flanked by sequences homologous to specific sequences in the targeted genomic DNA; replacement of a given sequence of genomic DNA is then dependent on the (relatively rare) occurrence of homologous recombination between sequences in the vector and those in the genome.

The efficiency of gene-targeting has been improved by the development of certain phage-encoded recombinase systems (LAMBDA (λ) RED RECOMBINATION and ET RECOMBINATION – see also RECOMBINEERING). Manipulation of nucleic acids *in vivo* (within cells) has also been facilitated by certain site-specific recombination systems (e.g. CRE–LOXP SYSTEM and FLP).

In GENE THERAPY, gene targeting may be used to attempt to correct a particular malfunctioning gene.

For the complete inactivation of a gene (i.e. a gene KNOCK-OUT MUTATION) it is possible to use e.g. methods such as the lambda (λ) Red system; methods involving the TARGETRON approach; methods involving ZINC-FINGER NUCLEASES; and other techniques such as the MIRAGE procedure.

gene therapy Any form of therapy that involves genetic modification of (particular) cells to correct a GENETIC DISEASE, to treat certain cancers (see e.g. HEPATOCELLULAR CARCINOMA SUPPRESSOR 1 and MELANOMA), or to fight certain infectious diseases; in this context, 'genetic modification' refers to any change(s) in specific gene(s) and/or changes in gene expression.

Currently, much of the work in this area is conducted on an experimental basis and includes (i) studies on the methods for introducing DNA into cells (cells in culture, and cells within live animals), and (ii) the use of animal (e.g. mouse) models for investigating the functionality, control and persistence of transgenes.

Public awareness of gene therapy has often focused mainly on the treatment of inherited disorders (e.g. CYSTIC FIBROSIS and ADENOSINE DEAMINASE DEFICIENCY) but much work is also being carried out on cancer – and on infectious diseases, e.g. those caused by hepatitis, influenza and human immunodeficiency viruses, and by *Mycobacterium tuberculosis*.

Two basic approaches in gene therapy are: (i) the *ex vivo* mode, in which cells are removed from the body, genetically modified *in vitro*, and then returned to the body; (ii) the *in vivo* mode, in which genetic modification is carried out *in situ*, the modifying agent being internalized.

Inherited disease, involving a single malfunctioning gene, may be remediable by installing the normal, i.e. functional, wild-type, gene in order to counteract the specific deficiency. For example, successful therapy for ADENOSINE DEAMINASE DEFICIENCY has been achieved by inserting the normal gene into STEM CELLS (q.v.) and introducing these cells into the patient; in this case, adenosine deaminase is formed by the T cells which derive from those stem cells.

Stem cells suitable for gene therapy may be isolated from donor samples by exploiting the $CD34^+$ marker in IMMUNOMAGNETIC ENRICHMENT methods. After appropriate *in vitro* modification, one technique selected the recombinant cells by means of fluorescent reporter genes, recombinant cells being detected and separated by FLUORESCENCE-ACTIVATED CELL SORTER.

Vectors in gene therapy

1. Viral vectors

Vectors (i.e. vehicles used for inserting genetic material into cells) include certain types of virus which have been made replication-deficient and from which pathogenic potential has been eliminated. A gene which is to be inserted into cells (a *transgene*) is first inserted into a modified viral genome and subsequently enters the cells during viral infection.

In some cases replication-*competent* virus vectors are used: see the section on RCR vectors in the entry RETROVIRUSES).

Viral vectors include e.g. lentiviruses (see LENTIVIRINAE), ADENOVIRUSES and AAVS; these viruses can infect dividing and non-dividing cells. [Imaging targeted lentiviral vectors *in vivo*: Gene Therapy (2009) 16:894–904.]

Gene therapy involving retroviral vectors has lead to cases of INSERTIONAL ONCOGENESIS. Reasons have been advanced for the use of avian sarcoma and leukosis virus-based vectors (ASLV vectors) in preference to other retroviral vectors, and a method has been suggested for preventing the methylation-directed silencing of these vectors when they integrate into the genome [J Virol (2008) 82(16):7818–7827].

An engineered adenovirus was used as a *cancer terminator virus* against human melanoma cells in a mouse model (see the entry MELANOMA).

Silica implants have been used for the extended release of oncolytic adenoviruses in anti-cancer gene therapy: see entry GENE-DELIVERY SYSTEM.

(See also ADEASY XL ADENOVIRAL VECTOR SYSTEM.)

AAVs are able to carry relatively small sequences (inserts), while adenoviruses can carry larger sequences.

Transfection of liver cells by AAVs may be enhanced by the use of a small regulatory sequence from chromosome 19 – see the entry AAVS.

Some viral vectors infect many types of cell. This can be a disadvantage because, to deliver a transgene to a *specific* type of cell, it may be necessary to administer the vector in high doses in order to allow for wastage (i.e. the take-up by other types of cell); the use of high doses of the given vector may increase the risk of adverse reactions. For other viral vectors the range of host cells may be too narrow. In either case, the problem may be solved by PSEUDOTYPING (q.v.).

Certain modified *phages* (see BACTERIOPHAGE) have been reported to act as vectors for gene delivery to mammalian cells – and a similar role has been reported for PHAGEMID particles. To fulfill this role, the coat proteins of phages or phagemid particles are modified to display certain targeting ligands which can facilitate uptake by eukaryotic cells. In the case of phagemid particles, coat proteins can be encoded by the *helper* phage so that more of the phagemid's sequence can be used for the gene of interest; phagemid particles with a specific ligand have been prepared by using a coat-modified strain of phage M13 as helper phage. For phagemid vectors, the reported low levels of transgene expression in recipient cells may reflect factors such as the requirement to convert ssDNA to dsDNA and the translocation of nucleic acid into the nucleus.

A phage displaying multivalent anti-CD30 scFv on its coat proteins was reported to mediate transfer of a eukaryotic gene cassette to Hodgkin cell lines with transient expression of the gene [BMC Mol Biol (2008) 9:37].

In cystic fibrosis, gene therapy is aimed at epithelial cells of the respiratory tract. Thus far, benefit has been achieved for only a limited period of time. Because extended benefit needs repeated dosing, this tends to exclude the use of viral vectors because repeat dosing with viral vectors may compromise the treatment as a result of a developing immune response to the vector. However, a non-primate (caprine) AAV (see CYSTIC FIBROSIS) may be useful because it shows high resistance to neutralization by human antibodies and has a marked tropism for (murine) lung tissue.

Virus-mediated transfer of genes into cells may be facilitat-

ed by certain peptides or chemicals: see *Chemical facilitators* (below).

2. Plasmid vectors

See e.g. DNA VACCINE and GENE-DELIVERY SYSTEM.

3. Bacterial vectors

Bacterial vectors are used much less often than viral vectors. However, bacteria transfer gene(s) to mammalian cells when the bacteria are internalized – lysis of the bacteria (within the mammalian cells) releasing the transgene(s) for transcription and translation [Nature Biotechnol (1998) 16(9):862–866].

Lactococcus lactis can deliver DNA to mammalian cells [Appl Environ Microbiol (2006) 72(11):7091–7097; Genet Vaccines Therapy (2009) 7:4], and *Escherichia coli* has been used e.g. to deliver siRNA-encoding DNA to HEK (human embryonic kidney) cells in culture [Nucleic Acids Res (2009) 37(1):e8].

4. Chemical facilitators, and other gene-delivery systems

Virus-mediated transfer of genes into cells may be enhanced by certain peptides, and also by certain types of polycationic compound. From *in vitro* studies on the effect of peptides and polycationic agents on the transfer of genes from lentiviral and adenoviral vectors into human tumor cells, it was concluded that protamine sulfate may be the most suitable agent for clinical use [BioTechniques (2006) 40(5):573–576].

For cystic fibrosis, attention has been focused on increasing the effectiveness of *liposome*-mediated delivery of the transgene, e.g. by facilitating transgene internalization, and escape from the endosome, as well as import into the nucleus.

See the entry GENE-DELIVERY SYSTEM for other methods of gene delivery – including nanoparticle-based systems.

(See also INTRAMER.)

[Biological gene-delivery vehicles (a review): Mol Therapy (2009) doi: 10.1038/mt.2009.41.]

Gene therapy for cells of the central nervous system (brain and spinal cord)

Until quite recently it was considered problematic to deliver transgenes across the blood–brain barrier; for example, AAV vectors were introduced into the brain cells of mice by using direct *intracranial inoculation* to study GM2 gangliosidoses [Proc Natl Acad Sci USA (2006) 103:10373–10378]. However, a simpler procedure, the *intravenous injection* of AAV vectors, has been found to be effective for delivering genes to brain and spinal cord (in mice), and it was suggested that this approach may offer the possibility of gene therapy for certain human neurodegenerative diseases [Nature Biotechnol (2008) 27:59–65].

Gene therapy in orthopedics

[See: Orthopedic gene therapy in 2008: Mol Therapy (2008) doi: 10.1038/mt.2008.265.]

Gene therapy against infectious diseases

With infectious diseases, the possibilities for intervention by genetic methods are greater – owing to the nature of infection and the diversity of infectious agents. For example, in some insect-borne diseases it may be possible to prevent the initial (human) infection (or to reduce the incidence of infection) by genetically modifying the insect vector population so that it is less able to carry (and transmit) the pathogen (see below).

The fight against mosquito-borne diseases (such as malaria) involves not only the search for an effective prophylaxis (e.g. a vaccine) for the *human* victim of disease – but also attempts to carry out genetic modification of *mosquito* vectors in order to control the spread of infection; for this purpose, efforts are being made to produce *genetically modified mosquitoes* that are unable to function as carriers for the malaria parasite (see the entry GMMS).

In general, however, the approach is to consider the use of gene therapy in the (human) host after infection has occurred. Some of the approaches that were, or are being, investigated include:

- Antisense oligonucleotides which bind to the pathogen's mRNAs, blocking transcription (see e.g. VITRAVENE).
- Suicide genes whose expression – leading to cell death – occurs only in *infected* cells.
- RIBOZYMES targeting specific sequence(s) in infected cells. A single vector can encode many different types of ribozyme. The ribozyme's double requirement for correct binding and cleavage sites offers good specificity. The (enzymic) activity of the ribozyme allows ongoing destruction of targets. To be useful, though, a ribozyme must be stabilized against intracellular degradation by RNases. Useful results were obtained, for example, in a study that used a simian immunodeficiency virus (SIV) model with an anti-SIV *pol* ribozyme.

[Catalytic nucleic acid enzymes for study and development of therapies in the CNS (review): Gene Ther Mol Biol (2005) 9A:89–106.]

- RNA INTERFERENCE (q.v.) *in vitro* can inhibit replication of e.g. HIV-1 and the viruses that cause hepatitis, influenza and poliomyelitis. Moreover, siRNA-mediated downregulation of a cell's co-receptor for HIV can block the initial *infection* of cells by HIV.

Synthetic siRNAs can be introduced into mammalian cells by appropriate transfection methods. In a different approach, cassettes *encoding* siRNAs – for intracellular expression – can be transfected by a suitable gene transfer vector.

(See also SHORT HAIRPIN RNA.)

Delivery of siRNA to mammalian cells has been carried out with an siRNA conjugated to an APTAMER via a streptavidin bridge (the aptamer having been previously demonstrated to bind to prostate tumor cells); this conjugate (i.e. the siRNA + aptamer) was internalized within 30 minutes of its addition to cells, and the siRNA produced an efficient inhibition of gene expression [Nucleic Acids Res (2006) 34(10):e73].

Some problems with RNAi in gene therapy

(a) Mutation in the target sequences can permit escape from RNAi – a consequence of the high-level specificity of RNAi cutting. Moreover, although RNAi may be effective against highly conserved sequences, it may be inhibited by mutations in the central, target-cleavage region.

Knockdown of RNAi-escape strains of HIV-1 (with one or two mutations) has been reported with miRNAs and shRNAs

[Nucleic Acids Res (2009) doi: 10.1093/nar/gkp644].

(b) Non-specific ('off-target') modification of gene expression can occur – particularly under high-dose conditions.

(c) There is a need for a highly efficient delivery system so as to avoid a residue of unprotected cells in which target viruses can replicate and mutate to RNAi-insensitivity. [The progress of RNAi therapeutics, and novel synthetic materials for the encapsulation and delivery of nucleic acids: Nature Rev Drug Disc (2009) 8:129–138.]

• Intrabodies (see entry INTRABODY). Intrabodies were used successfully e.g. against vIL-6 – a form of IL-6 (interleukin-6) encoded by human herpesvirus-8 (HHV-8) and linked to pathogenesis in e.g. Kaposi's sarcoma. (Subsequently, translation of vIL-6 was inhibited by the use of morpholino antisense oligomers [Mol Cancer Ther (2008) 7(3):712–720].)

• Transdominant negative proteins (TNPs): modified, inhibitory forms of a pathogen's normal (regulatory or structural) proteins. TNPs may block activity of wild-type proteins e.g. by forming inactive complexes or by competing for specific binding sites.

Harmful immune reactions against cells containing TNPs must be avoided.

A transdominant negative version of the human immunodeficiency virus-1 (HIV-1) regulatory REV PROTEIN (which e.g. mediates the export of certain viral transcripts from the nucleus to the cytoplasm) was found to inhibit replication of HIV-1 *in vitro*.

• DNA VACCINES have shown promise in animal models. In some early studies a DNA vaccine encoding the 65-kDa heat-shock protein of *Mycobacterium tuberculosis* (Hsp65) was found to act both therapeutically and prophylactically in mice – while a plasmid encoding interleukin-12 (IL-12) gave even better results.

DNA vaccines are being developed against e.g. papillomavirus type 16 (a virus associated with human cervical cancer); protozoa *Leishmania donovani* and *Plasmodium falciparum*; and the bacterium *Mycobacterium tuberculosis*.

gene transfer agent See CAPSDUCTION.

gene-trap vector See GENE TRAPPING.

gene trapping A procedure which is used e.g. for introducing random insertional mutations into *cultured* mouse embryonic stem cells (ES cells); the inserts used for mutagenesis include a selectable marker or reporter.

The insertional mutations are introduced by transfecting ES cells with a *gene-trap vector* that integrates randomly in the genome. Two common types of vector are used.

One type of gene-trap vector (used in 'promoter trapping') is a promoter-less selectable sequence/reporter gene flanked on one side (upstream) by a splice acceptor and on the other side by a polyadenylation signal. If this vector integrates into a gene in such a way that the host gene's promoter function can be used, the selectable marker or reporter gene may be expressed; the disrupted gene is likely to be non-functional. Cells that express the marker or reporter can be selected, and the site of the vector's insertion can be determined. (If a gene in the target cell is transcriptionally silent it will not express the marker/reporter in this type of gene-trap vector.)

The second type of gene-trap vector consists of a promoter-competent selectable sequence which lacks a polyadenylation signal but has a splice donor site. Chance integration of this vector in the genome, at a site upstream of a functional polyadenylation sequence, may permit expression of the reporter system. This type of gene-trap vector, compared with the first type, has the advantage that it does not depend on expression of the disrupted gene.

mRNA formed by transcription of either type of integrated vector can be used to determine the site of integration of the vector (and the identity of the disrupted gene).

ES cells which have been genetically modified in this way can be injected into an (isolated) BLASTOCYST which is then reimplanted in a mouse and allowed to undergo development. This procedure is a useful method for studying gene function, including studies on genetic modification of the germ line.

[Enhanced gene trapping in mouse ES cells: Nucleic Acids Res (2008) 36(20:e133.]

[Expression-independent gene-trap vectors for random and targeted mutagenesis in embryonic stem cells: Nucleic Acids Res (2009) doi: 10.1093/nar/gkp640.]

(See also TARGETED TRAPPING.)

GeneChip® array See MICROARRAY.

GeneChip® Exon 1.0ST array See EXON ARRAY.

GeneHogs® Electrocompetent cells: see ELECTROPORATION.

GeneMorph™ PCR mutagenesis kit A product (Stratagene, La Jolla CA) designed to generate random mutations during PCR-based amplification; the product includes an error-prone DNA polymerase (Mutazyme™) that is reported to generate from one to seven mutant bases per kilobase in a given PCR reaction.

Examples of use: generation of random mutations in studies on the laccase gene in a *Bacillus* sp [BMC Biotechnol (2009) 9:12]; and random mutagenesis in studies on optimization of activity of a homing endonuclease [Nucleic Acids Res (2009) 37(3):877–890].

(See also MUTAGENESIS.)

generalized transduction See TRANSDUCTION.

GeneTailor™ site-directed mutagenesis system A product (Invitrogen, Carlsbad CA, USA) which is used e.g. to introduce specific point mutations, deletions, or insertions of up to 21 bp in length. Circular, or circularized, DNA of up to 8 kb, from any bacterial source, can be mutagenized.

(See also SITE-DIRECTED MUTAGENESIS.)

Initially, methylation of the (circular) molecule of DNA is carried out with a DNA methylase and *S*-adenosylmethionine as methyl donor.

A point mutation can be introduced by synthesizing copies of the DNA with a pair of primers which bind to overlapping sites, one of the primers incorporating the required mutation. Progeny strands anneal to form (nicked) circular molecules which are inserted into an appropriate strain of *Escherichia coli*, by transformation, for replication. The *mcrBC*$^{+}$ strain of

E. coli digests the (methylated) parental DNA (see the entry MCRBC).

Examples of use: making (promoter) site-specific deletion variants [PLoS Genet (2009) 5(1):e1000334]; and studies on the cystathionine β-synthase gene [J Pediatr (2009) 154(3): 431–437].

genetic code The form of the information which enables a sequence of nucleotides in DNA (or, in some viruses, RNA) to direct the synthesis of a polypeptide – such direction usually involving the relay of information to an intermediate effector molecule: *messenger RNA* (mRNA); consecutive triplets of nucleotides in the mRNA (see CODON) encode consecutive AMINO ACID residues in the polypeptide, while certain of the codons (see the table) specify the termination of synthesis.

(In certain viruses with a *positive*-sense ssRNA genome the genomic RNA itself is translated into polyproteins – without the need for relaying information to an intermediate effector molecule.)

(Compare synthesis of an oligopeptide by NON-RIBOSOMAL PEPTIDE SYNTHETASE.)

In that a given amino acid may be specified by more than one codon (see e.g. Leu in the table), the genetic code is said to be *degenerate*.

A given amino acid may be specified preferentially by a certain codon in one species but by another codon in different species. (See also CODON BIAS.)

Although widely applicable, the code shown in the table – often called the 'universal' genetic code – is not invariable. In the mitochondria of at least some plants, animals and fungi the codon UGA specifies tryptophan (rather than acting as a stop codon), while codons UAA and UAG specify glutamine in ciliate macronuclei.

Among the arthropods, AGG is reported to specify serine in certain species, and either lysine or arginine in others – with some species not using the codon at all [PLoS Biol (2006) 4 (5):e127; PLoS Biol (2006) 4(5):e175].

Dual-function codons

An unusual example of a dual-function codon was reported in the ciliate *Euplotes crassus*, in which codon UGA can specify *either* selenocysteine *or* cysteine; which of these amino acids is inserted at a particular UGA codon in the mRNA depends on (i) a specific sequence in the 3′-UTR – called the *selenocysteine insertion sequence* (see SECIS) – and (ii) the distance from the given UGA codon to the SECIS element. Such a differential use of the UGA codon was found even within the same gene, and the ability of the *Euplotes* mRNA to exhibit this dual function for UGA was maintained when expressed in mammalian cells [Science (2009) 323(5911):259–261].

[See also the interesting and informative brief review of the subject *The genetic code has a 'shift' key*: Heredity (2009) doi: 10.1038/hdy.2009.40.]

Coexistence with a nucleosome-positioning code

In eukaryotes, the DNA may incorporate an additional code – based on the so-called A-tracts (poly(A) sequences) – which signals the location of nucleosomes in the genome [Nucleic Acids Res (2009) doi: 10.1093/nar/gkp689].

GENETIC CODE[a–e]

First base in codon (5′ end)	Second base in codon				Third base in codon (3′ end)
	U	C	A	G	
U	Phe	Ser	Tyr	Cys	U
	Phe	Ser	Tyr	Cys	C
	Leu	Ser	*ochre*	*opal*	A
	Leu	Ser	*amber*	Trp	G
C	Leu	Pro	His	Arg	U
	Leu	Pro	His	Arg	C
	Leu	Pro	Gln	Arg	A
	Leu	Pro	Gln	Arg	G
A	Ile	Thr	Asn	Ser	U
	Ile	Thr	Asn	Ser	C
	Ile	Thr	Lys	Arg	A
	Met	Thr	Lys	Arg	G
G	Val	Ala	Asp	Gly	U
	Val	Ala	Asp	Gly	C
	Val	Ala	Glu	Gly	A
	Val	Ala	Glu	Gly	G

[a]A = adenine; C = cytosine; G = guanine; U = uracil.
[b]Amino acid symbols are explained in the entry AMINO ACID (table).
[c]Codons are conventionally written in the 5′-to-3′ direction.
[d]The *ochre*, *opal* and *amber* codons are normally stop signals.
[e]See text for some variations in the code.

genetic complementation See COMPLEMENTATION.

genetic counseling Advice given to e.g. parents regarding the possibility/probability of genetically based problems in their existing or future offspring, or, in general, to others regarding their susceptibility to particular genetically based problem(s).

The availability of tests for genetic susceptibility to certain diseases has raised questions regarding the interpretation and use of the tests and their regulation [Eur J Hum Genet (2009) doi: 10.1038/ejhg.2008.246].

genetic disease A term generally used to refer specifically to a *heritable* disease or disorder, i.e. one which is determined by an individual's genotype.

Examples of genetic diseases are listed in the table. Others include e.g. BLOOM'S SYNDROME and WERNER SYNDROME.

A genetic disease may be detected pre-natally: see the entry GENETIC DISEASE (pre-natal diagnosis).

(See also ONLINE MENDELIAN INHERITANCE IN MAN.)

genetic disease (pre-natal diagnosis) For various approaches to the pre-natal diagnosis of genetic disease see entries DOWN'S SYNDROME, EPIGENETIC ALLELIC RATIO ASSAY and PGD.

genetic disorder *Syn*. GENETIC DISEASE.

GENETIC DISEASE (human): some examples

Adenosine deaminase deficiency	An autosomal recessive disorder, involving abnormal susceptibility to infectious diseases, caused by deficiency of a functional adenosine deaminase (EC 3.5.4.4) encoded by gene *ADA* (20q13.11); deficiency in this enzyme inhibits normal development of T lymphocytes so that the patient's immune system is inefficient
Angelman syndrome	A condition characterized by mental impairment, defective balance, abnormal behavior, and severe speech defects. At least four genetic mechanisms are believed to be capable of causing Angelman syndrome. Most cases appear to result from a deletion at 15q11.2–q13, a small number result from imprinting defects, and some are reported to be due to mutation in the gene for ubiquitin protein ligase E3A
Axenfeld–Rieger syndrome	An autosomal dominant disorder affecting eyes and certain systemic features. Associated with mutation e.g. in the *PITX2* gene (encoding a transcription factor). In some cases the syndrome is reported to involve aberrant splicing of pre-mRNA
Bardet–Biedl syndrome	A genetically heterogeneous disorder, involving at least 11 chromosomes, characterized by e.g. renal and retinal abnormalities
Bernard–Soulier syndrome	A rare autosomal recessive condition involving a tendency to abnormal/excessive bleeding. Relevant mutations are found on chromosomes 3, 17 and 22, affecting proteins which form the subunits of a receptor for von Willebrand factor. See entry BERNARD–SOULIER SYNDROME. See also Hemophilia and Wiskott–Aldrich syndrome, below, for further examples of abnormal/excessive bleeding disorders
Charcot–Marie–Tooth disease	Any of a group of heterogeneous inherited neuropathies that include demyelinating and axonal types; autosomal dominant, autosomal recessive and X-linked forms of the disease are recognized (see entry CHARCOT–MARIE–TOOTH DISEASE)
Citrullinemia	Deficiency in arginosuccinate synthetase, with accumulation of citrulline and ammonia in the plasma and damage to the nervous system. The *classic* (type I) form involves mutation in gene *ASS* (9q34) which encodes arginosuccinate synthetase The *adult-onset* (type II) form is caused by mutation in gene *SLC25A13* (7q21.3) encoding *citrin*; this leads to a deficiency of the enzyme arginosuccinate synthetase
Cockayne syndrome	Poor development, photosensitivity and (generally) failure to survive beyond early adulthood. Forms A and B, resulting from mutation in genes (encoding DNA-repair proteins) at locations 5q12 (*CSA*) and 10q11 (*CSB*)
Cystic fibrosis	Commonly: abnormal secretion in the respiratory tract, with risk of infection, associated with a defective gene for the cystic fibrosis transmembrane conductance regulator (CFTR) protein (7q31.2)
Down's syndrome	Mental impairment and characteristic facial features. It is associated with trisomy (chromosome 21)
Duchenne muscular dystrophy	Muscular abnormality due to absence or abnormality of the (normal) muscle protein dystrophin, encoded by gene *DMD* (Xp21.2)
Dyggve–Melchior–Clausen syndrome	A condition, involving e.g. skeletal abnormalities, associated with mutation in the *DYM* gene (18q12–q21.1)
Emphysema (one type)	Deficiency of the protease inhibitor α-1-antitrypsin (gene location 14q32.1). Deficiency in this protein is also associated with cirrhosis
Fabry disease	An X-linked condition involving deficiency in the activity of α-galactosidase A (gene location Xq22). The range of manifestations apparently include e.g. vascular-type skin lesions and cardiac symptoms; renal failure is reported to be a common cause of death Fabrazyme® is used for treating this condition (see the table in the entry BIOPHARMACEUTICAL)
Fragile X disease	An X-linked disorder, characterized e.g. by mental impairment, associated with instability of MICROSATELLITE DNA (the CGG triplet) near the promoter of the *FMR1* gene, which results in gene silencing

(*continued*)

Fucosidosis	An autosomal recessive disease involving a defective lysosomal enzyme, α-L-fucosidase (gene *FUCA1*, location 1p34) and the consequent accumulation of glycoproteins
Glanzmann's thrombasthenia	An autosomal recessive disease characterized by a deficiency in platelet agglutination; it is associated with abnormalities in the $\alpha_{IIb}\beta_3$ INTEGRIN
Glycogen storage disorder	Dysfunctional glycogen metabolism. Various types. Type I disease is characterized by a deficiency of the enzyme glucose 6-phosphatase (EC 3.1.3.9) (gene location 17q21); infants exhibit e.g. severe hypoglycemia and an enlarged liver
Hemophilia	Defective blood-clotting mechanism characterized e.g. by a tendency to hemorrhage, easy bruising etc. There is more than one type of hemophilia. *Hemophilia A* is an X-LINKED DISORDER resulting from a deficiency in the activity of clotting factor VIII; the genetic lesion is at Xq28. *Hemophilia B*, also an X-LINKED DISORDER, results from a deficiency in the activity of factor IX (the Christmas factor); the genetic lesion is at Xq27.1–q27.2. The term hemophilia has also been used to include the hemorrhagic condition von Willebrand disease – typically an autosomal dominant disorder associated with chromosomal location 12p13.3. See also Bernard–Soulier syndrome, above, and Wiskott–Aldrich syndrome, below, for some other examples of abnormal/excessive bleeding disorders. Various recombinant products are available for the therapeutic treatment of these disorders: see, for example, Kogenate® and NovoSeven® in the table in BIOPHARMACEUTICAL
Huntington's disease	An autosomal dominant disorder that involves chorea (i.e. involuntary movements) and progressive dementia; it is caused by expansion of MICROSATELLITE DNA (the CAG triplet) associated with a gene on chromosome 4
Hurler's syndrome	See Mucopolysaccharidosis type IH
Hyper-IgM syndrome	Immunodeficiency, with normal or elevated levels of IgM but a deficiency in IgG and other isotypes; patients with hyper-IgM syndrome are highly vulnerable to bacterial infections. There is an X-linked form and an autosomal recessive form. The *X-linked* form (HIGM1) is associated with mutation in the *CD40L* gene (affecting T cell–B cell interaction) The *autosomal recessive* form (HIGM2) may be due to mutation in the *CD40* gene or to a deficiency in the enzyme activation-induced cytidine deaminase (which may be involved in RNA editing)
ICF syndrome	Immunodeficiency, centromere instability and facial abnormality syndrome. A rare syndrome possibly caused by dysregulation of genomic methylation due to mutation in the *DNMT3B* gene (20q11.2)
Lesch–Nyhan syndrome	An X-linked syndrome, with e.g. neurologic and metabolic abnormalities, associted with mutation in the *HPRT* gene (Xq26–q27.2)
Leukocyte adhesion deficiency	An autosomal recessive disorder characterized by a low resistance to certain infectious diseases. Leukocytes are deficient in the β_2 INTEGRIN component of cell-surface adhesion molecules (involved in binding to pathogens)
Mucopolysaccharidosis type IH	An autosomal recessive disease (also called Hurler's syndrome) affecting various organs, including the heart. It results from mutation in the *IDUA* gene (encoding the lysosomal enzyme α-L-iduronidase).
Niemann–Pick disease	Mental retardation. Several types. Types A and B associated with lysosomal accumulation of sphingomyelin in the tissues due to deficiency in the enzyme sphingomyelin phosphodiesterase
Peutz–Jeghers syndrome	An autosomal dominant condition with intestinal hamartomatous polyposis, pigmentation in the oro-facial region, and an increased risk of gastrointestinal malignancy. Mutation in the coding region of a serine-threonine kinase gene, *STK11*/*LKB1* (chromosomal locus 19p13.3), appears to be present in most of the cases
Phenylketonuria	An autosomal recessive disease with e.g. mental retardation. Phenylpyruvate is found in the urine, and phenylalanine is found in blood and other tissues; the disorder is caused by inadequacy of the enzyme phenylalanine 4-mono-oxygenase (EC 1.14.16.1) (gene location: 12q24.1)

(*continued*)

GENETIC DISEASE (human) (*continued*)

Prader–Willi syndrome	Obesity, mental retardation etc. Error in genetic imprinting. In at least some cases it may result from a deficiency of paternal copies of certain genes, including the gene (*SNRPN*) encoding small nuclear ribonucleoprotein polypeptide N (chromosomal location: 15q12)
Rett syndrome	Neurodevelopmental disorder. Commonly: mutation in the *MECP2* gene (encoding a methyl-CpG-binding protein); the condition apparently can also result from mutation in gene *CDKL5* (which encodes a cyclin-dependent kinase)
Sandhoff disease	A condition related to Tay–Sachs disease (see below) but involving the β-subunit of hexosaminidase A
SCID	Severe combined immunodeficiency: a group of diverse diseases involving the failure of T lymphocytes to develop normally. One of these diseases is adenosine deaminase deficiency (see above). Another is an X-linked disorder involving a defective γc subunit in a cell-surface cytokine receptor which (e.g.) inhibits normal development of T cells; this disorder is called XSCID or SCIDX1
Severe combined immunodeficiency	See SCID, above
Sickle-cell anemia	Anemia, involving the β-chain of hemoglobin (see the entry SICKLE-CELL ANEMIA)
Smith–Lemli–Opitz syndrome	An autosomal recessive disorder, involving mental retardation, reported to involve defective cholesterol metabolism owing to mutation in the *DHCR7* gene (encoding 3β-hydroxysterol-Δ7 reductase)
Tay–Sachs disease	An autosomal recessive condition resulting from mutation in the gene for the α-subunit of hexosaminidase A (EC 3.2.1.52) located at 15q23–q24, a lysosomal enzyme required for the breakdown of gangliosides. This neurodegenerative disease is commonly seen early in life: infants show poor development, with mental impairment and blindness, and the condition is often fatal within the first few years of life. See also Sandhoff disease, above
Thalassemia	Anemia, resulting from a disorder involving the α-chain or β-chain of hemoglobin
von Willebrand disease	See Hemophilia, above, and Wiskott–Aldrich syndrome, below
Wiskott–Aldrich syndrome	An X-LINKED DISORDER: thrombocytopenia (and defective blood clotting), eczema, and a susceptibility to recurrent infection (see entry WISKOTT–ALDRICH SYNDROME). See also Hemophilia, above

genetic drift Ongoing changes in genotype resulting from the occurrence of chance mutations; the phrase is usually used to describe such changes in viruses which exhibit this tendency. Although any given mutation may have only a minimal effect on phenotype, genetic drift is regarded as a significant factor in the evolution of at least some viruses.

The factors involved in genetic drift are usually understood to include viral replication and/or inter-host transmission; in at least some cases, the occurrence of mutations reflects poor fidelity of RNA polymerases and/or a lack of proofreading capability.

In viruses, genetic drift is distinguished from *antigenic drift* – changes in the antigenic determinants of the viral capsid; antigenic drift can arise by selection of viruses that replicate successfully in their host organism – with counterselection of those viruses that do not replicate successfully in their normal hosts.

[Genetic drift of HIV populations (in culture): PLoS Genet (2009) 5(3):e1000431.]

(cf. GENETIC SHIFT.)

genetic engineering Manipulation of genetic material – DNA and/or RNA – in order to bring about any desired change, in genes or genetic systems, for any purpose, either *in vitro* or *in vivo*. Genetic engineering includes e.g. those *in vitro* methods used in the study of genes and their regulation; methods used in GENE THERAPY; and methods used for creating new strains of organisms for medical, agricultural or industrial purposes.

genetic footprinting A technique that can be used to examine gene function – e.g. to determine whether the expression of specific genes is required for growth under given conditions; the method is used in eukaryotic and prokaryotic microorganisms, including *Saccharomyces cerevisiae* and various Gram-negative bacteria.

Note. This technique is completely unrelated to the method, used for studying DNA–protein interactions, that is referred to as FOOTPRINTING.

One of the formats for genetic footprinting in bacteria is as follows. First, a population of bacteria is subjected to random

transposon-mediated mutagenesis. This could involve e.g. (i) transposon mutagenesis of DNA fragments *in vitro*, followed by insertion of the mutagenized fragments into cells by transformation, or (ii) the Transposome™ approach (see TN5-TYPE TRANSPOSITION), in which mutagenesis occurs *in vivo*. (The latter approach has certain advantages, such as the possibility of limiting insertions to one per cell (by using an appropriate Transposome–cell ratio), and the lack of a requirement for a special transposon-delivery vector.)

An aliquot of the (mutagenized) population is set aside for subsequent analysis; it is labeled T0 (for time zero).

The remainder of the mutagenized population is then grown for many generations under conditions in which further transposition is inhibited.

Finally, genomic DNA from (i) T0 cells, and (ii) cells from the cultured, mutagenized population is isolated, and nucleotide sequences from specific genes are determined, e.g. by a PCR-based approach.

If, under the given conditions of growth, a specific gene is essential for cell viability, then transposon insertions in that gene should be easily detectable in cells of the T0 population – but not in cells of the *cultured* population, because the non-growing (transposon-inactivated) cells in *this* population will be greatly outnumbered by cells which have the (active) gene and which will have formed numerous progeny.

Using this principle, studies were carried out to investigate the mutational background underlying bacterial susceptibility and tolerance to particular types of antimicrobial agent [PLoS ONE (2009) 4(5):e5629].

genetic immunization (*syn*. DNA immunization, also known as DNA vaccination) A phrase which refers to the use of a DNA VACCINE (q.v.).

[Uses of term (e.g.): Genet Vaccines Ther (2008) 6:15, and (DNA immunization) BMC Genomics (2009) 10:112.]

genetic imprinting (*syn*. genomic imprinting; parent-of-origin-specific gene expression) In certain (*imprinted*) genes: the repression of function of *one* allele (either the maternal *or* the paternal allele); such repression is mediated by methylation at one or more specific loci on the imprinted allele – see the entry IMPRINTING CONTROL REGION. Only *one* of the alleles of an imprinted gene (from either the male or female parent) is expressed under given circumstances.

Genetic imprinting is common among mammalian species, being found e.g. in mice, pigs (hogs) and sheep as well as in humans.

The differential methylation of an imprinted gene is established during gametogenesis and persists in the adult animal.

Errors in imprinting may cause serious disorders – e.g. the Prader–Willi syndrome (which is characterized by obesity, mental retardation and hypogonadism: see Eur J Hum Genet (2008) doi: 10.1038/ejhg.2008.232).

Imprinting is an epigenetic phenomenon, that is, it directs a heritable phenotype which is not dependent on the sequence of nucleotides in the given genes.

genetic screening (1) Any procedure for detecting or selecting individual cells or organisms with specific genetic characteristic(s) from among a population of those cells or organisms. (2) In a clinical context: any test(s) carried out on patients in order to detect the risk of possible/probable genotype-based disorders in the patients themselves or in their offspring.

genetic shift An abrupt, major change in genotype; the phrase is commonly used to describe such changes in viruses which exhibit such a tendency. For example, as a result of a genetic shift, a (reassortant) influenzavirus may have a significantly altered and previously unknown antigenic structure and may therefore be able to initiate a pandemic of influenza because of the lack of protective antibodies in the human population.

(See also REASSORTANT VIRUS and SEGMENTED GENOME.)

[The potential for segment reassortment in La Crosse virus in nature: Virol J (2008) 5:164.]

(cf. GENETIC DRIFT.)

genetically modified food (GM food) Any food (such as GM soya) which has been produced after specific and intentional alteration of the genome of a given type of plant or animal – see entry GMO.

One example of GM food is so-called *golden rice*: a form of rice engineered to produce β-carotene (provitamin A) in the grain; this can benefit those who would otherwise suffer from a deficiency of vitamin A (which has been linked e.g. to blindness). (See the website www.goldenrice.org for further information on this product.)

A review on genetically modified plants and human health has considered direct effects (e.g. the effects of modifications targeted at nutrition) and indirect effects (e.g. effects on the environment), and has also examined some of the opposition to genetically modified products [J Roy Soc Medicine (2008) 101(6):290–298].

genetically modified mosquitoes See GMMS.

genetically modified organism See GMO.

Geneticin® A commercial brand of G418 SULFATE (Invitrogen, Carlsbad CA, USA).

[Use (e.g.): studies on the involvement of Tat-SF1 in HIV-1 infection (500 μg/mL): PLoS ONE (2009) 4(5):e5710.]

genistein 4′,5,7-Trihydroxyisoflavone: a phytoestrogen (found e.g. in soya beans) which can bind e.g. to human estrogen receptors; it inhibits tyrosine kinases, acting as a competitive inhibitor of ATP at the binding site.

[Use as a TK inhibitor (e.g.): Int J Clin Exp Med (2009) 2(1):87–94.]

Genistein was reported to cause APOPTOSIS in the embryos of zebrafish [PLoS ONE (2009) 4(3):e4935], and to regulate quadruplex–duplex competition in (human) telomeric DNA [Nucleic Acids Res (2009) 37(8):2471–2482].

(cf. TYRPHOSTIN.)

(See also TYROSINE KINASE.)

genome (1) The genetic content of a cell or virus – in terms of either (a) the information (coding) content or (b) the physical content of nucleic acid (i.e. a cell's DNA or the DNA or RNA of a virus).

In this sense, DNA in mitochondria (including kinetoplast

DNA) and chloroplasts forms a part of the genome (together with chromosomal DNA). Phrases such as 'the mitochondrial genome' are nevertheless used in order to refer specifically to a given type of DNA within the cell.

In bacteria, which often contain plasmids, *genome* is sometimes used specifically to refer to *chromosomal* DNA; such usage of the term may be justified in that plasmids are often dispensable to the cell, and in that the cells of a given species may contain different assortments of (dissimilar) plasmids – which would correspond to a 'variable genome'. However, the term genome is sometimes used to include bacterial plasmids, particularly when cells of a given species commonly contain a stable complement of plasmid(s) – or the plasmid(s) have a particular relevance to the organism's biological activity or its pathogenicity; thus, the term 'genome' has been used in this way e.g. for *Bacillus anthracis* [J Bacteriol (2000) 182(10): 2928–2936], *Pseudomonas syringae* pv. tomato [Proc Natl Acad Sci USA (2003) 100(18):10181–10186], for *Rhizobium leguminosarum* [Genome Biol (2006) 7(4):R34], and also for *Rhodococcus* RHA1 [Proc Natl Acad Sci USA (2006) 103: 15582–15587].

(2) The *haploid* set of genes or chromosomes in a cell.

(3) The *diploid* set of genes or chromosomes in a cell.

(4) The complete set of *different* genes in an organism.

(5) The term 'genome' has also been used for *viroids* (single-stranded molecules of RNA that infect plants) [BMC Microbiol (2006) 6:24].

Genome Deletion Project (for *Saccharomyces cerevisiae*) A project that systematically replaced most of the open reading frames (ORFs) in the genome of *Saccharomyces cerevisiae* with a kanamycin-resistance cassette, creating a collection of strains in which a different ORF has been replaced in each strain. This collection of (GDP) strains is a comprehensive source of knock-outs for *S. cerevisiae*, facilitating studies on gene function.

In addition to their intended application, GDP strains can also be used e.g. to copy any desired *S. cerevisiae* promoter; a copy of this promoter can then be inserted into a strain of interest and used for controlling the expression of a given gene. For this purpose, one first identifies the particular GDP strain in which the desired promoter is flanked by a kanamycin knock-out immediately on its 5′ side. This promoter, together with its contiguous kanamycin-resistance cassette, is amplified by PCR, using primers with 5′ extensions homologous to sites of insertion in the strain of interest. The PCR products can then be inserted into the genome of the given strain by homologous recombination, replacing the existing promoter. [Method: BioTechniques (2006) 40:728–734.]

Genome Sequence 20^{TM} DNA sequencing system See GS FLX.

genome walking *Syn*. CHROMOSOME WALKING.

genomic cloning See CLONING.

genomic footprinting A phrase which has been used to refer to footprinting *in vivo* (i.e. within living cells): see the entry IN VIVO FOOTPRINTING.

(cf. GENETIC FOOTPRINTING.)

genomic imprinting *Syn*. GENETIC IMPRINTING.

genomic island In a bacterial genome: a discrete region of the chromosome, acquired from an exogenous source, that differs in GC% (and codon usage) from the rest of the chromosome and is often located near tRNA genes; genomic islands are often flanked by direct repeat sequences and may include any of various genes (including e.g. antibiotic-resistance genes) as well as mobile genetic elements.

Recombination may result in fragmentation of a genomic island with re-location of part(s) of the island to different regions of the chromosome.

Genomic islands are reported to be generated, for example, at the sites of insertion of transposon Tn*7* [J Bacteriol (2007) 189(5):2170–2173].

A 54-kb genomic island in the (Gram-negative) bacterium *Leptospira interrogans* serovar Lai (containing 103 predicted coding sequences) is reported to excise from the chromosome and to exist as a (replicative) plasmid [Infect Immun (2007) 75(2):677–683].

[Computational identification and visualization of genomic islands: Bioinformatics (2009) 25(5):664–665.]

genomic library See LIBRARY.

genomic masking See PHENOTYPIC MIXING.

genospecies Any group of (microbial) strains between which genetic exchange can occur.

genotoxin Any agent that damages nucleic acid.

genotype (of an organism) The genetic make-up of an organism – reflecting the actual sequence of nucleotides in its DNA or (for some viruses) RNA.

(cf. PHENOTYPE.)

genotype imputation A predictive approach used e.g. to facilitate the interpretation of genome-wide SNP data in *association studies*, i.e. studies in which an attempt is made to associate specific SNP(s) in the genome with phenotypic traits such as susceptibility to particular diseases – or the failure to respond to particular drugs. Imputed genotypes indicate the range of SNPs that could be evaluated in association studies and may also help to distinguish true associations from apparent ones.

A flexible and accurate genotype imputation method for the next generation of genome-wide association studies has been proposed [PLoS Genetics (2009) 5(6):e1000529].

GenoType MTBDR See LINE PROBE ASSAY.

genotypic susceptibility testing (antibiotic resistance) See the entry ANTIBIOTIC RESISTANCE TESTING.

genotyping Any of various procedures in which an organism is characterized on the basis of specific features of its genomic DNA or (for some viruses) RNA. One example is SNP GENOTYPING. Another is the determination of an individual's DNA profile using a system such as CODIS.

Further examples include the various 'genotypic' (i.e. DNA- and RNA-based) methods for TYPING bacteria – and methods for determining bacterial sensitivity to particular antibiotics; the latter include e.g. detection of the MECA GENE in strains of methicillin-resistant *Staphylococcus aureus* (MRSA) by the use of PCR with gene-specific primers.

(See also MLVA and SPOLIGOTYPING.)

gentamicin An AMINOGLYCOSIDE ANTIBIOTIC used e.g. as a selective agent for bacteria containing a vector which carries a gentamicin-resistance gene.

germ line cells Cells whose development leads to the formation of gametes (reproductive cells).

GFP See GREEN FLUORESCENT PROTEIN.

GFP reactivation technique An approach designed to facilitate the drug-assisted analysis of DNA methylation. (See also DEMETHYLATION.) This method identifies genes which have been activated by drug treatment – such activation resulting from the removal of methylation-induced repression. Use is made of a reporter gene whose activation (by the drug) can indicate 'global' drug-induced activation of genes in the cells. [Method: BioTechniques (2006) 41(4):461–466.]

The method uses transgenic heterozygous female mice with an X-linked reporter gene that encodes GREEN FLUORESCENT PROTEIN. In this method, the relevant GFP transgene is in the *inactivated* X chromosome (see entry X-INACTIVATION).

In the method, a population of cells, isolated from the transgenic mouse, undergoes several stages of screening – carried out in a FLUORESCENCE-ACTIVATED CELL SORTER.

Cells from the animal are first sorted to select those which contain an inactive, i.e. *non*-expressed, gene for GFP.

After drug treatment (with 5-AZA-2′-DEOXYCYTIDINE) these cells are again sorted by flow cytometry in order to separate those containing active GFP from those in which the GFP is still inactive. This final stage of screening therefore isolates those cells which exhibit *drug-induced* activation of the GFP reporter gene present on the *inactivated* X chromosome.

MICROARRAY-based gene expression profiling of the three populations of cells – i.e. (i) the GFP-inactive untreated cells, (ii) GFP-inactive drug-treated cells and (iii) GFP-active drug-treated cells – indicated a marked increase in gene expression in the drug-treated cells containing an active GFP gene on the inactivated X chromosome. Examination of the distribution of the activated genes showed a bias towards X-linked genes – i.e. genes on the X chromosome had an increased sensitivity to activation by this method.

giant chromosome See POLYTENIZATION.

GIFT Gamete intrafallopian transfer: see IVF.

Gigapack® Any of various products (Stratagene, La Jolla CA, USA) used for the packaging of some types of lambda-based vector molecule ('packaging' referring to the process in which each DNA vector molecule is enclosed in phage components to form an infective phage particle). Gigapack® products are used e.g. with the LAMBDA FIX II VECTOR.

gilvocarcins A group of glycoside anti-tumor agents produced e.g. by species of the Gram-positive bacterium *Streptomyces*. (See also MUTASYNTHESIS.)

GISA Strains of MRSA with decreased susceptibility to glycopeptide antibiotics (e.g. vancomycin); the abbreviation refers to 'glycopeptide intermediately susceptible *S. aureus*'.

Glanzmann's thrombasthenia An autosomal recessive disease that is characterized by a deficiency in platelet agglutination and linked to abnormalities of the $\alpha_{IIb}\beta_3$ INTEGRIN.

[Review of Glanzmann's thrombasthenia: Orphanet J Rare Dis (2006) 1:10.]

glargine See INSULIN GLARGINE.

glucose effect See CATABOLITE REPRESSION.

Glucuron™ See CHEMILUMINESCENCE.

glutathione *S*-transferase (GST) See GENE FUSION (uses). (See also GENE FUSION VECTOR.)

glycerol nucleic acid See the entry GNA.

glycoengineered fungi See GLYCOSYLATION.

glycogen storage disorder Any of various human diseases that involve dysfunctional glycogen metabolism.

The type I disease is characterized by deficiency of glucose 6-phosphatase in the liver. In infants it involves severe hypoglycemia and an enlarged liver containing abnormally high levels of glycogen.

The type VIII form of glycogen storage disorder, which is generally less severe, is characterized by inadequate levels of the enzyme phosphorylase kinase.

(See also GENETIC DISEASE (table).)

glycosylation (of recombinant proteins) Post-translational modification of recombinant proteins, carried out in cells capable of adding glycosyl residue(s) to the appropriate amino acid residue(s). In many cases such glycosylation is an essential requirement for normal activity of the recombinant protein.

Some types of cell (e.g. bacteria) do not normally carry out the patterns of glycosylation which are found in mammalian proteins; however, bacteria have been used for the production of certain mammalian proteins (e.g. interleukin-2) in which a lack of post-translational glycosylation apparently does not significantly affect biological activity. (See also Neupogen® in the entry BIOPHARMACEUTICAL (table).)

Glycoproteins synthesized in fungi generally have a high-mannose pattern of glycosylation (unlike that in mammalian cells); this may decrease persistence of recombinant proteins used therapeutically in humans, and such proteins may also be immunogenic. To avoid these problems, strains of *Pichia pastoris* have been 'glycoengineered' to modify their pattern of glycosylation, and a recombinant humanized and immunologically compatible form of the glycoprotein lactoferrin was produced in one of these strains [Glycoconj J (2008) 25(6): 581–593].

GM food GENETICALLY MODIFIED FOOD.

GM2 gangliosidoses Certain GENETIC DISEASES in which non-degraded gangliosides accumulate in lysosomes. Examples of such diseases: TAY–SACHS DISEASE and SANDHOFF DISEASE.

gMap A tool which can be used to compare genomic sequences from different bacteria. It can be found at: http://www.ncbi.nlm.nih.gov/sutils/gmap.cgi

GMDD A database of methods for the detection of genetically modified organisms (see entry GMO).

GMMs Genetically modified mosquitoes: mosquitoes that have been genetically engineered primarily to control the spread of mosquito-borne diseases – e.g. dengue, MALARIA and yellow fever. Research on GMMs is being conducted simultaneously

with other approaches – for example, attempts to improve the efficacy/availability of therapeutic drugs as well as ongoing efforts to produce effective vaccines. The use of this technology would require release of GMMs into the natural environment; schemes for using GMMs must therefore be subjected to rigorous examination (e.g. by 'contained field trials').

The GMM strategy involves two distinct approaches:

1. Engineered sterility – reducing the size of populations of mosquito vectors.

2. POPULATION REPLACEMENT – meaning the replacement of (existing) populations of mosquito vectors with mosquitoes that are unable to act as vectors (see later). (This approach is less likely to affect local biodiversity.)

1. Engineered sterility

In one scheme, *Anopheles gambiae* (a major vector of human malaria) was genetically engineered so that, during spermatogenesis, the mosquitoes expressed an enzyme that selectively cuts a sequence of DNA found only in certain essential genes in the X chromosome; this modification resulted in complete dominant embryo lethality in crosses with wild-type females.

An important feature of this scheme is that all spermatozoa which were produced by these *Anopheles gambiae* males had the specific 'X-cutting' enzyme and were active against the (maternal) X chromosome in an embryo. The presence of this enzyme also produced a strong negative bias against spermatozoa containing an X chromosome. [Method: PLoS Genet (2008) 4(12):e1000291.]

(See also ARBOVIRUSES.)

2. Population replacement

As mentioned, the basic idea in *population replacement* is to replace the natural wild-type populations of mosquito vectors with genetically modified, transgenic individuals that are not able to transmit specific pathogen(s). To do this, GMMs must contain – as well as anti-pathogen transgene(s) – a *gene drive system* (see later) that is able to bring about effective *spread* of the anti-pathogen genes into the wild-type populations of mosquito vectors. Such genetic elements can be introduced into wild-type populations through the reproductive system, i.e. via the gametes. Sufficient numbers of GMMs need to be released in order to achieve effective spread and to maintain the transgene(s) in wild-type populations. (See also RELEASE RATIO.)

Anti-pathogen transgenes are designed to make mosquitoes 'refractory' – i.e. unable to transmit specific pathogen(s); such transgenes have been called REFRACTORY GENES (q.v.).

Gene drive systems

A gene drive system (or simply 'drive system') is essentially one or more genetic elements/constructs with innate ability to spread between individuals of a given species – one familiar example being a TRANSPOSABLE ELEMENT ('jumping gene'). Hence, if a given refractory gene (see above) were linked to a suitable drive system, the latter should be able to drive that gene into natural populations of mosquitoes.

Transposable elements were initally considered as potential drive systems but are now regarded as insufficiently reliable, not least because of their tendency to lose DNA inserts. Other potential drive systems include:

Engineered underdominance constructs: a pair of constructs, each containing (in addition to an anti-pathogen gene): (i) a lethal gene and (ii) a gene encoding a product that suppresses the effect of a lethal gene on the *other* construct. Mosquitoes which lose one of the constructs are killed by expression of the lethal gene on the other construct. This engineered underdominance system has the advantage that both constructs can carry anti-pathogen gene(s).

Meiotic drive. In one proposal, certain engineered genes on a mosquito's male ('Y') chromosome block the maturation of any gametes that contain *a sensitive form* of a certain allele on the female ('X') chromosome. This Y-linked meiotic drive system could be exploited by releasing male mosquitoes that contain an insensitive form of the X-linked allele attached to an anti-pathogen gene. The post-introduction dynamics of a drive system of this type have been studied by mathematical models and computer simulations [Evolution (2007) 61(4): 717–726].

Medea system. *Medea* (maternal-effect dominant embryonic arrest) is a construct which encodes a maternally expressed toxin *and* a zygotically expressed antidote; it kills all progeny lacking *Medea*. The system selects for those offspring which contain *Medea*, making the *Medea* element a potential drive system. The *Medea* system is not available for mosquitoes at present, though it has been used successfully to drive population replacement in the fruitfly *Drosophila melanogaster* [Science (2007) 316:597–600].

Wolbachia. This Gram-negative bacterium (see WOLBACHIA) infects various invertebrates, in which it can give rise to *cytoplasmic incompatibility* (CI). CI refers to the phenomenon in which a cross between an *un*infected female and an infected male is generally sterile. *Wolbachia* is therefore a 'maternally inherited' organism because an unconditionally fertile cross requires an infected female. *Wolbachia*-infected females thus have a reproductive advantage which allows them to spread *Wolbachia* without further selective pressure; this has been thought to have potential as a gene drive mechanism. In the context of dengue, *Wolbachia* was shown to transfer genes to its mosquito host, *Anopheles aegypti* [BMC Genomics (2009) 10:3], suggesting possible transfer of anti-plasmodial genes to GMMs for malaria control.

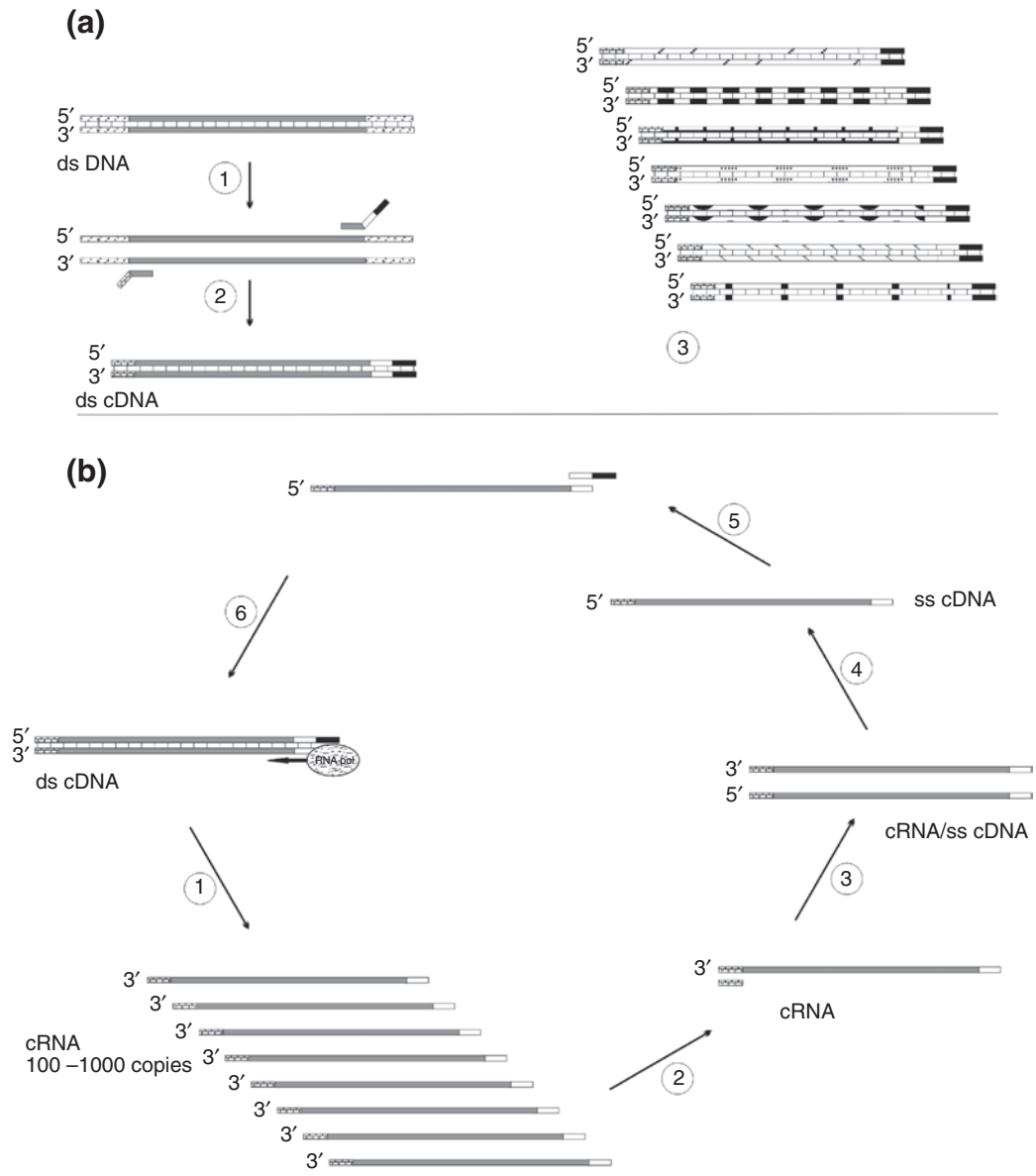
Fitness cost

In producing GMMs, one factor to be considered is that the introduction of transgenes, including both anti-pathogen and gene drive constructs, may have an important FITNESS COST.

[Guidance for contained field trials of mosquitoes engineered to contain a gene drive system (recommendations): Vector Borne Zoonotic Dis (2008) 8(2):127–166.]

[Malaria control/transgenic mosquitoes: PLoS Med (2009) 6(2):e1000020.]

GMO Genetically modified organism: any organism in which the genome has been specifically and intentionally altered for a given purpose – e.g. for the production of a GENETICALLY



GMO: amplification of the sample DNA by NAIMA (NASBA implemented microarray analysis) – a method which allows quantitative on-chip detection of GMOs (schematic representation).

(a) Template synthesis. Sample dsDNA is denatured (1), and 5′-tagged site-specific primers (one with a promoter tag) are extended by *Taq* polymerase to produce templates (2). The templates obtained from different samples have the same 5′ and 3′ ends (3).

(b) Multiplex amplification. Each DNA template is flanked by the promoter for T7 DNA-dependent RNA polymerase. Transcription by T7 polymerase (1) produces numerous copies of antisense cRNA which are then reverse transcribed into ss cDNA, forming RNA–DNA duplexes (2 and 3). The RNA strands are degraded by RNase H (4). The cDNA is then used as a template by reverse transcriptase after annealing a primer (5), yielding a second strand of DNA (6). This dsDNA can be used for repeated cycles of amplification. The final product is antisense cRNA. Only two types of primer are needed for the multiplex amplification.

Diagram courtesy of Dany Morisset, Department of Biotechnology and Systems Biology, National Institute of Biology, Ljubljana, Slovenia. The method was first published by Dany Morisset and colleagues in *Nucleic Acids Research* (2008) 36(18):e118.

MODIFIED FOOD (q.v.) such as *golden rice*. (See also GMMS.)

The construction of a GM organism is not necessarily the accelerated production of an organism that could be obtained by a naturally occurring process. This is because it is possible to modify genomes with sequence(s) of nucleotides found in totally unrelated organisms – or even sequences that were designed on a computer, i.e. sequences which do not necessarily reflect any that now occur, have ever occured, or may ever occur, in nature, in the given organism.

In some countries the law requires that GM products on the market be traceable, and in some cases the inclusion of a GM content above a given threshold must be indicated by appropriate labeling. Consequently, for monitoring purposes, there have been a number of suggested methods for detecting the presence of a GMO in a given product.

GMOs can be detected by a range of methods, some based on the detection of a protein target and some on the detection of particular nucleic acid sequences. One method, an adaptation of NASBA known as 'NASBA implemented microarray analysis' (NAIMA), permits quantitative on-chip detection of GMOs – this method is shown in the accompanying diagram.

GMOs have also been detected e.g. by the LAMP method [BMC Biotechnol (2009) 9:7].

[GMDD (a database of GMO detection methods): BMC Bioinformatics (2008) 9:260.]

[A comparison of different real-time PCR chemistries and their suitability for the detection and quantification of GMOs: BMC Biotechnol (2008) 8:26.]

[Detection of multiple (non-authorized) GMOs by a probe- and microarray-based approach using a single reaction: BMC Genomics (2008) 9:584.]

GNA (glycerol nucleic acid) A nucleic acid analog in which the backbone chain consists of phosphoglycerol units; the molecule can form stable duplexes.

(See also THERMINATOR DNA POLYMERASE; cf. PNA, TNA.)

goi Gene of interest.

gold standard Any well-established and trusted procedure (e.g. a particular form of test) regarded by workers in the field as being superior (in e.g. sensitivity/specificity/reliability) when compared to other (alternative or proposed) methods. In the diagnosis of infectious diseases a *combined* or *expanded* gold standard is generally taken to mean both culture *and* nucleic-acid-based methods.

Goldberg–Hogness box See CORE PROMOTER.

golden rice See GENETICALLY MODIFIED FOOD.

gp (1) GENE PRODUCT.

(2) A *glycoprotein* gene product encoded by retroviruses.

gp41 (in HIV-1) See HIV-1 and RETROVIRUSES.

***gp64* gene** (in baculoviruses) See BACULOVIRIDAE.

gp120 (in HIV-1) A major envelope glycoprotein of the HIV-1 virion: see HIV-1 and RETROVIRUSES.

(See also DC-SIGN.)

Grace's medium A liquid medium used for the culture of (i.e. growth of) insect cells. Grace's medium includes a number of inorganic salts, sugars (fructose, D-glucose and sucrose), and a variety of amino acids.

[Use of Grace's medium for growing Sf9 cells: PLoS ONE (2009) 4(2):e4369.]

(See also BACULOVIRUS EXPRESSION SYSTEMS.)

granulin A protein, encoded by GRANULOSIS VIRUSES, which forms the matrix of the intracellular *occlusion bodies* of these viruses: see BACULOVIRIDAE for details.

(cf. POLYHEDRIN.)

granulocyte See PMN.

granulosis viruses (GVs) A category of baculoviruses that are distinguished from NUCLEAR POLYHEDROSIS VIRUSES (q.v.) e.g. on the basis of characteristics of their occlusion bodies: see BACULOVIRIDAE for details.

Unlike the NPVs, granulosis viruses appear to be confined to lepidopteran hosts.

Viruses in this group include e.g. a pathogen of the cabbage looper (*Trichoplusia ni*) (TnGV), a pathogen of the codling moth (*Cydia pomonella*) (CpGV), and viruses isolated from a variety of other lepidopterans.

granzymes Agents – released e.g. by $CD8^+$ T cells – which are lethal for target cells. Granzyme A brings about single-strand damage to DNA and causes caspase-independent cell death; it targets protein KU70 [EMBO Rep (2006) 7(4):431–437].

In adenovirus-infected cells, granzyme H was reported to inactivate proteins needed for (i) the replication of viral DNA and (ii) the inhibition of granzyme B [EMBO J (2007) 26(8): 2148–2157].

GRAS Acronym for 'generally regarded as safe'.

gratuitous inducer (of an operon) Any compound which acts as an inducer of an OPERON but which has no metabolic role in the cell (see e.g. IPTG).

gray A unit of the quantity of radiation absorbed: 1 joule/kg. Symbol: Gy.

(cf. RAD.)

***GRB7* gene** See GENE AMPLIFICATION.

green fluorescent protein (GFP) A naturally fluorescent protein found e.g. in the jellyfish *Aequorea*; on irradiation with blue light ($\lambda = \sim 395$ nm) GFP emits green light ($\lambda = 508$ nm).

The GFP gene is widely used as a partner in GENE FUSION and it is also frequently chosen as a reporter gene in a variety of applications (e.g. investigations on histidine repeats in human proteins [PLoS Genet (2009) 5(3):e1000397], and detection of antibiotic resistance in *Mycobacterium tuberculosis* [PLoS ONE (2009) 4(3):e4870].).

(See also EGFP and REPORTER GENE.)

The GFP from *Aequorea* has been associated with a certain level of toxicity in mammalian cells in which it is expressed. GFP from *Renilla reniformis* (see HRGFP) has been used as a less-toxic alternative.

A modified form of the GFP gene has been used to study blockage and resumption of transcription following transient exposure of cells to genotoxic effects. Normal GFP is much too stable to act as a reporter for a *short* cessation of transcription. To reduce the stability of GFP, the gene (within a plasmid vector) was fused with (i) a nucleotide sequence en-

coding the PEST DOMAIN (which targets the fusion protein to a PROTEASOME, increasing the protein's turnover), and (ii) a MAR (MATRIX-ATTACHMENT REGION) sequence promoting rapid turnover of GFP mRNA [BioTechniques (2007) 43(2): 222–227].

For fluorescence microscopy with GFP, it was reported that the mercury arc lamp may be substituted with a light-emitting diode (LED) as this was said to give comparable results.

(See RED FLUORESENT PROTEINS, YELLOW FLUORESCENT PROTEIN; see also IN VIVO FLUORESCENCE.)

grepafloxacin See QUINOLONE ANTIBIOTICS.

gRNA (in trypanosomatids) See RNA EDITING.

grooves (in the DNA helix) In B-DNA, the *minor groove* is the helical groove between the two base-paired strands, while the *major groove* is the helical groove which runs between the overall turns of the double helix.

Certain dyes bind preferentially to the minor groove – see the entry DNA STAINING.

Grunstein–Hogness procedure See COLONY HYBRIDIZATION.

GS FLX A commercial system (Roche/454 Life Sciences) used for MASSIVELY PARALLEL DNA SEQUENCING; the GS FLX system superseded an earlier system: Genome Sequence 20™ DNA sequencing system (also known as the GS20 system). Sequencing by this method is often termed '454 sequencing'.

[The principle and details of this technology are described in Nature (2005) 437:376–380 (with an *erratum* published in Nature 441:120).]

The following is based on the method reported in the above paper. A library of random DNA fragments, of a few hundred base-pairs in length, is initially generated by the mechanical shearing of whole-genome DNA; adaptors are ligated to both ends of the fragments. These fragments are then captured on beads (each bead is approximately 28 μm in diameter) under conditions in which each bead binds one fragment.

The beads are then dispersed within an emulsion consisting of an aqueous phase (a PCR reaction mixture) in oil. Within this emulsion, each aqueous droplet (about 100 μm diameter) is intended to contain a single bead.

Each of the droplets acts as a microreactor – in which the DNA template undergoes PCR-based amplification (emPCR, emulsion PCR), yielding ~10^7 copies/bead. The beads, with DNA attached, are then released from the emulsion and are subjected to an enrichment process in which beads that lack DNA are eliminated.

The beads – carrying single-stranded amplified templates – are then centrifugally deposited into the wells of a fiber-optic slide. This slide contains ~1.6 million wells, each well being of approximate capacity 75 picoliters. (The slide is currently referred to as a PicoTiterPlate™.) The size of the beads (~28 μm) tends to ensure that only one bead occupies each well.

The slide, carrying the beads, is mounted in a flow chamber to create a vertical channel above the opening of each well; this allows entry and exit of reagents for PYROSEQUENCING. (The pyrosequencing enzymes, carried on smaller beads, are also loaded into the wells.) The light signals, which are generated during pyrosequencing, pass through the base of each well and are transmitted, via a fiber-optic imaging bundle, to a CCD sensor and are recorded in a computer.

The system described above was used for sequencing the genome of the bacterium *Mycoplasma genitalium* (580 kb), with a 96% coverage obtained from average reads of ~100 bases in a 4-hour run. The (subsequent) GS FLX system was reported to produce 100 Mb of sequence from average reads of 250 bases.

GS FLX has been used e.g. for metagenomic detection of pathogenic viruses in clinical specimens [PLoS ONE (2009) 4(1):e4219].

[Use of GS FLX for sequencing and *de novo* analysis of a coral larval transcriptome: BMC Genomics (2009) 10:219.]

GS20 sequencing system See GS FLX.

GST Glutathione *S*-transferase: see e.g. GENE FUSION (section: *Uses of gene fusion technology*).

(See also GENE FUSION VECTOR.)

GT. . .AG rule (GU. . .AG rule) A rule stating that the 5′ and 3′ ends of an intron (i.e. the donor splice site and acceptor splice site, respectively) have the dinucleotide sequences GT and AG respectively; the rule apparently applies in most (but not all) cases.

GTA Gene transfer agent: see CAPSDUCTION.

GTP Guanosine 5′-triphosphate.

GU. . .AG rule *Syn.* GT. . .AG RULE.

guanidine salts Agents that inhibit certain viruses (e.g. polioviruses) by inhibiting viral RNA synthesis. Resistant mutants develop readily.

guanosine A RIBONUCLEOSIDE.

guanosine nucleotide exchange factor (guanyl nucleotide exchange factor) See e.g. CYTOTRAP TWO-HYBRID SYSTEM.

Gubler–Hoffmann technique See the entry FIRST STRAND.

guide RNA (in trypanosomatids) See RNA EDITING.

guide sequence (external guide sequence) See SNRNAS.

guide strand (in ds siRNA) The antisense (effector) strand. (cf. PASSENGER STRAND.)

Guthrie card A card used for collecting and storing specimens of blood. Blood is spotted onto each of several circular areas and then allowed to dry. The card includes an area on which details of the patient etc. can be written.

GVs GRANULOSIS VIRUSES.

GW678248 A potent benzophenone NON-NUCLEOSIDE REVERSE TRANSCRIPTASE INHIBITOR. GW695634, the prodrug, was reported to be under development for treatment of HIV-1 [Antimicrob Agents Chemother (2005) 49(10):4046–4051].

GW695634 See GW678248.

gyrase See TOPOISOMERASE.

H

H (1) A specific indicator of ambiguity in the recognition site of a RESTRICTION ENDONUCLEASE or in another nucleic acid sequence; for example, in GDGCH↓C (enzyme SduI) the 'H' indicates A or C or T. (In RNA 'H' indicates A or C or U.) In the example, 'D' is A or G or T.

(2) L-Histidine (alternative to His).

H-DNA See TRIPLEX DNA.

H′-DNA See TRIPLEX DNA.

H-NS protein See (below) under H-NS PROTEIN.

H-*ras* See RAS.

H strand (of dsDNA) The 'heavy' strand: that strand found in the lower band when a given sample of denatured DNA is subjected to density gradient centrifugation; the other strand is the 'light' (L) strand.

The designation 'heavy' may also refer to a strand of DNA labeled with a heavy isotope.

H1 histone See HISTONE.

H2A, H2B histones See HISTONE.

H3, H4 histones See HISTONE.

H37ra An (avirulent) strain of *Mycobacterium tuberculosis*. In comparison with the (virulent) reference strain (H37rv), the genome of this strain is 8445 bp longer (due to 53 insertions and 21 deletions), many of the genetic changes resulting from variation in repetitive sequences such as IS*6110* [PLoS ONE (2008) 3(6):e2375].

H37rv Reference strain of MYCOBACTERIUM TUBERCULOSIS.

(cf. H37RA.)

Ha-*ras* See RAS.

HAART Highly active antiretroviral therapy, i.e. chemotherapy used in the treatment of AIDS which involves a *combination* of various ANTIRETROVIRAL AGENTS (q.v.). Such a regime should (theoretically) block all replication of HIV, but low-level viremia can occur and may be due e.g. to (i) replication in long-lived HIV-infected cells, and/or (ii) selection of new drug-resistant mutants of the virus.

[Management of HIV-1 infection in adults: Proc Bayl Univ Med Center (2009) 22(1):62–68.]

HAC HUMAN ARTIFICIAL CHROMOSOME (q.v.).

hadacidin ($CHO.NOH.CH_2.CO_2H$) A compound which acts as a structural analog of L-aspartic acid, inhibiting the enzyme adenylosuccinate synthetase and inhibiting the biosynthesis of adenine-containing nucleotides.

Hadacidin has antimicrobial and antitumor activity.

(See also AZASERINE and DON.)

haemophilia See HEMOPHILIA.

Haemophilus A genus of Gram-negative bacteria that occur as parasites (e.g. on mucous membranes) and opportunist pathogens in man and other animals. The cells are coccoid, small rods (e.g. ∼0.4–1.5 μm) or filaments, according e.g. to growth conditions. Growth requires rich media (e.g. *chocolate agar* – i.e. heat-treated blood agar), X FACTOR and/or the V FACTOR (according to species).

The GC% of genomic DNA in species of *Haemophilus* is in the range 37–44.

The type species of the genus is *H. influenzae*.

hair (as a source of DNA) See FORENSIC APPLICATIONS.

hairpin (*DNA technol.*) In a strand of nucleic acid: a region in which adjacent sequences have base-paired with each other to form a local double-stranded section.

Hairpins are formed e.g. in molecules of pre-miRNA and in tRNAs.

(See also PALINDROMIC SEQUENCE.)

halocins BACTERIOCINS produced by certain members of the domain Archaea.

hammerhead ribozyme A type of RIBOZYME that is found e.g. in certain kinds of VIROID and which has been subjected to extensive experimentation and engineering.

Essentially, the hammerhead ribozyme consists of a single, folded (structured) strand of RNA with a central, catalytic component that is flanked by two recognition sequences. The recognition sequences base-pair with (complementary) target sequences which bracket a central triplet – such as GUC, or a similar variant triplet with a central U; cleavage of the target occurs at the 3′ end of this central triplet.

Hammerhead ribozymes have been reported in the 3′ UTR region of mammalian mRNAs; atypically, each of *these* ribozymes occurs in the form of two fragments that are separated by a sequence of (unrelated) nucleotides [Nature (2008) 454: 899–902].

hand-made cloning See the entry SCNT.

handcuffing A proposed mechanism for inhibition of replication (and for control of copy number) in plasmids that contain ITERONS; a specific replication-control protein, which binds to iterons, couples the origins of replication of two plasmids and, in this way, blocks initiation of replication in both of the plasmids.

Handcuffing is one of the proposed mechanisms by which copy number control is achieved in plasmid R6K; in R6K, *dimers* of the (plasmid-encoded) π protein appear to bind to alternate – rather than adjacent – iterons [J Mol Biol (2008) 377(3):609–615].

Handcuffing may contribute to replication control in the P1 plasmid.

Hantavirus A genus of enveloped, ssRNA viruses of the family BUNYAVIRIDAE (q.v.).

haploid See PLOIDY.

haploinsufficiency An effect in which a loss of function in one of a pair of alleles causes a change in phenotype in a diploid cell – the type of change depending on gene and conditions.

[Identification of (human) haploinsufficient genes and their proximity to segmental duplications within the genome: Eur J Hum Genet (2008) 16:1350–1357.]

Haploinsufficiency associated with tumor suppressor gene *SMAD4* was reported in a substantial fraction of the juvenile

polyps in patients with germline *SMAD4* mutations. SMAD4 dosage-dependent transcriptional changes were reported in a subset of target genes in the intestinal tissues of patients with juvenile polyposis syndrome [Pathogenetics (2008) 1:2 (doi: 10.1186/1755-8417-1-2)].

haplotype A term which is used in various (distinct) ways by different authors in different contexts.

The meanings given to the term 'haplotype' include:

(1) The specific sequence of alleles in a given chromosome, part of a chromosome, or another, specified region of DNA; of these alleles, any given allele is from either the maternal or paternal parent.

(2) Alleles encoding the major histocompatibility complex, components of the immune system such as cell-surface antigens on leukocytes. In humans, these alleles (on chromosome 6) are designated HLA (human leukocyte antigen).

(3) The complete complement of MHC *antigens* encoded by genes derived from *a given* parent (i.e. of either maternal or paternal origin).

hapten (*immunol*.) A molecule which, on it own, does not elicit an antibody response – but which can be made to do so by linking it to another molecule or particle (referred to as the *carrier*). An isolated hapten (i.e. one without the carrier) is able to combine with antibodies produced in response to the hapten–carrier complex.

Hardy–Weinberg equilibrium In a given, large population, if the frequency of the alleles A and a be represented by p and q ($q = 1 - p$), respectively, then, in general, the expected ratio of the genotypes AA, Aa and aa is given by p^2:$2pq$:q^2. [Exact tests for Hardy–Weinberg proportions: Genetics (2009) doi: 10.1534/genetics.109.108977.]

HAT Histone acetyltransferase: see the entry ACETYLATION (of histones).

HAT medium See HYBRIDOMA.

HC2 (hc2) An abbreviation for the DIGENE HYBRID CAPTURE 2 ASSAY.

HCCS1 See HEPATOCELLULAR CARCINOMA SUPPRESSOR 1.

HCV pseudoparticle See PSEUDOPARTICLE.

HDA HELICASE-DEPENDENT AMPLIFICATION.

HDAC Histone deacetylase: any enzyme which removes acetyl groups from histones (see ACETYLATION (of histones)).

Inhibitors of HDACs, used e.g. for manipulating the acetylation status of histones, include 4-phenylbutyric acid, sodium butyrate, TRICHOSTATIN A and 2-propylpentanoic acid (also called valproic acid).

The use of specific HDAC inhibitors has been suggested as a means of addressing the problem of HIV latency – see entry AIDS.

headful mechanism See e.g. the entry PHAGE T4.

headloop suppression PCR A form of PCR used for selectively suppressing the amplification of one or more related sequences without affecting amplification of the (required) target sequence; it is reported to be particularly useful when a rare target sequence occurs among a large number of copies of a closely related, but unwanted, sequence.

Each primer has a 5′ tag (the 'head') which can bind to an internal region of the *unwanted* sequence. After extension of a primer, the head can loop over and hybridize with the complementary region in an unwanted sequence, thereby forming a hairpin-type structure which is not readily amplifiable. The head does not hybridize with the required target sequence as there is no complementary region; accordingly, the required sequence is amplified normally.

Headloop suppression PCR has been used with BISULFITE-treated DNA for the (selective) amplification of sequences containing methylated cytosines (the amplification of non-methylated sequences being suppressed) [Nucleic Acids Res (2005) 33(14):e127].

heat-activatable primers See HOT-START PCR.

heated-lid cycler See THERMOCYCLER.

heavy strand (of dsDNA) See H STRAND.

HEK cells Human embryonic kidney cells.

helicase (DNA helicase; unwinding protein) A type of protein which can separate (unwind) the strands of a DNA duplex, in an ATP-dependent fashion.

The *Escherichia coli* helicases include the Rep protein (*rep* gene product); helicase II (= UvrD protein), involved e.g. in DNA repair (see UVRABC-MEDIATED REPAIR); and the DnaB protein (*dnaB* gene product), a hexameric ring of which is involved in chromosomal replication. Another *E. coli* helicase is encoded by the RECG GENE (q.v.).

The F plasmid's helicase I (encoded by gene *traI*) may be used to nick the plasmid's *oriT* site and unwind plasmid DNA during bacterial conjugation.

Some phages use the host cell's helicase – while others, e.g. T4 (the *dda* gene product) and T7 (gp4), encode their own helicases.

Helicases of the RECQ PROTEIN family (q.v.) are found in a wide range of organisms. Studies on (human) RecQ1 helicase identified a putative 'strand-separation pin' [Proc Natl Acad Sci USA (2009) 106(4):1039–1044].

A thermostable helicase (from *Thermoanaerobacter tengcongensis*) is used in HELICASE-DEPENDENT AMPLIFICATION.

Some enzymes which are referred to as 'helicases' may not be able to unwind DNA. For example, the type IC restriction endonuclease EcoR124I, which has ATP-dependent translocation, is reported to be able to move along a DNA molecule even when there are interstrand cross-links (preventing strand separation); translocation in this enzyme appears to depend primarily on contact with the sugar–phosphate backbone and bases in the 3′–5′ strand – although contact with the 5′–3′ strand appears to be important for stabilizing the enzyme's motor on the DNA [EMBO J (2006) 25:2230–2239].

helicase I A HELICASE encoded by the F PLASMID (gene *traI*).

helicase II *Syn*. UvrD (see entry UVRABC-MEDIATED REPAIR).

helicase-dependent amplification (HDA) An isothermal DNA AMPLIFICATION method in which the need for temperature cycling is avoided by the use of a helicase (which separates the strands of dsDNA to reveal target sequence(s)). [Original description of method: EMBO Reports (2004) 5(8):795–800.]

The specificity and performance of HDA were improved by carrying out the process at higher temperatures (~60–65°C). This involved using a thermostable helicase from the thermophilic prokaryote *Thermoanaerobacter tengcongensis* [J Biol Chem (2005) 280(32):28952–28958].

HDA has been used for detecting the *nuc* and *mecA* genes in *Staphylococcus aureus* in order to detect methicillin resistance directly from blood cultures [J Clin Microbiol (2008) 46 (4):1534–1536].

Helicobacter pylori A species of Gram-negative bacteria; the cells are typically curved or spiral rods up to ~3 μm in length. This organism is catalase-positive, oxidase-positive and also (strongly) urease-positive.

H. pylori is associated e.g. with gastric disease in man – but the organism is apparently present in the stomach (without symptoms) in a significant part of the human population.

The (circular) genome of *H. pylori* is ~1.6 Mb.

The GC% of the genomic DNA is ~39.

helix (double helix) See B-DNA.

helix–coil transition (of dsDNA) Transition from an ordered, helical conformation to a 'random coil'; this generally refers to MELTING.

(See also THERMAL MELTING PROFILE.)

helix-destabilizing protein A term sometimes used to refer to a SINGLE-STRAND BINDING PROTEIN.

helix-to-coil transition *Syn*. HELIX–COIL TRANSITION (q.v.).

helix-turn-helix A structural feature found in certain types of DNA-binding protein: two helical regions linked by a β-turn.

helper phage Any phage (such as M13K07) which, on infecting a cell, can promote replication of a vector (e.g. a PHAGEMID) containing a phage f1 or similar origin of replication, leading to encapsidation of single-stranded copies of vector DNA in phage particles. Such *single-strand rescue* has been used e.g. to prepare samples of DNA for sequencing.

In the process of infecting a bacterial cell, the helper phage M13 (see PHAGE M13) first binds to the tip of the cell's PILUS (q.v.); internalization of phage DNA follows.

In some cases one type of helper phage may be found to be more efficient than another: see e.g. phage KM13 in the entry M13K07.

helper virus See e.g. AAVS.

hemagglutinin (viral) In certain viruses (e.g. influenza viruses A and B, some members of the family PARAMYXOVIRIDAE): a type of protein which projects from the viral envelope and which can bind to receptor sites on red blood cells (RBCs); such viruses can therefore *agglutinate* RBCs, hemagglutinins on the virions linking the RBCs together.

Some of the viruses which agglutinate RBCs (e.g. influenza viruses A and B) spontaneously elute from the RBCs after a period of time. This is due to the activity of the viral enzyme NEURAMINIDASE which destroys the virus-binding sites on the RBCs, thus breaking the virion–RBC links. (RBCs from which viruses have been eluted cannot again be agglutinated by the same strain of virus.)

Antigenically distinct forms of viral hemagglutinin are designated H1, H2, H3 etc. – see the entry INFLUENZAVIRUS.

hemiphosphorothioate site See e.g. SDA.

hemophilia Any of several distinct types of disorder involving a failure of the normal blood-clotting (coagulation) process: see GENETIC DISEASE (table). Recombinant proteins are used for treatment/prophylaxis (see e.g. NovoSeven® in the table in BIOPHARMACEUTICAL).

hemorrhagic colitis See EHEC.

hemorrhagiparous thrombocytic dystrophy The BERNARD–SOULIER SYNDROME.

HEPA filter See SAFETY CABINET (class I).

hepatitis C virus pseudoparticle See PSEUDOPARTICLE.

hepatocellular carcinoma suppressor 1 A gene which, when inserted into, and expressed in, human hepatocellular carcinoma cells, suppressed the cells' ability to produce tumors in nude mice. On insertion into an oncolytic ADENOVIRUS, the *HCCS1*-augmented virus was reported to inhibit cancer cells *in vitro* and *in vivo* and to bring about complete regression of 50% of established tumor xenografts in these mice [Cancer Gene Therapy (2008) doi: 10.1038/cgt.2008.52].

HER2 See EPIDERMAL GROWTH FACTOR RECEPTOR FAMILY; see also NEU.

Herceptin® See HER2 in entry EPIDERMAL GROWTH FACTOR RECEPTOR FAMILY.

hESC Human *embryonic* STEM CELL.

heteroauxin (auxin) See the entry AUXINS.

heterochromatin See the entry CHROMATIN.

heterodimer Any composite molecular entity consisting of two *non-identical* subunits. (cf. HOMODIMER.)

heteroduplex (1) A double-stranded nucleic acid in which each of the strands is derived from a different parent duplex – the strands may or may not be fully complementary.

(See also HOMOLOGOUS RECOMBINATION.)

(2) Any double-stranded nucleic acid in which the strands are not fully complementary.

heterogeneous nuclear RNA (hnRNA) Within a (eukaryotic) nucleus: various types of RNA that include pre-mRNA and the RNA components of ribonucleoproteins.

(See also HNRNA.)

heterogeneous test A nucleic-acid-amplification test in which the amplified product is transferred from the reaction mixture (e.g. to a gel) for detection/quantitation.

(cf. HOMOGENEOUS TEST.)

heterogenote See MEROZYGOTE.

heterologous From a source other than the organism, or entity, under consideration. For example, expression of a protein-encoding mammalian gene in a bacterium would involve the synthesis of a heterologous protein in the bacterial cell.

heteromerous See MACRONUCLEUS.

heteronuclear RNA See HNRNA, sense 2.

heteroplasmy Refers to the condition in which both wild-type and mutant copies of given mitochondrial gene(s) coexist in the same cell. (cf. HOMOPLASMY.)

The ratio of mutant to wild-type copies of mtDNA appears to be a major factor in determining the development or other-

wise of clinical disease; this ratio can vary between different tissues in the body. mtDNA recombination has been reported from studies on skeletal muscle in heteroplasmic individuals [Nature Genet (2005) 37:873–877]. [Heteroplasmy in *Caenorhabditis elegans*: BMC Genet (2007) 8:8.]

Heteroplasmic tandem repeat arrays in halibut fish (species of *Hippoglossus* and *Reinhardtius*) were believed to be maintained by SLIPPED-STRAND MISPAIRING and DNA recombination mechanisms [BMC Genomics (2008) 9:10].

Pathogenic mtDNA mutations are reported to be common in the general population [Am J Hum Genet (2008) 83(2):254–260].

(See also CYBRID.)

heteroploid (1) Refers to any cell (or organism) in which the number of chromosomes is different from that characteristic of the species. (cf. EUPLOID, sense 1.)

(See also ANEUPLOID.)

(2) Refers to any cell or organism which has a complement of chromosomes other than the haploid or diploid number.

heterosome (heterochromosome; also sex chromosome) A sex-linked chromosome – e.g. the human X or Y chromosome.

(cf. AUTOSOME.)

heterothallism (*fungal genetics*) The need for two (different) thalli in the process of sexual reproduction.

Morphological heterothallism (*dioecism*) refers to the need for separate male and female thalli.

Physiological heterothallism involves compatibility of the sexual partners which is based on MATING TYPE (q.v.).

heterozygous See the entry ALLELE.

hexadimethrine bromide See the entry POLYBRENE.

hexidium iodide A fluorescent (phenanthridinium) dye which binds to DNA. The cell membrane in mammalian cells – and in Gram-*positive* bacteria – is permeable to hexidium iodide.

(See also DNA STAINING and LIVE BACLIGHT™ BACTERIAL GRAM STAIN KIT.)

Hfq protein An RNA chaperone: see entry SRNAS (*Prokaryotic sRNAs*).

Hfr donor See the entry F PLASMID.

Hi-Res Melting™ A system for DNA melting analysis (Idaho Technology Inc., Salt Lake City UT, USA) which provides an enhanced level of information; it involves an instrument called the LightScanner®.

(See also HRM.)

high frequency of recombination donor See F PLASMID.

high-mobility group proteins (HMG proteins) See HMG BOX.

high-resolution melt See the entry HRM.

high-stringency conditions See e.g. PROBE and STRINGENCY.

(See also PCR.)

highly active antiretroviral therapy See the entry HAART.

HIGM1, HIGM2 See HYPER-IGM SYNDROME.

***Himar1* transposon** See the entry MARINER FAMILY.

***HimarFT* transposon** See the entry MARINER FAMILY.

hirudin A potent anticoagulant found in the leech, *Hirudo medicinalis*; it binds to, and inhibits, thrombin.

A recombinant form of hirudin, which is used therapeutically, is produced in the yeast *Saccharomyces cerevisiae* – see Refludan® in the entry BIOPHARMACEUTICAL (table).

***his* operon** See OPERON.

His tag See SIX-HISTIDINE TAG.

His-Cys box endonucleases See HOMING ENDONUCLEASES.

histidine kinase A kinase in which the active site contains a histidine residue. Histidine kinases are involved e.g. in TWO-COMPONENT REGULATORY SYSTEMS.

histidine operon See OPERON.

(See also AMES TEST.)

histone A type of small, basic protein present in CHROMATIN; there are three main groups: H1 (lysine-rich); H2A and H2B (slightly lysine-rich); H3 and H4 (arginine-rich).

(See also the entry ACETYLATION (of histones).)

Constitutive, non-allelic, histone *variants*, which can differ in different species, can alter both structure and dynamics of chromatin. Unlike other histones, these are formed throughout the cell cycle and they may affect e.g. DNA transcription, replication and repair [Genes Dev (2005) 19(3):295–316].

Genome-wide plotting of the yeast variant histone H2A.Z indicated global localization to *inactive* genes and a role in nucleosome positioning [PLoS Biol (2005) 3(12):e384].

Vertebrates contain two versions of variant H2A.Z protein, referred to as H2A.Z-1 and H2A.Z-2, which differ by three amino acids. [The evolutionary differentiation of the H2A.Z-1 and H2A.Z-2 variants: BMC Evol Biol (2009) 9:31.]

[High-resolution dynamic mapping of histone–DNA interactions in a nucleosome: Nature Struct Mol Biol (2009) doi: 10.1038/nsmb.1526.]

histone acetyltransferase See ACETYLATION (of histones).

histone deacetylase See HDAC, and see also ACETYLATION (of histones) and TRICHOSTATIN A.

histone-like nucleoid-structuring protein See H-NS PROTEIN.

HIV-1 Human immunodeficiency virus-1: a virus of subfamily LENTIVIRINAE (family Retroviridae – see RETROVIRUSES for general information); infection with HIV-1 can lead to AIDS. (For chemotherapy of AIDS see the entry ANTIRETROVIRAL AGENTS; see also HAART.)

HIV-1 was linked to AIDS in 1983. Later, another (distinct but related) retrovirus was associated with AIDS patients in West Africa; this virus was designated HIV-2.

The HIV-1 virion is approx. 100 nm in diameter (see entry RETROVIRUSES for basic structure and genome).

Interaction with (susceptible) cells occurs via the envelope glycoprotein gp120, which is located at the distal (outermost) end of a trans-envelope protein, gp41. The main high-affinity binding site for gp120 is cell-surface antigen CD4 (found e.g. on 'helper' T lymphocytes). However, the entry of HIV-1 into cells depends not only on gp120–CD4 binding but also on the presence of particular cell-surface receptors for chemokines (e.g. CXCR4 and CCR5) – with which gp120 also interacts; thus, the CXCR4 coreceptor appears to be required for entry of T-cell-tropic strains of HIV-1, while the CCR5 coreceptor is needed for entry of M-tropic strains of HIV-1 – M-tropic strains being those which are able to infect cells of both the T

lymphoid and macrophage/monocyte lineages. Some ('dual-tropic') strains of HIV-1 can use both of these coreceptors. Strains of HIV-1 have been categorized as X4, R5 or R5X4 on the basis of their use of these coreceptors. (Mutations have been induced in the human *CCR5* gene, and it was suggested that this technology may be used for producing HIV-resistant cells [Genome Res (2009) 19:1279–1288].)

In addition to the CXCR4 and CCR5 coreceptors, other co-receptors, such as CCR3 and FPRL1, have been investigated. Thus, e.g. HIV-1 envelope proteins from 66 patients infected with the major prevalent subtypes of HIV-1 were studied to determine whether virus uptake by NP-2 cell lines (which express most of the known alternative coreceptors for HIV-1) differed according to viral subtype; the results of this study suggested e.g. that the CCR3 coreceptor may be used as a substitute by HIV-1 subtype B, and that FPRL1 may be used as a substitute by the HIV-1 subtypes A and C [J Virol (2009) 83(17):8353–8363].

The (dual) involvement of gp120 and gp41 in the infection process is discussed in a proposed model of interactions at the gp120–gp41 interface [J Mol Biol (2008) 376:786–797].

[Interactions between HIV-1 and human proteins: Nucleic Acids Res (2009) 37(Database issue):D417–D422.]

[High-level recombination in HIV-1 during reverse transcription: Microbiol Mol Biol Rev (2009) 73(3):451–480.]

Following reverse transcription of viral RNA (see RETROVIRUSES) the dsDNA provirus is circularized before integration into the host's chromosome. Integration involves the viral integrase, an enzyme encoded by the *pol* region of the viral genome. (See also ELVITEGRAVIR.)

Once integrated in the host's chromosome the provirus may remain quiescent (see note in the entry AIDS). Alternatively, depending on circumstances (e.g. its location in the chromosome), it may be transcribed and give rise to progeny virions (see RETROVIRUSES). (See also URACIL-N-GLYCOSYLASE.)

In patients infected with *both* HIV-1 and *Mycobacterium tuberculosis*, replication of HIV-1 was reported to be subject to a strain-specific influence from the infecting strain of *M. tuberculosis* [PLoS ONE (2009) 4 (7):e6116].

The current classification of HIV-1 into subtypes A, B, C. . . and also 'circulating recombinant forms' (CRFs) may be artifactual [J Virol (2007) 81(16):8543–8551].

HIV-2 See the entry HIV-1.

HLA Human leukocyte antigen (see HAPLOTYPE, sense 2).

HMG box The DNA-binding region of the small, non-histone *high-mobility group proteins* of eukaryotes; it consists of a region of ~80 amino acid residues. Like the LrpC protein in the (prokaryote) *Bacillus subtilis* [Nucleic Acids Res (2006) 34(1):120–129], HMG proteins apparently bind to HOLLIDAY JUNCTIONS.

HMG proteins High-mobility group proteins: see HMG BOX.

hMUTYH See the entry 8-OXOG.

HNH endonucleases See HOMING ENDONUCLEASE.

hnRNA (1) HETEROGENEOUS NUCLEAR RNA.

(2) *Heteronuclear* RNA: used in the book *From Genes to Genomes*, 2nd edition [ISBN 978-0-470-01734-0] page 175, to refer to the *primary transcript*, i.e. PRE-MRNA (q.v.).

H-NS protein ('histone-like nucleoid-structuring protein') The small, multi-functional prokaryotic protein that is associated e.g. with condensation of DNA into the compact NUCLEOID [Nucleic Acids Res (2000) 28:3504–3510]. It also appears to be needed for transposition of IS*1* (insertion sequence 1) [J Bacteriol (2004) 186:2091–2098], for temperature-dependent repression of P fimbriae in *Escherichia coli* – and for transcriptional regulation, often repression, of various other genes. [Regulation of the sigma factor RpoS (by H-NS): J Bacteriol (2006) 188(19):7022–7025.] The H-NS protein also appears to repress transcription (from the P_Y promoter) of the transfer (*tra*) operon of the F plasmid [J Bacteriol (2006) 188(2):507–514].

Homologs of H-NS occur in many Gram-negative bacteria, and, although sequence homology may be rather low, these homologs appear to have analogous DNA-binding properties [J Bacteriol (2005) 187(5):1845–1848].

H-NS was reported selectively to silence foreign DNA with an [A + T] content higher than that of the resident genome [Genes Dev (2007) 21:1456–1471].

H-NS interacts with the transpososome and promotes transposition by Tn*5* [Nucleic Acids Res (2009) 37(2):309–321].

(cf. HU PROTEIN.)

HO endonuclease See MATING TYPE.

Hoechst 33342 A fluorophore, used for DNA STAINING, which binds to the minor groove at AT-rich sequences; it is able to permeate cells more rapidly than some related dyes, and is used e.g. in association with the FLUORESCENCE-ACTIVATED CELL SORTER.

Hoechst 33342 fluoresces blue; excitation can be achieved e.g. with the mercury-arc lamp or with the 325 nm spectral line of an He-Cd laser.

Hoechst 33342 is used e.g. in the (veterinary) technique for SEX SORTING of spermatozoa.

Hogness box See CORE PROMOTER.

Holliday junction (Holliday structure) A conformation formed by two DNA duplexes as an essential intermediate in the process of homologous recombination (which occurs e.g. during meiosis). In the Holliday junction, one strand in one duplex is covalently linked to (continuous with) a strand in the other duplex, and a strand in the other duplex is continuous with a strand in the first duplex; the two duplexes are therefore held together by two single-stranded regions of DNA. So-called *branch migration* of a Holliday junction leads to formation of stretches of heteroduplex DNA: duplex DNA containing one strand from each of the participating duplexes. Resolution of a Holliday junction (i.e. restoration of two separate duplexes) involves the introduction of a nick in each of the strands by a junction-resolving enzyme, followed by ligation – which thus completes recombination.

Holliday junctions are also formed during site-specific recombination mediated by the phage lambda-encoded integrase (Int) protein (involved in integration of the phage genome

into the bacterial chromosome) – as well as in other instances of site-specific recombination mediated by tyrosine recombinases.

Resolution of Holliday junctions in homologous recombination is clearly important for the integrity of the genome, and resolution of these structures in site-specific recombination mediated by tyrosine recombinases is also important because this kind of recombinational event is involved e.g. in certain kinds of regulation of gene expression.

Resolution of Holliday junctions can be inhibited by certain peptides (e.g. WRWYCR, KWWCRW) which antagonize the junction-processing enzymes; these peptides are also reported to be inhibitory for certain other recombinases, including the Cre protein (see CRE–LOXP SYSTEM) and the Flp protein (see FLP) [Nucleic Acids Res (2008) 36(16):5319–5334].

The WRWYCR peptide inhibitor has been used in a study on phage λ integrase [J Biol Chem (2008) 283(18):12402–12414].

(See also HMG BOX.)

Holliday structure *Syn*. HOLLIDAY JUNCTION.

homeobox A conserved (180-bp) sequence found within HOX GENES.

homeotic gene Any gene which is concerned with the identity of a given part of an organism.

(See also HOMEOBOX, HOX GENE and POLYCOMB-GROUP GENES.)

homing endonuclease (HE) Any of various highly site-specific endonucleases encoded by genes which are embedded within 'intervening sequences' (see INTRON HOMING and INTEIN for a brief account of their activities). Homing endonucleases are encoded by many types of microorganism, including bacteria, archaeans, bacteriophages and single-celled eukaryotes. (The homing endonucleases are sometimes called *meganucleases*.)

These enzymes recognize target sites of between 15 and 40 bp in length; some base variation may be tolerated.

There are at least five families of homing endonucleases on the basis of primary sequence homology. Two of the families (LAGLIDADG and His-Cys box) occur in eukaryotes and/or archaeans. The GIY-YIG and HNH endonucleases are found primarily in phages. HEs of the PD-(D/E)XK family occur in bacterial genomes. HEs in the LAGLIDADG family have the highest site-specificity, and are reported to recognize target sites of about 20–24 bp in length. Many of the LAGLIDADG (and His-Cys box) HEs are reported to occur within introns inside rRNA genes.

Single-chain versions of LAGLIDADG enzymes have been engineered from native homodimeric precursors; this should facilitate their use e.g. for gene targeting procedures [Nucleic Acids Res (2009) 37(5):1650–1662].

Homing endonuclease genes are currently being considered (as one approach) for the creation of transgenic mosquitoes in the control of malaria [PLoS Med (2009) 6(2):e1000020].

(See also I-SCEI.)

homodiaphoromixis See SECONDARY HOMOTHALLISM.

homodimer Any composite molecular entity consisting of two *identical* subunits.

(cf. HETERODIMER.)

homodimixis See SECONDARY HOMOTHALLISM.

homogeneous test Any nucleic-acid-amplification test which is carried out entirely in a single, closed vessel. REAL-TIME PCR in the LIGHTCYCLER apparatus is one example (the products being detected/quantitated during the reaction).

(cf. HETEROGENEOUS TEST.)

homogenote See MEROZYGOTE.

homoheteromixis *Syn*. SECONDARY HOMOTHALLISM.

homoiomerous See MACRONUCLEUS.

homologous (1) Refers to the similarity or 'relatedness' of given samples of nucleic acid in terms of nucleotide sequence; thus, if two samples of nucleic acid have identical sequences then they are said to be 100% homologous.

(2) *Syn*. Complementary.

(3) Refers to the *corresponding* sequence(s) of nucleotides in nucleic acids from different sources, i.e. sequences encoding similar functions – although not necessarily identical in terms of nucleotide sequence.

(4) Refers to the two chromosomes that form each pair (so-called 'homologous pair') in meiosis.

(5) (of an antibody) Refers to a specific antibody which binds to a given antigen.

homologous recombination A type of RECOMBINATION which requires an extensive region of sequence homology between the two participating duplexes. Homologous recombination is involved in events such as CROSSING OVER in meiosis and (in some cases) the insertion of a vector-borne gene or fragment into genomic DNA.

(cf. RECOMBINEERING.)

(The ability of *non-homologous DNA end-joining* (NHEJ), a normal DNA repair process, to compete with (i.e. interfere with) homologous recombination, was studied by a procedure involving inactivation of the KU70 protein (q.v.).)

A feature of homologous recombination is the development of a so-called *nucleoprotein filament*: a structure consisting of the free 3′ end of a strand of DNA coated with monomers of a protein which has the ability to promote the juxtaposition (*synapsis*) of homologous sequences of DNA; this leads to the formation of a hybrid duplex (*heteroduplex*) between the strand in the nucleoprotein filament and its complementary strand in the target duplex. In bacteria, the strand within the nucleoprotein filament is coated with RECA protein; in other types of cell the DNA strand may be coated with e.g. RADA or RAD51.

In *Escherichia coli*, the development of the nucleoprotein filament may be initiated by the RecBCD system or the RecF system – or by a hybrid system consisting of elements from both of these 'recombination machines'.

RecBCD is a large multisubunit protein which exhibits both helicase and nuclease functions; it also has a synaptogenic function. RecBCD appears to be involved e.g. in the repair of double-stranded breaks.

The RecF system, involving a number of separate proteins,

appears to be responsible e.g. for the repair of single-stranded gaps in circular molecules.

Following synapsis, and the formation of a heteroduplex, the process of homologous recombination proceeds to the development of an intermediate structure called a HOLLIDAY JUNCTION (q.v.) which is subsequently resolved to complete the process of recombination.

(Enzymes that mediate resolution of the Holliday junction can be inhibited by certain peptides which may be useful for experimentation in this area: see HOLLIDAY JUNCTION.)

In *Escherichia coli* homologous recombination is facilitated in certain regions of the chromosome. Thus, recombination is enhanced up to 10-fold in regions containing a so-called chi (χ) site:

5′-GCTGGTGG-3′

with which RecBCD has been found to interact. Chi sites are commonly found at or near the 3′ terminus in a nucleoprotein filament.

(See also entries INSERTION–DUPLICATION RECOMBINATION and INTEIN.)

homology (1) In general usage: similarity or relatedness. Thus, e.g. 'homologous chromosomes' refers to chromosomes, in a given nucleus, that contain a similar or identical sequence of genes. When comparing e.g. fragments or molecules of nucleic acid, sequences of nucleotides may be described as exhibiting a certain degree of homology (e.g. 'the sequences are 80% homologous') – meaning that, in this case, 80% of the nucleotides in one sample are the same as those, in the corresponding positions, in the other sample(s).

(See also HOMOLOGOUS.)

(2) In some cases the term is given a meaning which is not equivalent to the general concept of 'level of similarity'. This meaning refers specifically to the relatedness of two or more sequences which have evolved, divergently, from a common ancestral sequence.

homology arm One of the two terminal sequences that bracket a given fragment of DNA which is designed to be inserted, by recombination, into a given target molecule that contains the two corresponding sequences at the target site.

Homology arms can be quite short; for example, arms of 36–50 nucleotides in length are used in the protocol referred to in the entry LAMBDA (λ) RED RECOMBINATION.

homomixis *Syn.* HOMOTHALLISM.

homoplasmy Refers to the (normal) condition in which there is a single type of mitochondrial DNA in a given cell (i.e. not a mixture of wild-type and mutant mitochondrial genes).

(cf. HETEROPLASMY.)

homopolymer tailing See TAILING.

homothallism (homomixis) (*fungal genetics*) Self-fertility – in which there is no barrier to the fusion of gametes etc. from the same individual; MATING TYPES are not involved.

(cf. SECONDARY HOMOTHALLISM and HETEROTHALLISM.)

homozygous See ALLELE.

Hong Kong flu See INFLUENZAVIRUS.

Hoogsteen base-pairing Base-pairing in which the pattern of hydrogen bonding differs from that in Watson–Crick base-pairing. Hoogsteen base-pairing occurs e.g. in TRIPLEX DNA.

horizontal transmission (of genes) The transmission of genes from one individual to another, contemporary, individual. An example is the transmission of genes from one bacterium to another by the process of CONJUGATION.

(cf. VERTICAL TRANSMISSION.)

horseradish peroxidase (HRP) A hemin-containing peroxidase used e.g. for enzyme-amplified immunodetection procedures (immunoperoxidase assays) in which an antibody-conjugated peroxidase is used to locate a specific antigen e.g. on cells or tissues. HRP produces a brown precipitate with the substrate 3,3′-diaminobenzidine tetrahydrochloride dihydrate (DAB).

This peroxidase is also used e.g. in the TYRAMIDE SIGNAL AMPLIFICATION procedure and in an assay for pyrophosphate (see PIPER PYROPHOSPHATE ASSAY KIT).

(See also ZENON ANTIBODY-LABELING REAGENTS.)

hot-start PCR A form of PCR in which cycling does not begin until the temperature of the reaction mixture is, for the first time, sufficient to inhibit non-specific binding of primers; the object is to avoid *mispriming* – i.e. the binding of primers to sequences other than their fully complementary binding sites. Such avoidance (or minimization) of mispriming promotes specificity of the assay and reduces the background against which the actual target is to be detected. This approach may also enhance sensitivity by minimizing wastage of potential on the formation of non-specific products. Various types of procedure have been used for the hot-start approach.

One commercial system uses a modified form of the DNA polymerase that is inactive when added to the reaction mixture; activation of the enzyme occurs at the initial stage of denaturation when a temperature of ∼95°C is reached.

In another system, the polymerase is inactivated by its own antibody. Inactivation is lost as the temperature rises for the initial stage of denaturation.

(See also PLATINUM TAQ DNA POLYMERASE.)

Yet another method involves the use of a physical barrier of wax (AmpliWax™, Applied Biosystems) which – initially – separates some of the components of the mixture from others (above and below the layer of wax). Cycling begins once the wax has melted (at 75–80°C).

[AmpliWax™ use (examples): Nucleic Acids Res (2008) 36(10):e59; J Vision (2008) 8(2):5.1–5.9.]

Hot-start PCR can also be carried out with heat-activatable primers. In this study, the 3′-terminal and 3′-penultimate sites of each primer carried a (heat-labile) 4-oxo-1-pentyl (OXP) phosphotriester group which, at lower temperatures, inhibited primer extension. On reaching higher temperatures, the OXP moiety was lost, and the primer, now carrying an unmodified 3′ terminal region, was extended (as normal) by DNA polymerase [Nucleic Acids Res (2008) 36(20):e131].

(See also PRIMER SEQUESTRATION.)

housekeeping gene (constitutive gene, or reference gene) Any gene that is expressed constitutively in a given type of active/functional cell, such expression being essential for the cell's normal activity.

Examples of housekeeping genes include those that encode species of rRNA.

***HOX* gene** (*Hox* gene) A generic designation for a HOMEOTIC GENE (which contains a HOMEOBOX).

(See also POLYCOMB-GROUP GENES.)

hpi Hours post-infection, i.e. the number of hours since the start of infection.

HPr kinase See CATABOLITE REPRESSION.

HPr protein See PTS.

***hprK* gene** See CATABOLITE REPRESSION.

HPRT deficiency *Syn*. LESCH–NYHAN SYNDROME.

HPV16 A (high-risk) strain of human PAPILLOMAVIRUS.

(See also DIGENE HYBRID CAPTURE 2 ASSAY.)

hrGFP Humanized *Renilla* GFP: a HUMANIZED, recombinant form of GREEN FLUORESCENT PROTEIN from the invertebrate *Renilla reniformis* which is used (in mammalian cells) as a less-toxic alternative to the GFP from *Aequorea victoria*.

hrGFP is available commercially (Stratagene, La Jolla CA, USA).

Examples of use: studies on intercell mitochondrial transfer [Proc Natl Acad Sci USA (2006) 103(5):1283–1288]; the activation of the nuclear receptor FXR [Proc Natl Acad Sci USA (2006) 103(4):1006–1011]; and studies on adeno-associated virus (AAV) vectors [Virol J (2009) 6:3].

(See also PATHDETECT SYSTEMS.)

HRM™ High-resolution melt: DNA melting analysis (Corbett Life Science) in which the methodology employed permits an increased amount of information to be obtained in the analysis; the Rotor-Gene™ 6000 instrument is a real-time rotary analyzer used in this procedure.

(See also HIGH-RES MELTING.)

HRP HORSERADISH PEROXIDASE.

***hsd* genes** In *Escherichia coli*: genes encoding the functions of a type I RESTRICTION ENDONUCLEASE.

Gene *hsdR* is required for restriction activity; it is transcribed from the promoter P_{RES}.

Genes *hsdM* and *hsdS* are both transcribed from promoter P_{MOD}.

Gene *hsdM* is required for the modification function.

Gene *hsdS* is the specificity determinant for both *hsdR* and *hsdM*. A null mutation in *hsdS* abolishes both restriction and modification activities.

hSos A human guanosine nucleotide exchange factor used e.g. in the CYTOTRAP TWO-HYBRID SYSTEM.

HSV1716 A strain of herpes simplex virus modified e.g. for use as an ONCOLYTIC VIRUS (q.v.).

hTERT Human telomerase reverse transcriptase.

(See also TELOMERASE.)

***htpR* gene** *Syn*. *rpoH* gene: see ESCHERICHIA COLI (table).

HTS assay High-throughput screening assay.

Hu protein One of a number of RNA-binding proteins which are involved in neuronal differentiation and are associated (as targets) with human autoimmune disease.

Hu proteins were reported to be associated with alternative splicing in the NF1 gene – loss of function in which results e.g. in neurofibromatosis [Mol Cell Biol (2008) 28(4):1240–1251].

Note. There is a possibility of confusion with the HU protein: see next entry.

HU protein In various bacteria: a small DNA-binding, multifunctional protein which is associated with compaction of the NUCLEOID and events such as replication of the chromosome and site-specific recombination.

The HU protein is also involved in regulating a number of OPERONS which are associated with responses to (e.g.) acid stress, anaerobiosis and high osmolarity [PLoS ONE (2009) 4(2):e4367].

HU was reported to bind, and fold, ssDNA [Nucleic Acids Res (2008) 36(3):1026–1036].

The binding of the *Helicobacter pylori* HU protein to DNA apparently differs from that of the *E. coli* protein in that e.g. it does not exhibit a strong preference for DNA with mismatches [Biochem J (2004) 383(2):343–351].

In vitro, the HU protein is reported to dictate the morphology of the DNA condensates formed by crowding agents and polyamines – favoring a shift from toroidal to rod-like forms [Nucleic Acids Res (2007) 35(3):951–961].

Note. According to the discoverers of this protein, the name 'HU' derives from 'histone-like' and 'U93' (the designation of a bacterial strain).

human artificial chromosome (HAC) A construct, analogous to the YEAST ARTIFICIAL CHROMOSOME, which can replicate extrachromosomally in human cells; when used as a VECTOR, one major advantage is that it avoids the problem of insertional mutagenesis that may occur with other types of vector which insert into the genome.

Two basic approaches have been used in the preparation of HACs: (i) engineering an existing chromosome to a miniaturized version which retains essential biologic features, and (ii) the assembly of specific cloned chromosomal components, *in vitro*, to form a functional replicon. An essential part of any HAC is a centromere/kinetochore structure; this is necessary for mediating chromosome–spindle interaction in mitosis. To promote replication, some authors have included a fragment of genomic DNA greater than ~100 kb; this is based on the understanding that the mammalian genome has, on average, one origin of replication in every 100 or so kilobases. Linear HACs also require telomeric sequences (see TELOMERE), but this is not required for covalently closed circular HACs.

One report has described an HAC vector incorporating the entire (2.4 Mb) gene of human dystrophin (see DUCHENNE MUSCULAR DYSTROPHY). It was maintained stably in mouse cells – and in immortalized human mesenchymal stem cells; the authors suggested that it may have therapeutic potential for Duchenne muscular dystrophy [Mol Therapy (2008) doi: 10.1038/mt.2008.253].

(See also BACTERIAL ARTIFICIAL CHROMOSOME.)

Human Epigenome Project See EPIGENETICS.

human foamy viruses See SPUMAVIRINAE.

human immunodeficiency virus See HIV-1.

human insulin *crb* Recombinant human insulin (see INSULINS) which is synthesized as two separate chains (A and B) within bacterial cells. When isolated from the cells, the two chains are incubated under suitable conditions in order to promote the development of interchain disulfide bonds.

(cf. HUMAN INSULIN PRB.)

human insulin *prb* Recombinant human insulin (see INSULINS) which is synthesized in bacteria as a single molecule (i.e. the precursor form, *proinsulin*). When isolated from the cells, the proinsulin is processed (cut) enzymically to form the A and B chains – which are then incubated under suitable conditions for the development of interchain disulfide bonds.

(cf. HUMAN INSULIN CRB.)

Human Splicing Finder See SPLICING.

humanized (1) (of a non-human gene) Engineered in order to be expressed more readily in human cells – e.g. by changing certain codon(s) to avoid problems of CODON BIAS – and/or engineered in order that the gene product, when expressed in human cells, is better tolerated (e.g. less toxic) than another form of that product.

(2) (of a product) Suitable for use in humans or in human cells (that is, having appropriate biological properties) owing to synthesis within a specifically recombinant organism: see e.g. recombinant humanized lactoferrin in GLYCOSYLATION.

(3) (of an animal) Made recombinant by receipt of a *human* transgene with the object of providing a better animal model for predicting the human response to chemicals and/or other stimuli.

humanized *Renilla* GFP See HRGFP.

Huntington's disease A hereditary disorder, characterized by chorea and progressive dementia, which is associated with a gene on chromosome 4. The disorder involves expansion of MICROSATELLITE DNA.

Hurler's syndrome *Syn*. MUCOPOLYSACCHARIDOSIS TYPE IH.

HvrBase++ A database which includes details of *hypervariable regions* in the mitochondrial DNA (mtDNA) from a range of primate species, including humans. The source also includes details of 1376 complete mitochondrial sequences as well as sequences from autosomal and X chromosomes [HvrBase++: Nucleic Acids Res (2006) 34(Database issue):D700–D704].

hybrid-arrested translation A technique used in the identification of a protein encoded by a given mRNA. Essentially, mRNAs from the particular cell are translated *in vitro* in the presence of cloned, single-stranded copies of the cDNA that is complementary to the mRNA of interest. All the mRNAs which fail to hybridize with the cDNA yield proteins that incorporate a labeled amino acid; mRNA complementary to the cDNA will bind it and will not be translated. Translation is repeated without cDNAs; the mRNA previously blocked by hybridization to the cDNA can now be translated. The protein of interest is the *extra* protein produced in the second translation, and it can be identified by comparison of the two sets of proteins – e.g. by gel electrophoresis.

hybridization-based sequencing A method for sequencing a DNA fragment (e.g. 100 nt in length) by finding a set of short oligonucleotides which hybridize to the fragment in an overlapping pattern, and which – collectively – cover the whole fragment. This set of oligonucleotides is then arranged, by a computer, into an order that reflects the complete sequence of nucleotides in the sample fragment.

In practice, the short oligonucleotides are present within an array of probes on a solid support (a chip), the array covering all possible combinations of nucleotides in a probe of given length; for example, an 8-nt probe has 4^8 (65536) possible combinations of nucleotides, and this number of probes is present on the chip. The chip is exposed to (labeled) copies of the unknown sequence, and the particular set of probes to which they hybridize (under very stringent conditions) is determined and used in the computer analysis.

The use of this method is restricted to *short* sequences of nucleotides. It is unsuitable for sequencing genomes, or other long sequences, owing e.g. to the possible effects of any short repetitive sequences in the sample DNA (genomes frequently contain such sequences – see e.g. REP SEQUENCE); such short repetitive sequences could lead to difficulty in interpretation when attempting to assemble the data.

(See also DNA SEQUENCING and SHOM.)

hybridization protection assay See ACCUPROBE.

hybridoma The product (and progeny) of CELL FUSION (q.v.) between a tumor cell and a non-tumor cell. The purpose of making a hybridoma is to obtain a cell in which some/all of the properties of the non-tumor cell are exhibited on a long-term basis due to ongoing replication specified by the tumor cell.

A hybridoma formed by fusion between a B lymphocyte (a B cell) and a myeloma cell is a source of *monoclonal antibodies*.

During the preparation of a hybridoma the efficiency of cell fusion may be low – e.g. 1 in 10^4. Moreover, it is necessary to prevent overgrowth of the tumor cells during incubation/culture of fusogen-treated cells. Therefore, use is made of (mutant) tumor cells which are blocked in *one* of two normal pathways for nucleotide synthesis; these mutant cells lack the *ancillary* pathway and are therefore unable to use exogenous thymidine or hypoxanthine. Consequently, *non*-fused tumor cells can be suppressed by the use of an incubation medium (HAT medium) that contains *aminopterin* (which blocks the *main* pathway of nucleotide synthesis) together with hypoxanthine and thymidine. In HAT medium, the fused (hybridoma) cells are able to grow by using the ancillary pathway of the non-mutant partner cells.

hydrodynamic injection (of DNA, for gene delivery) In e.g. mice: high-pressure injection, through the tail vein, of a very large volume of (naked) DNA, in saline solution, within 10 seconds; this is reported to result in up to 40% of hepatocytes expressing the transgene 1 day post-injection.

Traumatic effects of this procedure are reported to include rupture of the animal's liver endothelium [Gene Ther (2007) 14(2):129–137].

(See also GENE-DELIVERY SYSTEM.)

hydroxyl radical (·OH) See e.g. FENTON REACTION.

hydroxylamine (as a mutagen) *In vitro*, this agent (NH_2OH) reacts primarily with cytosine residues in DNA, particularly in ssDNA, where it replaces the amino group with a hydroxyamino group (NHOH); adenine also reacts with the agent, but much more slowly. Bases modified in this way tend to undergo mispairing during the next round of DNA replication – leading e.g. to GC→AT mutations.

(See also NITROUS ACID.)

hydroxymethylcytosine (in phage DNA) See PHAGE T4.

hygromycin B An AMINOGLYCOSIDE ANTIBIOTIC that is active against bacterial, plant and mammalian cells. It is used e.g. as a selective agent for cells that are stably transfected with (and expressing) a gene encoding resistance to hygromycin B (e.g. the *hyg* gene of *Escherichia coli*).

[Uses (e.g.) (for selection): (at 100 μg/mL) BMC Microbiol (2009) 9:31; (50 mg/liter) Plant Physiol (2009) 149(2):916–928; (500 μg/mL) Retrovirology (2009) 6:11.]

(See also G418 SULFATE.)

hyperauxiny See AUXINS.

hyper-IgM syndrome (HIGM) A rare immunodeficiency that is characterized by high-level vulnerability to bacterial infections, normal (or elevated) levels of IgM, a lack of IgG (and other isotypes), and lymph node hyperplasia.

The X-linked form of the syndrome (= HIGM1) has been attributed to mutation in the *CD40L* gene. CD40L is a cell-surface molecule on T lymphocytes (T cells) which interacts with CD40 on B cells; interaction promotes the development of B cells in processes such as *class switching* (in which IgG and other isotypes are produced by the B cells as part of the normal antibody response).

The autosomal recessive form of the syndrome (= HIGM2) may result from mutation in the *CD40* gene or a deficiency in the enzyme *activation-induced cytidine deaminase* (AID), an enzyme found in germinal center B cells. AID is essential for the processes of affinity maturation and class switching in B cells, and its activity may involve RNA EDITING.

hypersensitivy reaction (*plant pathol.*) See GENE-FOR-GENE CONCEPT.

hyphenated dyad symmetry See PALINDROMIC SEQUENCE.

hypoxanthine The product formed by oxidative deamination of adenine.

(See also INOSINE.)

hypoxanthine guanine phosphoribosyltransferase deficiency *Syn*. LESCH–NYHAN SYNDROME.

I

I L-Isoleucine (alternative to Ile).

I-like pili See PILUS.

I-SceI (or ISceI, I*Sce*I, I-*Sce*I etc.) A HOMING ENDONUCLEASE (of the LAGLIDADG family) found in the yeast *Saccharomyces cerevisiae*. [Intermediates in the DNA cleavage pathway: Nucleic Acids Res (2008) 36(10):3287–3296.]

I-SceI has been used e.g. for engineering a reduced-genome strain of *Escherichia coli* [Microb Cell Factories (2009) 8:2]; for studies on reducing cytotoxicity of zinc-finger nucleases [PLoS Genet (2009) 5(2):e1000376]; and studies on DNA–protein associations within *E. coli* [Nucleic Acids Res (2009) 37(5):e37].

IAA Indole 3-acetic acid. (See also AUXINS.)

IAN Indole 3-acetonitrile. (See also AUXINS.)

IAN-PCR Inverse affinity nested PCR: see METAGENOME.

IAP INTRACISTERNAL A-PARTICLE.

IAP proteins Inhibitor of APOPTOSIS proteins – see (e.g.) entry SURVIVIN.

iatrogenic (*adj*.) Refers to adverse effects, in a patient, resulting from the treatment received.

ICAD A factor which complexes with – and inhibits – caspase-activated DNase: see APOPTOSIS.

ICE (*syn*. caspase 1) IL-1β-converting enzyme: see APOPTOSIS.

ICF syndrome Immunodeficiency, centromere instability and facial abnormality syndrome: a (human) autosomal recessive disorder caused by mutation in *DNMT3B*, the gene encoding DNA methyltransferase 3B. The normal methylation of cytosine residues in the satellite DNA (in certain loci) is almost completely lacking in patients with ICF syndrome. This syndrome is characterized by low levels of the serum immunoglobulins, and many of the patients contract fatal infectious disease before reaching adulthood.

(See also GENETIC DISEASE (table).)

ICM Inner cell mass: see BLASTOCYST.

icosahedron A geometric figure bounded by 20 plane faces, all of which are equilateral triangles of the same size. Icosahedral symmetry (also called 5-3-2 symmetry) is a characteristic of the CAPSID of some viruses, including some phages.

ICR See IMPRINTING CONTROL REGION.

ICSI Intracytoplasmic sperm injection: see IVF.

identifier DNA Essentially, DNA which is synthesized with a unique (i.e. totally specific) sequence and used to identify a particular object; the DNA is immobilized on a 'tag' which is secured to the object. The object can be rapidly identified e.g. by using sequence-specific molecular beacon probes.

idiogram A systematized, diagrammatic representation of a KARYOTYPE; a photographic version may be referred to as a *karyogram*.

idling In DNA replication: ongoing enzymic activity in which the conversion of dNTPs to dNMPs does not result in any *net* synthesis or degradation of DNA; it involves the addition and removal of nucleotides at a 3′ terminus.

Idling by DNA polymerase δ (delta) has been implicated in a mechanism for the maturation of OKAZAKI FRAGMENTS in lagging strand synthesis; idling may permit the maintenance of a 3′ terminus suitable for ligation to the previous Okazaki fragment.

iduronidase See MUCOPOLYSACCHARIDOSIS TYPE IH.

***ie-1* gene** See BACULOVIRIDAE.

iFRET (induced FRET) A modified form of FRET designed as an improved fluorescence system for DNA melting analyses [original description: Genome Res (2002) 12(9):1401–1407]. The system involves a conventional FRET acceptor, bound to the DNA, and a donor consisting of a dsDNA-binding intercalating dye (such as SYBR GREEN I); in this arrangement, a positive signal is generated in a duplex setting and is lost on strand separation.

(See also THERMAL MELTING PROFILE.)

IGR Intergenic region.

IHF INTEGRATION HOST FACTOR (q.v.).

IκB See the entry NF-KB.

IL-1β-converting enzyme (ICE) See the entry APOPTOSIS.

imatinib An agent which inhibits oncogenic TYROSINE KINASE activity (encoded e.g. by *bcr–abl* or *kit*); an anticancer agent. (cf. ERLOTINIB.) Mutant target molecules can inhibit binding of the drug.

immuno-PCR A PCR-based method used for the ultrasensitive detection of antigens. Monoclonal antibodies, specific to the antigen of interest, are immobilized on microtiter plate wells and exposed to the sample under test. Molecules of specific antigen, if present within the sample, are then captured by the monoclonal antibodies. Any bound molecules of antigen are subsequently detected by homologous antibodies which are conjugated – via a PROTEIN A–STREPTAVIDIN chimera – to a biotinylated DNA fragment. Amplification of a segment of the DNA by PCR, and analysis of the PCR products by gel electrophoresis, provides an extremely sensitive indication of the presence of the specific antigen.

(See also PROXIMITY LIGATION ASSAY.)

The method was simplified by using a recombinant phage which has a surface-displayed *single chain variable fragment* (see SCFV) that binds to the specific antigen; when the phage has bound to the antigen, phage DNA is exposed (by lysis of the phage particle) and *this* DNA serves as a template for the PCR. [Phage display-mediated immuno-PCR: Nucleic Acids Res (2006) 34(8):e62.]

An immuno-PCR approach has detected 10 pg/mL purified Shiga toxin 2 (Stx2) as compared with 1 ng/mL detected by a commercial enzyme immunoassay [J Clin Microbiol (2008) 46(4):1292–1297].

immunoblotting (Western blot analysis) A form of BLOTTING in which a Western blot (see WESTERN BLOTTING) is exposed to specific, labeled *antibodies* in order to detect/identify particular proteins by antibody–protein binding.

Multiplex detection of proteins on a Western blot requires the use of a range of different labels. Difficulties involved in

using multiple types of fluorescent dye may be overcome by the use of QUANTUM DOT PROBES.

Recombinant proteins may be detected or identified by an antibody-independent procedure called ENZYME FRAGMENT COMPLEMENTATION.

immunodeficiency, centromere instability, facial abnormality syndrome See ICF SYNDROME.

immunomagnetic enrichment Any procedure which involves the use of antibody-coated superparamagnetic beads, such as DYNABEADS, for isolating a given type of cell or molecule, which binds to the antibody, from a mixture of cells or molecules. Beads are added to the mixture and are allowed to bind ('capture') the cells/molecules of appropriate specificity; the beads, with their adsorbed cells or molecules, are retained (by a magnetic field) while other components of the mixture are removed by washing. The cells/molecules of interest are then eluted from the beads.

immunomagnetic separation (IMS) See e.g. DYNABEADS and IMMUNOMAGNETIC ENRICHMENT.

immuno-PCR See under IMMUNO- (above).

immunoperoxidase assay See HORSERADISH PEROXIDASE.

implantation In embryology: attachment of the BLASTOCYST (q.v.) to the wall of the uterus.

importins See KARYOPHERINS.

imprinting See GENETIC IMPRINTING.

imprinting control region (ICR) A specific region in *one* of the two parental alleles (either the maternal *or* the paternal allele) of a given *imprinted* gene (see GENETIC IMPRINTING); repression of the imprinted allele involves methylation at the ICR. (A given imprinted gene may have more than one ICR.)

Different ICRs may function in different ways. In some cases an ICR inhibits the allele's expression by functioning as a CHROMATIN INSULATOR which blocks promoter–enhancer interaction. However, in other cases, non-coding RNAs may be involved (as e.g. in X-INACTIVATION).

IMS Immunomagnetic separation: see e.g. DYNABEADS.

In-Fusion™ A system (Clontech, Mountain View CA, USA) which can be used e.g. to join two or more pieces of double-stranded DNA, or to insert a piece of (ds) DNA into a linearized vector; in each pair of dsDNA ends to be joined, the ~15 5′ terminal nucleotides in one molecule must be able to base-pair with (i.e. must be complementary to) the ~15 5′ terminal nucleotides of the other molecule. If such compatible termini are not present on the sample molecules, they can be included as 5′ tags in the PCR primers used to amplify the molecules.

The In-Fusion™ enzyme that mediates the reaction acts as a 3′→5′ exonuclease which removes nucleotides from the 3′ end of a strand in each of the pair of target termini; the resulting 5′ overhangs, being derived from compatible termini, can base-pair with each other – either fully, with nicks, or with short gaps of a few nucleotides. A vector with gaps or nicks can be inserted without prior ligation into host cells for repair and replication.

This method has been used e.g. for the production of fusion proteins and modular vectors – and also for mutagenesis, and for replacing a sequence in a circular molecule of DNA [BioTechniques (2007) 43(3):354–359].

Uses of In-Fusion™ (e.g.): studies on *Drosophila* SUMO protein [PLoS ONE (2009) 4(6):e5905]; studies on regulation of transcription factor genes [Nucleic Acids Res (2009) 37 (3):778–792]; and studies on protein degradation in the aging process [EMBO J (2009) 28(7):889–901].

(See also STICKY RICE.)

in silico By computer.

***in situ* hybridization** See ISH.

***in situ* solid-phase DNA amplification** See SOLID-PHASE PCR.

***in vitro* fertilization** See IVF.

***in vitro* mutagenesis** Any form of MUTAGENESIS which is conducted outside living cells. Various commercial products are in use: e.g. EXSITE SITE-DIRECTED MUTAGENESIS KIT, GENEMORPH PCR MUTAGENESIS KIT, GENETAILOR SITE-DIRECTED MUTAGENESIS SYSTEM, and QUIKCHANGE SITE-DIRECTED MUTAGENESIS KIT.

(See also MEGAPRIMER MUTAGENESIS and SITE-DIRECTED MUTAGENESIS.)

***in vitro* protein synthesis** See CELL-FREE PROTEIN SYNTHESIS.

***in vitro* transcription–translation** Alternative designation for CELL-FREE PROTEIN SYNTHESIS.

***in vitro* transposition** See the entry TN5-TYPE TRANSPOSITION.

(See also PHAGE MU.)

***in vivo* bioluminescence** (engineered) BIOLUMINESCENCE produced in LUCIFERASE-expressing (i.e. genetically engineered) living cells/animals following the provision of luciferin.

In one experiment, tumor cells that constitutively expressed luciferase were implanted in the forebrain and flank region of mice and allowed to develop into tumors. Subsequently, luciferin was injected intraperitoneally. In this experiment it was found that the intensity of bioluminescence varied with time following injection of luciferin, peak emission being recorded about 20 minutes after the injection and subsequently decaying slowly.

Bioluminescence in (live) BALB/c mice was used to study anti-smallpox treatment after lethal challenge with luciferase-expressing WR (recombinant) vaccinia virus [J Virol (2009) 83(20):10437–10447.

***in vivo* biotinylation** (engineered) BIOTINylation of a specified protein – within living cells – engineered by the action of the enzyme *protein–biotin ligase*.

Initially, the gene of the target protein must be modified so that, when expressed in the cell, the target protein exhibits a *biotin-binding tag* (referred to as a *biotin ligase recognition peptide* (BLRP)).

The protein–biotin ligase can be an endogenous enzyme but is often one expressed from an imported plasmid-borne gene – usually the *birA* gene encoding the protein–biotin ligase of *Escherichia coli*.

When the recombinant (BLRP-tagged) target protein is co-expressed in a cell with *E. coli* BirA ligase, BirA mediates covalent binding of biotin to the tag. This procedure has been carried out e.g. in mammalian cells as well as in bacteria and

the yeast *Saccharomyces cerevisiae*.

(See also BIOEASE.)

The BLRP tags used for engineering biotinylation *in vivo* are much shorter than the biotin-binding sequences present on naturally biotinylated proteins. The natural tags are long (>50 amino acid residues), and were these to be used in the preparation of a recombinant (tagged) protein they may affect the protein's properties. Moreover, these (natural) tags can be recognized by endogenous biotin ligases, making it difficult to regulate the biotinylation of a specific target protein. For these reasons, shorter, artificial tags have been prepared by carrying out multiple rounds of screening of peptide libraries for those peptides which exhibit optimal biotinylation by the BirA biotin ligase. The resulting tags, of e.g. ~23 amino acid residues, have an efficacy in biotinylation comparable to that of the natural tags.

Biotinylation of a specified protein *in vivo* has been used, for example, to facilitate subsequent detection, recovery and purification of the target protein from the cell lysate. This is often achieved by using STREPTAVIDIN-coated DYNABEADS; this procedure (in which the target protein is held securely by the Dynabeads) permits efficient washing and purification. In CHROMATIN IMMUNOPRECIPITATION, *in vivo* biotinylation is employed as an alternative approach to the use of antibodies, thus avoiding the potential problem of inefficient binding of antibodies to target proteins [BMC Mol Biol (2009) 10:6].

In general, *in vivo* biotinylation of proteins is valuable as an affinity purification method in the field of proteomics.

[Examples of *in vivo* biotinylation: J Biol Chem (2006) 281 (7):3899–3908, and Nature Protocols (2009) 4:506–517.]

(See also PROTEIN LABELING.)

***in vivo* DNA synthesis** (monitoring of) See DNA SYNTHESIS *in vivo* (monitoring).

***in vivo* electroporation** See ELECTROPORATION.

***in vivo* expression technology** See IVET.

***in vivo* fluorescence** Fluorescence produced in live, genetically engineered, cells/animals.

One factor which has facilitated *in vivo* fluorescence is the availability of modified forms of the GREEN FLUORESCENT PROTEIN with altered absorbance/emission characteristics. In comparison to wild-type GFP (major absorbance at ~395 nm, minor absorbance at ~475 nm), some genetically modified forms of this protein have enhanced absorbance at 475 nm, permitting excitation solely with blue light; the significance of this is that excitation can be achieved in the absence of ULTRAVIOLET RADIATION (thus avoiding a factor which is potentially harmful to the cells/animals under investigation). In one development, live transgenic mice expressing a fusion protein, GFP fused to the τ (tau) microtubule-binding protein, were seen as fluorescent animals when placed under radiation from five blue LEDs (light-emitting diodes) [BioTechniques (2003) 34(3):474–476].

Increased flexibility in the use of *in vivo* fluorescence has been achieved by the development of O^6-alkylguanine-DNA alkyltransferase (AGT) as a fusion partner: see GENE FUSION. (See also LUMIO.)

As mentioned above, one of the problems which can affect fluorescence imaging of living cells is the phototoxicity that results from exposure to ultraviolet radiation used to excite certain types of fluorophore. One approach to protecting cells from phototoxicity is the use of stroboscopically controlled excitation; thus, the use of 1 ms pulses of radiation enabled sharp images to be obtained of a rapidly moving flagellum – and also reduced the level of phototoxicity, as indicated by extended viability/activity of the cells used in the experiment [BioTechniques (2006) 41(2):191–197].

A similar approach to problems of phototoxicity has been employed for the fluorescent imaging of mouse and bovine preimplantation embryos: the use of a spinning-disk confocal microscope; embryos examined in this way were reported to maintain full developmenal potential [BioTechniques (2006) 41(6):741–750].

Dual-fluorescence reporter/sensor plasmids have been used to study the dynamics of microRNA formation in living cells of zebrafish and mouse [BioTechniques (2006) 41(6):727–732].

Fluorescent imaging of a mouse lymph node *in vivo* was carried out with glutathione-coated QUANTUM DOTS (480 nm excitation; 800 nm emission); one advantage of near-infrared emission for *in vivo* fluorescence is the low level of scattering (and absorption) of radiation in tissues at these wavelengths [Int J Mol Sci (2008) 9(10):2044–2061].

(See also CONFOCAL MICROENDOSCOPE.)

***in vivo* footprinting** A form of FOOTPRINTING in which DNA–protein interactions are studied in living cells. One form of *in vivo* footprinting studies DNA–protein interaction in a small, specific region, such as a given gene. Cells are treated briefly with (e.g.) 1% dimethylsulfate – which penetrates the nucleus and methylates various guanine residues; any given guanine resists methylation if it is protected (from the reagent) by a bound protein. DNA is extracted from the cells and cleaved – at methylated guanines – by piperidine. The sample DNA is then denatured, yielding ssDNA fragments; these fragments bind gene-specific primers which are extended to form blunt-ended products – to which linkers are ligated (at both ends). These products are PCR-amplified with a gene-specific primer and a linker-specific primer. The final products are labeled and are examined by gel electrophoresis using, as a control, products prepared *in vitro* from DMS-treated genomic DNA. The *absence* of particular bands of experimental products in the gel indicates site(s) of ligand binding.

Genome-wide footprinting has been carried out in the yeast *Saccharomyces cerevisiae* by mapping the cleavage sites of DNase I and using massively parallel DNA sequencing. This study revealed thousands of footprints of regulatory proteins and permitted the derivation of factor-binding motifs and the identification of many new binding sites for major regulators [Nature Methods (2009) 6:283–289].

***in vivo* ligation** (of polypeptides and proteins) See *split intein* in the entry INTEIN.

***in vivo* mode** (in gene therapy) See the entry GENE THERAPY.

***in vivo* mutagenesis** MUTAGENESIS within living cells.

(See e.g. RECOMBINASE-MEDIATED CASSETTE EXCHANGE; SIGNATURE-TAGGED MUTAGENESIS; SITE-DIRECTED MUTAGENESIS; TRANSPOSON MUTAGENESIS.)

(See also RECOMBINEERING.)

***in vivo* protein analysis** See e.g. entry TEV.

Inc groups (among plasmids) See PLASMID (*Compatibility*: *Inc groups*).

IncFI group See PLASMID (*Compatibility*: *Inc groups*).

IncFII group See PLASMID (*Compatibility*: *Inc groups*).

inclusion (inclusion body) In *Chlamydia* infections: an intracellular population of elementary bodies: see CHLAMYDIA.

inclusion bodies (*DNA technol.*) See entry OVEREXPRESSION.

IncN group See PLASMID (*Compatibility*: *Inc groups*).

incompatibility groups (among plasmids) See PLASMID (*Compatibility*: *Inc groups*).

IncP group See PLASMID (*Compatibility*: *Inc groups*).

IncX group See PLASMID (*Compatibility*: *Inc groups*).

indel An error – either a single-base insertion or a single-base deletion.

indinavir See PROTEASE INHIBITORS.

indirect suppression See SUPPRESSOR MUTATION.

induced FRET See IFRET.

inducer exclusion See CATABOLITE REPRESSION.

inducible recombinase (*in vivo*) See e.g. CRE–LOXP SYSTEM.

induction (of a lysogen) See LYSOGENY.

influenza virus A See INFLUENZAVIRUS.

influenza virus B See INFLUENZAVIRUS.

Influenzavirus A genus of enveloped viruses (family ORTHOMYXOVIRIDAE) in which the genome consists of eight molecules of linear negative-sense ssRNA that form a helical ribonucleoprotein within the capsid. The capsid is enclosed by a lipid envelope (derived from the cell), but projecting from the envelope are two types of protein – HEMAGGLUTININ (q.v.) and NEURAMINIDASE (q.v.); each type of protein can occur in different antigenic forms which are designated by numbers – e.g. H1, H2, H3 etc. and N1, N2, N3 etc., respectively.

Influenza virus A, or influenza virus B, can cause sporadic or epidemic influenza. Influenza virus A can cause epizootics in e.g. pigs and birds, and has caused human pandemics.

The surface proteins of the influenza viruses undergo small antigenic changes, over time, producing new strains that lead to fresh outbreaks of influenza. Rarely, however, an abrupt, major antigenic change occurs – producing a completely new strain which has the potential to cause an influenza pandemic in the (immunologically naive) human population. This kind of abrupt *antigenic shift* may occur, for example, when a cell is co-infected with two distinct strains of *Influenzavirus*; thus, as these viruses have multipartite genomes, co-infection may result in the formation of a novel REASSORTANT VIRUS. As influenza virus A infects other species (as well as humans), mixed infection, involving e.g. viruses from different species, may lead to hybrid viruses containing genome segments from each of the parent virions. A study of the recently circulating strains of human influenza virus reported that high levels of mixed infection had occurred, including some cases in which viruses of different types (A and B) had co-infected [J Virol (2009) 83(17):8832–8841].

[*Influenzavirus* sequences and epitopes: Nucleic Acids Res (2009) 37(Database issue):D423–D430.]

Nomenclature of influenza viruses

The designation of each strain of virus contains the following data: (i) the type of influenza virus (e.g. A, B); (ii) the animal from which the strain was initially isolated (this is omitted if the host was human); (iii) the location where the strain was first isolated; (iv) the (laboratory) strain number; (v) the year of isolation; (vi) the specific viral hemagglutinin (H) and the neuraminidase (N). Thus, for example:

A/duck/Ukraine/1/63(H3N8)

The 1957 pandemic of influenza (the 'Asian flu') was caused by strain:

A/Singapore/1/57(H2N2)

The 1968 pandemic of influenza ('Hong Kong flu') was due to strain:

A/Hong Kong/1/68(H3N2)

Other strains of influenza virus

The 1997 outbreak of 'bird flu' in Hong Kong resulted from avian-to-human transmission of an avian strain, H5N1. There was a small number of deaths, but no further deaths occurred following large-scale slaughter of poultry.

In 1999, strain H9N2 infected children in Hong Kong and also patients in Guangdong province.

In 2004–2005, H5N1 caused over 50 deaths in East Asia. In all viruses isolated, the genes were those of avian influenza viruses, suggesting that these strains did not arise by reassortment. Human-to-human transmission was described.

The 2009 outbreak ('swine flu') involved an H1N1 strain of swine origin. The outbreak originated in Mexico (with deaths in that country), the causal strain being first detected in April. In May 2009, viruses isolated from individuals in Mexico and the USA were reported to be resistant to both amantadine and rimantadine but were, thus far, sensitive to OSELTAMIVIR and to ZANAMIVIR.

Culture of influenza viruses

Influenza viruses can be propagated e.g. in chick embryo cell cultures or in embryonated hens' eggs. Some researchers use e.g. monkey kidney cells for primary isolation.

Inactivation of influenza viruses

Influenza viruses can be inactivated e.g. by lipid solvents, pH <5, heat (e.g. 60°C/30 min), detergents, phenolic disinfectants etc.

Influenza viruses are apparently resistant to drying at room temperature.

Vaccines

Anti-influenza vaccines include those containing inactivated viruses; vaccines containing only the purified components

of virions (such as hemagglutinin and neuraminidase); and those containing cold-adapted, 'live' viruses – so-called *attenuated* vaccines.

Influenza vaccines are reviewed regularly, and are changed frequently, to ensure efficacy against the prevailing, antigenically drifted strain(s) of *Influenzavirus*. Constant monitoring is also carried out in order to permit the earliest identification of potential pandemic strains. Efforts are ongoing to improve the performance of influenza vaccines.

A trivalent VLP (i.e. virus-like particle), consisting of hemagglutinin, neuraminidase and matrix 1 protein, was reported to elicit protective immunity in animals [PLoS ONE (2009) 4 (6):e6032].

It was suggested that a microRNA-based approach could be used for the attentuation of vaccine strains of influenza virus A [Nature Biotechnol (2009) 27:572–576].

[Pandemic vaccine preparedness: PLoS Pathog (2009) 5(6): e1000482.]

Genotyping of influenza viruses

See e.g. T5000 BIOSENSOR SYSTEM (reference).

informosome See MRNP.

In-Fusion™ See under '-IN-' (above).

INH Isonicotinic acid hydrazide: see ISONIAZID.

inhibitor of apoptosis (IAP) proteins See e.g. SURVIVIN.

INN International non-proprietary name – an internationally recognized name of a given drug/therapeutic agent etc.

(rINN has been used to mean *recommended* international non-proprietary name, i.e. when, concurrently, more than one non-proprietary name is in use.)

inner cell mass (ICM) See BLASTOCYST.

inner primer See NESTED PCR.

INNO-LiPA Mycobacteria® A PCR-based PROBE system from Innogenetics (Gent, Belgium) which is used for identifying medically important mycobacteria (cf. ACCUPROBE). Amplification of spacers in the 16S–23S rRNA genes of the test strains (by PCR) is followed by reverse hybridization of the products to a range of membrane-bound probes [see J Clin Microbiol (2001) 39:3222–3227 and J Clin Microbiol (2004) 42:3083–3088].

[INNO-LiPA Mycobacteria used in studies on the relevance of nontuberculous mycobacteria: Emerg Infect Dis (2009) 15 (1):53–55; Emerg Infect Dis (2009) 15 (2):292–294.]

[INNO-LiPA Mycobacteria used in studies on isolates from extrapulmonary specimens: BMC Clin Pathol (2009) 9:1.]

INNO-LiPA Rif TB See the entry LINE PROBE ASSAY.

inorganic pyrophosphate (PP_i) See PYROPHOSPHATE.

inosine A ribonucleoside in which the base is hypoxanthine; the monophosphate is found e.g. in the 'wobble position' in some tRNAs; in the wobble position, inosine may pair with C or U (see WOBBLE HYPOTHESIS).

(See also UNIVERSAL NUCLEOTIDE.)

inosine monophosphate See entries INOSINE, MYCOPHENOLIC ACID and UNIVERSAL NUCLEOTIDE.

INSDC The INTERNATIONAL NUCLEOTIDE SEQUENCE DATABASE COLLABORATION.

insect-cell-based expression systems Those systems used for the expression of recombinant proteins either in cultures of insect cells or in live insects: see BACULOVIRUS EXPRESSION SYSTEMS and DES.

insect cells (for baculovirus culture) See e.g. SF9 CELLS.

insertion–duplication recombination (Campbell-type integration) A form of RECOMBINATION in which the insertion of a sequence of nucleotides results in duplication of the insertion site *without* the involvement of new DNA synthesis – see e.g. (a) in the diagram. Insertion–duplication recombination thus gives rise to a pair of DIRECT REPEAT sequences either side of the inserted sequence.

Compare this with the insertion of a transposable element: see diagram in the entry TRANSPOSABLE ELEMENT. In this process DNA synthesis *does* occur, and it also gives rise to the formation of a pair of direct repeat sequences.

insertion sequence (IS; IS element) A type of TRANSPOSABLE ELEMENT (q.v.) which encodes only those functions that are necessary for transposition. (cf. TRANSPOSON.) The structural gene(s) of an insertion sequence are bracketed by a pair of INVERTED REPEAT sequences; usually, these inverted repeats form the terminal parts of an insertion sequence, but in some IS elements (e.g. members of the IS1111 FAMILY) the inverted repeats are subterminal.

A specific insertion sequence is commonly designated by 'IS' followed by an *italic* number (e.g. IS*1*, IS*2* etc.); in some cases the designation of an IS element includes an indication of the source organism – e.g. ISHP608, which is present in the Gram-negative bacterial pathogen *Helicobacter pylori*.

Insertion sequences exist in their own right – but they also form the terminal parts of class I (composite) transposons.

An insertion sequence may transpose by a conservative cut-and-paste mechanism or by a replicative mechanism, depending on the insertion sequence – see TRANSPOSABLE ELEMENT (figure) for an explanation of these terms; IS1071 (q.v.) is one example of an IS that transposes in a replicative way.

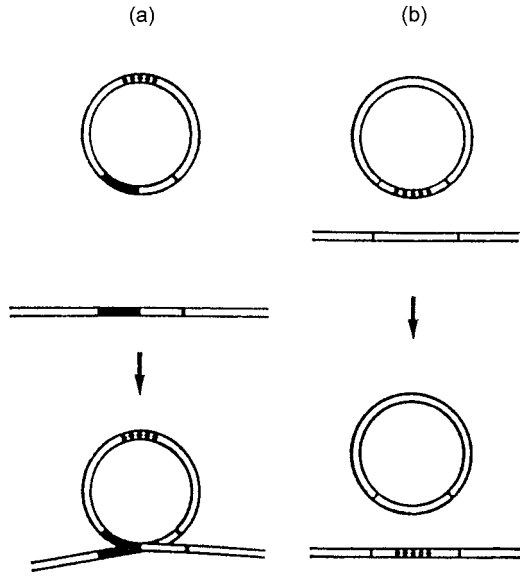
Some insertion sequences transpose with the formation of a circular intermediate (see the entry TRANSPOSABLE ELEMENT for details).

In bacteria, chromosomal REP sequences were reported to be common targets for insertion sequences [BMC Genomics (2006) 7:62].

Transposition frequency varies with insertion sequence, and with the physiological state of the host cell, but, in general, the frequency of transposition ranges from $\sim 10^{-9}$ to $\sim 10^{-5}$ per IS per cell division.

The presence of an insertion sequence may be indicated e.g. by the effects of its insertion. For example, the presence of IS*3* within the *finO* gene of the F plasmid inactivates the *finO* gene and results in the ability of this plasmid to behave as a constitutively derepressed conjugative plasmid (see the entry FINOP SYSTEM).

Some insertion sequences are used as markers for detecting or typing particular bacteria – e.g. IS*900* has been used as a



INSERTION–DUPLICATION RECOMBINATION: insertion of vector-borne DNA into a specific bacterial gene (diagrammatic).

(a) A circular plasmid, constructed *in vitro*, is carrying (i) a gene encoding resistance to a given antibiotic (*shown as five dots*, *top of plasmid*), and (ii) a fragment of DNA corresponding to part of the bacterial target gene (*bottom of plasmid*). The left-hand side of the fragment is shown as black, and the right-hand side as white – simply to help in describing the fate of each part of this fragment. Copies of this plasmid are introduced, e.g. by transformation, into a population of bacteria containing the target gene. Within the cells, recombination can occur between the fragment of DNA in the plasmid and the corresponding (homologous) target sequence in the bacterial gene. The target sequence in the bacterial chromosome is shown (also in black/white) below the plasmid; the nucleotide sequence of the plasmid's black/white fragment corresponds to the nucleotide sequence of the black/white target region in the chromosome. Recombination between the plasmid's fragment and the bacterial target sequence involves a single cross-over between the fragment and target gene (at the black/white junctions) and the joining of ends to form a continuous sequence, as shown below. Note that, in this case, (i) the entire plasmid has inserted into the target gene, and (ii) part of the target gene's sequence has been duplicated (black/white . . . . . black/white); for this reason, the process is sometimes called *insertion–duplication recombination* (or Campbell-type integration).

The above procedure is carried out with a *non*-replicating plasmid. Only if such a plasmid *integrates* with the chromsome can it be replicated (as part of the chromosome) when the cell divides. A medium containing the given antibiotic will therefore select for those cells in which the plasmid has integrated – such cells being able to form normal sized colonies because *both* daughter cells (at each division) will carry the plasmid's antibiotic-resistance gene in their chromosomes. (If a *replicating* plasmid had been used, copies of the plasmid would be passed to both daughter cells at each cell division – so that *all* cells would be selected by the antibiotic, regardless of whether or not the plasmid had integrated.)

If the fragment carried by the plasmid corresponds to an *internal* (coding) part of the target gene, then insertion of the plasmid would interrupt the coding sequence and most probably inactivate the target gene.

If the fragment carried by the plasmid is a mutant *end* sequence of the target gene, then insertion of the plasmid will replace the wild-type end of the gene with a mutant end. The final result of insertion would therefore be a complete gene with a mutant end-sequence – flanked (on one side) by plasmid DNA.

(b) This circular plasmid (*top*) contains a fragment of DNA corresponding to part of the bacterial wild-type target gene – but the fragment carries a mutagenized sequence (*shown as five dots*, *bottom of plasmid*). The (wild-type) chromosomal sequence corresponding to the plasmid's mutagenized fragment is shown below. In this case, interaction between the plasmid and chromosomal gene has involved a cross-over on *both* sides of the mutant region – leading to an *exchange* of DNA between the plasmid and chromosome. Compare with the insertion–duplication interaction, shown in (a), in which only a single cross-over is involved.

Figure reproduced from *Bacteria in Biology*, *Biotechnology and Medicine*, 6th edition, Figure 8.21, page 254, Paul Singleton (2004) John Wiley & Sons Ltd, UK [ISBN 0-470-09027-8] with permission from the publisher.

specific marker for *Mycobacterium paratuberculosis*. IS*6110* is regarded as specific for the *Mycobacterium tuberculosis* complex; different strains contain different numbers of copies of IS*6110* – e.g. the genome of the reference strain of *M. tuberculosis* (H37rv) contains 16 copies, while some strains apparently lack this sequence. Strains of *M. bovis* generally contain a single copy.

[Diversity of insertion sequences in the Archaea: Microbiol Mol Biol Rev (2007) 71(1):121–157.]

(See also entries for some individual insertion sequences under 'IS'.)

insertion vector A CLONING VECTOR into which target DNA can be inserted, at *a single* restriction site, without the need to remove non-essential DNA.

(cf. REPLACEMENT VECTOR.)

insertional inactivation (of a gene) Inactivation caused by the insertion of a DNA sequence, e.g. in the coding sequence or in the promoter of the gene, blocking the normal function of that gene.

In DNA technology, insertional inactivation can be useful as an indicator system. For example, in *blue–white screening* (see PBLUESCRIPT), insertion of the fragment of interest into the polylinker normally abolishes synthesis of a functional α-peptide, the presence of the fragment within the vector being indicated by the reaction of the bacterial colonies on suitable media.

insertional mutagenesis (in cells) (*DNA technol.*) A procedure in which DNA is introduced into cells in order to generate recombinant (mutant) cells through integration of the DNA into the genome.

In some cases, the insertion of DNA will generate *random* mutation in which the DNA has inserted at random sites into the genome (see e.g. TRANSPOSON MUTAGENESIS). Random insertion of DNA commonly involves inactivation of gene(s).

Random insertional mutagenesis in the fungus *Aspergillus fumigatus* was carried out by transfer to the fungus of T-DNA from the plant-pathogenic bacterium *Agrobacterium* [PLoS ONE (2009) 4(1):e4224].

The libraries of mutated murine embryonic stem (ES) cells, produced e.g. by insertional mutagenesis, were reported to be limited in scope by incomplete genome coverage; improved coverage was reported in a library prepared with Blm$^-$ cells mutagenized by the *piggyBac* transposable element [Genome Res (2009) 19:667–673].

Non-random, i.e. *targeted*, insertional mutagenesis can be carried out e.g. with an approach such as the LAMBDA (λ) RED RECOMBINATION system (see also entries RECOMBINEERING, TARGETRON, ZINC-FINGER NUCLEASE).

(See also entries INSERTION–DUPLICATION RECOMBINATION and INSERTIONAL ONCOGENESIS.)

insertional oncogenesis Oncogenesis resulting from insertion of a DNA sequence into the genome, causing e.g. inactivation of a tumor suppressor gene or activation of an ONCOGENE.

Insertional oncogenesis has resulted from retrovirus-based gene therapy in several patients being treated for SCIDX1 [J Clin Invest (2008) 118:3132–3142].

To avoid insertional oncogenesis by lentiviruses, lentiviral vectors have been provided with CHROMATIN INSULATORS – the object being to protect adjacent endogenous genes from the influence of the integrated viral sequence; however, while transgenes can be expressed from these engineered vectors, it was found that the titer of integrants was adversely affected (at a step preceding viral integration into the host's genome), emphasizing the need for a careful evaluation of the vector's function following the insertion of *cis*-acting elements such as chromatin insulators [BMC Biotechnol (2009) 9:13].

insertional targeting vector See VECTOR.

insulator See CHROMATIN INSULATOR.

insulin aspart A (non-commercial) designation of a genetically engineered short-acting form of human insulin: see INSULINS.

insulin glargine The (non-commercial) name for a genetically engineered long-acting form of human insulin in which the C-terminal residue in the A chain was replaced with a glycine residue, and the B chain has two additional arginine residues. The recombinant protein is produced in *Escherichia coli*.

(See also INSULINS.)

insulins (recombinant) Any of various forms of insulin – some 'short-acting', others 'long-acting' – produced by recombinant DNA techniques in various commercial organizations; some of these engineered insulins are synthesized in *Escherichia coli* while others are synthesized in the yeast *Saccharomyces cerevisiae*.

The native (human) insulin molecule consists of two polypeptide chains, A and B – of 21 and 30 amino acid residues, respectively – that are held together by two disulfide bridges. Even small changes in the architecture of this molecule can cause significant changes in biological activity. For example, replacement of proline at position B28 by aspartic acid gives rise to a short-acting insulin (insulin aspart: compare INSULIN GLARGINE).

(See also HUMAN INSULIN CRB and HUMAN INSULIN PRB.)

Porcine insulin differs from human insulin by only a single amino acid residue, and it has little immunogenicity in man; it can be converted enzymically/chemically to a form that has the same composition as human insulin.

Bovine insulin differs from human insulin by three amino acid residues, and it does trigger an immunological response in humans.

***int* gene** (phage λ) See PHAGE LAMBDA.

IntaRNA A (bioinformatics) approach to predicting targets for bacterial sRNAs: see the entry SRNAS.

integrase See RECOMBINASE.

integrase inhibitor (anti-HIV-1) An ANTIRETROVIRAL AGENT which inhibits the insertion of viral DNA into the host cell's genome, thereby inhibiting viral replication. Some integrase inhibitors are used in the treatment of AIDS.

Integrase inhibitors include e.g. ELVITEGRAVIR and raltegravir.

[Integrase and integration (biochemical activities of HIV-1

integrase): Retrovirology (2008) 5:114.]

[Emerging role of integrase inhibitors in the management of treatment-experienced patients with HIV-1 infection: Ther Clin Risk Manag (2009) 5:331–340.]

integration host factor (IHF) (in *Escherichia coli*) A multifunctional, heterodimeric protein with contributory roles that include e.g. expression of the F plasmid, on–off switching of type I fimbriae, promotion of sigma factor specificity, and integration of phage λ genome into the host cell's chromosome.

integrin Any of a specific category of cell adhesion molecules which form part of the cytoplasmic membrane in many types of eukaryotic (including mammalian) cell – and which have roles in various aspects of physiology. The integrin molecule is a heterodimer consisting of α and β protein subunits linked non-covalently. There are many types of α subunit, but only relatively few types of β subunit; the particular combination of α and β subunits in any given integrin determines its distribution and biological properties.

The β_1 integrins, which occur e.g. in epithelial cells and in endothelial cells, have important roles in cell–cell and cell–matrix interactions; components of the matrix to which these integrins bind include e.g. collagen, fibronectin and laminin.

As well as the maintenance of tissue structure, such binding is also involved in generating signals that influence a range of cell functions, including aspects of the cell cycle.

Integrins of the β_1 group may also act as receptors for the specific adhesins of invasive pathogenic bacteria.

The β_1 integrins have been targeted in a novel approach to anti-cancer therapy in which a DNAzyme downregulates the β_1 integrins in endothelial cells to inhibit neovascularization (i.e. angiogenesis): see the entry CANCER THERAPY.

The β_2 integrins can be expressed by leukocytes and may also serve as receptors for bacterial adhesins. Many integrins bind, via their α subunit, to ligands containing the RGD motif (arginine–glycine–aspartic acid motif). This motif is mimicked by the FHA (i.e. filamentous hemagglutinin) of the Gram-negative bacterial pathogen *Bordetella pertussis* – binding between FHA and the $\alpha_M\beta_2$ integrin (= Mac-1) on an activated macrophage facilitating uptake by the macrophage; the macrophage may then be killed by a toxin (cyclolysin) that is produced by *B. pertussis*.

Inadequate expression of β_2 integrins by leukocytes results in the failure of these cells to respond normally in the process of inflammation (a condition which is referred to as *leukocyte adhesion deficiency*, LAD); individuals having this condition are more susceptible than others to infectious diseases.

A β_3 integrin ($\alpha_{IIb}\beta_3$), found e.g. on blood platelets, binds to ligands which include fibrinogen, von Willebrand factor and fibronectin; deficiency of this integrin gives rise to defective platelet agglutination (GLANZMANN'S THROMBASTHENIA).

(See also PSEUDOTYPING.)

Nomenclature of integrins

Any given integrin may be denoted by *two* CD numbers, one number referring to the α subunit, the other to the β subunit – for example, LFA-1 (leukocyte function-associated antigen-1) is CD11a/CD18, and the Mac-1 integrin is CD11b/CD18.

integron A segment of DNA containing (i) a gene encoding a recombinase and (ii) site(s) (*att*) for the insertion or excision of one or more gene cassettes, the insertion or excision being mediated by the recombinase.

Resistance integrons characteristically contain genes which encode resistance to antibiotics and/or to disinfectants. These integrons are found within bacterial chromosomes and also in plasmids and transposons. They have been divided into three classes, primarily on the basis of structure and distribution; resistance integrons of class 2 are associated with transposons of the Tn*7* family and are found in Gram-negative species.

A *superintegron* is a species-specific element which may contain up to ∼100 gene cassettes encoding a wide variety of functions; they are chromosomal elements.

The functions encoded by integrons are now known to be more extensive than those associated with earlier examples. Thus, gene cassettes found in organisms from an estuary with a history of industrial pollution include genes that encode the types of enzyme likely to be useful for catabolism of certain forms of industrial waste [PLoS ONE (2009) 4(4):e5276].

intein In some proteins: an *internal* sequence of amino acids which is able to catalyze its own excision (self-splicing), and to join the two flanking regions (*exteins*) to form a functional protein; a branched polypeptide (having two N terminals and one C terminal) is formed as an intermediate product in the splicing process.

At least some excised inteins have the properties of a site-specific endonuclease (i.e. a HOMING ENDONUCLEASE). This type of intein can (in e.g. the *VMA1* gene of *Saccharomyces cerevisiae*) mediate insertion of an intein-encoding sequence into an intein-less copy of the given gene; in this process, the gene is cleaved, and an intein-encoding sequence – which is copied from the intein-containing allele – is inserted. The ability of an intein to mediate insertion of its coding sequence into an intein-less copy of the gene is called *intein homing*.

(See also INTRON HOMING.)

The sequence-specific nature of the homing endonucleases – which are also called *meganucleases* – has been exploited e.g. in promoting homologous recombination in the vicinity of their specific target sequences. It was suggested that these enzymes might be useful for inserting transgenes at specific (rather than random) locations in the context of gene therapy. However, given the limited repertoire of naturally occurring homing endonucleases, it would probably be difficult to find an enzyme that could promote the insertion of any particular, chosen transgene; for this reason, a combinatorial approach has been used to create artificial homing endonucleases that cleave specific, chosen sequences [Nucleic Acids Res (2006) doi: 10.1093/nar/gkl720].

First discovered in the yeast *VMA1* gene, inteins also occur e.g. in the bacterium *Mycobacterium tuberculosis* (*recA* gene [Nucleic Acids Res (2003) 31(14):4184–4191]), the archaean

Thermococcus litoralis (DNA polymerase gene), and in some algal viruses (DNA polymerase gene [Appl Environ Microbiol (2005) 71(7):3599–3607]). They have also been reported in the alga *Chlamydomonas reinhardtii* [BMC Biol (2006) 4: 38].

Some applications of inteins

An intein-based molecular switch, with activity regulated by a small, cell-permeable molecule (4-hydroxytamoxifen), was used to regulate transcription factor activity in a mammalian setting [J Am Chem Soc (2006) 128(27):8939–8946].

Split inteins, i.e. inteins which were split (either naturally or by technology) into two (N-intein and C-intein) parts, have various uses – e.g. site-specific protein modification; ligation of polypeptides (including those unrelated to the protein from which the intein was derived); and also the control of protein activity *in vivo* (i.e. within living cells). In the latter context, this approach was used for *in vivo* ligation of two fragments of firefly luciferase [Nature Chem Biol (2007) 3(1):50–54].

The essential feature of a split intein is that it can mediate so-called *trans* splicing of polypeptides or proteins, i.e. if one polypeptide includes a terminal N-intein and another includes a terminal C-intein, then these two polypeptide chains can be ligated (as though they were the original exteins separated by the intein), with elimination of the intein sequence. Hence, if short versions of intein fragments were available, they could be fused to proteins (as 'ligation tags') to facilitate ligation of the (tagged) proteins.

One approach to the preparation of short (i.e. minimized) ligation tags was to remove the sequence which encodes the homing endonuclease as that part of the intein is not involved in the *trans* splicing process.

Work on DnaE inteins from the bacteria *Synechocystis* and *Nostoc* reported an engineered C-intein of only 15 amino acid residues; this was more effective (at *trans* splicing) than were the naturally occurring split DnaE inteins [PLoS ONE (2009) 4(4):e5185].

intein homing See INTEIN.

intercalating agents Molecules with a planar chromophore that can insert (intercalate) between adjacent base pairs in double-stranded nucleic acids; a *bifunctional intercalating agent* has two planar chromophores in each molecule. The intercalation of these agents may cause e.g. local unwinding of the duplex.

In linear dsDNA, such intercalation results in an increase in viscosity and a reduction in the sedimentation coefficient. In ccc dsDNA, the intercalation results e.g. in a local increase in pitch and changes in the viscosity of the DNA solution – and in the sedimentation coefficient of the DNA. Nicked circular (or linear) molecules of DNA can bind more molecules of an intercalating agent than can an equivalent ccc dsDNA molecule (see also ETHIDIUM BROMIDE).

When one molecule of an intercalating agent inserts at a given site it inhibits the insertion of another molecule in the adjacent two or three base pairs (the *neighbor exclusion principle*).

In DNA technology, intercalating agents are sometimes of use e.g. in DNA COMBING.

Intercalation *in vivo* can inhibit transcription, and replication, and may also induce mutations. Lower concentrations of e.g. some acridines, or of ethidium bromide, may selectively block replication of small bacterial plasmids (see CURING).

Intercalating agents include (e.g.) ACTINOMYCIN D, anthracyclines, ETHIDIUM BROMIDE, PSORALENS, QUINOXALINE ANTIBIOTICS and tilorone.

intergenic repeat unit *Syn*. ERIC SEQUENCE.

intergenic spacer region (ISR) See RRN OPERON.

intergenic suppression See SUPPRESSOR MUTATION.

internal resolution site (IRS) See TN3.

internal ribosomal entry site See IRES.

International Nucleotide Sequence Database Collaboration A joint (collaborative) project involving the DNA DataBank of Japan (DDBJ), European Molecular Biology Laboratory (EMBL) and GenBank (NCBI: National Center for Biotechnology Information).

interrupted gene *Syn*. SPLIT GENE.

interrupted mating An early method used to study the transfer of genes, from donor to recipient, in bacterial CONJUGATION. At time zero, a population of (Hfr) donor cells is mixed with a population of recipients; samples are then removed from the mating mixture at regular intervals and are plated in order to select recombinants. Transfer of donor genes to recipient cells occurs sequentially (according to their positions on the chromosome) – so that a particular donor gene will appear in recombinants after a characteristic interval from time zero.

In *Escherichia coli* the entire chromosome may be transferred in ~100 minutes, although, as time progresses, there is an increasing chance of strand breakage – so that, in many cases, only a part of the donor's chromosomal DNA is transferred to the recipient.

Interrupted mating has been used for genetic mapping.

interstitial junction sequence See TRANSPOSABLE ELEMENT.

intervening sequence *Syn*. INTRON. (This phrase is also used as a synonym of INTEIN.)

intrabody An engineered molecule, based on part(s) of an antibody, which is synthesized intracellularly and which remains within the cell in which it was synthesized.

(See also SINGLE-CHAIN VARIABLE FRAGMENT.)

Intrabodies are used e.g. in studies on GENE THERAPY and for down-regulating the expression of cell-surface molecules by sequestering them intracellularly.

intracisternal A-particle (IAP) A RETROTRANSPOSON, copies of which are dispersed throughout the mouse genome.

As CpGs in the IAP region are heavily methylated, the IAP has been exploited in a method for assessing global genomic demethylation (by agents such as 5-AZA-2′-DEOXYCYTIDINE).

In one approach, following exposure of the cells to a given demethylating agent, genomic DNA is extracted and digested with a (methylation-sensitive) RESTRICTION ENDONUCLEASE (HpaII). An attempt is then made to amplify the HpaII cutting sites – using primers flanking the HpaII site. Given extensive demethylation, unmethylated HpaII cutting sites are cleaved

by the enzyme, thus preventing PCR amplification; however, failure of demethylation leaves methylated (blocked) cutting sites, allowing amplification by (quantitative) PCR.

[Microarray-based mapping of integration sites for IAP in the mouse genome: Nucleic Acids Res (2008) 36(10):e59.]

intracytoplasmic sperm injection (ICSI) See IVF.

intragenic suppression See SUPPRESSOR MUTATION.

intramer An (intracellular) APTAMER (sense 1) which has been transcribed within a cell.

[Use of an intramer to inhibit the activity of T-cell factor 1 (TCF-1): Mol Cancer Ther (2006) 5(9):2428–2434.]

intron (intervening sequence) A sequence that occurs *between exons* in a SPLIT GENE (q.v.); split genes (containing introns) are found in eukaryotes, archaeans, bacteria and viruses. Split genes are common in eukaryotes but are atypical in bacteria.

Normally, the expression of a given split gene requires (i) transcription, followed by (ii) excision from the transcript of those sequences that correspond to the intron(s), and ligation (joining together) of the exons to form a continuous coding sequence (called a *mature* mRNA).

In human pre-mRNA, the excision of introns occurs by a complex process involving a SPLICEOSOME (see SPLICING).

At least some of the introns in bacteria and bacteriophages are *autocatalytic*, i.e. self-splicing; an autocatalytic intron has the ability to cut the RNA at both ends of its own sequence and simultaneously ligate the two flanking sequences.

Introns are categorized into different groups on the basis of e.g. nucleotide sequence, structure, and, where applicable, the mechanism involved in translocation (i.e. INTRON HOMING).

In some cases an intron encodes a functional molecule: see, for example, INTRON HOMING. Moreover, many MICRORNAS are reported to be encoded by sequences within introns.

[Survey of introns of groups I and II in bacterial genomes: Nucleic Acids Res (2008) 36(14):4529–4548.]

(cf. INTEIN.)

Removing introns from genomic DNA in vitro

For an *in vitro* method of eliminating introns from samples of genomic DNA see the entry SPLICE.

intron homing (in bacteria) The replicative spread of a given intron to an intron-less allele – or to an ectopic target site in a DNA molecule. For group I introns, homing involves a cut in the intron-less allele made by a *homing endonuclease* (which is encoded by the intron). During repair, the intron is used as a template, and a copy of the intron is inserted into the intron-less site.

The group II introns are generally appreciably larger (up to several kilobases in length), and more complex, than group I introns; moreover, in at least some cases, the intron-encoded protein (IEP) can act as a REVERSE TRANSCRIPTASE (as well as an endonuclease), and the mechanism of intron homing in group II introns differs from that in the group I introns.

In group II introns, the homing mechanism involves a ribonucleoprotein particle (RNP) that consists of the IEP and the excised lariat RNA copy of the intron. The RNP mediates the insertion of the RNA copy of the intron into one strand of the DNA target, this being referred to as *reverse splicing*; the IEP cuts the other strand of target DNA and then uses the free 3′-end as a primer to reverse transcribe the inserted RNA strand. After normalization in the recipient genome (e.g. by a repair process), the end product of homing is therefore an inserted dsDNA copy of the intron at the target site.

This process is called *retrohoming* because it involves an RNA intermediate.

Studies using atomic force microscopy indicated that DNA capable of acting as a target sequence in retrohoming tends to be more flexible than random genomic sequences – and that the target DNA is bent during retrohoming; it also appears that the 5′ and 3′ exons flanking the intron have roles to play in retrohoming [Biochemistry (2006) 45(41):12424–12435].

Recognition of the target sequence (by the RNP) depends on the nucleotide sequence of the lariat RNA and (partly) on the IEP. Targeting of the RNP can be adjusted by changing the nucleotide sequence of the RNA; this has been exploited for targeted mutagenesis in various bacteria and has also been demonstrated to have potential for making targeted changes in eukaryotic cells: see the entry TARGETRON.

(cf. *intein homing* in the entry INTEIN.)

intron retention See ALTERNATIVE SPLICING.

***inv–spa* genes** See PATHOGENICITY ISLAND (*Salmonella*).

Invader® A system (Third Wave Technologies Inc., Madison WI, USA), used e.g. in various types of diagnostic test, which involves isothermal reactions in which two types of probe are cleaved to produce a target-specific fluorescent signal.

In one reaction, two oligonucleotides, a so-called Invader® oligo and a primary probe, both bind to the specific target on sample DNA; when both oligos are bound, the primary probe overlaps the Invader® oligo at the specific nucleotide being interrogated – leaving the remainder (unbound, 5′ end) of the probe as a free oligonucleotide *flap*. An enzyme, CLEAVASE, cleaves the primary probe at the point of overlap, releasing the 5′ flap together with the adjacent nucleotide. This reaction is repeated many times, the number of flaps released from the primary probes reflecting the amount of target in the sample.

Each of the flaps released during the primary reaction binds to a different oligo: the FRET probe. On binding, the 3′ end of the flap is overlapped by the end region of the FRET probe (which folds back on itself); this end of the probe contains an inactive fluorophore. Cleavage of the overlapping probe then releases the fluorophore, which can therefore fluoresce under suitable excitation. Repeated cleavage of FRET probes gives an amplified target-specific signal.

[Comparison of the Invader® human papillomavirus assay with the Digene Hybrid Capture® 2 assay for the detection of high-risk HPV DNA: J Clin Microbiol (2008) 46(5):1641–1646.]

invasive cleavage A phrase which has been used to refer e.g. to the endonucleolytic action of certain types of bacterial DNA polymerase: see STRUCTURE-SPECIFIC ENDONUCLEASE.

inverse affinity nested PCR (IAN-PCR) See METAGENOME.

inverse PCR A form of PCR which is used e.g. for amplifying the (unknown) DNA that *flanks* a known sequence (on both sides), using a circularized template molecule. Inverse PCR is useful e.g. for investigating the insertion site of a transposon in a plasmid or chromosome (see also ENHANCER TRAPPING).

In inverse PCR the primers are complementary to the two terminal (end) regions of the known sequence. However, unlike the normal arrangement in PCR, when these primers are bound to their complementary sequences they are extended in *opposite* directions, i.e. outwards, into the unknown flanking sequence on each side (in the *circularized* template).

If the known sequence (e.g. transposon) has inserted into a long molecule (e.g. a bacterial chromosome), the initial task is to cut the molecule (with a RESTRICTION ENDONUCLEASE) and obtain the known sequence, together with both flanking regions, on a small, linear restriction fragment. This fragment is then circularized (through hybridization and ligation of the STICKY ENDS). (Circularization can be promoted by carrying out the ligation with low concentrations of the fragments; in this way, the sticky ends on a given fragment are more likely to hybridize with each other than with sticky ends on another fragment.)

Circularized fragments are PCR-amplified, with the primers described, forming circular, dsDNA nicked amplicons; after ligation, these can be linearized by cutting through the known sequence, forming amplicons with known end sequences and the (original) flanking sequences as the central region in each amplicon. The amplicons can be sequenced etc.

Other uses of inverse PCR include preparation of deletion mutations (see e.g. the entry SITE-DIRECTED MUTAGENESIS) and the preparation of full-length cDNAs (see CAPFINDER).

[Example of use of inverse PCR: BMC Plant Biol (2009) 9: 77.]

TAIL-PCR (q.v.) is an alternative approach to copying an unknown sequence adjacent to a known sequence.

inverted repeat (IR) In a given, double-stranded molecule of nucleic acid, either of two sequences such as:

5′ . . . . CCATC . . . . . GATGG . . . . . 3′
3′ . . . . GGTAG . . . . . CTACC . . . . . 5′

i.e. sequences which are identical, with the same polarity, but opposite in orientation. The two sequences may or may not be contiguous.

(cf. DIRECT REPEAT; see also PALINDROMIC SEQUENCE.)

In single-stranded nucleic acids an IR is equivalent to one of the two strands in a double-stranded IR; such IRs can e.g. form hairpin structures.

Typically, a pair of inverted repeats form the two terminal sequences in each TRANSPOSON and INSERTION SEQUENCE, although in some transposable elements (see e.g. IS1111) the inverted repeats are in subterminal locations.

The reported instability of inverted repeats within genomic DNA underpins their use in a proposed method for *in vivo* mutagenesis: see the entry MIRAGE.

invertible DNA Any DNA segment which can be inverted by a site-specific recombinase (see e.g. FLP).

IpaB A protein, secreted by *Shigella flexneri*, which can bring about APOPTOSIS in macrophages.

iPS cells Induced pluripotent STEM CELLS (q.v).

IPTG Isopropyl-β-D-thiogalactoside: a synthetic inducer of the LAC OPERON, i.e. an agent that inhibits the repressor system of that operon.

As IPTG is not a substrate for β-galactosidase (an enzyme encoded by the *lac* operon) it is called a *gratuitous* inducer.

(See also ALLOLACTOSE.)

IR INVERTED REPEAT.

IRES Internal ribosomal entry site: a specific segment of RNA (first obtained from the encephalomyocarditis virus, EMCV) which, when present in mRNAs, promotes high-level, CAP-independent synthesis of proteins in mammalian cells (and in *in vitro* systems) from monocistronic and bicistronic vectors. (The designation IRES is also used to refer to segments from other organisms/cells which promote cap-independent translation in eukaryotes.)

Versions of the EMCV IRES which differ in sequence have been widely used in a range of studies. These versions do not behave in the same way – e.g. they can exhibit a requirement for different types of translation factor – and so may behave differently in different types of cell. Hence, results reported in studies with EMCV IRES should include precise details of the particular IRES sequence used. [Translational efficiency of EMCV IRES: BioTechniques (2006) 41(3):283–292.]

The 5′-UTR region of mRNA of the eukaryotic p27 protein was reported to include a sequence that functions as an IRES [Nucleic Acids Res (2007) 35(14):4767–4778].

Some authors have drawn a distinction between viral forms of IRES, which they regard as typically structure-dependent, and cellular counterparts, whose function was thought to rely more on short motifs and *trans*-acting factors [Nucleic Acids Res (2007) 35(14):4664–4677]. However, studies on natural variant forms of the IRES of the hepatitis C virus concluded that the primary sequence of IRES subdomain IIId is likely to play a key role in cap-independent translation [Nucleic Acids Res (2009) 37(3):957–971].

An IRES sequence was used in studies on murine leukemia virus for the development of (replicating forms of) retroviral vectors [BMC Mol Biol (2009) 10:8].

[Viral IRES RNA structures: Trends Biochem Sci (2008) 33(6):274–283; IRESite tool: Nucleic Acids Res (2009) doi: 10.1093/nar/gkp981.]

iRNA/DNA See OKAZAKI FRAGMENT.

iron oxide nanoparticles (used for gene delivery) See the entry GENE-DELIVERY SYSTEM.

IRS Internal resolution site: see the entry TN3.

IS INSERTION SEQUENCE.

IS element *Syn.* INSERTION SEQUENCE.

IS*1* A 768-bp INSERTION SEQUENCE present e.g. in some R plasmids (which encode antibiotic-resistance), in the chromosome of (some) strains of *Escherichia coli* K12, and in the

genome of phage P1.

Target sites for IS*1* appear to occur preferentially in AT-rich regions, and insertion may generate 9-bp or 8-bp target-site duplication.

IS*3* An INSERTION SEQUENCE found e.g. in the *finO* gene of the F plasmid (see FINOP SYSTEM).

IS*5* A 1195-bp INSERTION SEQUENCE found in the chromosome of strains of *Escherichia coli* K12.

IS*10* (IS*10*L, IS*10*R) See entry TN10.

IS*50* (IS*50*L, IS*50*R) See entry TN5.

IS*101* A 209-bp defective (gene-less) INSERTION SEQUENCE; transposition of IS*101* depends on functions encoded by the insertion sequence IS*1000*. (The inverted repeat sequences of IS*101* are similar, but not identical, to those of IS*1000*.)

IS*900* An INSERTION SEQUENCE that has been used as a marker of the species *Mycobacterium paratuberculosis*.

IS*911* An INSERTION SEQUENCE which forms a circular intermediate ('minicircle') between excision and insertion into the target sequence. During circularization of IS*911* juxtaposition of the inverted repeats (separated by a short sequence) brings together a –35 box and a –10 box which are reported to be at an optimal distance to provide an efficient promoter function for gene(s) in the insertion sequence.

IS*1000* (γδ) A 5.8-kb INSERTION SEQUENCE that has extensive homology with the DNA in transposon Tn*3*.

IS*1071* A 3.2-kb INSERTION SEQUENCE often associated with xenobiotic-degrading genes. The (110-bp) terminal inverted repeats and the transposase gene are related to those of class II transposons. Transposition involves the formation of a co-integrate, and a 5-bp target-site duplication is produced at the site of insertion.

[Functional analysis of IS*1071*: Appl Environ Microbiol (2006) 72(1):291–297.]

IS*1111* An (atypical) INSERTION SEQUENCE characterized by subterminal inverted repeats, a lack of target-site duplication on transposition, and the formation of a circular intermediate between excision and insertion into the target site.

(See also ISEC11.)

IS*6110* An insertion sequence apparently specific to members of the *Mycobacterium tuberculosis* complex – see INSERTION SEQUENCE for further details.

(See also IS900.)

I-SceI A HOMING ENDONUCLEASE: see earlier entry under 'I-'.

ISEc*11* An INSERTION SEQUENCE – related to members of the IS1111 family – present in the chromosome (and the virulence plasmid pINV) in a strain of enteroinvasive *Escherichia coli* (EIEC) [J Bacteriol (2006) 188(13):4681–4689].

ISH *In situ* hybridization: a PROBE-based procedure for detecting a target sequence *in situ*, at a normal or aberrant location within cells or tissues, or e.g. within an *in vitro* preparation of chromosomes. A probe used for ISH, or a conjugate which is subsequently bound to a hybridized probe, may exhibit any of various types of label. Radioactive labels were widely used in early protocols, but fluorescent labels (see FISH) are currently popular.

ISH has various applications in the study of structure and function in the biological sciences. In medicine it can be used e.g. to detect certain pathogens in clinical samples by using probes that bind to pathogen-specific sequences. In certain cases (e.g. *Mycobacterium tuberculosis*) a pathogen may be detected by ISH significantly earlier than it could be detected by culture. (ISH may also be able to distinguish virulent from non-virulent strains of an organism by the use of probes for specific virulence factors.)

ISHp*608* An INSERTION SEQUENCE, present within a plasmid in *Helicobacter pylori*, which excises as a circular element; the break in the parent molecule is resealed. Transposition is reported to be site-specific (5′-TTAC-3′) [EMBO J (2005) 24(18):3325–3338].

island rescue PCR A form of PCR which has been used e.g. for the amplification of CPG ISLAND sequences which are present in fragments of genomic DNA. This approach depends on the presence, in a fragment, of (i) an ALU SEQUENCE (Alu sequences are common in the human genome), and (ii) a GC-rich recognition site for a RESTRICTION ENDONUCLEASE such as BSSHII – GC-rich cleavage sites are commonly found to be associated with CpG islands.

Sample genomic DNA is cut with a restriction enzyme, and the linear fragments are ligated to *vectorette* linkers (see VECTORETTTE PCR). PCR is then carried out using a primer specific for an Alu sequence and a vectorette-specific primer.

isochore In human genomic DNA: a long sequence of nucleotides (of the order of hundreds of kilobases in length) that is characterized by a uniform (or near-uniform) content of the bases guanine + cytosine (G + C); the G + C content differs among different isochores. Isochores can be classified in five groups according to their G + C content.

Isochore mapping of the human genome has revealed some 3200 isochores which cover the entire genome [Genome Res (2006) 16(4):536–541].

[A web-based tool (GC-profile) for analyzing the variation in G + C content in genomic DNA: Nucleic Acids Res (2006) 34(Web Server issue):W686–W691.]

[G + C content, exon density, and recombination rate in the (human) genome: BMC Bioinformatics (2009) 10(Suppl 1): S66.]

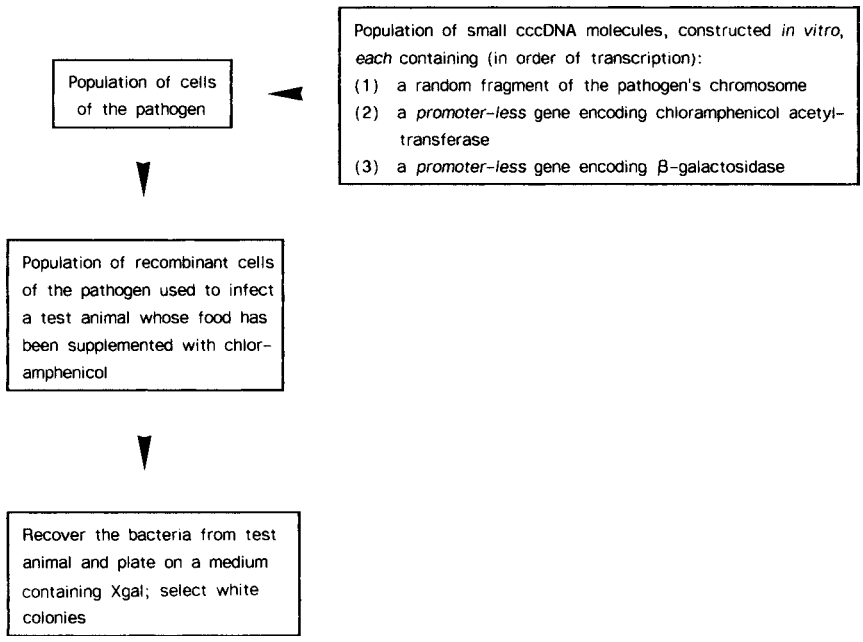
[The evolution of isochore patterns in vertebrate genomes: BMC Genomics (2009) 10:146.]

isoform Any one of two (or more) forms of a given protein that differ in structure but which have the same function.

isolation of DNA See DNA ISOLATION.

isoniazid (INH) Isonicotinic acid hydrazide: a synthetic antibiotic, used as an anti-tuberculosis drug, which is bactericidal for actively growing cells of e.g. *Mycobacterium tuberculosis* (and some other mycobacteria). The antibacterial function of isoniazid depends on its intracellular activation by a catalase (the *katG* gene product). The catalase-activated drug appears to inhibit synthesis of the essential cell-wall mycolic acid.

Resistance to isoniazid may result from mutation in any of several genes, including *katG* and *ahpC*.



IVET (*in vivo* expression technology): a method used e.g. for detecting those genes (in a bacterial pathogen) which are expressed *only during infection of the host animal*. The figure shows, diagrammatically, one of several versions of IVET; other versions are described at the end of the legend.

Vector molecules (see *top*, *right*) are inserted, by transformation, into a population of cells of the pathogen. In each cell the (random) fragment of chromosome in the vector inserts into the corresponding part of the pathogen's chromosome by an insertion–duplication mechanism; thus, the genes that are present in the vector are inserted into a specific part of the chromosome. Because the vector molecules contain different fragments of the chromosome, they will insert into different chromosomal sites in different cells – forming a heterogeneous population of recombinant cells.

In some of the recombinant cells the vector's two promoter-less genes will have inserted 'in frame' with an upstream promoter; in such cells both of these genes are transcribed *if the promoter is active*.

The recombinant cells are used to infect a test animal whose food contains the antibiotic chloramphenicol. Under these conditions, a recombinant cell can grow if it produces chloramphenicol acetyltransferase (CAT), i.e. if the CAT gene (in the vector) is controlled by an active promoter in the cell's chromosome; thus, if the given recombinant cell *grows* in the test animal this indicates that its CAT gene is controlled by a promoter *which is active within the test animal*. The cells that produce CAT can form large populations which will greatly outnumber those cells which do not form CAT: a cell that does not produce CAT will be inhibited by the chloramphenicol present in the animal's food.

We need to know whether the promoter controlling a CAT gene is active *only* when the pathogen is infecting a test animal – or whether it is also active when the pathogen is cultured (e.g. on agar media). If it is active only in the test animal, this indicates that the gene *normally* controlled by that promoter is induced ('switched on') during infection; such a gene is of interest because of its possible association with virulence.

To resolve this question we need to study promoter activity further. The bacteria recovered from the test animal are plated on a medium lacking chloramphenicol but containing the agent Xgal; all the cells can grow on this medium. If the promoter regulating the vector's genes is active in culture, then β-galactosidase will be formed in the cell, giving rise to a *blue-green* colony; hence, such a colony indicates that activation of the given promoter does not occur *only* in the test animal. By contrast, a *white* colony (a *lac*$^-$ colony) indicates that β-galactosidase and CAT are formed *only* within the test animal (not in culture) – i.e. the relevant promoter is active only during infection of the animal. The gene which is normally controlled by this promoter can be identified, cloned and sequenced, and its role in virulence can be investigated. (*continued*)

(See also LINE PROBE ASSAY.)

isonicotinic acid hydrazide See ISONIAZID.

isopeptidase *Syn*. SUMO protease (see entry SUMOYLATION).

isopropyl-β-D-thiogalactoside See entry IPTG.

isopsoralen See entry AMPLICON INACTIVATION.

isoschizomer Any RESTRICTION ENDONUCLEASE whose recognition site is identical to that of another restriction enzyme. In general, the restriction enzyme which is first shown to recognize a particular sequence is called the *prototype* (in relation to that sequence); a prototype may have many isoschizomers (which can occur in various species).

An isoschizomer whose *cutting* site (*within the recognition site*) differs from that of the prototype is referred to as a *neoschizomer*.

isotachophoresis (ITP) A form of ELECTROPHORESIS in which fragments of nucleic acid *of different sizes* can be focused into a single band, i.e. the polynucleotides behave in a way which is independent of size. DNA has been isolated from biological fluids by ITP and analyzed by PCR. As this may reveal cancer-associated genes, it has been suggested that ITP may be useful e.g. for detecting/monitoring cancer.

[Use (e.g.): Nucleic Acids Res (2002) 30:1688–1694.]

isothermal amplification (of DNA and RNA) Amplification at a fixed temperature (as opposed to amplification that depends on temperature *cycling* – as in PCR).

Some examples of isothermal amplification are given in the table in the entry NUCLEIC ACID AMPLIFICATION.

isotopic labeling (of probes) See PROBE LABELING.

ISR Intergenic spacer region: see RRN OPERON.

iteron One of a series of tandemly repeated (i.e. reiterated) sequences of nucleotides found e.g. in the origin of replication of certain types of PLASMID.

The term 'iteron' (singular) has also been used to refer to a *series* of these sequences.

(See also HANDCUFFING.)

ITP ISOTACHOPHORESIS.

ITR Inverted terminal repeat – see e.g. ADENOVIRUS.

***iutA* gene** In *Escherichia coli*: the gene which encodes an outer membrane protein that acts as a receptor for the siderophore aerobactin (and for the colicin cloacin DF13).

IVET *In vivo* expression technology: a method used originally for detecting those genes (in a pathogen) which are expressed *only during infection* of the host animal; any gene expressed only during infection of the host *may* be a virulence gene and, once identified, it can be followed up with animal tests using strains of the pathogen which have mutation(s) in the relevant gene.

The principle of IVET is shown (diagrammatically) in the figure (page 179).

The IVET principle is now exploited in different contexts. For example, the method was used to identify those genes of *Erwinia amylovora* which are induced only during infection of immature pear tissue [J Bacteriol (2005) 187(23):8088–8103], and was also used to study *in vivo* gene expression in *Lactobacillus reuteri* during the fermentation of sourdough [Appl Environ Microbiol (2005) 71(10):5873–5878].

[IVET for exploring niche-specific gene expression: Microbiol Mol Biol Rev (2005) 69(2):217–261.]

An IVET approach is also used to determine whether or not a given gene is expressed – under specified conditions – by using a reporter system in which expression of the given gene triggers a recombinational event that has a detectable effect; this protocol is referred to as recombination-based IVET or resolvase-based IVET (RIVET).

RIVET has been used e.g. for studying QUORUM SENSING in *Salmonella enterica*. To determine whether the gene *srgE* is activated by quorum sensing, a cassette (*srgE–tnpR–lacZY*) was inserted into a strain of *S. enterica* that carries a chromosomal cassette *res1–tetRA–res1*. Activation of *srgE* is linked to expression of the resolvase TnpR which (by acting on the *res1* sites) excises *tetRA*; hence (as *tetRA* encodes resistance to tetracycline), activation of the *srgE* gene is indicated by the loss of tetracycline resistance in the test organism [PLoS ONE (2008) 3(7):e2826].

IVF *In vitro* fertilization: a procedure used e.g. to circumvent problems of infertility, or to carry out preimplantation genetic

IVET (*in vivo* expression technology): (*continued*) A different version of IVET uses an *auxotrophic* population of the pathogen. These cells can (i) grow on agar media supplemented with the given growth requirement, but (ii) cannot grow within the test animal unless they contain the relevant – and functional – gene. Each vector molecule includes (in addition to the random fragment of chromosome) the genes encoding (i) the specific growth requirement and (ii) β-galactosidase – both of these genes being promoter-less. The rationale is analogous to that given above.

In another version of IVET, each vector molecule includes (i) a tetracycline-resistance transposon and (ii) promoter-less genes encoding a transposase and β-galactosidase. If the relevant promoter is active in the test animal, transposase is synthesized and excises the transposon; the excised transposon does not replicate – leading to a clone of tetracycline-*sensitive* cells. Tetracycline-sensitive cells can be detected by replica plating using media which (i) lack, and (ii) contain tetracycline. The β-galactosidase has the same function as in other versions of the method.

Figure reproduced from *Bacteria in Biology, Biotechnology and Medicine*, 6th edition, Figure 11.3, pages 360–361, Paul Singleton (2004) John Wiley & Sons Ltd, UK [ISBN 0-470-09027-8] with permission from the publisher.

diagnosis (in order to detect an inherited disorder – see PGD).

Following ovarian stimulation, oocytes are collected, by an ultrasound-guided instrument, just before ovulation. After the oocytes are fertilized, *in vitro*, they are briefly cultured to the (e.g.) four-cell stage and subsequently inserted into the uterus for development; usually only one embryo is involved.

Zygote intrafallopian transfer (ZIFT) refers to the transfer of the *fertilized* oocyte, at the zygote stage, to the recipient's oviduct.

Gamete intrafallopian transfer (GIFT) refers to transfer of oocyte and spermatozoa, separately, into the recipient's oviduct – fertilization and subsequent development occurring *in vivo*.

If infertility is due to insufficient numbers of spermatozoa, *intracytoplasmic sperm injection* (ICSI) can be used; in this procedure, an individual spermatozoan is injected, via a fine glass pipet, directly into the oocyte.

J

J biosynthesis Synthesis of BASE J (q.v.).

Jacob–Monod model An early model (~1960) for the bacterial system now known as an OPERON.

JAK kinase A member of the family of JANUS KINASES.

janiemycin A polypeptide with antibacterial activity; it inhibits transglycosylation during synthesis of the cell-wall polymer peptidoglycan.

Janus kinases A family of cell-membrane-associated tyrosine kinases: see e.g. SIGNAL TRANSDUCERS AND ACTIVATORS OF TRANSCRIPTION.

(See also TYROSINE KINASE.)

Java codon adaptation tool See CODON BIAS.

JC virus An icosahedral virus, with a ccc dsDNA genome, that belongs to the POLYOMAVIRUS group. The JC virus infects humans and other mammals; it can be oncogenic in newborn rodents, can transform certain types of cell in culture – and in immunocompromised humans, e.g. AIDS patients, it can give rise to a fatal demyelinating condition known as progressive multifocal leukoencephalopathy (PML). Adaptive mutations in the JC virus protein capsid have been associated with PML [PLoS Genet (2009) 5(2):e1000368].

DNA (but not proteins) of JC virus was reported in various regions of normal human brain tissue, including oligodendrocytes. As the virus can gain access to the brain in immuno-*competent* hosts, it was suggested that an impaired immune system (as e.g. in AIDS patients) may permit an endogenous JC virus to initiate a lytic infection in oligodendrocytes, i.e. a characteristic feature of PML [Ann Neurol (2008) 64(4):379–387].

JC virus is similar to another human polyomavirus, the BK virus. Detection/differentiation of these two viruses (in CSF (cerebrospinal fluid), serum and urine) may be achieved with a commercial hybridization assay [J Clin Microbiol (2006) 44(4):1305–1309].

It was suggested that co-infection with Epstein–Barr virus and the JC virus might be relevant to pathogenesis of a rare adrenal B cell lymphoma [Infect Agent Cancer (2009) 4:1].

JCat See CODON BIAS.

JM109 A strain of *Escherichia coli* used e.g. for cloning; it has mutations in various genes (e.g. *endA*, *recA*, *relA* and *supE*), and is suitable for blue–white screening and for single-strand rescue.

juglone A compound (5-hydroxy-1,4-naphthoquinone) found in leaves and roots of the walnut (*Juglans*); it has antibacterial and antifungal activity.

[Transcription-blocking activity: Nucleic Acids Res (2001) 29:767–773; possible immunosuppressant action: PLoS ONE (2007) 2(2):e226.]

Juglone was used to assess oxidative stress in the nematode *Caenorhabditis elegans* in a study involving deletion of the organism's superoxide dismutase (*sod*) genes [PLoS Genet (2009) 5(2):e1000361].

jumping gene *Syn.* TRANSPOSABLE ELEMENT.

junction-resolving enzyme See HOLLIDAY JUNCTION.

junk DNA That part of a (eukaryotic) genome which does not encode proteins or structural RNAs; the human genome was reported to contain >98% of such 'junk' DNA. This part of the genome specifies e.g. various types of small non-coding RNA molecule that are essential for gene regulation [see e.g. Mol Syst Biol (2008) 4:231].

The kind of organization in which only a small proportion of the genome encodes proteins or structural RNAs is found throughout the multicellular eukaryotes (but not in the lower eukaryotes or in the prokaryotes) [Nucleic Acids Res (2009) 37(4):1011–1034].

K

K (1) A specific indicator of ambiguity in the recognition site of a RESTRICTION ENDONUCLEASE or in another nucleic acid sequence; for example, in CMG↓CKG (enzyme NspBII) the 'K' indicates G or T. (In RNA 'K' indicates G or U.) In this example, 'M' is A or C.
(2) L-Lysine (alternative to Lys).

K-*ras* See RAS.

K1 toxin (of *Saccharomyces cerevisiae*) See KILLER FACTOR.

Kaiso protein A protein that binds to DNA containing methyl-CpGs; the binding involves a characteristic zinc finger motif. The protein also binds to the sequence CTGCNA.

Mice with a null mutation in the *Kaiso* gene were found to show resistance to intestinal cancer; this suggested the possibility that Kaiso might be a potential target for therapeutic intervention [Mol Cell Biol (2006) 26(1):199–208].

Kaiso-like proteins designated ZBTB4 and ZBTB38, unlike Kaiso, are able to bind to *single* methylated CpGs [Mol Cell Biol (2006) 26(1):169–181].

(See also MBD PROTEINS.)

kanamycin An AMINOGLYCOSIDE ANTIBIOTIC that is used e.g. as a selective agent for bacteria containing a vector carrying a kanamycin-resistance gene.

Kaposi's sarcoma A form of SARCOMA found e.g. in immunosuppressed patients, including AIDS patients; this sarcoma has been associated with human herpesvirus 8 (HHV-8).

In vitro activation of latent HHV-8 was achieved by promoter demethylation using the reagent tetradecanoylphorbol acetate (TPA); a relationship was suggested between methylation status, transactivation, and HHV-8-mediated disease.

vIL-6 – a form of IL-6 (interleukin-6) encoded by HHV-8, and apparently associated with pathogenesis – was targeted by an INTRABODY [J Virol (2006) 80(17):8510–8520].

karyogram See IDIOGRAM.

karyolysis (*histopathol.*) Degradation of a eukaryotic nucleus – with loss of affinity for basic dyes.

(cf. KARYORRHEXIS.)

karyomere See MACRONUCLEUS.

karyonide (*ciliate protozool.*) A clone of cells whose macronuclei are derived from the same parental macronucleus.

karyopherins In eukaryotic cells: agents involved in transport of proteins/nucleoprotein complexes between the nucleus and the cytoplasm; they include the *exportins* and the *importins*. Exportin 1 is also called CRM1.

karyoplast An isolated eukaryotic nucleus enclosed within a sac of cytoplasmic membrane that includes a small amount of cytoplasm.

karyorrhexis (*histopathol.*) The fragmentation of a eukaryotic nucleus.

(See also KARYOLYSIS and PYKNOSIS.)

karyosome A nucleolus-like body, or a nucleolus.

karyotic Refers to cells that contain a nucleus, i.e. most of the cells of eukaryotes, but not e.g. erythrocytes (red blood cells).

karyotype A representation of the chromsomes in the nucleus of a given type of (eukaryotic) cell – showing their size and morphology (usually) at metaphase.

(cf. IDIOGRAM.)

kasugamycin An AMINOGLYCOSIDE ANTIBIOTIC that has antibacterial activity against mainly Gram-positive species. It has also been used to control rice blast disease due to the fungus *Pyricularia oryzae*.

(See also ANTIBIOTIC.)

***katG* gene** See ISONIAZID and LINE PROBE ASSAY.

kb Kilobase: a length of 10^3 bases in a single-stranded nucleic acid, or 10^3 base-pairs in a double-stranded nucleic acid. In double-stranded nucleic acids a length may be given in *base pairs* (bp) or in *kilobase-pairs* (kbp).

kbp Kilobase-pairs: see KB.

kDNA See KINETOPLAST.

***kdpABC* operon** See TWO-COMPONENT REGULATORY SYSTEM.

Kemptide sequence The sequence of amino acid residues:

leucine-arginine-arginine-alanine-serine-leucine-glycine

(also written as LRRASLG). This sequence is used e.g. as an isolated substrate in a test for the enzyme cAMP-dependent protein kinase A (PKA), which phosphorylates the Kemptide sequence.

The Kemptide sequence is also used as a tag that is suitable for labeling recombinant proteins expressed in *E. coli*; thus, an expressed protein carrying a Kemptide tag can be labeled (by the PKA enzyme) with ^{32}P (see e.g. pCAL vectors in the entry AFFINITY PROTEIN EXPRESSION AND PURIFICATION).

[Examples of use: Mol Cell Biol (2008) 28(11):3804–3816; BMC Biochem (2008) 9:18; J Clin Invest (2008) 118:3966–3979; BMC Cancer (2009) 9:63.]

Kepivance® See BIOPHARMACEUTICAL (table).

Ki-67 A nuclear antigen – used e.g. as a marker of cell proliferation – which is expressed only in cycling cells; for example, in a population of cells, it is found that expression of Ki-67 correlates strongly with cells in the S phase of the cell cycle (DNA replication).

A quantitative assay of (stained) Ki-67 in tumor sections is used to estimate the *proliferation index* of the tumor.

Ki-*ras* See RAS.

Kid See PARD SYSTEM.

(See also R1 PLASMID.)

killer factor (in *Saccharomyces cerevisiae*) In 'killer' strains of *S. cerevisiae*: any of several secreted protein toxins encoded by a (cytoplasmically inherited) dsRNA element. Type 1 (or K1) killer strains form the K1 toxin – which binds initially to the cell wall of a sensitive cell and subsequently disrupts the cytoplasmic membrane.

(cf. KILLER PLASMIDS.)

killer plasmids (in yeasts) In certain strains of *Kluyveromyces marxianus*: two linear, multicopy dsDNA plasmids (pG*l*1 and pG*l*2); cells containing these plasmids produce a glycoprotein

toxin that kills (sensitive) strains of e.g. *Candida*, *Kluyveromyces* and *Saccharomyces*.

Killer plasmids can be transferred to other yeasts, e.g. by protoplast fusion; the recipients then exhibit the killer phenotype.

(cf. KILLER FACTOR.)

kilobase See KB.

Kineret® See BIOPHARMACEUTICAL (table).

kinetochore In eukaryotic chromosomes: either of two multiprotein structures that assemble on both sides of the centromere at cell division; they are involved e.g. in attachment to the spindle and in regulating alignment of chromosomes prior to anaphase.

The outer part of a vertebrate's kinetochore includes microtubule-binding proteins (such as CENP-E), while the inner part includes e.g. CENP-A, -B, -C, -H and -I proteins, which keep their positions throughout the cell cycle.

A knock-out mutation of CENP-A was reported to result in misaggregation of the kinetochore.

Stability of the genome during cell division is dependent on normal functioning of the kinetochore, including its correct attachment to microtubules of the spindle apparatus. Specific microtubule-depolymerizing kinesins have been reported to operate in certain phases of mitosis to correct mal-orientation in kinetochore–microtubule association(s) [Nature Cell Biol (2008) 11:27–35].

kinetoplast A complex network of catenated DNA molecules (kDNA) within the mitochondrion in protozoa of the order Kinetoplastida (e.g *Trypanosoma*). There are ~20–50 large circular molecules (*maxicircles*) and several thousand small circles (*minicircles*).

Maxicircles appear to be equivalent to mitochondrial DNA in other eukaryotes.

Minicircles encode most of the guide RNAs (gRNAs) that are involved in RNA EDITING [kDNA: Eukaryotic Cell (2002) 1:495–502].

The mitochondrion (including the kinetoplast) divides prior to cell division.

Division of the kinetoplast can be inhibited by certain intercalating agents, e.g. some phenanthridine derivatives such as ethidium bromide.

Kis See PARD SYSTEM.

kit (*KIT*) An ONCOGENE present in a strain of feline sarcoma virus. The c-*kit* product is a transmembrane protein with tyrosine kinase activity in the cytoplasmic region; it binds to, and enhances the activity of, *stem cell factor* (= kit ligand; mast cell growth factor; steel factor): an agent which can e.g. stimulate the proliferation of mast cells and regulate the growth/survival of hemopoietic precursors.

c-*kit* is strongly expressed in certain types of cancer.

kit ligand See KIT.

Kleinschmidt monolayer technique A method used for examining molecules of nucleic acid by electron microscopy. The nucleic acid is first coated with e.g. cytochrome *c* and it then forms extended molecules within a protein monolayer at an air–liquid interface; part of this monolayer is transferred to a grid, shadowed with heavy metal, and then examined under the electron microscope.

Klenow fragment See the entry DNA POLYMERASE I.

KM13 A HELPER PHAGE (q.v.). (See also the entry M13K07.)

knock-in mutation Any engineered change in a nucleotide sequence which is designed to promote the expression of a given gene or genes – including e.g. gene(s) previously inactivated by a KNOCK-OUT MUTATION.

Such a procedure may be used e.g. to insert a gene, or to modify an existing gene and/or its control system.

knock-out mutation (a knock-out) Any engineered change in a nucleotide sequence designed to eliminate particular gene(s) – either by their removal or by disruption of essential coding and/or control sequences.

(Compare GENE SILENCING and GENE TARGETING; see also the entry CONDITIONAL KNOCK-OUT.)

Knock-out can be effected by various procedures involving e.g. HOMOLOGOUS RECOMBINATION (see figure legend in the entry INSERTION–DUPLICATION RECOMBINATION); LAMBDA RED RECOMBINATION system; and ZINC-FINGER NUCLEASES.

(See also entries KNOCK-IN MUTATION and KNOCKDOWN.)

knockdown A form of GENE SILENCING which is frequently mediated by RNA INTERFERENCE in studies on mammalian systems, and by MORPHOLINO ANTISENSE OLIGOMERS e.g. in studies on the zebrafish (*Danio rerio*); however, there is no clear-cut division of methodology in knockdown studies – for example, antisense morpholino oligomers have been used in studies on mammalian (kidney) cells [BioTechniques (2008) 44(4):547–549].

[Defining the optimal parameters for hairpin-based knockdown constructs: RNA (2007) 13(10):1765–1774.]

[Morpholino antisense oligomers in knockdown studies on zebrafish: J Cell Biol (2008) 181(2):381–394 and PLoS ONE (2008) 3(9):e3114.]

'Knockdown' is also used to refer to the *organism* in which one or more genes have been silenced – or even to the genes themselves.

Note. 'Knockdown' has also been used to refer to the process of blocking the activity of MICRORNAS by means of antisense oligomers; thus, when the miRNA's function is *inhibitory*, the knockdown process produces an effect which is the opposite of gene silencing.

knockdown plasmid Any plasmid containing a nucleotide sequence which, when transcribed intracellularly, gives rise to a SHORT HAIRPIN RNA designed for the KNOCKDOWN of a particular gene.

Kogenate® Recombinant human factor VIII (Bayer) which is used e.g. for therapy in cases of hemophilia A.

[Use (in the *in vitro* assay of factor VIII): Biochem J (2006) 396(2):355–362.]

(See also the table in the entry BIOPHARMACEUTICAL.)

Kornberg enzyme *Syn*. DNA POLYMERASE I.

Kozak sequence A sequence of nucleotides, flanking the initiator codon, reported to optimize initiation of translation by

eukaryotic ribosomes. In the original paper, the sequence was reported as ACCATGG.

[Examples of use: Clin Vaccine Immunol (2007) 14(1):28–35; Acta Vet Scand (2008) 50(1):44; Mol Biol Cell (2009) 20(1):389–399.]

The Kozak sequence is also used e.g. in some vectors of the EXCHANGER SYSTEM.

KpnI A type IIP RESTRICTION ENDONUCLEASE (see the table in the entry).

KpnI family (of LINEs) See LINE.

Ku70 A protein essential e.g. for the NON-HOMOLOGOUS DNA END-JOINING (NHEJ) pathway of DNA repair – in which it forms part of a (heterodimeric) end-binding complex. (In this complex, Ku70's partner protein is designated Ku80 by some authors but Ku86 by others.)

Ku70 has been studied to determine the ability of NHEJ to compete with (i.e. interfere with) homologous recombination (HR) in a way that inhibits HR-dependent insertion of vector-borne sequences into genomic DNA. One study (which used human somatic cells) generated knock-out mutations in *Ku70* and looked at the effect of these mutations on the targeting success of AAV vectors; $Ku70^{+/-}$ cells were found to exhibit a gene targeting frequency 5- to 10-fold higher than that found in wild-type cells [Proc Natl Acad Sci USA (2008) 105(25): 8703–8708].

(See also GRANZYMES.)

KWWCRW A hexameric peptide that can e.g. inhibit enzymes involved in the resolution of HOLLIDAY JUNCTIONS (q.v.).

L

L L-Leucine (alternative to Leu).

L-complex The EDITOSOME in trypanosomatids.

L-RNA (Spiegelmer) A mirror-image form of naturally occurring RNA.

L-SIGN *Syn.* CD209L (q.v.).

L strand (of dsDNA) See the entry H STRAND.

labeling (of antibodies) See e.g. ZENON ANTIBODY-LABELING REAGENTS.

labeling (plasmids) See e.g. NICK TRANSLATION. Many of the methods described in PROBE LABELING are also applicable to the labeling of plasmids.

labeling (primers) See PRIMER.

labeling (probes) See PROBE LABELING; see also ACCUPROBE.

labeling (proteins) See PROTEIN LABELING.

***lac* operon** (lactose operon) In e.g. *Escherichia coli* (and some other bacteria): an OPERON that encodes proteins involved in the transport (uptake) and use of β-galactosides (e.g. lactose); in *E. coli* this operon is located at ∼8 minutes on the chromosome map.

This operon contains (i) three regulated genes: *lacZ*, *lacY* and *lacA*; (ii) a promoter (P), (iii) an operator (O), between the promoter and the first structural gene (*lacZ*), and (iv) the regulator gene (*lacI*); *lacI* is upstream of the promoter, and it encodes a repressor protein. The *lacI* gene is transcribed constitutively, at a low rate, from its own independent promoter.

The *lacZ* gene encodes the enzyme β-galactosidase, which cleaves lactose into glucose and galactose; this reaction also yields a minor product, ALLOLACTOSE, which is the (*in vivo*) inducer of the *lac* operon.

The *lacY* gene encodes β-galactoside permease – involved in the transport (i.e. uptake) of β-galactosides into the cell.

The *lacA* gene encodes thiogalactoside transacetylase, an enzyme which catalyzes the transfer of an acetyl group from acetyl-CoA to a β-galactoside. The function LacA in lactose metabolism is apparently unknown; one suggestion was that it might be used to detoxify any non-metabolizable sugars by acetylating them and facilitating their excretion.

In the absence of inducer, the regulator protein, LacI, binds to the operator and minimizes transcription of the operon by the RNA polymerase. A low level of transcription appears to occur in each cell under these conditions.

In the presence of lactose, the enzyme β-galactosidase converts some of the lactose to *allolactose*; allolactose acts as an inducer of the *lac* operon by binding to – and inactivating – LacI (the regulator protein). Once the regulator protein has been inactivated, the genes of the operon can be transcribed. The (single) mRNA transcript (i.e. a *polycistronic* transcript) includes an initiator codon and a stop codon for each of the genes.

Note that the *lac* operon is regulated not only by the LacI repressor protein but also by CATABOLITE REPRESSION (q.v.).

When the available lactose has been used up, transcription of the operon is no longer needed; the inducer (allolactose) is not formed, and the (now active) regulator protein switches off the operon. The mRNA is degraded.

When used in DNA technology, the *lac* operator is switched on (i.e. 'induced'), as required, by adding the compound isopropyl-β-D-thiogalactoside (IPTG), which blocks the LacI repressor protein. (IPTG is a *gratuitous* inducer, i.e. although it induces the operon it is not metabolized by the cell.) This switching mechanism can also be used when the expression of *other* genes is regulated via the *lac* operator sequence – in which case the *lac* operator sequence is inserted upstream of the gene(s) to be controlled. (Another approach to controlling transcription is described in OVEREXPRESSION (*Optimization of transcription*).)

Although this account describes the essential operation of the *lac* operon, the actual process is rather more complex. For example, the regulator protein can bind to more than one site: it also binds to sites designated O-2 and O-3, albeit with a lower affinity. During repression of the operon it appears that a tetramer of regulator protein binds at the main operator (O) and at O-2 (or O-3) – thus forming a loop of DNA – and that this may stabilize binding between the regulator and operator. Gene repression in the *lac* operon has been investigated in the context of DNA looping [PLoS ONE (2006) 1(1):e136].

As well as the main promoter (designated P1), the operon includes various overlapping promoter-like sequences. Some of these promoter-like sequences can be modified by a single base-pair substitution to yield a Lac^+ phenotype independent of the catabolite activator protein (CRP – see CATABOLITE REPRESSION).

Mutants with defective regulation of the lac operon

Constitutive transcription of the operon occurs in those $lacI^-$ strains which fail to produce a repressor or which produce an inactive one.

The $lacI^s$ mutants form a super-repressor which binds to the operator but which has little or no affinity for the inducer; as a result, the repressor binds to the operator even when an inducer is present – so that the operon is non-inducible.

In $lacO^c$ (= o^c) mutants the operator has an altered affinity for the repressor; if the affinity is too low the *lac* genes are expressed constitutively.

Cells with a $lacI^q$ gene produce higher-than-normal levels of repressor protein. This allele is included in some expression vectors in order to obtain 'tight' control of the expression of gene(s) that are regulated via a *lac* operator, and $lacI^q$ is also included in some strains of bacteria (e.g. XL1-BLUE) for the same reason.

Lac plasmid LACTOSE PLASMID.

***lacI*^{q}** A mutant repressor gene of the *lac* operon associated with high-level production of the repressor protein. It is included in some expression vectors (and cells, e.g. XL1-BLUE) so that tighter control can be exercised over *lac*-regulated gene(s).

***lacI*^{s} mutant** See LAC OPERON.

***lacO*^{c} mutant** See LAC OPERON.

LacR An alternative designation of the LacI repressor/regulator protein of the LAC OPERON.

β-lactam antibiotics A class of antibiotics in which the molecule is characterized by the 4-membered, nitrogen-containing β-lactam ring; all the antibiotics in this group kill, or inhibit, bacteria by targeting the so-called *penicillin-binding proteins* (PBPs) – enzymes involved in the synthesis and maintenance of the bacterial cell wall polymer peptidoglycan.

A given β-lactam antibiotic may be more effective against Gram-positive species than against Gram-negative species, or vice versa; in general, a given β-lactam is ineffective against a number of bacterial species.

Bacterial resistance to these antibiotics commonly involves (i) the production of β-LACTAMASES and/or (ii) synthesis of a (*mecA*-encoded) PBP (designated PBP 2a, or PBP2′) which is resistant to a range of β-lactams. Resistance may also be due e.g. to reduced permeability of the cell envelope to particular antibiotic(s).

These antibiotics include CEPHALOSPORINS, carbapenems, CEPHAMYCINS, clavams, nocardicins and various penicillins (see PENICILLIN).

β-lactamases Enzymes, produced by certain bacteria, which inactivate susceptible β-LACTAM ANTIBIOTICS by cleaving the β-lactam ring at the C–N bond. Some β-lactam antibiotics are inherently resistant to some β-lactamases; in some cases the resistance of a given β-lactam antibiotic to specific β-lactamases may be attributed to particular structural features (such as the 7-α-methoxy group in cephamycins and the cysteamine side-chain of thienamycin).

The β-lactam antibiotic clavulanic acid (a clavam) has only weak antibacterial activity but it strongly inhibits a number of β-lactamases; it is used therapeutically e.g. in combination with certain β-lactamase-sensitive β-lactams that have greater antibacterial activity (see e.g. AUGMENTIN).

Use of β-lactamase genes in DNA technology

A β-lactamase gene (*bla*) is employed (as a reporter gene) in various commercial fusion vectors (GeneBLAzer™) that are marketed by Invitrogen: see pcDNA™6.2/GeneBLAzer™ in the entry GATEWAY SITE-SPECIFIC RECOMBINATION SYSTEM (table) for details of the reporter mechanism.

Detection of β-lactamases

The detection of β-lactamases can be facilitated e.g. by using certain substrates which give a color reaction when they are cleaved by these enzymes: see, for example, NITROCEFIN and PADAC. A nitrocefin substrate was used e.g. for a β-lactamase reporter system selecting high-producing clones of the yeast *Pichia pastoris* [BioTechniques (2008) 44(4):477–484].

lacticin 3147 See LANTIBIOTIC.

lactoferrin A SIDEROPHILIN found e.g. in milk and tears, and in neutrophils. A humanized (and also immunologically compatible) recombinant form was produced in 'glyco-engineered' *Pichia pastoris* [Glycoconj J (2008) 25(6):581–593].

lactose operon *Syn*. LAC OPERON.

lactose plasmid (Lac plasmid) In certain types of bacteria (e.g. some strains of lactococci): a plasmid that encodes the uptake and metabolism of lactose.

lacZ A gene that encodes β-galactosidase (EC 3.2.1.23), an enzyme that hydrolyzes lactose to glucose and galactose; *lacZ* is found in the LAC OPERON in *Escherichia coli* and in some other organisms.

***lacZ* fusion** Any fusion (e.g. between two genes) in which *lacZ* is a partner (see also GENE FUSION).

***lacZΔM15* mutation** See e.g. the entry PBLUESCRIPT.

ladder (DNA ladder) See GEL ELECTROPHORESIS.

Laemmli electrophoresis See SDS–PAGE.

LAGLIDADG endonucleases See HOMING ENDONUCLEASE.

LAL test *Limulus* amebocyte lysate test: a procedure used for detecting or quantitating the pyrogen LIPOPOLYSACCHARIDE (LPS). [Example of use: Genet Vaccines Ther (2007) 5:2.] In DNA technology it is used for detecting LPS in recombinant products synthesized in Gram-negative bacteria.

The LAL test is based on the ability of LPS to bring about gelation of a lysate of amebocytes (blood cells) of the horseshoe crab (*Limulus polyphemus*). A quantity as small as 10^{-9} gram of LPS/mL can be detected in this way; the test is read within 1 hour – the formation of a clot (a coagulated lysate) being recorded as a positive test (i.e. LPS present).

The capacity to cause gelation (clot formation) in the LAL test differs in different types of LPS.

Some substances other than LPS, such as lipoteichoic acid and polynucleotides, are reported to cause a clot in the LAL test.

As an alternative procedure, the product under test is introduced parenterally into three rabbits. The presence or absence of pyrogen in the product is subsequently inferred from the presence or absence of a specified rise in temperature in the animals. This test is useful for pyrogens in general (not only LPS). However, the test is precluded (in favor of the LAL test) if the product itself (apart from any contaminating LPS) has known pyrogenic properties.

lambda exonuclease See EXONUCLEASE.

lambda FIX® II vector A CLONING VECTOR (Stratagene, La Jolla CA), derived from PHAGE LAMBDA, which is able to accomodate inserts of 9–23 kb. Lambda Fix® II is a replacement vector in which the gene/insert of interest is used to replace an excised portion of the lambda genome.

A central (∼14 kb) sequence of the genome is removed by digestion with the RESTRICTION ENDONUCLEASE XhoI. This produces two terminal regions of the genome (20 kb, 9 kb), each having an outer *cos* sequence, which are subsequently to be ligated on either side of the gene or insert of interest. The vector–insert–vector construct can then be packaged (e.g. by a GIGAPACK product) to form an infective phage particle.

The initial digestion of the genome (with XhoI) produces 5′-TCGA overhangs on the 20-kb and 9-kb fragments. These overhangs are subjected to a *partial fill-in* procedure, using DNA polymerase and the deoxyribonucleoside triphosphates dTTP and dCTP, to produce 5′-TC overhangs; these overhangs do not base-pair, so this avoids subsequent self-ligation

between the two vector fragments.

The sample DNA is cut with one of the restriction enzymes BamHI, MboI, BglII or Sau3AI – see table in RESTRICTION ENDONUCLEASE for recognition sequences. A partial fill-in is carried out with dGTP and dATP, leaving 5′-GA overhangs; as these fragments cannot base-pair, this avoids the problem of subsequent ligation of multiple inserts to the vector ends.

Binding can now occur between the partially filled-in ends of the vector and the partially filled-in ends of the fragments.

Prior to ligation, the fragments of sample DNA, produced by restriction, can be size-fractionated in a sucrose gradient, allowing selection of fragments within a given range of sizes; this permits use to be made of those fragments which, when inserted into the cloning vector, will yield products of a size suitable for packaging.

The Lambda FIX® II vector includes T7 and T3 promoter sequences on either side of the cloning gap to permit transcription, if required.

The *unwanted* central part of the lambda genome (containing e.g. the Red recombination genes) – which is formed by XhoI digestion – is selected against by plating the recombinant λ phages on a host strain of *E. coli* that is lysogenic for phage P2; λ phages in which the (recombinant) genome lacks the Red genes are able to form plaques on this host strain.

lambda phage See PHAGE LAMBDA.

lambda (λ) Red recombination A type of facilitated homologous recombination (in *Escherichia coli* or other organisms) involving the Red recombinase function of PHAGE LAMBDA (genes γ, β and *exo* – encoding products Gam, Bet and Exo, respectively); this permits *linear* fragments of DNA (whose terminal sequences are homologous to the target sequences) to be inserted into cells (e.g. using ELECTROPORATION or by transformation) and then undergo recombination at sites in the chromosome (or plasmid etc.) homologous to the terminal sequences on the fragment. The homologous sequences on a fragment can be as short as (e.g.) ∼50 bp; the fragments may be prepared by PCR using primers which carry the terminal (homologous) sequences as 5′ tags.

Lambda Red recombination is a form of RECOMBINEERING (q.v.).

The Exo product is a 5′-to-3′ exonuclease which produces 3′ single-stranded overhangs in fragments of linear dsDNA.

The Bet product is a *single-strand-annealing protein* which pairs an Exo-produced overhang to its (complementary) sequence in the target DNA.

Linear fragments of DNA are usually degraded in the cells of *E. coli* by exonuclease activity, but the Gam component of the Red system inhibits (*recD*-encoded) exonuclease activity of the host cell's RecBCD recombination system. (However, it has been reported that Gam may not be an absolute requirement for recombineering – particularly for small fragments [Proc Natl Acad Sci USA (2008) 105(5):1626–1631].)

The λ Red system can be used in either a plasmid-encoded or chromosome-encoded format, i.e. the genes encoding this system can be incorporated in either plasmid or chromosome.

This recombination system permits targeted replacement, inactivation or mutagenesis of genes – *in vivo* – using linear fragments of DNA. For example, PCR-derived products have been used to inactivate a range of genes in *E. coli*; each of the targeted genes was initially substituted with an antibiotic-resistance gene (allowing selection of the recombinant cells), and the antibiotic-resistance gene was subsequently excised [Proc Natl Acad Sci USA (2000) 97(12):6640–6645]. In this experiment, the antibiotic-resistance gene was first amplified by PCR. For this reaction, one primer carried a 5′ extension homologous to the terminal sequence of *the gene on one side of the target gene*; the other primer carried a 5′ extension that was homologous to the terminal sequence of the gene on the *other* side of the target gene. Also, the copy of the antibiotic-resistance gene was bracketed by FRT sites – recognized by a site-specific recombinase (see the entry FLP).

Following transformation of cells with these PCR products, the two end sections of an amplicon undergo λ-Red-mediated recombination with the genes *flanking* the target gene – thus replacing the target gene with an antibiotic-resistance gene. Selection of recombinant cells is facilitated by their ability to grow in the presence of the given antibiotic.

Finally, the antibiotic-resistance gene is removed by inserting a helper plasmid encoding the Flp recombinase – which excises the gene via the FRT sites.

In this protocol, the 5′ extensions on the PCR primers were only 36–50 nucleotides long, yet this allowed effective site-directed insertion. Moreover, while it is possible to transform *recD* mutants with linear fragments, the Red protocol can be carried out in wild-type cells.

A lambda (λ) Red recombination system has also been used for inserting TEV protease cleavage sites into the *putA* gene of *Salmonella enterica* in order to carry out probing of protein structure/function *in vivo* [BioTechniques (2006) 41(6):721–724].

The efficiency of the (λ) Red recombination system may be improved by co-expressing the RecA protein with the Gam, Bet and Exo proteins [Mol Biotechnol (2006) 32:43–53].

[Preparation of an improved cassette, encoding resistance to gentamicin, for use in chromosomal gene replacement in *Escherichia coli* by the lambda Red recombination technique: BioTechniques (2006) 41(3):261–264.]

The Red recombination system has been used for the rapid generation of reporter gene fusions, at defined locations, in bacterial genomes [Appl Environ Microbiol (2007) 73(13): 4234–4242].

In *Caenorhabditis elegans*, the Red system has been used for installing fosmid-based reporter genes [PLoS ONE (2009) 4(3):e4625].

Red recombination has been used for introducing mutations in EPEC strains of *Escherichia coli* [BMC Microbiol (2009) 9:30].

Red recombination has been used for genetic engineering in *Pantoea ananatis* [BMC Mol Biol (2009) 10:34].

Red recombination has been used for gene deletions and for

constructing *lacZ* fusions [BMC Gemomics (2009) 10:165].

lambdoid phages A class of temperate phages characterized by the ability of their genomes to undergo recombination; these phages include λ, 21, φ80 and P22. While the phages have a similar sequence of genes, the gene products are often phage-specific; moreover, some of the phages (e.g. λ, 21, φ80) have a unique (i.e. phage-specific) immunity region – hence, a cell lysogenized by one of these phages does not exhibit SUPERINFECTION IMMUNITY in respect of any of the others.

lamin kinase See NUCLEAR LAMINA.

lamins See NUCLEAR LAMINA.

lamivudine The (–) enantiomer of 2′-deoxy-3′-thiacytidine: an antiviral agent active against retroviruses (see NUCLEOSIDE REVERSE TRANSCRIPTASE INHIBITOR) – and also e.g. against some DNA viruses.

LAMP (loop-mediated isothermal amplification) A method for NUCLEIC ACID AMPLIFICATION (DNA or RNA) developed by the Eiken Chemical Co Ltd of Japan.

For amplifying DNA the reaction mixture contains sample (template) nucleic acid, four types of primer, nucleoside triphosphates, and a DNA polymerase which has STRAND DISPLACEMENT activity.

Incubation of the reaction mixture is carried out at ~65°C; at such a temperature dsDNA is in dynamic equilibrium with ssDNA, and primers are able to bind to their complementary sites.

The reaction proceeds in two phases, the second following on automatically from the first.

In the initial phase, a primer binds to target DNA and is extended (copying the target sequence). A *bumper primer* (see the entry STRAND DISPLACEMENT), which binds upstream of the first primer, is extended and displaces the complementary copy of the target sequence. The 5′ end (primer segment) of the displaced strand forms a hairpin structure by binding to a complementary sequence a short distance downstream on the displaced strand.

A primer binds near the 3′ end of the displaced strand; it's extended to the 5′ end of the displaced strand. The resulting (complementary) copy of the displaced strand is displaced by a bumper primer to form an isolated, single-stranded piece of DNA; the 5′ end of this ssDNA forms a hairpin – as does the 3′ end.

The dumb-bell-shaped ssDNA product (which has a hairpin structure at both ends) serves as the starting structure for the cyclic amplification phase, which follows.

The 3′ end of the dumb-bell structure is extended fully (disrupting the hairpin at the 5′ end).

A primer binds to the existing single-stranded (loop) region and is extended, displacing the strand which had been newly synthesized by extension of the 3′ terminus; the 3′ end of *this* displaced strand forms a hairpin. Extension from this 3′ end leads to the formation of products, each of which consists of a chain of copies of the target sequence.

(RNA amplification by this technique involves the use of a reverse transcriptase.)

Further details of LAMP (including diagrams) is available at:

http://loopamp.eiken.co.jp/e/lamp/anim.html

This method has been used e.g. for the detection of nucleic acid sequences from genetically modified organisms [BMC Biotechnol (2009) 9:7].

landmark enzyme The enzyme (usually NotI) which is used in RESTRICTION LANDMARK GENOMIC SCANNING.

lantibiotic Any of the (typically pore-forming) BACTERIOCINS in which the molecule includes certain unusual residues, such as D-amino acids and lanthionine (a sulfur-containing amino acid).

Lantibiotics are synthesized by Gram-positive bacteria (e.g. species of *Bacillus* and *Lactococcus*); they are active against other Gram-positive bacteria.

The food preservative *nisin* is one example of a lantibiotic.

Lacticin 3147 is a lantibiotic which consists of two peptide components – both of which are necessary for activity; each component requires modification by a separate enzyme.

Lantus® See BIOPHARMACEUTICAL (table).

large T antigen (of SV40 virus) See T ANTIGEN.

lariat See SPLICING. (See also INTRON HOMING.)

latency (viral) Refers to the situation in which a virus is infecting a cell but is not producing virions – although it retains the potential to produce virions. For example, in some cells, the genome of HIV-1 remains integrated in the host's genomic DNA without giving rise to virions (see *Quiescent HIV-1* in the entry AIDS).

Latency among viruses is not uncommon: see, for example, EPSTEIN–BARR VIRUS, KAPOSI'S SARCOMA (HHV8), NASBA (cytomegalovirus) and SV40.

(See also BORTEZOMIB and EBER.)

lawn plate A Petri dish containing an agar gel medium whose surface bears a continuous (confluent) growth of microorganisms (usually bacteria) of one species or strain.

(See also FILAMENTOUS PHAGES.)

LB broth A liquid medium used for the growth/maintenance of *Escherichia coli* and other bacteria. LB broth contains (grams per liter): tryptone 10, yeast extract 5, sodium chloride 5; pH 7.

LCN DNA test Low-copy-number DNA test: a PCR-based test used for obtaining a DNA profile by using *minute* amounts of sample DNA – i.e. less than the quantities normally used for standard PCR-based assays; an LCN assay is carried out on a sample containing less than 100 pg [BMC Genomics (2009) 10:159], or (for STR assays) <200 pg [Croat Med J (2009) 50(3):207–217]. In LCN tests, more cycles of amplification are required than in the standard method (this being a direct consequence of the smaller amount of sample DNA used).

This test is even more susceptible than the standard method to contamination. To provide DNA-based evidence in a legal context, contamination must be excluded by ensuring the use of suitable methods for collecting the sample DNA.

Opinion has differed on whether or not the scientific basis of this test is sound. In the United Kingdom, use of the LCN

approach was suspended temporarily as a result of the critical comments on the test from a judge in a legal case. A review, commissioned by the UK government, examined the fitness of the test (for forensic purposes), and, in 2008, the review found that the method is essentially sound – but made certain recommendations regarding e.g. the manner in which sample DNA should be collected. However, a different view is that – while LCN typing is appropriate for certain purposes such as identification of missing persons, or human remains etc. – its application to other tasks should be made with caution until improvements in the method make it a more reliable form of test [Croat Med J (2009) 50(3):207–217].

(See also FORENSIC APPLICATIONS.)

LCR LIGASE CHAIN REACTION.

LCx® assays Diagnostic assays (Abbott Diagnostics), based on the LIGASE CHAIN REACTION, that are carried out on clinical samples for detecting certain types of pathogen (such as e.g. *Chlamydia trachomatis*).

leader peptide See *5′*-UTR.

leader sequence (leader) (in mRNA) See *5′*-UTR.

(See also TRANS SPLICING.)

LEE (pathogenicity island) See PATHOGENICITY ISLAND (locus of enterocyte effacement).

***lef* genes** (in baculoviruses) Late expression factor genes: see BACULOVIRIDAE.

left end (of phage Mu DNA) See the entry PHAGE MU.

Lentivirinae A subfamily of viruses (of the family Retroviridae – see RETROVIRUSES) which replicate in, and kill, host cells; members of this group include the human immunodeficiency viruses (see HIV-1 and AIDS), feline immunodeficiency virus, visna virus (cause of a progressive, fatal meningoencephalitis in sheep) and the zwoegerziekte virus.

At least some lentiviruses – including e.g. HIV-1 and visna virus – can replicate in cell cultures.

Atypically, for retroviruses, lentiviruses infect non-dividing as well as dividing cells.

Engineered lentiviruses are used in GENE THERAPY.

(See also REDUCED-GENOME ESCHERICHIA COLI.)

lentiviruses Viruses of the subfamily LENTIVIRINAE (q.v.).

lepirudin See Refludan® in BIOPHARMACEUTICAL (table).

leptin A signaling molecule encoded by the OB GENE.

Abnormal leptin function may be seen e.g. in the BARDET–BIEDL SYNDROME.

Lesch–Nyhan syndrome (*syn*. HPRT deficiency) An X-linked syndrome, involving neurologic and metabolic abnormalities, associated with mutation in the gene encoding hypoxanthine guanine phosphoribosyltransferase (at chromosomal location Xq26–q27.2).

***letA* gene** See entry CCD MECHANISM.

***letB* gene** See entry CCD MECHANISM.

leucine-responsive regulator protein (Lrp) In *Escherichia coli* and in other bacteria (and archaeans): a homodimeric protein that contributes to the regulation of a wide range of genes and operons, either as an activator or repressor of transcription; in some case, the activity of Lrp is inhibited by leucine, while in other cases it is stimulated.

Although usually regarded as a natural dimer, Lrp has been reported to occur as higher-order oligomers; for example, an octameric assembly was reported for the *E. coli* Lrp protein when co-crystallized with a short duplex oligodeoxynucleotide [J Mol Biol (2007) 366(5):1589–1602]. [Crystal structure of the leucine-responsive regulator protein of the bacterium *Mycobacterium tuberculosis*: Protein Sci (2008) 17(1): 159–170.]

leucine zipper In certain DNA-binding (and other) proteins: a region in which leucine residues are found at every seventh residue in the chain.

leukemia See the entry CANCER.

leukemia inhibitory factor (LIF) A protein encoded by gene *LIF* on (human) chromosome 22 and formed by various types of cell; it is a member of the interleukin-6 (IL-6) family of cytokines which bind to the cell-membrane signal transducer gp130.

In cultures of mouse embryonic STEM CELLS the leukemia inhibitory factor promotes pluripotent growth – i.e. it inhibits differentiation of the ES cells.

leukocyte adhesion deficiency See GENETIC DISEASE (table).

levofloxacin See QUINOLONE ANTIBIOTICS.

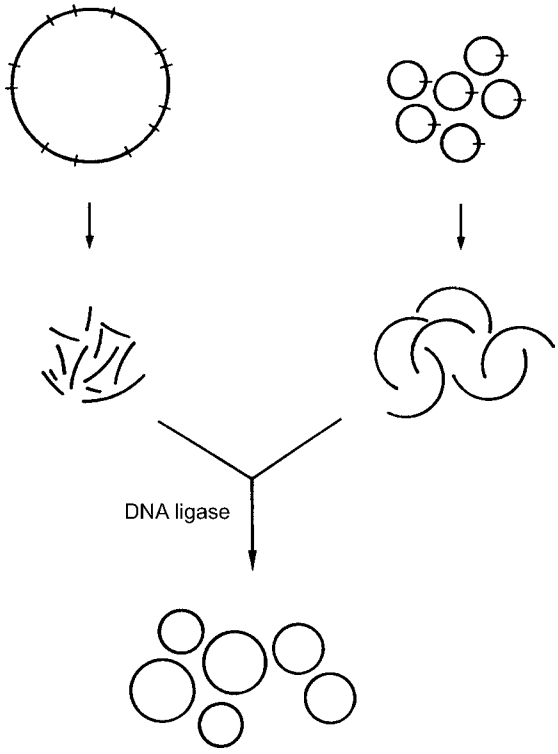
LexA protease See SOS SYSTEM.

LFA-1 *Syn*. CD11a/CD18 (q.v.).

library (1) (genomic library) A set of DNA fragments (usually in cloned form within vector molecules inside a population of cells) which – collectively – may represent the entire genome of a given organism. Such a library is prepared by isolating genomic DNA, cutting it into fragments with a RESTRICTION ENDONUCLEASE, inserting the pieces into VECTOR molecules (e.g. plasmids), and then transforming a population of cells with the recombinant plasmids. Replication of these plasmids in the cells generally gives rise to a clone of a given fragment in each transformed cell and its progeny; hence, *if each cell forms a separate colony* (on a suitable medium), each colony will consist of cells that contain copies of the same fragment. A library prepared in this way from a METAGENOME includes fragments from a range of uncharacterized genomes.

During the preparation of a library, conditions are chosen such that the properties of the cloned fragments produced are appropriate to the purpose(s) for which the library has been constructed; this involves attention to factors such as choice of restriction enzyme, digestion time (with the enzyme), and the nature of the vector. For example, a limited (rather than extended) period of digestion with a given restriction enzyme (a so-called *partial digest*) may be used in order to avoid the production of an excessive number of very small fragments; in this type of protocol, cleavage occurs at only a small proportion of the potential restriction sites in the genome. After digestion, the choice of vector will depend, for example, on the average size of the fragments obtained and/or chosen (by size fractionation) for cloning: see the entry CLONING for the sizes of insert accomodated by various types of vector.

In preparing a genomic library, one general requirement is



LIBRARY: making a bacterial genomic library (diagrammatic). Chromosomes are first isolated from a population of bacteria of a given strain and exposed to a particular restriction endonuclease. The enzyme cuts at specific (recognition) sites in each chromosome (*top*, *left*), forming many fragments of different sizes (*center*, *left*). A population of a given plasmid (*top*, *right*) provides the vector molecules. This plasmid is one which can be cut by the same enzyme as that used to cut the chromosomes; however, the plasmid has only *one* cutting site for the enzyme – so that cutting simply linearizes the plasmid molecules (*center*, *right*).

The chromosome fragments and linearized plasmids are then mixed and subjected to the action of DNA ligase. Some plasmids (and fragments) may simply re-circularize via their sticky ends, but concentrations can be chosen such that, in many cases, plasmids and fragments will join together – randomly –via their sticky ends to form a large number of recombinant molecules; many of these recombinant molecules will consist of a plasmid circularized with one chromosomal fragment. Because the fragments are of different sizes, the recombinant plasmids will also be of different sizes (*bottom of diagram*). Collectively, the recombinant plasmid molecules should carry all the DNA in the chromosome, and this collection of molecules is therefore called a *genomic library*. By using e.g. transformation, or electroporation, the library of molecules can be inserted into a population of bacteria – *with different cells receiving different fragments of the chromosome*. These bacteria can be used to inoculate an agar plate in such a way that each bacterium gives rise to a single colony; each colony will then contain a cloned copy of a particular fragment of chromsomal DNA.

Figure reproduced from *Bacteria in Biology*, *Biotechnology and Medicine*, 6th edition, Figure 8.7, page 218, Paul Singleton (2004) John Wiley & Sons Ltd, UK [ISBN 0-470-09027-8] with permission from the publisher.

that the products of enzyme digestion consist of *overlapping* fragments; this is needed e.g. when the purpose of the library is to determine the nucleotide sequence of the given genome by a procedure such as CHROMOSOME WALKING.

A library may be *screened* for a particular gene or sequence by various methods: see e.g. COLONY HYBRIDIZATION and RECACTIVE system.

(2) (cDNA library) A collection of cDNA molecules prepared from mRNAs obtained from a given type of cell under given conditions. Unlike a genomic library (see above), this type of library reflects only the *actively expressed* genes of the given organism under the particular set of conditions; a different set of cDNA molecules are likely to be formed from the same type of cell if the library is prepared under different conditions or in a different phase of growth or differentiation. Moreover, different patterns of cDNAs will be formed from different types of cell in a given mammal because cells with different functions naturally express different sets of genes.

cDNAs from the mRNAs of mammalian/eukaryotic SPLIT GENES contain INTRON-less coding sequences; this is useful if any of the genes are to be expressed in a bacterial host – in which introns would prevent the normal expression of a gene.

Compared with cDNAs from eukaryotes, the preparation of bacterial cDNAs is more problematic because e.g. (i) poly(A) tails are scarce on bacterial mRNAs; (ii) bacterial mRNAs are rather labile, i.e. they typically have short half-lives; (iii) many bacterial genes occur in an OPERON – transcription of which can give rise to a long POLYCISTRONIC MRNA that may be difficult to isolate and convert into a full-length cDNA.

For further information on cDNA and its synthesis see the entries CDNA, FIRST STRAND, CAPFINDER, CAPSELECT. (See also EXPRESSED SEQUENCE TAG.)

(3) A population of shRNAs (see SHORT HAIRPIN RNA), or other molecules (e.g. siRNAs), consisting of selected pieces of engineered nucleic acid. (A procedure for selecting potent siRNAs from an siRNA library is described in the entry RNA INTERFERENCE.)

(4) (phage library) A population of phages, of a given type, in which different individuals have different coat proteins – the population as a whole exhibiting an extensive range of types of coat protein; this kind of population is prepared by *in vitro* mutagenization and is used in PHAGE DISPLAY technology.

(5) Any population of diverse synthetic molecules prepared for a given purpose or function (often by a COMBINATORIAL method).

(6) (transposon library) In general, a population of cells that has been mutagenized with a particular type of TRANSPOSON, the transposons having inserted randomly into various genes in different cells. (See also TRANSPOSON MUTAGENESIS.)

licensing (of origins of replication) See RE-REPLICATION.

LIF LEUKEMIA INHIBITORY FACTOR.

***lif* gene** (in *Staphylococcus simulans*) A gene that encodes an enzyme which modifies the organism's peptidoglycan; such modification makes the organism resistant to the effects of its own BACTERIOCIN.

***LIF* gene** (in humans) See LEUKEMIA INHIBITORY FACTOR.

ligand-activated recombinase Any recombinase which, within cells, can be activated, as and when required, by treating the cells or organism with a suitable ligand. Thus, for example, temporally-controlled SITE-SPECIFIC RECOMBINATION in the zebrafish (*Danio rerio*) was carried out by using the ligand tamoxifen [PLoS ONE (2009) 4(2):e4640].

ligase (early name: synthetase) Any enzyme of EC class 6 that forms a C–O, C–S, C–N or C–C bond; this activity is energy-dependent.

An example is DNA LIGASE. The ATP-dependent phage T4 DNA ligase catalyzes the bond in a series of steps: a ligase–AMP intermediate complex transfers AMP to the terminal 5′-phosphate on the strand to be ligated, forming a –P–P– bond; the subsequent formation of a phosphodiester bond involves cleavage of the –P–P– bond with release of AMP. (See also BLUNT-END LIGATION.)

A NAD^+-dependent ligase is encoded by *Escherichia coli*.

Another example of a ligase is protein–biotin ligase: see e.g. IN VIVO BIOTINYLATION.

ligase chain reaction (LCR) A (probe-based) method for DNA AMPLIFICATION involving thermal cycling and the activity of a heat-stable DNA ligase. (Compare PCR.)

Essentially, in the basic scheme, sample dsDNA is initially heat-denatured to expose the target sequence (the *amplicon*) on one strand and a sequence complementary to the amplicon on the other strand. On lowering the temperature, two target-specific probes anneal, end-to-end, on the target sequence, the 3′ end of one probe being juxtaposed to the 5′ end of the other – i.e. the two probes jointly cover the precise sequence of the amplicon. With the probes bound to the amplicon, a heat-stable DNA LIGASE can join the two probes covalently.

Another two probes bind to the sequence complementary to the amplicon – on the other strand of the sample DNA – and are similarly ligated.

In this scheme, the 5′ juxtaposed nucleotide in one of the probes must be phosphorylated in order for ligation to occur.

Note that a pair of *ligated* probes forms a product to which another pair of probes can bind following a further phase of denaturation; thus, the quantity of product (i.e. ligated probe-pairs) rises exponentially with ongoing temperature cycling.

When probe–target hybridization occurs under sufficiently stringent conditions, the ability to amplify the target sequence argues for the presence of that specific sequence in the DNA sample; this is the basic principle underlying the use of the ligase chain reaction in diagnostic tests.

In the commercial form of the ligase chain reaction (LCx® assays; Abbott Diagnostics) the two target-hybridized probes are separated by a gap of one to several nucleotides (instead of being hybridized in a contiguous, end-to-end fashion); the gap between one pair of bound probes is staggered in relation to the gap between the other pair of bound probes. After the probes have bound to the target sequence, a DNA polymerase

fills in the gap and ligation then follows.

Detection of products

In LCx® assays the products of LCR (i.e. the pairs of ligated probes) are detected by a *microparticle enzyme immunoassay* (MEIA). For the purposes of this assay, the *outer* terminal of each probe (that end of the probe which is *not* ligated during cycling) carries a so-called *hapten*: a molecule which binds to a specific antibody. The hapten on one probe of a pair differs from that on its partner probe – i.e. the two haptens bind to different types of antibody.

After cycling, any probe pair that has been ligated consists of a single strand bearing hapten 1 at one end and hapten 2 at the other end. Unligated probes carry only a single terminal hapten (hapten 1 *or* 2, depending on probe).

The ligated probe-pairs bind, via hapten 1, to microparticles (<1 μm in diameter) that are coated with antibodies specific to hapten 1. The microparticles are subsequently immobilized on a solid surface; these microparticles will have bound (i) ligated probe-pairs and (ii) any individual (i.e. non-ligated) probes which carry a terminal hapten 1.

To the immobilized microparticles is then added a conjugate that is derived from the enzyme ALKALINE PHOSPHATASE (AP) linked to an antibody specific for hapten 2. This conjugate thus binds to hapten 2 on probe-pairs that are bound to the microparticles. On addition of the substrate for AP (MUP: 4-methylumbelliferyl phosphate), the MUP is cleaved by AP to a compound, 4-methylumbelliferone, which fluoresces under appropriate excitation. Monitoring the fluorescence provides quantitation of the LCR product. Note that *unligated* probes bearing hapten 2 are eliminated by washing during the assay; unligated probes bearing hapten 1 bind to the microparticles but, because they lack hapten 2, they do not bind conjugate and, hence, do not contribute to the measured fluorescence.

Anticontamination measures

As in other methods for *in vitro* DNA amplification (such as PCR), LCR can give false-positive results if extraneous DNA (e.g. from previous assays) contaminates the specimen and/or reagents. In general, anti-contamination measures include (i) use of anti-aerosol pipet tips, (ii) use of dedicated laboratory areas, and (iii) inactivation of products in the reaction vial (after the final reading) by automatic addition of a chelated metal complex and an oxidizing agent which, jointly, bring about almost total degradation of the DNA.

ligation (in DNA) Formation of a phosphodiester bond, usually mediated by a DNA LIGASE, between suitably juxtaposed 3′ and 5′ terminals.

Ligation can also be mediated by a topoisomerase – see the entry TOPOISOMERASE I CLONING.

ligation-mediated PCR (LMPCR) A procedure used e.g. for analyzing products from *in vivo* footprinting, i.e. helping to define the sites at which proteins bind to DNA in living cells.

In *in vivo*. footprinting, cells are treated briefly with certain agents such as dimethylsulfate (DMS), the enzyme DNase I, or ultraviolet light, each of which can modify or cut genomic DNA at sites that are *not* protected by a bound protein; hence, any sites that are *not* modified, or cut, by these agents may be ligand-binding sites. After exposure of cells to such an agent, the DNA is isolated and examined *in vitro*.

In e.g. DNA from DMS-treated cells, piperidine is used for strand cleavage at all methylated guanines. Denaturation then yields fragments of ssDNA of various lengths. Gene-specific primers are bound to these fragments and extended to form blunt-ended products – to which linkers are ligated (at both ends). The products are PCR-amplified with gene- and linker-specific primers, and labeled in a further round of PCR. The products are examined by gel electrophoresis, with a control prepared *in vitro* from DMS-treated genomic DNA.

One of the problems in LMPCR aries from an insufficient degree of specificity at primer extension in PCR, this making it difficult, for example, to distinguish single-nucleotide polymorphisms in the sample. This problem was approached by employing PYROPHOSPHOROLYSIS-ACTIVATED POLYMERIZATION (PAP-LMPCR) for primer extension in PCR [Nucleic Acids Res (2008) 36(3):e19].

light-activated RNAi RNA INTERFERENCE (q.v.) in which the activity can be controlled (toggled) by light. In one approach, photo-cleavable, diazo-based groups were added to *terminal* positions in 27-mer siRNA precursors; this was found to give a better performance than previous non-terminal substitutions [Nucleic Acids Res (2009) doi: 10.1093/nar/gkp415].

light strand (of dsDNA) See H STRAND.

Light Upon eXtension primer See LUX PRIMER.

LightCycler™ An apparatus (Roche) which is used for temperature cycling and monitoring in REAL-TIME PCR. Samples, each in a sealed capillary tube, are rapidly heated and cooled by a forced air current, permitting protocols to be completed in a short period of time.

Progress in a PCR assay is monitored by means of a probe-based system, as follows. When a BINARY PROBE binds to the target sequence on an amplicon (at the *annealing* stage of the cycle) it fluoresces by a FRET-based mechanism. Later, in the primer extension stage, DNA synthesis from the primer displaces – but does not degrade – both parts of the binary probe (cf. TAQMAN PROBE); fluorescence then stops, and the probes are free to take part (at the annealing stage) in the next cycle.

The probe system can use e.g. fluorescein as a FRET donor and LightCycler Red 640 as an acceptor; the latter compound fluoresces at 640 nm.

LightCycler™ Red 640 A fluorophore employed in the LIGHTCYCLER system: see REAL-TIME PCR.

LightScanner® See HI-RES MELTING.

LightUp® probe A commercial probe (LightUp Technologies, Huddinge, Sweden) used e.g. for monitoring amplification in REAL-TIME PCR. The probe consists of a PNA oligomer and an associated cyanine dye, thiazole orange, which is tethered to the oligomer; when the PNA binds to target DNA the ability of the dye to fluoresce is greatly increased.

[Examples of use: J Virol (2005) 79(3):1966–1969; J Clin Microbiol (2005) 43(8):4057–4063.]

limited enrichment method See AUXOTROPHIC MUTANT.

***Limulus* amebocyte lysate test** See LAL TEST.

LINE Long interspersed element: any of a number of repetitive sequences, typically >1 kb in length, that are found dispersed in at least some types of mammalian genome (including the human genome); they are capable of retrotransposition. The human genome includes ~50000 copies of the KpnI family of LINE elements

(cf. SINE; see also CPG ISLAND METHYLATOR PHENOTYPE.)

line probe assay A PROBE-based assay used e.g. for detecting bacteria of the *Mycobacterium tuberculosis* complex and for (simultaneously) detecting mutations which confer resistance to specific antibiotic(s).

One commercial line probe assay, the INNO-LiPA Rif TB (Innogenetics, Gent, Belgium), detects certain mutations in a specific region of the pathogen's *rpoB* gene (the gene which encodes the β-subunit of RNA polymerase); mutations which are detected in this assay are those generally responsible for resistance to rifampin (rifampicin) – a major anti-tuberculosis drug (see RIFAMYCINS). This assay uses a set of probes that are immobilized, in a line, on a membrane.

To detect the mutations, the 'resistance region' of the *rpoB* gene (in which these mutations commonly occur) is initially amplified by PCR in a reaction in which one of the primers is BIOTINylated. The resulting amplicons are denatured, and the single-stranded (biotinylated) products are exposed to the line of immobilized probes under stringent conditions.

Five probes – the S probes – consist of wild-type sequences which, collectively, cover the entire resistance region.

Another four probes – the R probes – cover sections of the resistance region associated with common mutations. Each of the four probes carries a specific mutation.

Any amplicons that bind to a given probe are subsequently detected by a STREPTAVIDIN–alkaline phosphatase conjugate which gives a purple coloration with an added reagent.

Coloration of all the S probes, with no staining of R probes, indicates a wild-type strain (sensitive to rifampin).

Amplicons that contain one of the four specific mutations covered by the R probes will bind to the appropriate R probe; coloration of this probe will then indicate a particular mutant genotype.

There are opportunities in the assay for both false-positive and false-negative results.

False-positive results may derive from contamination with amplicons from a previous assay and/or from using a temperature below 62°C for the hybridization step. Also, a *silent mutation* may (falsely) indicate resistance by preventing the hybridization of amplicons to relevant S probe(s).

False-negative results may arise if the temperature used for hybridization is too high (>62°C) or the concentration of the amplicons is too low. A false-negative result may also occur if a specimen contains a wild-type strain *and* a strain with a mutation that is not covered by the R probes. In this case the wild-type strain would cause coloration of all the S probes, indicating rifampin sensitivity; however, amplicons from the mutant strain would not bind to any of the R probes – so that the mutant strain would give no indication of its presence.

Another commercial line probe assay, GenoType MTBDR (Hain Lifescience, Nehren, Germany), detects resistance to both rifampin and ISONIAZID. Resistance to isoniazid is often due to mutation in the *katG* gene (often in codon 315), and this is reflected in the probes used in this assay. [GenoType MTBDR assays (a meta-analysis): Eur Resp J (2008) 32(5): 1165–1174.]

Evaluation of the line probe assays referred to above found that both are useful for rapid screening, the GenoType assay having an advantage in detecting resistance to two antibiotics simultaneoulsy [J Clin Microbiol (2006) 44(2):350–352].

[GenoType assay in studies on relevance of nontuberculous mycobacteria in Oman: Emerg Infect Dis (2009) 15(2):292–294.]

[World Health Organization. Molecular line probe assays for rapid screening of patients at risk of multidrug-resistant tuberculosis (MDR-TB). 2008. Policy statement. Available at http://www.who.int/tb/dots/laboratory/lpa_policy.pdf]

linear cloning vector Any type of CLONING VECTOR which is either produced in a linear form or produced by linearization of a circular vector.

(See, for example, BIGEASY LINEAR CLONING SYSTEM and USER CLONING.)

linearization Cleavage of a covalently closed circular molecule of nucleic acid to form a linear molecule.

(cf. CIRCULARIZATION.)

linezolid See OXAZOLIDINONE ANTIBIOTICS.

linkage (of eukaryotic genes) The degree of association of two genes on a chromosome; the shorter the distance between two genes (i.e. the closer the linkage) the more likely they are to be transmitted together to a daughter cell during cell division. Linkage can be estimated from the recombination frequency.

linker (*DNA technol.*) A short, synthetic, double-stranded fragment of DNA which is designed to be ligated (i.e. covalently joined) to the end(s) of a given dsDNA molecule in order e.g. to facilitate manipulation of that molecule. For example, the sequence of nucleotides in a linker may include the recognition site of a particular RESTRICTION ENDONUCLEASE; hence, when the (bound) linker is cleaved by the given restriction enzyme it can provide a desired STICKY END e.g. to facilitate insertion of the target molecule into a vector. (Before using a particular linker for this purpose it is necessary to ensure that the linker's cutting site does not occur also within the target DNA itself; if cutting site(s) occur within the target molecule they may be methylated, prior to ligation, in order to protect the site(s) from a (methylation-sensitive) restriction enzyme.)

A (blunt-ended) linker may be joined to the target molecule by BLUNT-END LIGATION.

In a different scenario, the linker may include the *binding* site of a restriction endonuclease whose *cutting* site is outside the linker: see e.g. the use of restriction enzyme MmeI in the entry SHORT HAIRPIN RNA.

A linker may also be ligated to a target molecule in order to provide a primer-binding site for amplification of the target

sequence by PCR (see e.g. ANCHORED PCR).

Note. The term 'linker' is sometimes used for a *spacer arm*: a sequence used to link a reactive reagent to a particular position on a molecule; see e.g. BIOTIN for the meaning of spacer arm.

(cf. ADAPTOR; see also NOTI.)

linker DNA (*in vivo*) See CHROMATIN.

linking number (α, *Lk*, topological winding number) In a given molecule of double-stranded cccDNA: the number of times one strand winds around the other.

Lipofectamine™ Any of several reagents (Invitrogen, Carlsbad CA, USA) used for LIPOFECTION.

Lipofectamine™ 2000 is a cationic lipid reagent designed for high-efficiency transfection; this reagent is reported to be suitable for nucleic acid delivery to a wide range of types of eukaryotic cell.

[Use of Lipofectamine™ (examples): Genome Res (2008) 18(4):597–609; PLoS ONE (2009) 4(5):e5710; PLoS Pathog (2009) 5(5):e1000450; J Biol Chem (2009) 284(22):14838–14848.]

Lipofectamine™ 2000 CD transfection reagent is a chemically defined reagent which is 100% free of components of animal origin; it may be used e.g. in any protocol designed to ensure that the downstream product is free of animal-derived entities – such as prions or viruses.

lipofection Lipid-assisted insertion of nucleic acids into cells: an alternative to the use of a viral vector. In some procedures, a mixture of cationic and neutral lipids in correct proportions forms liposome complexes with nucleic acids and facilitates internalization.

(cf. ELECTROPORATION, NUCLEOFECTION and also MAGNETOFECTION.)

Specific reagents are available, commercially, for particular types of application and/or particular types of cell. Thus, for example, Lipofectamine™ 2000 CD reagent is intended for use when there are concerns about the presence of animal-derived components (such as viruses) in the downstream product. Again, the 293fectin™ reagent has been optimized for transfection of FreeStyle™ 293-F cells.

Oligofectamine™ (Invitrogen, Carlsbad CA, USA) is used e.g. for transfecting cells with antisense oligos [for example, see Biochem Pharmacol (2008) 76(5):582–588].

A brief heat shock was reported to increase stable integration of lipofected DNA [BioTechniques (2005) 38(1):48–52].

Lipofection has been used e.g. in studies on the activation of T cells [J Clin Invest (2006) 116(10):2817–2826]; in GENE THERAPY for cystic fibrosis [Biochem J (2005) 387(1):1–15]; for studying (mammalian) mRNA synthesis [PLoS Biology (2006) 4(10):e309]; for studying the design of siRNAs [PLoS Genetics (2006) 2(9):e140]; for studies on retinal gliogenesis in *Xenopus* [Neural Dev (2009) 4:1]; and studies on the transfection of tumor cells [Genet Vaccines Ther (2009) 7:5].

(See also LIPOFECTAMINE.)

lipoplex A nucleic acid (e.g. a plasmid) complexed with a cationic lipid.

lipopolysaccharide (LPS) A component of the outer membrane – the outermost layer of the cell envelope – in Gram-negative bacteria (such as *Escherichia coli*).

LPS is a macromolecule consisting of three linked regions: (i) lipid A, (ii) core oligosaccharide, and (iii) the O-specific chain; the O-specific chains in the outer membrane form the outermost part of the cell envelope, i.e. that part which is in direct contact with the environment. The (glycolipid) lipid A is associated with endotoxic properties.

LPS is a PYROGEN that must be removed from recombinant protein products synthesized in Gram-negative bacteria. This can be achieved by the chromatographic separation processes involved in downstream processing of the product.

(See also LAL TEST and LPXC GENE.)

Listeria A genus of Gram-positive, facultatively anaerobic bacteria; the organisms range from coccobacilli and rods (~0.5 μm in width to ~0.5–2 μm in length) to filaments. Most of the strains are catalase-positive. Oxidase-negative. Species occur e.g. in silage, and decaying vegetation, and also in some dairy products.

L. monocytogenes is a human pathogen which can cross the blood–brain barrier; it causes (sometimes fatal) disease which ranges from gastroenteritis to meningitis and abortion.

The GC% of genomic DNA in *L. monocytogenes* is 38.

LIVE *Bac*Light™ bacterial Gram stain kit A commercial kit (Molecular Probes Inc., Eugene OR, USA) which is used for determining the Gram reaction of *living* bacteria by the use of a pair of DNA-staining fluorescent dyes: SYTO®9 (see SYTO DYES) and HEXIDIUM IODIDE. When treated with a mixture of these two reagents, Gram-positive bacteria are preferentially stained with the (orange-fluorescent) hexidium iodide – even though both dyes can penetrate the cell membrane. Gram-negative bacteria are stained by (green-fluorescent) SYTO®9 because hexidium iodide is excluded by the cell envelope in Gram-negative species.

live cell fluorescence See the entries GENE FUSION and IN VIVO FLUORESCENCE.

LIVE/DEAD® *Bac*Light™ bacterial viability kit A product (Molecular Probes Inc., Eugene OR, USA) designed as a one-step staining procedure for distinguishing between live and dead bacteria. This technique makes use of two fluorescent dyes, SYTO®9 and propidium iodide, which stain genomic DNA. SYTO®9 (green fluorescence) can enter living as well as dead cells, while propidium iodide (red fluorescence) can enter only cells with damaged/permeable cell membranes (a characteristic of dead cells). In this procedure, the living cells fluoresce green. In the dead cells, propidium iodide competes with SYTO®9 for binding sites on the DNA, and these cells fluoresce red.

Use of method in flow cytometry [Appl Environ Microbiol (2007) 73(10):3283–3290], and studies on the Lyme disease pathogen (*Borrelia burgdorferi*) [PLoS Pathog (2009) 5(3): e1000326].

Lk LINKING NUMBER.

LMPCR LIGATION-MEDIATED PCR.

LNA LOCKED NUCLEIC ACIDS.

locked nucleic acids (LNAs) Nucleic acids in which the bases bind more strongly, and with greater specificity, to their complementary bases – owing to the effects of *in vitro* chemical modification [see: Tetrahedron (1998) 54:3607–3630].

(cf. ZIP NUCLEIC ACID.)

Locked nucleic acids are used e.g. in PROBES. Moreover, if used in *unlabeled* probes, they are able to improve mismatch discrimination in melting analyses (T_m).

[Improvement of mismatch discrimination in LNA probes: Nucleic Acids Res (2006) 34(8):e60.]

LNAs enhance activity of triplex-forming oligonucleotides (TFOs) [Pu motif: J Biol Chem (2005) 280:20076–20085; Py motif: Nucleic Acids Res (2005) 33(13):4223–4234].

In some cases, the inclusion of LNAs in 2′-*O*-methyl oligonucleotides has been found to enhance stability of RNA/RNA duplexes (which may permit e.g. improved design for probes in Northern blot analysis) [Nucleic Acids Res (2005) 33(16): 5082–5093].

[Position-dependent effects of LNAs on DNA sequencing and PCR primers: Nucleic Acids Res (2006) 34(20):e142.]

In real-time PCR, LNA technology was reported to be useful in the design of probes for quantitative assays of genetically modified organisms (GMOs) – particularly when high specificity is required and the design of a TaqMan probe is problematic [BMC Biotechnol (2008) 8:26].

locus of enterocyte effacement See LEE pathogenicity island in the entry PATHOGENICITY ISLAND.

***lon* gene** (in *Escherichia coli*) A gene encoding the Lon heat-shock protein: an ATP-dependent protease.

At the end of the SOS response to damaged DNA – when normal growth is resumed – Lon inactivates the SulA protein, preventing it from blocking septation in the growing cells.

The Lon protease is one of several proteases involved in *trans* translation (see TRANS TRANSLATION).

Strains of *E. coli* with an inactivated *lon* gene are used e.g. in the production of fusion proteins (for a rationale, see the table in the entry ESCHERICHIA COLI; see also *The problem of proteolysis* in the entry OVEREXPRESSION).

long extension-PCR *Syn*. LONG PCR.

long non-coding RNAs A category of transcripts which appear to have various roles. (Compare small non-coding RNAs: see the entry SRNAS.) For some of the long non-coding RNAs it seems that transcription, *per se*, may play a part in regulating adjacent gene(s). Others may function e.g. as a 'scaffolding' on which subcellular structures are assembled or organized. [Long non-coding RNAs (review article): Genes Dev (2009) 23:1494–1504.]

long PCR (long-range PCR) A form of PCR used for copying long target sequences – e.g. up to ~40 kb.

PCR carried out with a DNA polymerase (such as the *Taq* DNA polymerase) which lacks PROOF READING capability is generally unsuitable for target sequences of more than a few thousand nucleotides; the reason for this appears to be that the incorporation of an incorrect nucleotide (an uncorrected error) in the nascent strand can lead to premature termination of DNA synthesis.

Early studies [e.g. Proc Natl Acad Sci USA (1994) 91(12): 5695–5699] found that the concomitant use of a second DNA polymerase, with proof reading ability, results in correction of errors and permits ongoing synthesis of the nascent strand; this approach therefore exploits the strengths of both types of polymerase: the synthetic capacity of a highly PROCESSIVE ENZYME such as the *Taq* DNA polymerase and the accuracy/fidelity of the proof-reading enzyme. PCR carried out in this way is referred to as *long PCR*.

Interestingly, one study [Nucleic Acids Res (2000) 28(11): e50] reported that the amplification of mildly or moderately damaged DNA is improved by the inclusion of *Escherichia coli* enzyme EXONUCLEASE III in the reaction mixture.

The efficiency and/or specificity of long PCR is reported to be enhanced by the inclusion of CARBON NANOTUBES (q.v.) in the reaction mixture.

(See also entries ELONGASE AMPLIFICATION SYSTEM and RTTH DNA POLYMERASE XL.)

long-range PCR *Syn*. LONG PCR.

long terminal repeat (LTR) See RETROVIRUSES.

LongSAGE See SAGE.

loop (*microbiol*.) A simple instrument, used e.g. in bacteriology and mycology, for manipulating small quantities of liquid or solid material (e.g. during the inoculation of a medium); it consists of a metal rod (the handle) to which is attached, at one end, a piece of platinum or nickel-steel wire (of about 5–8 cm) formed into a closed loop (~2–3 mm diam.) at the free end.

A loop (the wire portion) is flamed before and after use in order to sterilize it.

loop-mediated isothermal amplification See LAMP.

looping (in DNA) The phenomenon in which separate sites in a DNA molecule are physically linked, through DNA-binding proteins, such that the intervening sequence of DNA forms a loop between the sites.

Looping occurs e.g. when operator sequences are linked by regulatory proteins during the transcriptional control of some genes/operons (e.g. the LAC OPERON). (See also ENHANCER.) Gene repression in the *lac* operon was analyzed in the context of DNA looping [PLoS ONE (2006) 1(1):e136].

[DNA looping by restriction endonucleases: Nucleic Acids Res (2006) 34(10):2864–2877. Real-time observation of the DNA looping dynamics of the type IIE RESTRICTION ENDONUCLEASES NaeI and NarI: Nucleic Acids Res (2006) 34: 167–174.]

Although DNA looping has been proposed in the activity of the type III restriction endonucleases, other studies concluded that looping is both unlikely and unnecessary in the type III restriction endonucleases [Proc Natl Acad Sci USA (2009) 106(6):1748–1753].

[Dissecting protein-induced DNA looping dynamics in real time: Nucleic Acids Res (2009) doi: 10.1093/nar/gkp570.]

low-copy-number DNA test See LCN DNA TEST.

low-stringency conditions See e.g. PROBE and STRINGENCY.
(See also PCR.)

loxP See CRE–LOXP SYSTEM.

***lpp* gene** (in *Escherichia coli*) A gene encoding the Braun lipoprotein, molecules of which link the outer membrane to the underlying layer of peptidoglycan in the cell envelope.

LPS LIPOPOLYSACCHARIDE.

LPSiNPs Biodegradable NANOPARTICLES (q.v.).

***lpxC* gene** (in *Escherichia coli*) A gene encoding deacetylase, a zinc-containing enzyme essential for the biosynthesis of lipid A in the outer membrane LIPOPOLYSACCHARIDE.

LPXTG The cell-wall sorting signal in certain secreted proteins in *Staphylococcus aureus*: see the entry SORTASE.

LR Clonase™ An enzyme that is used in the GATEWAY SITE-SPECIFIC RECOMBINATION SYSTEM; it can mediate *in vitro* exchange between a segment flanked by *att*L sites and one flanked by *att*R sites; both of the segments may be in circular molecules.

[Uses of LR Clonase (e.g.): PLoS ONE (2009) 4(3):e4780; Plant Physiol (2009) 149(2):1196–1204.]

(cf. BP CLONASE.)

LR-PCR Long-range PCR – see LONG PCR.

Lrp (in *Escherichia coli*) LEUCINE-RESPONSIVE REGULATOR PROTEIN.

LRRASLG The KEMPTIDE SEQUENCE (q.v.).

LTR (long terminal repeat) See RETROVIRUSES.

luciferase An oxidoreductase involved in *in vivo* generation of light (BIOLUMINESCENCE) e.g. in certain species of bacteria, fungi and insects. The (ATP-dependent) luciferase of the firefly (*Photinus pyralis*) (EC 1.13.12.7) generates light when it oxidizes *luciferin*: 4,5-dihydro-2-(6-hydroxy-2-benzothiazolyl)-4-thiazolecarboxylic acid.

In experimental systems, the intensity of the light generated from a luciferase-based system may be monitored by means of e.g. a charge-coupled device camera.

Experimentation may be facilitated by the use of CAGED LUCIFERIN.

Luciferase is used in techniques for detecting and/or quantitating ATP (e.g. PYROSEQUENCING), and in reporter systems for studying a variety of phenomena – e.g. promoter activity; the efficacy of antisense oligomers; the dynamics of cell-free protein synthesis; studies on the prediction of miRNA targets in human and murine transcriptomes [PLoS ONE (2009) 4 (5):e5745]; and studies on the use of miRNA for controlling the activity of wild-type adenoviruses [PLoS Pathog (2009) 5(5):e1000440].

(See also PATHDETECT and REPORTER GENE.)

luciferin See BIOLUMINESCENCE and LUCIFERASE.

luciferin (caged) See CAGED LUCIFERIN.

***lukF* gene** See PANTON–VALENTINE LEUCOCIDIN.

***lukS* gene** See PANTON–VALENTINE LEUCOCIDIN.

Lumio™ A six-amino-acid-residue tag, encoded by some types of expression vector (Invitrogen, Carlsbad CA, USA), which binds a fluorescent substrate and allows fluorescence-based detection (and localization) of proteins in living mammalian cells (see e.g. pcDNA™6.2/nLumio™-DEST™ in the entry GATEWAY SITE-SPECIFIC RECOMBINATION SYSTEM (table)).

The Lumio™ tag also permits real-time detection of protein synthesis in a CELL-FREE PROTEIN SYNTHESIS system.

[Uses (e.g.): Curr Biol (2008) 18(6):401–409; PLoS ONE (2009) 4(2):e4509; Cell Commun Signal (2009) 7:16.]

(See also IN VIVO FLUORESCENCE.)

Lux™ (LUX) primer Light Upon eXtension primer: a hairpin-structured primer (Invitrogen, Carlsbad CA, USA), used in REAL-TIME PCR, which has a fluorophore close to the 3′ end; if the primer is *not* bound to its target, then the fluorophore is quenched by interaction with the 5′ end of the primer. If the primer binds to its target, the 3′ and 5′ ends are separated so that the fluorophore is no longer quenched.

[Comparison of probes and primers (including LUX) used in real-time PCR: BMC Biotechnol (2008) 8:26.]

LX-PCR LONG PCR.

lymphoid lineage Cells which include e.g. B lymphocytes (B cells), plasma cells (derived from B cells) and T lymphocytes (T cells).

(See also MYELOID LINEAGE.)

lysogen Any cell, or strain of bacteria, in which a BACTERIOPHAGE genome has established a stable, non-lytic presence – as in LYSOGENY.

lysogenic bacteria See BACTERIOPHAGE and LYSOGENY.

lysogenic conversion (*syn.* phage conversion) See LYSOGENY.

lysogeny A stable, non-lytic relationship between a BACTERIOPHAGE genome and a bacterial host cell in which no progeny virions are formed; in most cases (e.g. PHAGE LAMBDA) the phage genome (the *prophage*) is maintained by integration into the bacterial chromosome, but in some cases (e.g. PHAGE P1) the prophage persists as an extrachromosomal 'plasmid'. A cell containing a prophage is termed a *lysogen*; a lysogen continues to grow and divide, and replication of the prophage is coordinated with that of the host cell.

Commonly, the prophage retains a lytic potential, i.e. under certain conditions it is able to initiate the lytic cycle with the production of progeny virions and lysis of the host cell; in many cases this may occur spontaneously in a small number of cells in a lysogenic population. *Induction* of the lytic cycle in most or all cells in a lysogenic population may be achieved experimentally by certain agents – particularly by agents that damage DNA, e.g. ULTRAVIOLET RADIATION or MITOMYCIN C (both of which can trigger the SOS response in bacteria). In at least some cases, induction involves cleavage/inactivation of a phage-encoded repressor protein whose activity mediates the lysogenic state.

(See also SUPERINFECTION IMMUNITY.)

***Bacteriophage conversion* (*phage conversion*)**

In certain cases the prophage alters the phenotype of the host cell (a phenomenon termed *bacteriophage conversion*, *phage conversion* or *lysogenic conversion*). This may involve (i) the inactivation of host cell genes during integration of the prophage and/or (ii) expression of phage genes within the host cell. For example, integration of phage L54a into the chromo-

some of *Staphylococcus aureus* causes a loss of lipase activity. Phage conversion in some Gram-negative pathogens, e.g. species of *Salmonella* and *Shigella*, has been shown to alter cell-surface antigens; such modification of serotype has to be borne in mind when developing vaccine strains of these organisms.

Phage conversion involving the expression of *phage* genes is responsible e.g. for toxin production in the causal agents of cholera (the Gram-negative *Vibrio cholerae*) and diphtheria (the Gram-positive *Corynebacterium diphtheriae*). In EHEC strains of *Escherichia coli* – including O157:H7 – the shiga-like toxins are also encoded by a phage.

(See also STREPTODORNASE.)

lysostaphin A BACTERIOCIN, produced by *Staphylococcus simulans*, which acts against strains of *Staphylococcus aureus* – cleaving peptide bridges in the (cell wall) polymer peptidoglycan).

Lysostaphin has been used e.g. to isolate PROTEIN A (q.v.).

lysozyme An enzyme (*N*-acetylmuramidase; EC 3.2.1.17) that hydrolyzes *N*-acetylmuramyl-(1→4)-β-links in the (bacterial) cell-wall polymer peptidoglycan. The enzyme is found e.g. in saliva, tears and egg-white and is also found in macrophages. Lysozyme is used e.g. for the preparation of SPHEROPLASTS from Gram-positive cells.

lytic bacteriophage Any BACTERIOPHAGE which lyses the host bacterial cell.

M

M (1) A specific indicator of ambiguity in the recognition site of a RESTRICTION ENDONUCLEASE or in another sequence of nucleic acid; for example, in GT⌊MKAC (restriction enzyme AccI) the 'M' indicates A or C. In the example, 'K' is T or G. (2) L-Methionine (alternative to Met).

M plasmid See MICROCIN.

M-tropic strains (of HIV-1) See the entry HIV-1.

M13 phage See PHAGE M13.

M13K07 A HELPER PHAGE (q.v.) used e.g. for the single-strand rescue of target DNA encoded by PHAGEMIDS containing an f1 origin or replication; one study reported that phage KM13 was more efficient as a helper phage in the conditions of the experiment.

[Use of M13K07 (e.g.): PLoS ONE (2009) 4(2):e4348; use of helper phage KM13 (reported to have greater efficiency as a helper phage): BMC Biotechnol (2009) 9:6.]

mAbs MONOCLONAL ANTIBODIES.

Mac-1 See (β_2) INTEGRIN.

MacConkey's agar A selective, AGAR-based MEDIUM containing peptone, lactose, sodium chloride, bile salts (e.g. sodium taurocholate) and the pH indicator neutral red. The bile salts inhibit non-enteric bacteria but permit the growth of enteric species such as *Escherichia coli* and *Salmonella typhi*.

Species which form acid from the fermentation of lactose give rise to red colonies (the pH indicator, neutral red, is red below pH 6.8); these species include *E. coli* and some other members of the family Enterobacteriaceae. (Following overnight incubation at 37°C, the colonies of *E. coli* are normally round, red, ∼2 mm in diameter.) The non-lactose-fermenting species, e.g. most strains of *Salmonella*, do not form acid and therefore give rise to colorless colonies.

macroarray A term used for a test system which is similar in principle to a MICROARRAY but which employs fewer – often (significantly) longer – probes. Macroarrays have been used e.g. with probes 200–650 nt [Clin Diag Lab Immunol (2004) 11(4):691–698], 200–600 nt [J Virol (2005) 79:5315–5325], and up to 3 kb [PLoS ONE (2008) 3(5):e2259].

'Macroarray' also refers to a PROTEIN MACROARRAY (q.v.); this is used for detecting protein–protein interactions between 'target' proteins, which are spotted onto a membrane, and the proteins present in the samples being tested.

macronucleus (*ciliate protozool.*) The larger of the two types of ciliate nucleus (cf. MICRONUCLEUS); a cell may have more than one macronucleus, depending on species. Macronuclei (which contain a nucleolus) are generally polyploid, but are typically diploid in some ciliates – e.g. *Loxodes*; they encode the cell's somatic functions and generally disintegrate during conjugation, a replacement arising from the micronucleus.

A macronucleus is *heteromerous* if it is divided into two dissimilar *karyomeres* (parts) – the *orthomere* and *paramere*. Most ciliate macronuclei are *homoiomerous*, i.e. they are not divided into karyomeres.

magnetic resonance imaging (for monitoring gene expression) See the entry MRI.

magnetic separation See e.g. DYNABEADS.

magnetic tweezers A technique, used in SINGLE-MOLECULE DNA MANIPULATION/VISUALIZATION, in which one end of a DNA molecule is fixed to the surface in a glass flow cell, and the other end is attached to a magnetic bead which can be controlled by an external source of magnestism; this permits extension and twisting of the DNA, allowing observations on e.g. interaction with protiens. (cf. OPTICAL TWEEZERS.)

Magnetic tweezers have been used e.g. for studying translocation, on DNA, of enzyme EcoR124I [J Mol Biol (2008) 384(5):1273–1286], and for studying the activity of the type III restriction endonucleases [Proc Natl Acad Sci USA (2009) 106(6):1748–1753]. [A high-resolution magnetic tweezer for measurement of single molecules: Nucleic Acids Res (2009) doi:10.1093/nar/gkp725.]

magnetofection A process in which DNA – coated on superparamagnetic NANOPARTICLES – is inserted into cells (both *in vitro* and *in vivo*) by a strong magnetic field.

[Enhancement of the efficiency of non-viral gene delivery by application of a pulsed magnetic field: Nucleic Acids Res (2006) 34(5):e40.]

Magnetofection has been used e.g. for the transfection of endothelial cells for Dicer RNA knockdown [BMC Cell Biol (2008) 9:61] and for transfecting human neuroblastoma cells with a lentiviral construct [Neurotox Res (2008) 14(4):367–382].

Lentiviral vectors bound to magnetic nanoparticles (MNPs) were introduced into cells by a magnetic field. Moreover, use of external magnets permitted control of the biodistribution of systemically administered MNP-coupled lentiviral vectors (in mice), indicating organ-specific targeting of these vectors by magnetic fields *in vivo* [Proc Natl Acad Sci USA (2009) 106(1):44–49].

(See also LIPOFECTION.)

^{M}AGT See *Uses of gene fusion* in the entry GENE FUSION.

MaizeGDB Maize genetics and genomics database, a repository for biological information on maize (*Zea mays*): [Int J Plant Genomics doi: 10.1155/2008/496957].

Online: http://www.maizegdb.org/

major groove (in DNA) See the entry GROOVES.

malaria (human) An acute or chronic disease caused by certain species of PLASMODIUM and normally transmitted by the bite of a mosquito of the genus *Anopheles* – e.g. *A. gambiae* or *A. arabiensis*. (Transmission may also occur e.g. via transfusion with infected blood.)

Various drugs are available for antimalarial chemotherapy (and prophylaxis). Efforts are being made to reduce the transmission of malaria by seeking suitable technology for genetic modification of *mosquitoes*, either to bring about a reduction in the population size of mosquito vectors or to replace existing populations of mosquito vectors with mosquitoes that are unable to act as vectors: see the entry GMMS.

MALDI Matrix-assisted laser desorption/ionization: a method which is used e.g. for accurate determination of the relative masses of nucleic acid and protein molecules.

The analyte is initially embedded in a *matrix* which often consists of an aromatic carboxylic acid containing group(s) that can absorb laser radiation of wavelength approximately 340 nm. Pulses of laser, directed onto the embedded analyte, cause desorption (i.e. volatilization and ionization) of matrix material and of at least some analyte molecules. (Without the matrix, laser energy would be more likely to cause fragmentation of the analyte molecules.)

The ions (in gas phase) are first accelerated with an electric field and then pass into a region of constant field at the end of which is the detector; the detector records the time taken by each species of ion to travel a specific distance, and a so-called time-of-flight (TOF) analyzer provides data based on the relationship between travel time and mass.

Analytes with low solubility may precipitate on the wall of the flight tube; such analytes may give better results when examined by ESI.

This approach was used e.g. for studying acetylation by the p300 histone acetyltransferase (HAT) [Biochem Biophys Res Commun (2005) 336(1):274–280] and has also found use e.g. in SNP genotyping [Nucleic Acids Res (2006) 34(3):e13].

MALDI–TOF MALDI–time-of-flight: see the entry MALDI.

malignant melanoma See MELANOMA.

manipulation/visualization of DNA (at molecular level) See SINGLE-MOLECULE DNA MANIPULATION/VISUALIZATION.

map unit (1) A CENTIMORGAN.

(2) Any unit which is able to indicate the locations of specific markers in a genome – e.g. the chromosome of *Escherichia coli* is divided into a number of *minutes* (used in interrupted mating experiments).

MAPH MULTIPLEX AMPLIFIABLE PROBE HYBRIDIZATION.

MAR A MATRIX-ATTACHMENT REGION (q.v.).

(See also GREEN FLUORESCENT PROTEIN.)

***MAR2xT7* transposon** See MARINER FAMILY.

maraviroc A CHEMOKINE CORECEPTOR ANTAGONIST (q.v.).

maribavir An antiviral agent (a benzimidazole L-riboside) with activity against the UL97 kinase encoded by cytomegalovirus (see NUCLEAR LAMINA); maribavir reduces nuclear egress of nascent virions.

***mariner* family** A family of TRANSPOSONS that are widespread in animals, from humans and flatworms to arthropods; related transposons have been found in plants and fungi. The first transposon in the category to be isolated (*Mos1*) was obtained from a species of *Drosophila* (fruitfly).

In appropriate constructs, certain *mariner* transposons, or their derivatives, can insert into the chromosomes of various species of bacteria.

These transposons are widely used for insertional studies.

Characteristically, the *mariner* transposon inserts into TA-dinucleotide target sites with a cut-and-paste mechanism and with little or no requirement for host-specific factors.

The *mariner* transposon *Himar1* was used for insertional mutagenesis in *Leptospira interrogans* [J Bacteriol (2005) 187(9):3255–3258] and *Rickettsia prowazekii* [Appl Environ Microbiol (2007) 73(20):6644–6649]. It has also been transferred, in a plasmid, from *Escherichia coli* to *Leptospira* by conjugation [Appl Environ Microbiol (2008) 74(1):319–322].

mariner-based transposons have been used e.g. for studying *Bacillus subtilis* (Tn*YLB-1*) [Appl Environ Microbiol (2006) 72(1):327–333], *Francisella tularensis* (*HimarFT*) [Appl Environ Microbiol (2006) 72(3):1878–1885], *Pseudomonas aeruginosa* (*MAR2xT7*) [Proc Natl Acad Sci USA (2006) 103(8):2833–2838], and also *Listeria monocytogenes* [Appl Environ Microbiol (2007) 73(8):2758–2761].

marker rescue The re-location of a given marker (e.g. a gene) from an inactivated (or a defective) replicon to a functional replicon (through recombination) and, hence, acquisition of the capacity to be expressed or to replicate.

maspin A serine protease inhibitor and potent metastasis/tumor suppressor.

Maspin was reported to be upregulated via the interferon-inducible protein IFIXα [Mol Carcinog (2008) 47(10):739–743].

The gene encoding maspin, *SERPINB5*, was exploited in an EPIGENETIC ALLELIC RATIO ASSAY.

massively parallel DNA sequencing A new approach to DNA SEQUENCING – distinct from the DIDEOXY METHOD – which is now used in a range of techniques. To sequence a *genome*, whole-genome DNA is sheared or cut into a large number of small fragments that are then sequenced, simultaneously (i.e. 'in parallel'), in the same assay. The fragments are often a few hundred base-pairs in length, and, depending on the method, the average read length (i.e. the average number of sequenced nucleotides per fragment) has been reported to range from approximately 25 to hundreds of bases. When the sequencing data – from a single assay – are assembled (via contigs) they may yield an *overall* sequenced template of the order of Mb (10^6 bases) or Gb (10^9 bases). This type of sequencing is now called 'new-generation' or 'second-generation' sequencing.

Some of these massively parallel sequencing systems have been commercialized. Although all involve the simultaneous sequencing of a large number of short target molecules, they differ in the way in which the sequencing is conducted – see e.g. GS FLX, SOLEXA SEQUENCING and SOLID TECHNOLOGY.

The importance of calibrating the fragment library

The optimum performance of any massively parallel method depends critically on the use of a sequencing library that has an appropriate concentration of fragments (i.e. an appropriate number of fragments per unit volume in the sample). If the concentration is low (too few fragments in the sample) then – in e.g. the GS FLX method – some of the beads will not carry a fragment of DNA, so that the sequencer will not be running at full capacity (and may not provide an adequate number of reads). Conversely, if the concentration is too high (too many fragments in the sample), some of the beads will carry more than one fragment; this will cause at least some poor-quality reads.

To avoid the effects of over- or underestimating the number of fragments in the sample, one approach is to run a titration with the sequencer in order to assess the concentration in a given sample; this can be costly and also time-consuming. An alternative approach is to use a method of assessment that is based on DIGITAL PCR (q.v.).

Errors in massively parallel methods

The types of error found in one method may differ from those found in other methods, and the *combined* use of two of these methods (on a given sample) can yield sequences with fewer errors; for example, one study found useful complementarity when using both the GS FLX and Solexa methodology [BMC Genomics (2008) 9:603].

Applications other than sequencing genomic DNA

As well as sequencing genomic DNA, these new approaches have been used for analyzing transcriptomes; in this context, it was pointed out that the absence of introns and intergenic regions enhances the information content of the data and also eases the interpretation [Plant Physiol (2007) 144(1):32–42].

[High-resolution analysis of the 5′-end transcriptome using a (next-generation) DNA sequencer: PLoS ONE (2009) 4(1): e4108.]

[Massively parallel sequencing of the polyadenylated transcriptome of *Caenorhabditis elegans*: Genome Res (2009) 19: 657–666.]

[Construction of a eukaryotic transcriptome (*ab initio*) by massively parallel sequencing of mRNA (cDNA) [Proc Natl Acad Sci USA (2009) doi: 10.1073/pnas.0812841106.]

Massively parallel sequencing has also been used for the examination of DNA in CHROMATIN IMMUNOPRECIPITATION and for diagnosing aneuploidy (see DOWN'S SYNDROME). In virology, a massively parallel system was used for characterizing genetic diversity in the hepatitis B virus [J Virol (2009) 83(4):1718–1726].

[Next-generation sequencing: applications beyond genomes (overview): Biochem Soc Trans (2008) 36(5):1091–1096.]

mast cell growth factor See KIT.

MAT Mating-type locus: see the entry MATING TYPE.

mate-paired library See SOLID TECHNOLOGY.

maternal inheritance A type of CYTOPLASMIC INHERITANCE in which particular characteristics are inherited only from the female parent – for example, only the maternal mitochondrial genes are inherited when paternal mitochondria are excluded from the zygote.

In the *poky mutant* of *Neurospora*, a (defective) gene that affects growth rate is transmitted from the female parent cell to sexually derived progeny; the defective gene is apparently a mitochondrial gene.

(See also *Wolbachia* in the entry GMMS.)

mating type (*fungal genetics*) In general, any strain which can mate (interact sexually) only with a strain of the same species which exhibits a genetically distinct phenotype. Mating types have been studied especially well in the yeast *Saccharomyces cerevisiae*.

Haploid somatic (i.e. vegetative) cells of *S. cerevisiae* give rise to the gametes. Each gamete is of one of the two mating types: **a** and α. Two gametes can fuse only if they belong to *different* mating types. The mating type is determined by the particular allele which occupies the *MAT* (mating) locus: the *MAT***a** allele or the *MAT*α allele; the *MAT***a** allele occupies the *MAT* locus in cells of the **a** mating type, while the *MAT*α allele occupies the *MAT* locus in cells of the α mating type.

All the cells (of both mating types) contain α-specific genes (which specify the α phenotype) *as well as* **a**-specific genes (which specify the **a** phenotype). These genes are controlled from the *MAT* locus: with *MAT*α in the *MAT* locus, α-specific genes are expressed and the cell exhibits the α phenotype; in cells of the **a** phenotype the *MAT* locus is occupied by *MAT***a**.

*MAT*α encodes two products: the α1 and α2 proteins. The α1 protein is required for the expression of α-specific genes; products of these genes include a pheromone – the α-*factor*. The α2 protein acts as a repressor of the **a**-specific genes.

*MAT***a** encodes only a single product (the **a**1 protein) which appears not to be active in *vegetative* cells. With this allele in the *MAT* locus, the **a**-specific genes are expressed; products of the **a**-specific genes include a pheromone, the **a**-*factor*.

On mixing gametes of **a** and α mating types, the pheromone of a given mating type promotes fusion with a gamete of the other mating type. Fusion between a gamete of the **a** mating type and one of the α mating type gives rise to a (diploid) zygote which, itself, has no mating potential; it is designated **a**/α. Within the zygote, the α2 and **a**1 proteins (jointly) switch off both the α-specific and **a**-specific genes, and also switch off the so-called 'haploid-specific genes' which are normally expressed in cells of both mating types.

The (diploid) zygote gives rise to (diploid) vegetative cells. Later, the diploid cells undergo meiotic division, each of the cells producing four *ascospores* within a structure called an *ascus*; of these four ascospores, two are of the **a** mating type and two are of the α mating type.

S. cerevisiae frequently exhibits a spontaneous, reversible change from one mating type to the other. Such switching of mating type involves replacement of the resident allele (in the *MAT* locus) by a *copy* of an allele of the other mating type. This mechanism depends on the presence of 'silent' copies of the *MAT***a** and *MAT*α alleles which are kept transcriptionally inactive by the structure of their chromatin. The switching involves a double-stranded cut in the *MAT* locus (by the HO endonuclease) and removal of the resident allele; this allele is replaced by an allele copied from the silent gene of the other mating type (*gene conversion*).

matrix-assisted laser desorption/ionization See MALDI.

matrix-attachment region (MAR) Any of a number of (typically AT-rich) regions of genomic DNA that interact with the *nuclear matrix* (part of the nuclear envelope) in a eukaryotic nucleus; such interactions are involved e.g. in the formation of CHROMATIN loops which are characteristic features of the nuclear architecture and which are thought to be indicative of activity in terms of transcription and/or replication.

Part(s) of an interacting MAR may contain separated base-

pairs, reflecting unwinding stress; the *base-unpaired regions* (BURs) may be centered on the sequence ATATAT (a *BUR nucleation sequence*).

The base-unpairing associated with MAR–matrix interaction has been linked e.g. with enhancement of local promoter activity. Modulation of gene expression appears to involve a protein, the special AT-rich sequence-binding protein 1 (see the entry SATB1).

Inter-MAR association was reported in the human β-globin gene cluster [PLoS ONE (2009) 4(2):e4629].

Exploitation of MARs in DNA technology

MARs have been used in attempts to avoid gene silencing of transgenes; thus, transgenes were flanked by MAR sequences prior to transfection. However, although increased transgene expression was observed in some cases, no effect was seen in others; decreased expression of transgenes was seen in some cases.

MAR sequences were used specifically to limit expression of GREEN FLUORESCENT PROTEIN (q.v).

maturation inhibitor (HIV-1 maturation) See e.g. BEVIRIMAT.

Maxam–Gilbert method (of DNA sequencing) (*syn*. chemical cleavage method) A method of DNA SEQUENCING involving four separate reactions – in each of which an aliquot of end-labeled sample DNA is subjected to a chemical agent which cleaves at *one* of the four types of nucleotide; conditions are arranged so that, ideally, each fragment of the sample DNA in a given reaction is cleaved only once, i.e. at only one of the typically many residues of the relevant nucleotide within the fragment.

As each of the fragments is labeled at one end, the products of a reaction include a number of end-labeled subfragments of various lengths – each extending from the labeled end to one of the positions at which cleavage occurred, at the given nucleotide, within the fragment.

The reaction is repeated for each of the other three nucleotides – using an appropriate chemical cleavage agent in each case.

The four sets of subfragments (one from each reaction) are then subjected to gel electrophoresis in four separate lanes. As the subfragments are separated according to their lengths (i.e. according to the number of nucleotide residues in each) – and as the products from a given reaction indicate cleavage at the particular nucleotide at different locations in the fragment – the position of a given band in the gel indicates the identity of the nucleotide at a specific location in the fragment; hence, the overall sequence of nucleotides within the fragment can be read directly from the gel.

Analogous procedure for sequencing RNA

A procedure analogous to the Maxam–Gilbert method – but using base-specific endoribonucleases (rather than chemical agents) – has been employed for sequencing single-stranded RNA. Endoribonucleases used: T_1 (Gp↓N), U2 (Ap↓N), Phy M (Ap↓N, Up↓N) and an enzyme from the (Gram-positive) bacterium *Bacillus cereus* (Up↓N, Cp↓N).

maxicell A *recA*, *uvrA* mutant bacterium, subjected to ultraviolet radiation, in which chromosomal DNA has been extensively damaged and degraded. Some of the small, multicopy plasmids within such cells may escape damage and may be able to direct the synthesis of proteins.

(cf. MINICELL.)

maxicircles (in trypanosomatids) See KINETOPLAST.

maxizyme An engineered, allosterically controllable, dimeric, Mg^{2+}-dependent NUCLEIC ACID ENZYME that is able to carry out cleavage of mRNA, *in vivo*, in a highly sequence-specific manner [original paper: Mol Cell (1998) 2(5):617– 627].

A maxizyme has been used e.g. in studies on Alzheimer's disease [Gene Ther Mol Biol (2005) 9A:89–106].

(See also BINARY DEOXYRIBOZYME LIGASE.)

MBD proteins Methyl-CpG-binding domain proteins: a family of proteins that bind preferentially to a symmetrically methylated CpG motif in DNA.

These proteins can e.g. repress *in vitro* transcription.

(See also CPG ISLAND and KAISO PROTEIN.)

Mcc plasmid See MICROCIN.

McKusick's Online Mendelian Inheritance in Man See entry ONLINE MENDELIAN INHERITANCE IN MAN.

***mcrA* gene** (*syn*. *rglA*) In *Escherichia coli*: a gene encoding a RESTRICTION ENDONUCLEASE reported to cleave DNA e.g. at the sequence 5′-C^{m}CGG-3′ (in which the *second* cytosine residue is methylated).

Inactivation of *mcrA* apparently permits greater efficiency when cloning DNA methylated in the above sequence.

mcrBC (*syn*. *rglB*) In *Escherichia coli*: a sequence encoding a restriction system which is reported to cleave DNA at 5′-GC-3′ loci in which the cytosine residue is methylated.

Inactivation of *mcrBC* apparently permits greater efficiency when cloning DNA methylated in the above sequence.

An *mcrBC*$^{+}$ strain (i.e. one with an active form of this restriction system) is used in the GENETAILOR SITE-DIRECTED MUTAGENESIS SYSTEM.

Note. This sequence is also written *mcrCB*.

MCS (multiple cloning site) *Syn*. POLYLINKER.

MDA See MULTIPLE DISPLACEMENT AMPLIFICATION.

MDA-7/IL-24 A protein which, when expressed intracellularly by an adenoviral vector, induces cancer-specific APOPTOSIS. Moreover, when secreted, this protein can induce apoptosis in nearby, non-infected *tumor* cells (though not in normal cells) [suggested mechanism: Proc Natl Acad Sci USA (2008) 105 (28):9763–9768].

(See also MELANOMA.)

MDR-TB Multidrug-resistant tuberculosis.

MDS42 A strain of REDUCED-GENOME ESCHERICHIA COLI.

***me*53 gene** In *Autographa californica* NPV: the gene encoding a product that is necessary for efficient production of budded virions (BVs): see the entry BACULOVIRIDAE.

measles virus A virus of the genus *Morbillivirus* – within the family PARAMYXOVIRIDAE (q.v.).

***mecA* gene** In strains of the Gram-positive bacterium *Staphylococcus aureus* (see MRSA): a gene which encodes a cell-wall-synthesizing enzyme (PBP 2a or PBP 2′ – penicillin-binding

protein 2a or 2′) that is resistant to methicillin (and to other β-lactam antibiotics); this gene can confer high-level resistance to β-lactam antibiotics.

The *mecA* gene occurs in the SCCMEC element (q.v.).

Medea (gene drive system) See GMMS.

media (microbiological) The plural of MEDIUM.

medium (microbiological) Any solid (gel) or liquid preparation formulated specifically for the growth, maintenance (storage) or transportation of microorganisms (particularly bacteria or fungi).

Solid media are typically prepared with AGAR or with gelatin (cf. GELLAN GUM) and are usually dispensed into Petri dishes, glass screw-cap bottles or similar containers.

Liquid media are generally used in test tubes, screw-cap containers or flasks. *Any* medium must be sterile prior to use, and it must be inoculated and incubated with strict regard to avoidance of extraneous contamination.

(See also ASEPTIC TECHNIQUE and SAFETY CABINET.)

(Entries for specific types of medium: COMPLETE MEDIUM, MACCONKEY'S AGAR, NUTRIENT AGAR, SOC MEDIUM and e.g. TERRIFIC BROTH.)

meganuclease *Syn*. HOMING ENDONUCLEASE (q.v.).

megaprimer mutagenesis A PCR-based technique for SITE-DIRECTED MUTAGENESIS involving e.g. a small primer, containing the mutant site, which binds to an *internal* site on one strand of target DNA. Extension of this primer to the 5′ end of the template strand forms the *megaprimer*. Ongoing temperature cycling – with a conventional flanking primer to copy the *other* strand of the template – produces a large population of the megaprimer *and complementary copies* of the megaprimer sequence. The *complementary* copies are isolated and are used as one of the two types of primer to copy the target DNA in further rounds of PCR, leading to dsDNA amplicons containing the mutation. (Isolating complementary copies of the megaprimer can be facilitated if the conventional flanking primer is biotinylated; the complementary copies can then be isolated from denatured, single-stranded amplicons by the use of a streptavidin-based system.)

[Examples of megaprimer mutagenesis: J Bacteriol (2006) 188:5055–5065; J Virol (2006) 80(16):7832–7843; Mol Cell Biol (2006) 26(18):6762–6771; Biochemistry (2008) 47(33): 8590–8599; J Biol Chem (2008) 283(44):30351–362; J Mol Biol (2008) 376(4):1091–1099; Plant Physiol (2009) 149 (2): 1076–1086; and Retrovirology (2009) 6:3.]

MEIA Microparticle enzyme immunoassay: see LIGASE CHAIN REACTION.

meiotic drive (gene drive system) See GMMS.

melanocyte See MELANOMA.

melanoma (also known as malignant melanoma) A malignant tumor of melanocytes (cells that form the pigment *melanin*); metastases (spread of the tumor to other regions of the body) are common. Melanomas are usually darkly pigmented; they occur in the skin (often following over-exposure to sunlight) and, less frequently, e.g. in the conjunctiva of the eye. In the *amelanotic* melanomas the tumor cells lack melanin, and, as a consequence, such tumors may be detected late as they are less easily seen.

(See also ULTRAVIOLET RADIATION and XERODERMA PIGMENTOSUM.)

DNA technology has been used to search for methods of targeted therapy against this type of tumor. In one approach, a specific microRNA was used to interfere with cell-cycle molecules in melanoma cells, inhibiting progression through the cell cycle (see microRNA *let-7b* in MICRORNAS).

Another approach to therapy involves a *cancer terminator virus* (CTV). Thus, an ADENOVIRUS was engineered so that (i) expression of its E1A gene (which is necessary for viral replication) occurs *specifically* in cancer cells, and (ii) during replication, the CTV produces large amounts of MDA-7/IL-24 protein. Injection of this CTV into human melanoma cells demonstrated (i) adenoviral replication, (ii) expression of the MDA-7/IL-24 protein, (iii) inhibition of cell growth, and (iv) induction of apoptosis. In a mouse model that used athymic nude mice, injection of the CTV into xenografts derived from MeWo human metastatic melanoma cells eliminated not only the primary (treated) tumors but also the distant, non-treated tumors, effecting a cure. This has suggested the potential for treating human patients who have advanced melanomas with metastases [Cancer Gene Therapy (2008) 15:293–303].

(See also 5-AZA-2′-DEOXYCYTIDINE.)

melittin A cationic, 26-amino acid, membrane-permeable peptide. When bound (covalently) to poly(ethyleneimine) (PEI), this peptide forms complexes with DNA in which the DNA condenses into discrete particles (<100 nm diameter) which have enhanced entry into the cytoplasm during transfection; this approach also appears to promote entry of DNA into the nucleus (indicated by enhanced expression of transgenes) and may have potential as a non-viral gene-delivery system.

A sequence encoding melittin is included in some versions of the BACULODIRECT vector in order to promote secretion of the expressed protein (see e.g. BaculoDirect™ vector in the table in GATEWAY SITE-SPECIFIC RECOMBINATION SYSTEM).

Melittin (at 15 μg/mL) was used in an assay of endogenous reverse transcription of HIV-1 [PLoS Pathog (2009) 4(12): e1000231].

(See also ENDO-PORTER.)

melting (of dsDNA) Strand separation, caused e.g. by elevated temperature.

Analysis of DNA samples, based on their melting characteristics, can be carried out by using commercial systems such as HIGH-RES MELTING and HRM. (cf. TM-SHIFT PRIMERS.)

(See also HELIX–COIL TRANSITION and THERMAL MELTING PROFILE.)

membrane anchor sequence See SIGNAL SEQUENCE.

menadione (2-methyl-1,4-naphthoquinone) An agent which is able to induce production of ROS (reactive oxygen species) in cells; menadione is used e.g. in studies on oxidative stress. (Other agents, e.g. potassium bromate, are also used for this purpose.)

-mer Part of a designation used to refer to the number of sub-

units in a (short) linear polymer; thus, for example, an 8-mer oligonucleotide contains eight nucleotide residues.

2-mercaptoethanol ($HO(CH_2)_2SH$) A reducing agent present e.g. as a constituent in buffers used for storing enzymes (e.g. many restriction endonucleases); it maintains sulfhydryl (SH) groups in the reduced state.

merodiploid A partial diploid: a term used e.g. for a (haploid) bacterial cell which has two copies of some gene(s) – extra copies having been acquired e.g. by transformation.

merozygote A prokaryote which has received exogenous DNA (e.g. by conjugation or transformation) and which, as a consequence, has become diploid in respect of certain allele(s). The recipient's genetic complement is the *endogenote*, while the imported DNA is termed the *exogenote*.

If the sequence in the exogenote is the same as that in the endogenote, the merozygote is a *homogenote*; if not identical, the merozygote is a *heterogenote*.

MESG A designation used, in a number of papers, to refer to 2-amino-6-mercapto-7-methylpurine ribonucleoside, as used in the ENZCHEK PYROPHOSPHATE ASSAY (q.v.). In other papers, MESG is used for '7-methyl-8-thioguanosine'.

MessageAmp™ aRNA amplification kit A commercial kit (Ambion, Austin TX) used for the high-level amplification of RNA (for example, for microarray gene expression profiling) when only limited quantities are available. Essentially, total – or poly(A) – RNA is used for FIRST STRAND synthesis of cDNA using an oligo(dT) primer which has a 5′ promoter sequence for T7 RNA polymerase. The hybrid RNA/DNA duplexes are treated with RNase H (fragmenting the RNA) and the RNA fragments are used as primers for synthesis of the second strand. Synthesis of the second strand provides a double-stranded template for the T7 polymerase and permits ongoing *in vitro* transcription to produce large amounts of antisense RNA (aRNA).

The (aRNA) product may be labeled during, or after, synthesis. For example, aminoallyl UTP may be incorporated at the time of synthesis; this allows a choice of the dye which is subsequently coupled to the reactive amino group. (The range of available amine-reactive fluorophores includes a number of Alexa Fluor® dyes.)

[Uses (e.g.): Plant Physiol (2009) 149(3):1505–1528; PLoS ONE (2009) 4(2):e4486.]

messenger ribonucleoprotein See MRNP.

messenger RNA See MRNA.

meta operon (in the TOL plasmid) See TOL PLASMID.

metacentric Refers to a CHROMOSOME whose CENTROMERE is located at, or close to, the center of the chromosome.

(cf. ACROCENTRIC.)

metagenome The totality of genomes of all microflora found in nature, or in a particular environmental location (e.g. the 'ground-water metagenome') – including the genomes of those species that have been cultured and those that have not been cultured or identified.

[MEGAN: computer-based analysis of metagenomic data: Genome Res (2007) 17(3):377–386.]

The genomes of uncultured species are believed to form the major part of genetic information in the natural world and to constitute a rich source of information on e.g. novel products. Accordingly, various methods have been used for the *direct* screening of a metagenome with the object of isolating novel genes without prior isolation of the organisms from which the genes are derived.

One type of screening detects and amplifies target genes by means of probes or PCR primers based on known, *conserved* sequences of nucleotides.

A different approach starts with the preparation of a metagenomic library and uses it to screen for novel phenotypes by expression in an *Escherichia coli* (or other) host species. (See also SIGEX.)

One difficulty in studies on a metagenome is the inefficient amplification of rare sequences, i.e. sequences that occur in samples in low-copy-number fragments. One approach to this problem is to use inverse affinity nested PCR (IAN-PCR). In this method, the first step is a standard inverse PCR with one primer carrying an affinity tag (such as BIOTIN); the resulting PCR products are affinity purified (thus enriching the wanted product and removing the background products) and are then subjected to a nested PCR to recover target-flanking regions. Recovery of target-flanking regions facilitates the application of a genome walking strategy in respect of a particular target gene [BioTechniques (2006) 41(2):183–188].

Metagenomic detection of viral pathogens has been carried out by using new-generation sequencing [PLoS ONE (2009) 4(1):e4219].

metastasis See CANCER. (See also gene *NM23-H1* in the entry GENE-DELIVERY SYSTEM.)

[MicroRNAs in metastasis: Nature Reviews Cancer (2009) 9:293–302.]

methicillin 6-(2,6-dimethoxybenzamido)penicillanic acid – a semi-synthetic PENICILLIN which is resistant to cleavage by a range of β-lactamases. Bacterial resistance to methicillin can be due e.g. to penicillin-binding protein 2a (PBP 2a) (see the entries MRSA and MECA GENE).

methicillin-resistant *Staphylococcus aureus* See MRSA.

methotrexate *Syn.* AMETHOPTERIN.

methoxyamine (as a mutagen) An agent (NH_2OCH_3) with activity similar to that of HYDROXYLAMINE.

2-methyl-1,4-naphthoquinone See MENADIONE.

methyl-binding PCR A method for detecting methylation of CpG sequences in genomic DNA. Essentially, genomic DNA is fragmented with the restriction endonuclease MseI and any fragment of interest – e.g. a CpG island associated with the promoter of a given gene – is allowed to bind to immobilized molecules of a recombinant protein that has a high affinity for methylated CpGs; in the tube containing the immobilized proteins (and bound DNA fragments) PCR is carried out with gene-specific primers.

[Method: Nucleic Acids Res (2006) 34(11):e82.]

methyl-CpG-binding domain See the entry CPG ISLAND. (See also MBD PROTEINS.)

***N*-methyl-*N*′-nitro-*N*-nitrosoguanidine** See DNA REPAIR.

***N*-methyl-*N*-nitrosourea** See DNA REPAIR.

3-methyladenine DNA glycosylase II See DNA REPAIR.

methylation (of DNA) Addition of a methyl group ($-CH_3$) to specific bases in DNA: a normal *in vivo* process which has major implications for gene expression, especially in eukaryotes (see e.g. CPG ISLAND). The methyl donor is commonly *S*-adenosylmethionine.

For notes on methylation in prokaryotes and eukaryotes see the entry DNA METHYLTRANSFERASE.

(Relevant entries – BISULFITE, CPG ISLAND, CPNPG SITES, EPIGENETICS, GENETIC IMPRINTING, METHYL-BINDING PCR and METHYLATION-SPECIFIC PCR, RESTRICTION ENDONUCLEASE and RESTRICTION LANDMARK GENOMIC SCANNING.)

Methylation and gene expression in mammals

Much attention has been focused on the fact that methylation of normally unmethylated cytosine residues within CpG-rich sequences in the promoter region of certain genes can inhibit transcription of those genes; this may lead to disease if, for example, it blocks the expression of tumor suppressor genes. (Such aberrant methylation has been targeted by methylation-inhibitory chemotherapy which uses DNA methyltransferase inhibitors (such as 5-AZA-2′-DEOXYCYTIDINE) – see the entry DEMETHYLATION.)

Aberrant *hypo*methylation (under-methylation) of an ONCOGENE which is normally silenced by methylation can lead to activation of that oncogene [PLoS ONE (2009) 4(3):e4961].

De-repression (i.e. activation) by *hyper*methylation occurs in the gene encoding SURVIVIN – a gene often overexpressed in tumors. This gene is normally repressed by the binding of P53; hypermethylation of the gene inhibits the binding of p53 and allows gene expression. Experimental demethylation of the survivin gene (using the agent decitabine) was reported to enable resumption of p53-dependent repression [Oncogene (2009) doi: 10.1038/onc.2009.62].

Laboratory investigation of methylation

Methylation of specific bases may be achieved *in vitro* and *in vivo* e.g. by TAGM. It was reported that targeted methylation can also be achieved with mutant methyltransferases that are linked to zinc-finger arrays – the latter giving specificity; the pattern of methylation imparted to the DNA (*in vivo*) was transmitted through successive cell divisions [Nucleic Acids Res (2007) 35(3):740–754].

Methylation in specific alleles has been studied by ALLELE-SPECIFIC DNA METHYLATION ANALYSIS (see also METHYLATION-SPECIFIC PCR).

[Microarray-based DNA methylation profiles – technology and applications: Nucleic Acids Res (2006) 34:528–542.]

Detection of C5-methylated cytosine residues in genomic DNA was reported in a microarray-based protocol employing monoclonal antibodies [DNA Res (2006) 13(1):37–42].

PCR-based analysis of DNA methylation has been carried out with BISULFITE-treated – but initially non-desulfonated – DNA (SafeBis DNA) that is resistant to uracil-*N*-glycosylase (UNG), i.e. the enzyme used in a standard anti-contamination procedure; in this method, the desulfonation of SafeBis DNA occurs (in PCR) during an initial prolonged (30-minute) stage of denaturation at 95°C [Nucleic Acids Res (2007) 35(1):e4].

Methylation in various genomes can be compared/analysed by REDUCED REPRESENTATION BISULFITE SEQUENCING.

Genome-scale DNA-methylation maps have been prepared for various types of mammalian (mouse) cell, including both ES (embryonic stem) cells and differentiated cells. The maps show, for example, that patterns of methylation are correlated more closely with histone status than with genomic sequence. Moreover, CpG methylation exhibits extensive changes when cells differentiate. [DNA methylation maps of pluripotent and differentiated cells: Nature (2008) 454:766–770.]

methylation (of RNA) See e.g. CAP.

methylation-deficient cells Mutant cells which are defective in methylation of DNA bases. Thus, for example, strain SCS110 of *Escherichia coli* (Stratagene, La Jolla CA) lacks the *dam* (adenine methylase) and *dcm* (cytosine methylase) functions; such strains are useful e.g. for producing cloned DNA which can be cut by methylation-sensitive RESTRICTION ENDONUCLEASES (such as EcoRII and XbaI).

methylation-specific PCR (MSP) A form of PCR which is used e.g. for monitoring the pattern(s) of methylation of cytosine residues at CpG sites in DNA; it depends on the ability of BISULFITE to convert any unmethylated (but not methylated) cytosine residues to uracil residues. Following treatment with bisulfite, a nucleotide sequence that contains CpG site(s) will therefore change to one containing UpG site(s) if the cytosine residues had been unmethylated; in subsequent amplification of such a sequence by PCR, any U bases in the template will pair with A (adenine) during extension of the primers.

After bisulfite treatment, the target sequence is subjected to PCR with several different sets of primers, including a primer pair that can anneal to binding sites in which cytosine(s) have been converted to uracil(s), and another primer pair that can anneal to binding sites in which the given cytosine(s) have remained as cytosine(s). Analysis of the PCR products gives an indication of the methylation status of the cytosines in the target sequence.

[Uses (e.g.): PLoS ONE (2009) 4(4):e5011; BMC Cancer (2009) 9:101; Proc Natl Acad Sci USA (2009) 106(13): 5076–5081.]

methylation-specific single-base extension (MSBE) A method used for monitoring methylation at CpG sites in genomic DNA. Genomic DNA is initially treated with BISULFITE to convert *non*-methylated cytosine to uracil. Bisulfite-modified DNA is then amplified by PCR.

A *specific* cytosine which was *non*-methylated before treatment with bisulfite will, after PCR, appear as thymine in the resulting amplicons. That is, after conversion to uracil, it will base-pair with adenine during primer extension in the initial cycle of amplification – so that, in later rounds of PCR, the position originally occupied by this cytosine will be occupied by thymine in the amplicons.

The presence of T (or A) in the amplicons is determined by

using a primer whose 3′ end can be extended by an A (or a T) which base-pairs with the base *at the given site*. This primer is extended, for a single base only, by using a labeled, chain-terminating dideoxyribonucleoside triphosphate; extension of the primer by ddA or ddT (signaled by the labeled products) indicates that the given CpG is non-methylated in the original sample of genomic DNA.

***N*,*N*′-methylene-*bis*-acrylamide** (Bis) A cross-linking agent which is used in the preparation of POLYACRYLAMIDE gels from ACRYLAMIDE.

In polyacrylamide gels used for the separation of fragments of nucleic acid by GEL ELECTROPHORESIS, the ratio of acrylamide to Bis is often ∼19:1.

methylmethane sulfonate (MMS) An alkylating agent that can e.g. methylate guanine residues in DNA.

[Use (e.g.): Eukaryotic Cell (2008) 7(5):800–813.]

methylome The totality of DNA methylation, in a given cell or organism, and its inherent information. [Human methylome in embryonic stem cells and fetal fibroblasts: Nature (2009) doi: 10.1038/nature08514.]

methylotrophic fungi Those fungi – e.g. species of the yeasts *Hansenula* and *Pichia* – which are able to carry out oxidative metabolism of C_1 compounds (such as methanol) with the assimilation of formaldehyde as a major source of carbon. An example is *Pichia pastoris*.

methyltransferase See DNA METHYLTRANSFERASE.

***mfd* gene** A gene (*mutation frequency decline*) which encodes the *transcription repair coupling factor* (TRCF): a protein which facilitates the release of a stalled RNA polymerase at a lesion on a DNA template.

While, as the name indicates, the Mfd protein is associated with a reduction in mutation frequency, the loss of Mfd has been reported to depress stationary-phase mutagenesis in the Gram-positive bacterium *Bacillus subtilis* [J Bacteriol (2006) 188(21):7512–7520].

MicF A bacterial sRNA: see the entry SRNAS.

microarray (DNA chip; chip) Typically: thousands or millions of short, single-stranded DNA molecules, having diverse *but known* sequences, attached (at one end) to specific, known locations on the surface of a small piece of glass or silica etc.; acting as PROBES, the immobilized molecules have a range of uses (see below). (A microarray can also be constructed with peptide nucleic acid (PNA) (see later).)

(cf. MACROARRAY.)

Construction of a microarray

A microarray can be made in two main ways. In one method, a mechanical device is used to attach the oligonucleotides to a prepared solid support – either non-covalently (as in earlier chips) or covalently (see also ATMS and DENDRICHIP).

In the second method (so-called 'on-chip synthesis'), a high-speed robotic device is used for the *in situ* synthesis of probes at specific locations (see FEATURE) on the chip; each probe can be synthesized with any desired sequence of nucleotides.

Probe synthesis may involve *random access combinatorial chemistry* – for details see:

http://www.affymetrix.com/technology/chemistry.affx

On-chip synthesis is used for manufacturing GeneChip® arrays (Affymetrix Inc.). In GeneChip® arrays the probes, in at least some cases, are 25 nucleotides long. Probes in other microarrays reported in the literature have sizes ranging from ∼10 to ∼70 nucleotides; a certain minimum length is needed for adequate specificity in probe–target binding.

Basic ideas in the use of microarrays

Using a high-density microarray, the nucleotide sequence in a given labeled fragment of DNA may be determined by allowing the fragment to bind, i.e. hybridize, to the particular probe in the array that has a *complementary* sequence; the location of this probe, on the chip, indicates the probe's sequence, and therefore indicates the sequence of the (complementary) fragment. (Hybridization must be carried out with an appropriate degree of stringency.) However, while this approach may be useful for short sequences it would not be suitable when considering e.g. the sequencing of a large genome.

The use of *multiple copies* of each probe, at its given *spot* on the chip, increases the sensitivity of a microarray and also allows a comparison of populations of (differentially labeled) test samples (see below).

Microarrays are used (e.g.) to study gene expression under different conditions. In one method, two samples of mRNA – obtained from the same cells under different conditions – are each converted to (*differentially labeled*) cDNAs; these two, labeled populations are then simultaneously hybridized to a microarray that includes probes from many, most or all of the coding sequences in the cell's genome. After hybridization, a comparison of the intensity of the two labels at each *spot* – i.e. cluster of identical probes – is used to indicate the relative levels of transcripts of the given gene in the test samples; this may or may not reflect the actual rate of *transcription* of the gene in the two populations of cells (see below). Detection of only *one* label at a particular spot would suggest that the gene is expressed under only one of the (two) sets of conditions being examined. The labels used are often fluorescent dyes – e.g. one which fluoresces red and another which fluoresces green; this is the so-called two-color strategy. Hence, if only one label binds to a given spot, that spot would display either red or green fluorescence. A *yellow* spot is produced by the binding of *both* labels at similar levels – although the actual visual signal (color) at a given spot varies according to the relative levels of the two types of label. However, instead of relying on a visual signal, a microarray 'reader' measures the fluorescent signals.

The above experiment compares the levels of *transcripts* in the test samples, rather than assessing levels of transcription. The presence of a particular level of a given transcript in a test sample is influenced not only by the rate of transcription but also e.g. by the stability of the mRNA under conditions in the test population.

If there is only limited material for gene profiling it may be possible to address the problem by carrying out a preliminary

amplification (see e.g. MESSAGEAMP ARNA AMPLIFICATION KIT).

A further use for a microarray is the expression profiling of MICRORNAS; when hybridized to a microarray, labeled cDNA copies of miRNAs from a sample of tissue may indicate the pattern of expression of miRNAs under given conditions. For example, one study demonstrated a particular set of miRNAs under hypoxic conditions (i.e. low concentrations of oxygen) [Mol Cell Biol (2007) 27(5):1859–1867].

A microarray-based assay for finding C5-methylated cytosine residues in genomic DNA (without BISULFITE treatment *or* DNA amplification) made use of monoclonal antibodies (mAbs) that specifically recognize C5-methylated cytosine in strands which are hybridized to oligonucleotide probes. After probe hybridization, the mAbs were bound to target (i.e. C5-methylated) cytosine residues; this was followed by exposure of the array to a (dye-labeled) anti-immunoglobulin antibody which binds to mAbs and which is detected by fluorescence of its cyanine dye. [Method: DNA Res (2006) 13(1):37–42.]

(See also EXON ARRAY, THREE-PRIME ARRAY and TILING ARRAY.)

The widespread use of 'custom' microarrays has prompted an evaluation of design features with a view to maximizing efficiency, reducing cost, and reducing the amount of useless data produced [Nucleic Acids Res (2009) 37(6):1726–1739].

Optimizing a microarray

The opportunity for target–probe hybridization was reported to be optimized (for accurate and reproducible data) by active mixing [BioTechniques (2005) 38:117–119].

Stem–loop probes may be used to optimize the specificity of hybridization [BioTechniques (2008) 44(1):119–126].

Following hybridization, the intensity of a fluorescent label (and the ease with which it can be detected) may depend on the *position* of the label in the target sequence [Nucleic Acids Res (2005) 33(19):e166].

Factors potentially disruptive to probe–target hybridization, and to the interpretation of signals, include the effects of the so-called 'dangling ends' – i.e. non-hybridized parts of target strands which project beyond the free ends of the probes; the dangling ends may affect binding between probes and their legitimate targets, and they may also bind to other strands – leading to cross-hybridization signals [Appl Environ Microbiol (2007) 73(2):380–389].

Microarray studies on microorganisms

Among microorganisms, microarrays have been used e.g. to identify *Mycobacterium* spp (and *simultaneously* to test for antibiotic resistance); to investigate attenuation of virulence in *Pseudomonas aeruginosa* by the quorum sensing inhibitor furanone; and for profiling mutant strains of *Staphylococcus aureus*. In this type of study, target sequences are commonly generated by PCR-based amplification of specific regions of the genomic DNA; the resulting amplicons are then examined (in single-stranded form) by hybridization to the microarray.

A microarray of 14283 probes was developed for sequence-*independent* identification of organisms from samples of their genomic DNA. Essentially, DNA from a given sample was digested with restriction enzymes, labeled, denatured, and then applied to the microarray. The patterns of hybridization of the restriction fragments from different species were able, reproducibly, to differentiate the test organisms on the basis of their species-dependent hybridization signatures.

PNA-based microarray

The probes in a microarray can be made from a DNA analog, peptide nucleic acid (see PNA). In one early study, a microarray of PNA probes was used for detecting mutations in the gene for hepatocyte nuclear factor-1α that are responsible for one type of maturity-onset diabetes in the young. Initially, the relevant mutation-prone region in the genome was amplified by PCR using allele-specific primers; the resulting amplicons were examined as follows.

Mutant nucleotides in these amplicons were detected by the use of various primers, each *type* of primer having (i) a 5′ binding site for a *specific* PNA probe in the microarray, and (ii) a 3′ end that is capable of SINGLE-BASE EXTENSION at *one particular type* of mutant nucleotide. The SBE primers were extended by BIOTIN-labeled ddNTPs. Subsequently, when the (extended) primers were bound to PNA probes in the microarray they were detected by (dye-linked) STREPTAVIDIN. The binding of a given (extended) primer to a *particular* PNA probe thus indicated *one* type of mutation that was present in the sample DNA [Nucleic Acids Res (2005) 33(2):e19].

Non-optimal probe–target binding

One problem with a microarray is that the binding of a *range* of target sequences to their complementary probes is carried out at one temperature, even though the various probe–target pairs may have a range of different T_m values (see THERMAL MELTING PROFILE); therefore, a given operating temperature is likely to be optimal, or near-optimal, for some probe–target pairs but suboptimal for others. This temperature problem has been addressed with a ZIPCODE ARRAY.

microbial forensics See FORENSIC APPLICATIONS.

microbicidal (*adj.*) Able to kill (at least) certain types of microorganism under given conditions.

(cf. MICROBISTATIC.)

microbiota See the entry MICROFLORA.

microbistatic (*adj.*) Able to arrest the growth of certain microorganisms under given conditions.

(cf. MICROBICIDAL.)

microcin Any of a range of low-molecular-weight, typically thermostable BACTERIOCINS produced by enterobacteria and active e.g. against related bacteria. Unlike COLICINS, the synthesis of microcins is not induced by conditions that trigger the SOS response.

Microcins are usually synthesized in the stationary phase of growth; rich media may depress their synthesis.

Microcins range in size from microcin A15 (a single, modified amino acid – apparently a derivative of methionine) to peptides of up to ~50 amino acid residues.

Microcin B7 inhibits DNA gyrase, and microcin C7 blocks protein synthesis. Microcin A15 inhibits the enzyme homo-

serine succinyltransferase in the pathway for methionine biosynthesis.

Microcin-encoding plasmids include both conjugative and non-conjugative types. These plasmids are called M plasmids or (preferably) Mcc plasmids.

microflora (*microbiol.*) Collectively, the microorganisms that are normally associated with a given environment or niche; the term is sometimes abbreviated to 'flora'.

Some authors have used the term 'microbiota', in place of microflora, with the apparent aim of excluding the 'plant-like' connotation. However, 'microflora' is a long-established and widely accepted term which is generally understood to have the meaning given above.

microinjection The direct injection of nucleic acid into a cell, a purely physical alternative to chemically assisted or vector-assisted methods of internalizing nucleic acid. Microinjection was chosen as the preferred method e.g. in a study on the fate of internalized DNA within cytoplasm and nucleus; methods such as LIPOFECTION were regarded (by the authors) as unsuitable for this particular type of study because they result in the deposition of large amounts of DNA at the *surface* of the cell [Nucleic Acids Res (2005) 33(19):6296–6307].

A ribonucleoprotein particle, used for targeted mutagenesis in bacteria, was microinjected into cells of *Xenopus laevis* to demonstrate the potential of this approach for gene targeting reactions in eukaryotic cells [PLoS ONE (2008) 3(9):e3121].

micron (μ) A trivial name for the micrometer (μm, 10^{-6} m).

micronucleus (*ciliate protozool.*) The smaller of two types of ciliate nuclei (cf. MACRONUCLEUS); each cell may contain one, two or more micronuclei, depending on species. Micronuclei are usually diploid; they lack a nucleolus and appear to be involved mainly in recombination and the development of new macronuclei.

microparticle enzyme immunoassay See entry LIGASE CHAIN REACTION.

Microprocessor A complex which includes Drosha (an RNase III-like enzyme) and the dsRNA-binding protein DGCR8; it is commonly involved in the biogenesis of MICRORNAS.

(See also MIRTRON.)

microRNAs (miRNAs) Small, genomically encoded molecules of RNA, typically ∼19–22 nucleotides in length, whose roles, where known, are regulatory; the miRNAs bind to particular targets, commonly mRNAs, typically blocking their function and causing post-transcriptional GENE SILENCING (see entry PTGS). The miRNAs (like the siRNAs) work in conjunction with proteins of the ARGONAUTE family (q.v.).

miRNAs fall within the broader category of (eukaryotic and prokaryotic) *small non-coding RNAs* (see the entry SRNAS).

In at least some cases, gene silencing by miRNAs involves degradation of a target transcript by nucleolytic cleavage (as is characteristic of RNA INTERFERENCE); although this may appear to be a purely degradative process, it was suggested that it might also be a positive mechanism for the generation of biologically functional but non-coding fragments of RNA [Nature Rev Mol Cell Biol (2009) 10:141–148]. It was also suggested that the various pathways of the small, silencing RNAs, including both miRNAs and siRNAs, may form an interconnected and coordinated regulatory system which has a protective role [Nature Rev Genetics (2009) 10:94–108].

A given miRNA may target more than one type of transcript.

A given gene may be regulated by more than one miRNA.

Synthetic molecules, termed ANTAGOMIRS, have been used e.g. for *in vivo* silencing of endogenous miRNAs.

Biogenesis of miRNAs

In eukaryotic cells, the *maturation* of miRNAs involves the activity of a number of enzymes; however, it appears that the biogenesis (maturation) pathway of miRNAs is not the same in all cases – miRNA-specific differences have been reported [Nature Cell Biol (2009) 11:228–234]. [See also Nature Rev Mol Cell Biol (2009) 10:126– 139.]

Most mammalian miRNA-encoding sequences are reported to occur in the *introns* of primary transcripts – *pri-miRNAs* – synthesized by RNA polymerase II; these primary transcripts apparently include some that are protein-encoding and others that are non-protein-encoding.

The sequences that encode miRNAs frequently occur in clusters – a given cluster being transcribed (in an OPERON-like way) from a common promoter, yielding a polycistronic primary transcript. In the nucleus, the pri-miRNAs may be processed by the Microprocessor complex (involving e.g. the RNase III-like Drosha enzyme and the dsRNA-binding protein DGCR8) to ∼70 bp hairpin-containing structures called *pre-miRNAs*. However, in some cases, formation of pre-miRNAs involves the direct splicing of pri-miRNAs (thus bypassing the Microprocessor) – *these* pre-miRNAs being termed *mirtrons*. Pre-miRNAs are transferred to the cytoplasm (via the exportin-5 pathway), and a further stage of processing is carried out by the DICER enzyme, yielding the mature dsRNA miRNA – one strand of which is the effector molecule.

Studies on a cluster of 46 miRNA genes on chromosome 19 (designated C19MC) suggested that they are expressed from placenta-specific, non-protein-encoding transcripts [Nucleic Acids Res (2009) doi: 10.1093/nar/gkp205].

Functional overlap was reported between two miRNA gene clusters [Nucleic Acids Res (2009) 35(5):1672–1681].

Characteristics of miRNAs

miRNAs commonly bind to the 3′-UTR region of mRNA, but such hybridization generally appears to involve an inexact matching of sequences; in humans, miRNAs were reported to bind preferentially to AT-rich 3′-UTRs [Proc Natl Acad Sci USA (2005) 102(43):15557–15562]. One suggestion was that the restriction of target sites to UTRs in mammalian mRNAs may result from the (mechanistic) inability of these miRNAs to bind to coding regions owing to the occurrence of active translation [Nature Struct Mol Biol (2009) 16:144–150].

In addition to a binding site in the 3′ UTR of mRNAs (that interacts largely with the 5′ end of the miRNA), many human miRNAs were reported to have another binding site in the 5′

UTR region; moreover, simultaneous interaction between an miRNA and *both* ends of (a given) mRNA was demonstrated in model systems. It was also reported that genes that respond significantly to the overexpression or absence of the relevant miRNAs form mRNAs which have the predicted interaction sites in both the 3′ and 5′ UTR regions [Genome Res (2009) 19:1175–1183].

At least some miRNAs in plants were reported to be almost perfectly matched to their (known) target sequences.

[Artificial miRNA-mediated antiviral activity in plants: J Virol (2007) 81(12):6690–6699.]

Roles of miRNAs in vivo

Many miRNAs have been identified, but not all have known function(s). Early papers reported that some have major roles in differentiation and/or development of cells – for example, in *Drosophila*, cardiac differentiation was reported to be influenced by miRNA1 [Proc Natl Acad Sci USA (2005) 102 (52):18986–18991], and maternal miRNAs were reported to be essential for the development of the mouse zygote [Genes Dev (2007) 21:644–648].

miRNAs are now linked to various functions – e.g. they are major players in self-renewal and differentiation in stem cells [Nature Rev Mol Cell Biol (2009) 10:116–125]. However, as well as roles in normal development/growth, they have been associated with disease [e.g. acute myeloid leukemia: Brit J Cancer (2009) 101:743–748]. (See also ONCOMIR.)

miRNAs encoded by viruses

The EPSTEIN–BARR VIRUS expresses a number of miRNAs in latently infected B cells. The level of expression of a given miRNA depends on factors such as the type of cell and the stage of latency [PLoS Pathogens (2006) 2(3):e23]. miRNA miR-BART2 down-regulates the (viral) DNA polymerase; in the viral lytic phase there is a reduction in the levels of this miRNA – consistent with the notion that miR-BART2 may be involved in inhibiting transition from the latent to the lytic state [Nucleic Acids Res (2008) 36(2):666–675].

[Epstein–Barr virus miRNAs in nasopharyngeal carcinoma: J Virol (2009) 83(5):2357–2367.]

As well as EBV, various viruses – including herpesviruses, polyomaviruses (such as the BK and JC viruses) and HIV-1 – encode miRNAs; host–virus interactions involving miRNAs have been reviewed [Nucleic Acids Res (2009) 37(4):1035–1048].

Studies on miRNAs, and miRNAs used as indicators

The formation of miRNAs in (living) cells of zebrafish and mouse was investigated by dual-fluorescence reporter/sensor plasmids [BioTechniques (2006) 41(6):727–732].

Expression profiling of miRNAs is often carried out with a MICROARRAY, and the identification of miRNA functions has been studied by combining expression profiling with target prediction [Nucleic Acids Res (2006) 34(5):1646–1652]. One study, involving expression profiling, found that a particular spectrum of miRNAs is associated with hypoxic conditions (conditions of low levels of oxygen); it was pointed out that many of the miRNAs expressed under these conditions are also overexpressed in various human tumors [Mol Cell Biol (2007) 27(5):1859–1867].

[miRNA profiling from low-input total RNA: RNA (2007) 13(1):151–159.]

Expression profiling was used for the identification of host target mRNAs for the miRNAs encoded by Kaposi sarcoma-associated herpesvirus [Nature Genetics (2008) 41:130–134].

MicroRNA *let-7b* was reported to target cell-cycle molecules in malignant MELANOMA cells; it was suggested that this may lead to possible targeted therapy of this tumor [Cell Res (2008) 18:549–557].

[MicroRNA-10b (positive regulator of metastasis in breast cancer): Nature (2007) 449:682–688.]

[miRNAs in metastasis: Nature Reviews Cancer (2009) 9: 293–302.]

The patterns of expression of miRNAs (detected in samples of serum) were used to discriminate between serum samples from cancer patients and those from normal individuals (with no need to use amplification techniques) [PLoS ONE (2009) 4(7):e6229].

MicroRNA-9 was reported to direct the development of the brain in embryonic zebrafish [Nature Neuroscience (2008) doi:10.1038/nn.2115].

In situ hybridization (ISH) studies on mouse tissue found a significant loss of miRNAs in formaldehyde-fixed samples; this loss was avoided by fixation with the reagent 1-ethyl-3-(dimethylaminopropyl) carbodiimide (EDC) which immobilizes miRNAs at their 5′ phosphate [Nature Methods (2009) doi: 10.1038/nmeth.1294]. However, studies using formalin-fixed, paraffin-embedded (FFPE) material (melanocytic nevi) for miRNA microarray profiling reported a high correlation in miRNA expression between paired FFPE and fresh-frozen material [J Investig Dermatol (2009) 129:1219–1224].

As multiple miRNAs may be involved in the regulation of a particular condition or event, one approach for studying such regulation is to use multiple-target antisense molecules, each able to bind a range of relevant miRNAs. Hence, a number of anti-miRNA oligonucleotides (AMOs) were incorporated in a single antisense fragment which was able to target (bind to) a number of the (relevant) miRNAs [Nucleic Acids Res (2009) doi: 10.1093/nar/gkn1053].

Online source of information on microRNAs

Details of miRNAs may be obtained at miRNAMap: mirnamap.mbc.nctu.edu.tw

MicroSAGE See SAGE.

microsatellite DNA In various genomes (e.g. human, bacterial, fungal): regions of DNA which consist of tandem repeats of *short* sequences of nucleotides; *short* has been interpreted (by different authors) as 1–5 bp [BMC Res Notes (2009) 2:17], <6 bp, 2–10 bp, 2 bp and 1–8 bp etc.

Microsatellite sequences vary in length (i.e. in total number of repeated subunits), even in closely related subjects, and are useful e.g. as genetic markers and in forensic profiling.

A given microsatellite sequence may exhibit *expansion* (i.e. increase in the *number* of tandemly repeated units) or *con-*

traction (involving deletion of nucleotides). Variation in the number of repeated units may arise e.g. by SLIPPED-STRAND MISPAIRING. *Instability* in microsatellite DNA may give rise to various types of disorder, such as HUNTINGTON'S DISEASE, MYOTONIC DYSTROPHY, FRAGILE X DISEASE and cancer. An analogous event causes the phenomenon of PHASE VARIATION in the OPA GENES of the bacterial pathogen *Neisseria gonorrhoeae*. [Advances in mechanisms of genetic instability in relation to hereditary neurological diseases: Nucleic Acids Res (2005) 33(12):3785–3798.]

(See also REPEAT-EXPANSION DETECTION.)

[Novel technique for identifying microsatellite markers in genomic DNA: BioTechniques (2007) 42(4):479–486.]

Microsatellite instability testing

Certain microsatellite sequences (such as BAT-26) are used as targets for monitoring microsatellite instability (MSI): see the entry MSI ANALYSIS.

microsatellite expansion See MICROSATELLITE DNA.

microsatellite instability analysis See MSI ANALYSIS.

Mimic^{TM} Sf9 insect cells A cell line, derived from SF9 CELLS and marketed by Invitrogen (Carlsbad CA), in which the cells have been genetically engineered to enable them to express various mammalian glycosyltransferases – thereby permitting production of recombinant proteins which are more similar to those produced in mammalian systems.

minibody A construct, engineered from fragments of antibody, prepared by fusing a SINGLE-CHAIN VARIABLE FRAGMENT (scFv) to the C_H3 domain of human IgG1; a minibody therefore has the structure scFv–C_H3. For the prototype minibody [see Cancer Res (1996) 56(13):3055–3061] two designs were evalulated for the link between scFv and C_H3 – (i) a 2-amino-acid linker, and (ii) the hinge region of human IgG1 together with an extra 10-mer peptide. The prototype minibody was shown to exhibit bivalent binding to carcinoembryonic antigen (CEA) and was used to demonstrate xenografts in mice.

minicell (1) The product of an abnormal cell division process in certain mutant rod-shaped bacteria (such as *Escherichia coli*); it consists of the polar part (end) of a cell lacking a chromosome. Minicells cannot grow. Nevertheless, they contain the machinery for transcription and translation and have been of use in genetic studies.

The minicell phenotype can be obtained experimentally e.g. with engineered *minC* mutations (which affect the position of the septum during the process of bacterial cell division) [e.g. J Bacteriol (2005) 187(8):2846–2857].

(cf. MAXICELL.)

(2) (minicell; MiniCell; Minicell etc.) A device (also called a French press) used for disrupting cells – for example, for preparing whole-cell lysates.

minichromosome (1) The intracellular form of the genome of certain DNA viruses (in animals and plants); the circular viral genome complexes with histones to form a structure similar to that of the host's histone-associated DNA. Minichromosomes formed by simian virus 40 (SV40) each contain about 25 *nucleosomes* (see entry CHROMATIN). Minichromosomes are also formed e.g. by the cauliflower mosaic virus and the Epstein–Barr virus.

(2) In *Tetrahymena thermophila*: a stage in the development of the macronuclear rRNA-encoding sequence following its excision from the micronucleus [term used in: Nucleic Acids Res (2006) 34(2):620–634].

(3) A plasmid in which the sole origin of replication is one derived from a bacterial host.

minichromosome maintenance protein 1 See the entry ARS.

minicircles See KINETOPLAST.

minigene An *in vitro* construct (commonly part of a plasmid vector) which contains certain features of a typical eukaryotic gene, such as a promoter/polyadenylation region, exonic and intronic sequences, and a polylinker (multiple cloning site) at which a fragment of sample DNA can be inserted.

Minigenes are used e.g. in EXON TRAPPING.

minimal medium See the entry AUXOTROPHIC MUTANT.

mini-Mu A defective PHAGE MU (q.v.) in which the DNA lacks certain sequences but which has intact terminal regions (the sites used for transposase binding); these phage particles can infect susceptible cells, and may be replicated (and matured into progeny particles) with intracellular help from a normal, wild-type phage Mu (which may e.g. supply a transposase).

In a mixed infection of bacteria – with mini-Mu and wild-type Mu – DNA fed into some of the particles (by the headful mechanism) contains a greater proportion of *bacterial* DNA; these particles will have a greater ability to mediate *generalized* TRANSDUCTION.

minor groove (in DNA) See GROOVES.

minus strand (of a gene) See CODING STRAND.

MIP genotyping See MOLECULAR INVERSION PROBE GENOTYPING.

miR-17–miR-92 cluster See the entry ONCOMIR-1.

MIRAGE Mutagenic inverted-repeat-assisted genome engineering: SITE-DIRECTED MUTAGENESIS, *in vivo* (within cells), with a mutagenesis cassette containing an INVERTED REPEAT flanked by sequences which mediate homologous recombination at the desired site within the genome of *Saccharomyces cerevisiae*.

The central, major part of the ~2.2 kb mutagenesis cassette consists of two copies of the *URA3* gene back-to-back (i.e. in opposite orientations), forming the inverted repeat. (The role of *URA3* in the yeast *Saccharomyces cerevisiae* is described in the entry YEAST TWO-HYBRID SYSTEM.) The two (45 bp) flanking regions of the cassette direct the cassette's insertion, by homologous recombination, into the required site in the target genome. One of the flanking regions includes the sequence which is to replace a specific sequence in the genome.

Insertion of the cassette into the target genome replaces the unwanted genomic sequence – and converts the (auxotrophic) yeast strain to prototrophy, allowing selection of (cassette-containing) prototrophs. Subsequently, the relative instability of the inverted repeat in the genome leads to its self-excision, leaving the replacement sequence in the genome, and causing the yeast to revert to auxotrophy. The (recombinant) auxo-

trophs can then be selected.

[Method: Nucleic Acids Res (2009) 37(1):e9.]

miRNAs See the entry MICRORNAS.

mirtron An intermediate in the biogenesis of *some* MICRORNAS which is formed, directly, by splicing out from the primary miRNA (pri-miRNA); thus, unlike other pre-miRNAs, formation of a mirtron does not involve the Microprocessor (the Drosha-DGCR8 complex). Like other pre-miRNAs, mirtrons undergo further processing to form mature miRNAs.

Mirtrons occur in mammals as well as in various invertebrates [Mol Cell (2007) 28(2):328–336].

MIRU Mycobacterial interspersed repetitive unit.

(See also W-BEIJING STRAIN.)

mismatch extension (of a primer) The inefficient extension of a primer with a 3′ terminal mismatched base. The nature of the mismatch influences the readiness with which extension can occur; a G/T mismatch is reported to be one of the least inhibitory mismatches.

The above refers to primer extension by routine (replication or 'high-fidelity') polymerases. Extension from mismatched bases is reported to be facilitated by the use of a BYPASS DNA POLYMERASE.

mismatch repair (MMR) A system involved in the response to (i) mismatched nucleotides, incorporated during replication of DNA, and (ii) (in e.g. humans) DNA damage (as well as mismatched nucleotides) – see below.

Mismatched nucleotides which escape proofreading by the DNA polymerase can be corrected by a mismatch repair system that operates in both prokaryotes and eukaryotes.

MMR in Escherichia coli

In *Escherichia coli*, the incorporation of a mismatched nucleotide during DNA replication can be detected by the MutS protein. (Using a direct, electrochemical method, MutS was found to interact with a mismatch more strongly than with a normally base-paired site [Nucleic Acids Res (2006) 34(10): e75].) During interaction between MutS and a mismatch, a hydrogen bond is formed between a glutamate residue in MutS and the mismatched base; this interaction promotes the triggering of downstream events [EMBO J (2006) 25(2):409–419].

Detection of a mismatch by MutS is followed by activation of an endonuclease, MutH. MutH cleaves the new daughter strand, identified by its initial state of undermethylation, at an unmethylated GATC, a so-called 'Dam site' (see DAM GENE); the cleavage site may be some distance from the mismatch, and co-ordination between the mismatch and cleavage sites seems to involve the MutL protein. [Identification of a site of interaction between MutH and the (C-terminal) of MutL: Nucleic Acids Res (2006) 34(10):3169–3180.] (In that Dam sites are involved in MMR in *E. coli*, this system has been referred to as *Dam-directed mismatch repair*, DDMR, in this organism.)

After strand cleavage by MutH, the sequence of nucleotides between the cleavage site and a site on the other side of the lesion is removed (apparently in a process involving helicase II) and replaced by a sequence that is newly synthesized on the template.

A strain defective in MMR is associated with a mutator phenotype.

Interestingly, *dam* mutants of *E. coli* (cells unable to carry out Dam methylation of their DNA) are killed by an MMR-related mechanism when treated with 2-aminopurine – a base analog which pairs with thymine and cytosine, triggering the MMR response; in these cells the DNA becomes totally degraded [J Bacteriol (2006) 188(1):339–342].

The role of MutS in humans

In humans, MutS has an *additional* role: it is involved in the initiation of cell death following damage to DNA; a defect in this role can promote carcinogenesis. In the presence of DNA damage the role of MutS differs from its role in the context of mismatched nucleotides, and the mechanism by which DNA damage is recognized has been studied with MutS homologs [Nucleic Acids Res (2006) 34(8):2173–2185].

(See also DNA REPAIR.)

mis-sense mutation Any mutation in which a codon specifying a given amino acid is changed to one which specifies a different amino acid.

mis-sense suppressor See SUPPRESSOR MUTATION.

mitochondrial DNA (mtDNA) DNA associated with the mitochondrion (as opposed to the nucleus). Human mtDNA (∼16 kbp) derives from the female parent [diversity of mtDNA in humans: Nucleic Acids Res (2007) 35(9):3039–3045].

(See also DNA BARCODING and NUMTS.)

Details of the mitochodrial DNA have been reported from various other species: e.g. the complete sequence of mtDNA in zebrafish (*Danio rerio*) (∼16.6 kbp) [Genome Res (2001) 11(11):1958–1967] and in the teleost fish *Scleropages formosus* (Osteoglossidae) (∼16.6 kbp) [BMC Genomics (2006) 7:242].

mtDNA is examined e.g. in studies on population genetics and in certain FORENSIC APPLICATIONS of DNA technology. Optimization of the primers used for analyzing mtDNA has been suggested on the basis of known polymorphisms which are common in various racial groups [BioTechniques (2008) 44(4):559–562].

Information on mtDNA (e.g.): a US government database (the mtDNA Popstats Population Database); and HVRBASE++. [Enhanced MITOMAP with global mtDNA mutational phylogeny: Nucleic Acids Res (2007) 35(Database issue): D823–D828.]

(See also CYBRID and HETEROPLASMY.)

mitochondrial pseudogenes See NUMTS.

(See also FORENSIC APPLICATIONS.)

mitomycin C An antibiotic, produced by *Streptomyces* spp, in which the molecule is characterized by a quinone group and an aziridine ring. *In vivo* activity depends on intracellular reduction to a hydroquinone derivative which e.g. forms covalent cross-links between the strands of dsDNA.

Mitomycin C has been used e.g. for inducing a lysogenic population (see LYSOGENY) and for CURING plasmids.

***MLN64* gene** See GENE AMPLIFICATION.

MLPA See MULTIPLEX LIGATION-DEPENDENT PROBE AMPLIFICATION.

MLST (multilocus sequence typing) A method of TYPING, used for various (commonly) pathogenic microorganisms, which is designed to allow the long-term and global tracking of (e.g.) antibiotic-resistant and virulent strains – with dissemination of information on the Internet.

In this approach, strains of a given pathogen are characterized/classified according to nucleotide sequences in *specific* alleles, this type of categorization generally being regarded as stable for long periods of time.

Originally used for typing *bacterial* pathogens, MLST has also been used for the fungi *Candida* [Eukaryotic Cell (2008) 7(7):1075–1084; J Clin Microbiol (2008) 46(2):652–664] and *Batrachochytrium* [PLoS Pathog (2009) 5(5):e1000458].

[Molecular evolution of enteropathogenic strains of *E. coli* (EPEC) (clonal analysis by MLST): J Bacteriol (2007) 189 (2):342–350.]

MLST has been used for typing *Streptococcus pneumoniae* [Microbiology (2006) 152:367–376]; *Staphylococcus aureus* [Emerg Infect Dis (2009) 15:766–769, PLoS ONE (2009) 4(7):e6216]; *Helicobacter pylori* [BMC Microbiol (2009) 9: 126] and *Escherichia coli* [BMC Genomics (2009) 10:296].

[MLST typing of the bacterium *Wolbachia pipientis*: Appl Environ Microbiol (2006) 72(11):7098–7110; see database at http://pubmlst.org/wolbachia/]

MLST has also been used to study the population genetic structure of strains of (non-pathogenic) nodulating rhizobia [J Bacteriol (2006) 188(15):5570–5577].

MLV Murine leukemia virus.

MLVA Multiple loci VNTR analysis (in which 'VNTR' means VARIABLE NUMBER OF TANDEM REPEATS). In this method – which is used e.g. for TYPING bacteria – strains are compared on the basis of loci containing tandemly repeated sequences; strains may differ in the number of tandem repeats at a given locus and/or in the number of such loci.

The number of repeated sequences at a given locus can be conveniently investigated by PCR, the size of products from a given locus indicating how many copies of the basic unit are present at that locus.

(cf. REPEAT-EXPANSION DETECTION.)

Strains of *Bacillus anthracis* – indistinguishable by routine MLVA procedures – were differentiated by single-nucleotide repeats (SNRs) [J Clin Microbiol (2006) 44(3):777–782].

[MLVA used in the molecular characterization of *Coxiella burnetii*: BMC Microbiol (2006) 6:38.]

[MLVA used for the subtyping of *Neisseria meningitidis*: BMC Microbiol (2006) 6:44.]

[MLVA analysis of clusters of *Clostridium difficile* strains: J Clin Microbiol (2008) 46(3):954–960.]

[Typing *Clostridium difficile* on the basis of tandem repeat sequences: BMC Microbiol (2009) 9:6.]

[MLVA typing of *Staphylococcus aureus* (comparison with PFGE and *spa*-typing): PLoS ONE (2009) 4(4):e5082.]

[Genotyping, and phylogenetic analysis, of *Yersinia pestis* (the causal agent of plague) by MLVA: PLoS ONE (2009) 4(6):e6000.]

MmeI A type IIG restriction endonuclease from *Methylophilus methylotrophus* (details of the recognition/cutting sites are in the table in the entry RESTRICTION ENDONUCLEASE).

MmeI finds use e.g. in the preparation of shRNA libraries (see entry SHORT HAIRPIN RNA) and in LongSAGE (see entry SAGE).

[MmeI: Nucleic Acids Res (2008) 36(20):6558–6570.]

[The MmeI family (RM enzymes that employ single-strand modification for host protection): Nucleic Acids Res (2009) doi: 10.1093/nar/gkp534.]

MMF Mycophenolate mofetil: a prodrug which is conveted to the active agent MYCOPHENOLIC ACID.

MMLV reverse transcriptase Moloney murine leukemia virus reverse transcriptase – an enzyme used e.g. in FIRST STRAND synthesis.

MMR (1) MISMATCH REPAIR.

(2) Measles–mumps–rubella (vaccine).

MMS Methylmethane sulfonate: an alkylating agent which can e.g. methylate guanine residues in DNA.

MNNG *N*-methyl-*N*′-nitro-*N*-nitrosoguanidine (see entry DNA REPAIR).

MNU *N*-methyl-*N*-nitrosourea (see entry DNA REPAIR).

mobile genetic element Any nucleic acid sequence which can re-locate to a different site, or insert a copy of itself at a new location; such elements occur e.g. in chromosomes, plasmids and viral genomes.

Examples of mobile genetic elements: insertion sequences, some types of intron (see the entry INTRON HOMING), retrotransposons, transposons (including conjugative transposons).

(See also TRANSPOSABLE ELEMENT.)

mobilization (of chromosomes) See e.g. CONJUGATION and F PLASMID.

modification (*bacteriol.*) The *in vivo* process in which bacteria routinely add a methyl (CH_3–), or other, group to particular nucleotides in specific sequences of their DNA; modification is usually carried out shortly after synthesis of the DNA. The methylation is carried out by specific DNA METHYLTRANSFERASES, one example of which (in *Escherichia coli*) is the product of the DAM GENE (q.v.).

Modification apparently protects DNA from the cell's own RESTRICTION ENDONUCLEASES; most of these enzymes cut sequences that lack the specific methylation pattern produced by the cell's own methyltransferases. Restriction enzymes can usually degrade 'foreign' (i.e. imported) DNA (such as phage DNA) which lacks the cell's specific pattern of methylation.

Cooperation between these two types of enzyme is referred to as the *restriction–modification system* (R–M system). Note that, in some cases, both the methylating and restriction functions are carried out by the same enzyme.

Extensive data are available on components of the restriction–modification system [REBASE – enzymes and genes for

DNA restriction and modification: Nucleic Acids Res (2009) doi: 10.1093/nar/gkp874].

modification methylase A trivial name used for any enzyme that methylates particular bases in DNA.

MODOMICS A database of RNA MODIFICATION PATHWAYS – q.v.

modulator protein See the entry TN21.

MOI MULTIPLICITY OF INFECTION.

molecular beacon probe A PROBE which consists of a single-stranded linear oligonucleotide in which the complementary *ends* hybridize, forming a stem–loop structure. The sequence in the *loop* is complementary to the probe's target sequence. Within the stem, one of the strands carries a fluorescent dye and the other contains a quencher of fluorescence.

The *unbound* probe is non-fluorescent.

If the *loop* region binds to a target sequence, the two ends of the probe separate and the quencher ceases to be effective – so that the dye fluoresces.

These probes were reported to have greater *specificity* than other probes of equivalent length [Proc Natl Acad Sci USA (1999) 96:6171–6176]. They are used for real-time monitoring of nucleic acid amplification, and also (e.g.) for detecting specific nucleic acids in solution by flow cytometry [Nucleic Acids Res (2005) 33(2):e13].

A novel molecular beacon probe was developed by the use of the *intrinsically* fluorescent nucleotides 2-aminopurine and pyrrolo-dC [Nucleic Acids Res (2006) 34(6):e50].

To improve the sensitivity of detection of molecular beacon probes, use has been made of a signal amplification scheme involving target-dependent cleavage of molecular beacons by a DNA nicking enzyme. In this scheme, a probe is cut by the nicking enzyme when it binds to the target sequence; when the probe is cut the fluorophore remains unquenched. Hence, as a large number of probes can be cut – one after the other – when they bind to a single target sequence, this leads to high-level amplification of the fluorescent signal [Nucleic Acids Res (2008) 36(6):e36].

[Detection of Lyme spirochetes (*Borrelia burgdorferi*) with sensitive and specific molecular beacon probes: BMC Microbiol (2009) 9:43.]

(See also PYRENE BINARY PROBE and REAL-TIME PCR.)

molecular cloning See CLONING.

molecular combing See DNA COMBING.

molecular epidemiology Epidemiology involving nucleic-acid-based TYPING of isolates of a given pathogen or pathogens.

molecular inversion probe genotyping (MIP genotyping) A technique for SNP GENOTYPING or, in general, for obtaining information on targeted genomic loci.

The molecular inversion probe (MIP) technique is based on the use of a type of linear, oligonucleotide probe whose *end*-sequences are complementary to two *adjacent* genomic sequences that are separated by only a single nucleotide; that is, when bound to the target, the probe is circularized except for a one-nucleotide break. For a given probe, the one-nucleotide break corresponds to the exact site of *one* of the SNPs that are being assayed. In any given assay, thousands of different oligonucleotide probes are used – each type of probe being highly specific for one of the SNPs being assayed.

The one-nucleotide break in each (bound) probe is filled in by polymerase/ligase activity during one of the *four separate reactions* – each reaction involving *one* of the four types of dNTP. In a given reaction, all those probes whose break can be filled by the particular dNTP supplied will be completed and ligated, i.e. circularized.

The probes are released from genomic DNA, and the fully (covalently) circularized probes are amplified by PCR. All of the probes – regardless of their specificity – contain identical primer-binding sites for PCR, so that a universal primer pair can amplify all the circularized probes.

In addition to primer-binding sites, each probe includes a *unique* 'tag' sequence *which indicates its target specificity*, i.e. a given tag indicates a specific SNP at a given locus in the genomic DNA.

The PCR products are digested with restriction enzymes in order to release the (labeled) tag sequences. The tags are then detected by an array of probes (see MICROARRAY) which are complementary to the tag sequences. Detection of a given tag indicates the presence of the corresponding SNP in the genomic DNA.

The system of unique tag sequences, described above, has been termed *barcoding* (compare with ZIPCODE ARRAY). The large number (thousands) of unique sequence tags used in a given assay are designed to have minimal similarity, and tag design also aims to avoid problems associated with secondary structures. Furthermore, the tags are designed so that the overall range of melting temperatures is narrow.

[Use of method for the multiplexed detection of pathogens: PLoS ONE (2007) 2(2):e223.]

A MIP approach was also used for genome-wide studies on SNPs in (bovine) genomes [BMC Genomics (2009) 10:176], and for studying genotype data from FFPE (formalin-fixed, paraffin-embedded) samples [BMC Med Genomics (2009) 2: 8].

molecular zipper A construct used in a method for the *in vitro* isothermal amplification of DNA.

The sequence of nucleotides to be copied is initially present as an ssDNA circle. A forward primer binds to the circle and is extended by the ROLLING CIRCLE mechanism, producing a long strand of DNA consisting of many tandem repeats of the target sequence.

Multiple molecular zippers bind to the ssDNA multimer. Each molecular zipper consists of a 48-nt 'positive' strand of DNA, labeled with a quencher of fluorescence, hybridized (at its 5′ end) to a 25-nt 'negative' strand of DNA labeled with a fluorophore; in this construct, fluorescence from the fluorophore is quenched. The 23-nt, single-stranded 3′ terminal part of the positive strand acts as a reverse primer and binds to the ssDNA target multimer.

The 3′ ends of the (bound) molecular zippers are extended by DNA polymerases – resulting in the displacement of other

(bound and extended) *downstream* molecular zippers. Then, *forward* primers bind to the displaced zippers. Extension of a forward primer displaces the negative strand from a zipper, thus releasing its fluorophore from quenching by the positive strand. Ongoing amplification thus gives rise to an increasing level of fluorescence as there is an ongoing increase in the number of displaced negative strands (with active, i.e. non-quenched, fluorophores).

[Method: Nucleic Acids Res (2006) 34(11):e81.]

Moloney murine leukemia virus reverse transcriptase See e.g. FIRST STRAND.

monoclonal antibodies (mAbs) Any population of antibodies derived from a *specific clone* of B lymphocytes (B cells); all the antibodies obtained from a given clone of B cells recognize the same antigenic determinant(s).

Monoclonal antibodies can be obtained from a HYBRIDOMA and are used e.g. for the study and identification of specific macromolecules.

(See also APTABODY.)

monomeric red fluorescent protein (mRFP1) A monomeric derivative of DsRed – a (naturally tetrameric) red fluorescent protein derived from the coral *Discosoma* [Proc Natl Acad Sci USA (2002) 99(12):7877–7882]. Excitation and emission wavelengths for mRFP1 are 584 and 607 nm, respectively.

mRFP1, tagged with a nuclear localization signal, NLS, has been used to check for intracellular stress: localization of the mRFP1 to the cytoplasm, rather than uptake into the nucleus, indicates a problem with nuclear import.

Monomeric red fluorescent protein has been used e.g. as a sensor in the DFRS PLASMID (q.v.).

[Other uses of mRFP1 (e.g.): PLoS ONE (2009) 4(4):e5041 and PLoS ONE (2009) 4(5):e5727.]

(See also RED FLUORESCENT PROTEINS; cf. GREEN FLUORESCENT PROTEIN and YELLOW FLUORESCENT PROTEIN.)

mononucleotide repeat *Syn*. single-nucleotide repeat: see SNR.

monopartite 'One-piece' – refers e.g. to the genome of a virus that consists of a single piece of nucleic acid. Viruses with a monopartite genome include e.g. the herpesviruses (dsDNA), parvoviruses (ssDNA), paramyxoviruses (ssRNA, negative-sense) and picornaviruses (ssRNA, positive-sense).

(cf. BIPARTITE and MULTIPARTITE.)

monosomic See CHROMOSOME ABERRATION.

Morbillivirus A genus of negative-sense ssRNA viruses in the family PARAMYXOVIRIDAE (q.v).

morgan A unit of length on a eukaryotic chromosome; it is the length within which (on average) one crossover occurs in each meiosis.

morpholino antisense oligomer Any of a range of (synthetic) non-ionic, oligonucleotide analogs in which the pyrimidine and purine bases are carried on a backbone of morpholine residues linked by phosphorodiamidate bonds (morpholine = tetrahydro-1,4-oxazine). These oligomers are typically 10–20 morpholine residues in length; they are resistant to nucleases and are used e.g. for GENE SILENCING (see KNOCKDOWN).

Phosphorodiamidate morpholino oligomers (PMOs) bind to mRNAs by Watson–Crick base-pairing and block translation; this duplex is resistant to RNase H.

[Antisense PMOs (the effect of length etc. on gene-specific inhibition): Antimicrob Agents Chemother (2005) 49(1):249–255.]

PMOs have been used e.g. for studying antiviral activity [J Virol (2007) 81(11):5637–5648]. They have also been used to induce exon skipping in the *DMD* gene (which is defective in Duchenne muscular dystrophy); this yielded an internally truncated, but functional, gene product [Mol Therapy (2009) doi: 10.1038/mt.2008.287].

Morpholino oligomers have been used e.g. for knockdown in zebrafish [Blood (2009) 113(19):4754–4762] and *Xenopus* [PLoS ONE (2009) 4(5):e5742].

morula (*embryol*.) A compact ball of cells formed by cell division) in a fertilized oocyte; on further development it gives rise to a BLASTOCYST (q.v.).

Mos1 See MARINER FAMILY.

mosaic *Syn*. CHIMERA.

mosquitoes (genetically modified) See GMMS.

(See also ARBOVIRUSES.)

moxifloxacin See QUINOLONE ANTIBIOTICS.

MPA MYCOPHENOLIC ACID.

***mpl* gene** In *Escherichia coli*: the gene encoding a ligase involved in the recycling of tripeptide units during turnover of the cell wall polymer peptidoglycan.

M_r RELATIVE MOLECULAR MASS.

MreB protein See NUCLEOID.

mRFP1 See MONOMERIC RED FLUORESCENT PROTEIN.

MRI (magnetic resonance imaging) An approach based on the phenomenon of *nuclear magnetic resonance* in which certain types of atomic nuclei – placed in a strong magnetic field and exposed to suitable radio-frequency electromagnetic radiation – emit characteristic, informative signals; the nature of these signals varies according e.g. to the molecular environment in which the nuclei occur – and to the concentration of the given nuclei.

MRI was used e.g. for monitoring gene expression in the yeast *Saccharomyces cerevisiae* [Nucleic Acids Res (2006) 34:e51].

[Visualization of transgene expression in experimental tumors *in vivo* with optical imaging and MRI: Gene Ther (2009) 16:830–839.]

mRNA (messenger RNA) A molecule of RNA in which the nucleotide sequence specifies the sequence of amino acids in a given polypeptide or protein (see CODON); the codons in an mRNA guide the sequential assembly of amino acids during protein synthesis on a ribosome.

In *prokaryotic* organisms the RNA transcribed from a DNA template is commonly functional, as mRNA, with no need for post-transcriptional maturation. In *eukaryotic* organisms, the PRIMARY TRANSCRIPT is commonly a precursor, pre-mRNA, which has to undergo *maturation*, including: (i) capping (see CAP), (ii) POLYADENYLATION, and (iii) SPLICING (to remove intron(s): see SPLIT GENE). (The development of certain eu-

karyotic mRNAs is atypical: for example, most or all mRNAs encoding HISTONES derive from intron-less genes, and they are not polyadenylated.)

Monocistronic mRNA contains the coding sequence of *one* gene. (cf. POLYCISTRONIC MRNA.)

Bacterial mRNAs typically have a short half-life; however, in *Escherichia coli* the half-life of at least some mRNAs was reported to be significantly longer in slow, anaerobic growth. Some bacterial (and archaean) mRNAs carry a 3′ poly(A) tail which is synthesized by a poly(A) polymerase [details of *E. coli* poly(A) polymerase: Nucleic Acids Res (2000) 28:1139–1144]; it has been suggested that such polyadenylation may facilitate degradation of the molecule.

Eukaryotic mRNAs are synthesized in the nucleus by RNA polymerase II and pass to the cytoplasm for translation. In the cytoplasm, mRNAs are usually complexed with proteins (see MRNP) and (compared with bacterial mRNAs) are relatively stable.

(See also SELENOPROTEIN MRNAS.)

mRNAs in DNA technology

The presence of mRNAs in a cell is a convenient and detectable signal that specific gene(s) are being (or were) transcribed; hence, mRNAs are frequently assayed to obtain a GENE EXPRESSION PROFILE – often by using a MICROARRAY in which the probes are designed to detect all, or a subset, of the mRNAs from a given type of cell under given conditions.

mRNAs are natural targets for gene-silencing systems (see e.g. MICRORNAS and RNA INTERFERENCE), and they are also targets for the *engineered* knockdown/gene-silencing systems (see e.g. MORPHOLINO ANTISENSE OLIGOMERS and SHORT HAIRPIN RNA).

(Some relevant entries: ALTERNATIVE SPLICING, CAP, CAP TRAPPER, CDNA, EXPRESSED SEQUENCE TAG, FIRST STRAND, FMRP, LIBRARY, MICRORNAS, PRIMER EXTENSION ASSAY, RIBONUCLEASE PROTECTION ASSAY, RIBOSWITCH, SAGE, 3′-UTR, 5′-UTR.)

mRNP (informosome) Messenger ribonucleoprotein: a body, found in the cytoplasm of eukaryotic cells, which consists of a molecule of mRNA complexed with proteins – one of the proteins being Y-box-binding protein 1 (see YB-1 PROTEIN).

***mrr* gene** In *Escherichia coli*: a gene which encodes a type IV RESTRICTION ENDONUCLEASE reported to cleave DNA at the sequences 5′-C^{m}AG-3′ and 5′-G^{m}AC-3′ (in which the adenine residue is methylated).

Cells containing inactivated *mrr* apparently permit greater efficiency when cloning DNA containing methylated adenine residues in the above sequences.

(See also ESCHERICHIA COLI (table).)

MRSA Methicillin-resistant *Staphylococcus aureus*: the various strains of *S. aureus* that are resistant to at least METHICILLIN – and are generally resistant to a range of other antibiotics (in some cases including vancomycin); MRSA causes problems in a clinical setting owing to the reduced range of effective chemotherapeutic agents that are available for the treatment of infections involving these strains. [Evolutionary history of MRSA: Proc Natl Acad Sci USA (2002) 99(11):7687–7692.]

In *S. aureus*, resistance to methicillin can arise in different ways. Most attention has been paid to resistance conferred by the MECA GENE; *mecA* encodes a penicillin-binding protein (PBP 2a) with low affinity for methicillin (and other β-lactam antibiotics), allowing ongoing synthesis of the cell wall polymer peptidoglycan (and hence, growth) in otherwise inhibitory concentrations of these antibiotics.

A mobile genetic element SCC*mec*, which incorporates the *mecA* gene, is used for TYPING isolates of MRSA.

[Modeling of the bacterial mechanism of methicillin resistance by a systems biology approach: PLoS ONE (2009) 4(7): e6226.]

The resistance encoded by *mecA* depends on the formation of normal products from the *fem* genes (*fem* refers to *f*actor *e*ssential for *m*ethicillin resistance). Mutations in the *femAB* operon can result in the formation of abnormal cross-links in peptidoglycan; these abnormal cross-links are poor substrates for the enzymic role of PBP 2a but are better substrates for the normal, methicillin-sensitive PBPs – so that cells with the mutant *fem* genes are sensitive to methicillin.

Community-acquired MRSA (CA-MRSA) refers to strains of MRSA that are distinct from the 'hospital-acquired' MRSA associated with a stay in hospital. CA-MRSA appeared some years ago, within communities, and these strains can now be found world-wide; they may cause severe disease in various groups of individuals (including the young and healthy) who have had no apparent connection with a hospital or a healthcare environment.

(See also PANTON–VALENTINE LEUCOCIDIN.)

The prospect that CA-MRSA may become entrenched, or dominant, within hospitals has raised concern; a model of this scenario has been presented [modeling invasion of hospitals by CA-MRSA: Clin Infect Dis (2009) 48(3):274–284].

(See also GISA, PFGE and SCCMEC.)

MRSE Methicillin-resistant *Staphylococcus epidermidis*.

mRuby A RED FLUORESCENT PROTEIN with an excitation maximum of 558 nm and an emission maximum of 605 nm [characteristics: PLoS ONE (2009) 4(2):e4391].

MSBE See METHYLATION-SPECIFIC SINGLE-BASE EXTENSION.

msDNA See RETRON.

MSI analysis Microsatellite instability analysis: comparison of DNA from a tumor with DNA from 'normal' (control) tissues – of the same individual – with respect to a number of microsatellite marker sequences. (See MICROSATELLITE DNA.) The object is to determine any *instability* (changes in the number of repeats in a given sequence) in each of the microsatellite markers in the tumor DNA. If only one of the (usually five or more) markers in the tumor DNA is found to be unstable, the tumor is designated MSI-L. If two or more markers are found to be unstable, the tumor is designated MSI-H.

Unstable microsatellite DNA sequence(s) may indicate e.g. defective component(s) of the MISMATCH REPAIR system

(see also entry SLIPPED-STRAND MISPAIRING). Microsatellite instability may suggest e.g. the need to sequence MMR genes.

Essentially, in the test, the selected microsatellite sequences (e.g. BAT-26) in each sample of (tumor and control) DNA are amplified by PCR, and the products examined by gel electrophoresis. Each microsatellite sequence from the tumor is then compared with its corresponding sequence in normal DNA, and the results from all the microsatellite loci are scored.

[Example of MSI analysis: Mol Cancer (2008) 7:68.]

MSI-H See MSI-ANALYSIS.

MSI-L See MSI ANALYSIS.

MSP See METHYLATION-SPECIFIC PCR.

MSRE Methylation-sensitive restriction enzyme.

MSSA Methicillin-sensitive *Staphylococcus aureus*.

M.SssI See the entry SSSI.

MTase DNA METHYLTRANSFERASE.

mtDNA MITOCHONDRIAL DNA.

mtDNA Popstats Population Database See entry FORENSIC APPLICATIONS.

Mu phage See PHAGE MU.

mucolytic agent Any agent used for liquefying specimens of (mucoid) sputum. Mucolysis facilitates manipulation of the specimen and also serves to expose cells, including bacteria. Mucolytic agents – e.g. *N*-acetyl-L-cysteine and dithiothreitol – are used e.g. to prepare samples of sputum for nucleic-acid-based tests for *Mycobacterium tuberculosis* (e.g. the AMTDT test).

mucopolysaccharidosis type IH (Hurler's syndrome) An autosomal recessive disease resulting from a mutant allele in the *IDUA* gene, encoding the lysosomal enzyme α-L-iduronidase. The disease affects various organs, including the heart; death may result from congestive heart failure.

(See also GENETIC DISEASE (table).)

multicopy inhibition (Tn*10*) The increased level of inhibition of transposition of a (single-copy) chromosomal Tn*10* in the presence of a multicopy plasmid which contains the insertion sequence IS*10*. Such inhibition appears to be due to the effect of an antisense RNA, encoded by IS*10*R (see the entry TN10), which binds to the mRNA of the transposase, thus blocking the expression of the transposase at the level of translation. It seems that only the outer 180 base pairs of IS*10*R are needed for this inhibitory effect.

Tn*10* transposition is also regulated e.g. by adenine methylation (see DAM GENE).

multicopy plasmid Any PLASMID whose COPY NUMBER (in a given type of cell) – e.g. >10 copies per cell – is in contrast to that of low-copy-number plasmids (such as the F PLASMID) in which the copy number may be 1 or 2.

Examples of multicopy plasmids include ColE1 (typically 10–30 copies per cell) and the PUC PLASMIDS.

The copy number of ColE1 was reported to be decreased in *Escherichia coli* host cells with a mutant *pcnB* gene – which encodes the enzyme poly(A) polymerase I (PAP I); in these cells, the absence of polyadenylation of the (ColE1-encoded) antisense molecule RNA I has the effect of stabilizing this molecule, enhancing its inhibitory action on RNA II and thus inhibiting replication of the plasmid.

multicopy single-stranded DNA (msDNA) See RETRON.

multilocus sequence typing See MLST.

multilocus VNTR analysis *Syn*. multiple loci VNTR analysis: see the entry MLVA.

multipartite Refers e.g. to the genome of a virus that consists of more than two pieces of nucleic acid.

Viruses with a multipartite genome include the orthomyxoviruses (ssRNA, negative-sense), such as the influenzavirus types A and B (in which the genome consists of eight pieces). (cf. MONOPARTITE and BIPARTITE.)

multiple cloning site (MCS) *Syn*. POLYLINKER.

multiple displacement amplification Isothermal amplification of DNA (~30°C) in WHOLE-GENOME AMPLIFICATION. Many sites are primed by random (exonuclease-resistant) hexamers, and the replication is mediated by a highly processive DNA polymerase from bacteriophage φ29 (omitting an initial stage of heat denaturation) [Proc Natl Acad Sci USA (2002) 99: 5261–5266]. Average product length: >10 kb. Amplification bias was reported to be less than that in PCR-based methods, and the products are said to be useful for genomic studies and for SNP analysis.

Under-representation of terminal sequences in linear DNA may be overcome by ligation (forming a circular template) [BioTechniques (2005) 39(2):174–180].

Multiple displacement amplification has been used e.g. for amplifying low-levels of DNA in sources such as samples of plasma stored at −40°C for ~10 years [BioTechniques (2005) 39(4): 511–515]. MDA has also been used to assess the value of whole-genome amplification for preparing sufficient DNA for use in forensic SNP genotyping [BMC Genomics (2009) 10:159].

multiple loci VNTR analysis See MLVA.

multiplex amplifiable probe hybridization (MAPH) A technique used for assessing the copy number of target sequences and detecting deletion mutations. Essentially, genomic DNA is denatured and exposed to a range of probes. The probes are complementary to various target sequences in the genome – but all the probes have identical terminal (5′ and 3′) primer-binding sites.

The probes are allowed to hybridize with the genomic DNA (which is immobilized on a filter). Any unbound probes are removed by washing. The bound probes are then recovered (separated from genomic DNA) and amplified by PCR using a primer pair which is common to all the probes. The quantity of PCR product obtained from a given type of probe is assumed to correlate with the number of such probes present in the reaction mixture, thereby giving an indication of the copy number. A probe which is complementary to a deleted sequence in the genome is likely to yield no product by PCR; similarly, no products are likely to be obtained if the probe's target sequence has undergone a mutation which prevents binding of the probe.

[Use of MAPH for assay of copy number of the β-defensin gene: PLoS ONE (2009) 4(3):e4725.]

multiplex ligation-dependent probe amplification (MLPA) A technique used for *relative* quantitation of specific sequences of nucleotides in genomic DNA. The applications of MLPA include detection of, for example, Down's syndrome (trisomy 21), gene duplications, exon deletions and SNPs.

Probes are added to a sample, each probe consisting of two oligonucleotides that bind *adjacently* on the target sequence. If both oligonucleotides hybridize correctly on the target sequence they can be *ligated* and the (now complete) probe can be amplified by PCR (each probe includes two primer-binding sites). Quantitation of the PCR products can be facilitated by using labeled primers in the PCR reaction and measuring the signal. Products obtained from the target sequence, in a given sample, can be compared, quantitatively, with products from the corresponding target sequence in other samples; this can be done by comparing the signals obtained from the different samples.

MLPA can be used (in a multiplex mode) for simultaneous assessment of copy number of a range of different genes by using multiple types of probe (e.g. up to forty different types of probe) in a single assay.

[Example of use (assay of gene amplification (i.e. increased copy number) of HER2/*neu*): BMC Cancer (2009) 9:4.]

[MLPA used for detecting copy number differences: BMC Bioinformatics (2008) 9:261.]

(cf. COMPARATIVE GENOMIC HYBRIDIZATION.)

multiplex PCR A form of PCR for simultaneously amplifying two or more target sequences *in the same assay*; the reaction mixture contains primers for each target. PCR with several targets can be monitored by (target-specific) probes labeled with different types of fluorescent dye (i.e. dyes with different emission spectra).

As multiple primers are used, extra care is required in order to avoid the formation of PRIMER–DIMERS.

In some cases, one primer can be shared by two targets. For example, a sequence in the 16S rRNA gene in *Bacteroides forsythus* and *Prevotella intermedia* has been amplified with one FORWARD PRIMER (a BROAD-RANGE PRIMER common to both species) and two species-specific reverse primers.

Multiplex PCR has been used e.g. for detecting *mecA* and *coa* genes in *Staphylococcus aureus* [J Med Microbiol (1997) 46:773–778]; for diagnostic virology [Clin Microbiol Rev (2000) 13:559–570]; for detecting toxin genes in *Clostridium difficile* [J Clin Microbiol (2004) 42:5710–5714]; for studies on *Campylobacter jejuni* isolates [BMC Microbiol (2007) 7: 11]; for studies on mtDNA haplogroups [PLoS ONE (2009) 4(7):e6423]; for developing an assay for human papillomavirus-32 [Virol J (2009) doi: 10.1186/1743-422X-6-90]; and for detecting multiple types of respiratory infection [Virol J (2009) 6:89]

Multiplex PCR is also used e.g. for amplification of STRs in human DNA profiling (e.g. in CODIS).

One alternative technique is SELECTOR-BASED MULTIPLEX PCR.

multiplex QEXT See QEXT.

multiplex real-time PCR REAL-TIME PCR (q.v.) in which more than one target sequence is amplified in a given assay.

multiplicity of infection (MOI) (1) In any given system: the ratio of infectious virions to susceptible cells.

(2) The number of viral genomes in each infected cell.

MultiSite Gateway® technology A GATEWAY SITE-SPECIFIC RECOMBINATION SYSTEM in which a range of vector molecules, each having *att* sites with distinct, *modified* sequences and specificities, offers the choice of diverse recombination partners and directional insertion of the sequence of interest.

Example of use

MultiSite Gateway® technology was used e.g. for inserting a sequence (simultaneously) into two different locations in one plasmid.

Initially, a target site was inserted into the plasmid at each of the two locations. Each target site contained a copy of the *ccdB* gene (see entry CCD MECHANISM) flanked by a pair of *asymmetric att*P sequences; the pair of asymmetric sequences flanking one target site was different to the asymmetric pair flanking the other target site.

Copies of the given sequence to be inserted were prepared by PCR in two separate reactions.

In one PCR reaction, the primers introduced terminal *att*B sequences corresponding to one pair of asymmetric *att*P sites, i.e. the *att*P sites at one of the required locations.

A separate PCR used primers that introduced terminal *att*B sequences corresponding to the other pair of asymmetric *att*P sites.

Thus, as a result of the two PCR reactions, there were two sets of inserts, each prepared for introduction into one of the two target sites.

The (recipient) plasmid was then incubated with the two sets of inserts (for 1 hour at 25°C) together with the Int and IHF proteins. During incubation, each insert was exchanged, by site-specific recombination, with its corresponding target sequence in the plasmid. (Note that elimination of each target sequence from the plasmid resulted in the removal of a *ccdB* gene from the plasmid.)

All the above steps were carried out *in vitro*.

The recombinant plasmids were then used to transform a CcdB-sensitive strain of bacteria. As the plasmid included an ampicillin-resistance gene, selection of the dual recombinants was obtained by using an ampicillin-containing medium. Any cell which contained a plasmid in which *both* of the (CcdB-encoding) target sites had *not* been exchanged for inserts was killed by the CcdB toxin.

In this experiment, all copies of the insert were identical, except for their *att* sites, but it seems likely that this protocol can be easily adapted for the insertion of non-identical donor sequences at required sites.

[Approach used in the above experiment: BioTechniques (2005) 39(4):553–557.]

[Conversion of Gateway® vectors to MultiSite Gateway®

vectors by means of a single recombination event: BMC Mol Biol (2006) 7:46.]

[Uses of Multisite Gateway® technology (some examples): for studying herpes simplex virus helper function for AAVs [PLoS Pathog (2009) 5(3):e1000340]; studying tagged-gene expression in *Drosophila* [BMC Cell Biol (2009) 10:8]; gene knockout in protozoan *Trypanosoma* [BMC Microbiol (2009) 9:90]; and studies on mouse transgenesis [Nucleic Acids Res (2009) 37(7):e55.]

mumps virus A virus of the genus *Paramyxovirus* – within the family PARAMYXOVIRIDAE (q.v.)

mung bean nuclease A SINGLE-STRAND-SPECIFIC NUCLEASE. Mung bean nuclease differs from e.g. endonuclease S_1 in that it cannot cleave the strand opposite a nick in a DNA duplex.

(See also RIBONUCLEASE PROTECTION ASSAY.)

mural trophectoderm See BLASTOCYST.

murine (*adj.*) Of, or pertaining to, mice.

mutagen Any chemical (or physical) agent which can promote MUTAGENESIS. Some mutagens are CARCINOGENS.

Mutagens typically interact with, and alter/damage, nucleic acid. In some cases a mutagen may alter a base such that, in a subsequent round of replication, incorrect base-pairing leads to the insertion of a different nucleotide.

Some mutagens are more effective with ssDNA than with dsDNA.

Physical mutagens include ionizing radiations and ULTRAVIOLET RADIATION. Heat apparently increases the level of spontaneous mutation.

Chemical mutagens include compounds of a wide range of types – alkylating agents, base analogs, bisulfite, hydroxylamine, nitrous acid etc. Others include e.g. the AFLATOXINS (associated with some instances of hepatocellular carcinoma) and β-naphthylamine (implicated in bladder cancer).

mutagenesis (*in vitro*) The creation of mutation(s) in one or more target sequences of nucleic acid. Mutations are created for various reasons: e.g. (i) to inactivate a gene, (ii) to modify the coding sequence of a gene, (iii) to modify the regulation of a gene's expression. A gene may be inactivated e.g. to investigate its involvement (or otherwise) in a given function – cells with an inactivated gene being compared to those with a normal, wild-type gene. Modification of a gene's coding sequence, or its regulation, has widespread applications in medicine and industry as well as in biological research.

There are two basic approaches to mutagenesis: (i) random mutagenesis (i.e. creating mutation(s) at random sites within a given target sequence) and (ii) site-directed (= site-specific) mutagenesis – in which mutations are created at known, predetermined sites. (See also INSERTIONAL MUTAGENESIS.)

Random mutagenesis

Random mutagenesis may be carried out in order to produce specific types of mutant (e.g. temperature-sensitive mutants) for use in various types of study; such mutants are recovered by the use of selective procedures.

Random, transposon-mediated mutagenesis has been used for the detection of virulence-associated genes in pathogenic bacteria: see e.g. SIGNATURE-TAGGED MUTAGENESIS.

Transposition with Tn*5* can be carried out in wholly *in vitro* systems, insertions occurring randomly in e.g. a population of plasmids (see TN5-TYPE TRANSPOSITION).

Random mutations in bacteria can be created by exposing the cells to a *mutagen*, i.e. a physical or chemical agent such as ultraviolet radiation, X-rays, alkylating agents, bisulfite or hydroxylamine. Mutations develop in different genes (and/or non-coding sequences) in different cells. Those cells which have developed the required type of mutation may be isolated by using appropriate *selective* conditions – such as particular types of selective medium. (See AUXOTROPHIC MUTANT for some approaches to the methodology for auxotrophs.)

Random mutations can be introduced into an insert/gene by replicating it in a so-called MUTATOR STRAIN; one advantage of this is that mutagenesis can be achieved without the use of mutagens.

Random mutations can be introduced into a target sequence *in vitro* by a PCR-based method involving the use of an error-prone DNA polymerase (see, for example, GENEMORPH PCR MUTAGENESIS KIT).

***Site-directed mutagenesis* (*site-specific mutagenesis*)**

Mutation at specific site(s) is used e.g. for studies on gene expression, and also for modifying the products encoded by genes – for example, increasing the efficiency of an enzyme used in an industrial process by changing specific nucleotides in the coding region of the gene.

(See also SITE-DIRECTED MUTAGENESIS.)

Site-directed mutations can be single-nucleotide mutations (point mutations) or may involve insertion/deletion of short or long sequences of nucleotides.

For examples of methods used in site-directed mutagenesis see e.g. the EXSITE SITE-DIRECTED MUTAGENESIS KIT, the GENETAILOR SITE-DIRECTED MUTAGENESIS SYSTEM and the LAMBDA RED RECOMBINATION approach; see also MIRAGE, the QUIKCHANGE SITE-DIRECTED MUTAGENESIS KIT and the TARGETRON and ZINC-FINGER NUCLEASE.)

(See also entries CHIMERAPLAST, MEGAPRIMER MUTAGENESIS and TRANSPOSON MUTAGENESIS.)

mutagenic inverted-repeat-assisted genome engineering See MIRAGE.

mutagenic repair See SOS SYSTEM.

mutasynthesis The synthesis of new type(s) of end-product by a mutant organism in which a normal biosynthetic pathway is blocked. (Synthesis may involve the use of an abnormal substrate.)

Mutasynthesis has been exploited e.g. for the synthesis of novel products.

Examples of products derived from mutasynthesis include a number of modified gyrase inhibitors prepared by supplying synthetic analogs of a molecular component to mutant strains of the producing organisms. The structure–activity characteristics of these new compounds were investigated with the object of developing improved forms of derivative antibiotics [Antimicrob Agents Chemother (2008) 52(6):1982–1990].

New analogs of gilvocarcin (anti-tumor) agents were made by inactivating the *gilU* gene in the cell's biosynthetic pathway [Chembiochem (2009) 10(2):278–286].

Mutatest *Syn.* AMES TEST.

mutator strain Any strain of bacteria that is deficient in DNA repair function(s) and which is used e.g. for creating random mutations (see MUTAGENESIS) in target sequences replicated within such strains.

One advantage of using a mutator strain is that mutagenesis can be achieved without the use of mutagens.

One example of a mutator strain is the XL1-Red strain of *Escherichia coli* (Stratagene, La Jolla CA, USA); this strain lacks several of the normal repair functions and is reported to generate mutations at a rate which is several thousand times higher than is possible with a corresponding wild-type strain. The genotype of XL1-Red is *endA1*, *gyrA96*, *thi-1*, *hsdR17*, *supE44*, *relA1*, *lac*, *mutD5*, *mutS*, *mutT*, Tn*10* (Tet^r).

Examples of use of XL1-Red: generation of mutations in the *lolC* gene of *Escherichia coli* [J Bacteriol (2006) 188(8): 2856–2864] and investigation of the effect of mutation in the *RPB1* gene of *Saccharomyces cerevisiae* [Genetics (2006) 172(4):2201–2209].

A derivative of XL1-Red, containing the $lacI^q$ element, was used in studies on the *E. coli* ribonuclease E gene to control expression of *rne* [Genetics (2006) 172(1):7–15].

Preparation of a mutator strain

A mutator strain (of *Escherichia coli*) was generated by using antisense RNAs to silence genes *mutS*, *mutD* and *ndk*; it was reported to exhibit spontaneous mutation frequencies of over 2000-fold higher than those exhibited by wild-type strains of *E. coli* [Nucleic Acids Res (2009) doi: 10.1093/nar/gkp498].

Mutazyme® See entries GENEMORPH and COMPETENT CELLS.

MutH See MISMATCH REPAIR.

MutL See MISMATCH REPAIR.

MutM See 8-OXOG.

MutS See MISMATCH REPAIR.

MutT See 8-OXOG.

MutY See 8-OXOG.

Mx162 See RETRON.

myb (*MYB*) An ONCOGENE which was first identified in avian myeloblastosis virus (AMV), a virus that causes rapidly fatal leukemia in chickens.

The product of v-*myb* occurs in the nucleus. The product of c-*myb* is a transcription factor, with roles in cell proliferation and APOPTOSIS, which is strongly expressed e.g. in *immature* hemopoietic cells, although expression falls markedly during differentiation.

In chickens, insertional activation of c-*myb* is seen e.g. in lymphomas and myeloid leukemias.

myc (*MYC*) An ONCOGENE, first detected in avian myelocytomatosis virus. The v-*myc* product (found in the nucleus) is a *gag–myc* fusion protein which has no protein kinase activity. Human c-*myc* normally occurs on chromosome 8; it encodes a transcription factor which may regulate activities such as proliferation, differentiation and APOPTOSIS. Translocation of c-*myc* may be causally connected with Burkitt's lymphoma.

The oncogenic activity of c-*myc* appears to arise mainly by overexpression.

In mice, the experimental expression of c-*myc* results in the formation of tumors.

(See also ADENOVIRUS.)

mycobacteriophage Any BACTERIOPHAGE that is able to infect strain(s) of *Mycobacterium*.

Mycobacterium tuberculosis A species of Gram-positive, acid-fast bacteria; a causal agent of tuberculosis.

[Genome: Nature (1998) 393:537–544.]

Detection of *M. tuberculosis* in (smear-positive) sputa may be achieved e.g. by the AMTDT.

Typing: e.g. SPOLIGOTYPING (q.v.). Some typing schemes have been based on the insertion sequence IS*6110*, which is apparently specific to the '*M. tuberculosis* complex' (which includes *M. tuberculosis*, *M. bovis* and *M. africanum*). The number of copies of IS*6110*, per genome, varies according to strain, some strains apparently containing none; H37rv, the reference strain of *M. tuberculosis*, contains 16 copies. In at least some strains of *M. tuberculosis* the value of long-term IS*6110*-based typing may be compromised by a high rate of transposition of the element within the genome.

One commonly used experimental strain is H37RA (q.v.).

The detection of antibiotic resistance in clinical isolates of *M. tuberculosis* can be achieved e.g. by several commercial systems: see e.g. LINE PROBE ASSAY. (See also ANTIBIOTIC RESISTANCE TESTING and W-BEIJING STRAIN.)

mycophenolate mofetil (MMF) A prodrug which is converted, intracellularly, to the active agent MYCOPHENOLIC ACID.

mycophenolic acid An agent, with antitumor and antimicrobial activity, produced by *Penicillium brevicompactum*. It inhibits the enzyme inosine monophosphate dehydrogenase, blocking synthesis of xanthosine monophosphate from inosine monophosphate; this, therefore, blocks the formation of guanosine monophosphate from xanthosine monophosphate – depleting the pool of guanine nucleotides.

The prodrug, mycophenolate mofetil (MMF), is converted, intracellularly, to mycophenolic acid.

Examples of use: a pilot study in which MMF was used for treating HIV-positive patients [AIDS Res Ther (2006) 3:16]; and studies on the accumulation of RNAs of hepatitis delta virus (HDV) [J Virol (2006) 80(7):3205–3214].

[Screening yeast non-essential knock-outs for genes which confer resistance to mycophenolic acid: PLoS Genet (2009) 5(2):e1000364.]

mycovirus Any (commonly dsRNA) virus which can infect one or more species of fungus; at least some mycoviruses encode an RNA-dependent RNA polymerase.

Transmissibility of mycoviruses seems to depend primarily on processes such as hyphal fusion and mating (as opposed to the transmission of virions from one individual to another via an extracellular route).

Some mycoviruses do not form conventional virions – their nucleic acid being encapsulated in host-encoded vesicles.

myelodysplastic syndromes A group of disorders affecting the hematopoietic stem cells; the disorders, characterized e.g. by anemia, can develop into acute myeloid leukemia.

myeloid lineage Cells that include e.g. monocytes and macrophages, granulocytes (e.g. neutrophils, basophils and eosinophils) and erythrocytes (red blood cells).
(See also LYMPHOID LINEAGE.)

myotonic dystrophy A type of adult-onset muscular dystrophy affecting e.g. skeletal muscle, heart and the central nervous system. The type I form is due to a CTG expansion in the 3′ UTR (untranslated region) of the *DMPK* gene. Transcripts with this repeat accumulate in nuclei and apparently increase the levels of at least one type of RNA-binding protein.
The type II disease is caused by a repeat expansion of the CCTG sequence in intron 1 of the (zinc-finger) gene *ZNF9*.
(See also REPEAT-EXPANSION DETECTION.)

myristylation membrane localization signal A function used e.g. in the CYTOTRAP TWO-HYBRID SYSTEM.
(cf. NUCLEAR LOCALIZATION SIGNAL.)

myxophage Any bacteriophage that infects one or more species of gliding bacteria (bacteria of the order Myxobacterales).

N

N (1) An indicator of ambiguity in the recognition sequence of a RESTRICTION ENDONUCLEASE (or in any other sequence of nucleic acid); 'N' indicates A or C or G or T (in RNA: A or C or G or U), i.e. any nucleotide.

(2) L-Asparagine (alternative to Asn).

***N*-acetyl-L-cysteine** See MUCOLYTIC AGENT.

***N*-acetylmuramidase** See LYSOZYME.

***N*-acetylneuraminic acid** See NEURAMINIDASE.

***N*-acyl-homocysteine thiolactone** See QUORUM SENSING.

***N*-acyl-L-homoserine lactone** See QUORUM SENSING.

N-degron See N-END RULE.

N-end rule The observation that the half-life of a given protein is influenced by the identity of its N-terminal amino acid.

The N-end rule applies in prokaryotes and eukaryotes.

The features which determine early degradation are signals (*degrons*) that are incorporated in the protein itself. One of these is the *N-degron*: a destabilizing N-terminal amino acid recognized by the cell's specific degradation machinery. For example, in *Escherichia coli*, proteins with arginine or lysine at the N-terminal have characteristically short half-lives (see AAT GENE).

Eukaryotic signals for protein degradation include internal lysine(s); such proteins bind a polyubiquitin chain, by multi-enzyme activity, and are degraded in a 26S PROTEASOME.

[Auxin-based degrons in non-plant cells for the control of protein stability: Nature Methods (2009) 6:917–922.]

N15 phage See PHAGE N15.

NAAT Nucleic-acid-amplification test: any of various types of test (involving e.g. PCR) used in a clinical diagnostic setting.

NAD (former names: coenzyme I; cozymase; diphosphopyridine nucleotide, DPN) Nicotine adenine dinucleotide: a carrier of energy (and hydrogen) in many reactions; it is also involved in ADP-RIBOSYLATION. The oxidized form is written NAD^+, the reduced form is written NADH.

NAD^+ is a required cofactor for an *Escherichia coli* ligase.

NAD^+-dependent ligase See DNA LIGASE.

NAIMA NASBA implemented microarray analysis: a method used e.g. for detecting genetically modified organisms – see the entry GMO for details of the method.

nalidixic acid See QUINOLONE ANTIBIOTICS.

NANA *N*-acetylneuraminic acid: see entry NEURAMINIDASE.

nano- A prefix meaning 10^{-9}. (For other prefixes see the 'Ready reference' section at the front of the dictionary.)

nanocircle DNA See DNA NANOCIRCLE.

Nanog (transcription factor) See STEM CELL.

nanoparticles (*DNA technol.*) Various types of nanoparticle have been reported to enhance specificity and/or efficiency in PCR; they include nanogold and nanosilver as well as carbon nanopowder (see e.g. CARBON NANOTUBES).

Superparamagnetic nanoparticles are used as gene-delivery vehicles: see MAGNETOFECTION. [Magnetic nanoparticles for gene and drug delivery (review): Int J Nanomedicine (2008) 3(2):169–180.]

Nano-scale particles of colloidal silica and iron oxide nanoparticles have also been used for gene delivery: see the entry GENE-DELIVERY SYSTEMS.

Nanogels of polyethyleneimine have been used to introduce plasmids into tumor cells [Cancer Gene Therapy (2009) doi: 10.1038/cgt.2009.11].

(See also DENDRIMER and QUANTUM DOT.)

Biodegradable nanoparticles

The accumulation, *in vivo*, of nanoparticles used for medical (e.g. therapeutic) purposes might lead to toxicity, and this has suggested the use of 'self-destruct' nanoparticles which form residues that can be eliminated from the body on completion of the therapeutic (or other) function. One suggestion is for luminescent porous silicon nanoparticles (LPSiNPs); these particles (which can incorporate a drug) can be monitored *in vivo* by their intrinsic photoluminescence under near-infrared excitation. Unlike e.g. carbon nanotubes or quantum dots, the LPSiNPs were reported to break down into components that were cleared (in a relatively short period of time) – via the kidneys – in mice [Nature Materials (2009) 8:331–336].

nanopore (experimental) A nanometer-scale pore prepared e.g. from certain proteins or by a synthetic process. A nanopore-based method was suggested some years ago as a procedure for the sequencing of DNA [kinetics of translocation of DNA through synthetic nanopores: Biophys J (2004) 87(3):2086–2097]. (See also DNA SEQUENCING.)

NASBA Nucleic acid sequence-based amplification: a method used primarily for the (isothermal) amplification of specific sequences in RNA (at a temperature of e.g. 41°C/42°C).

Amplification of an RNA target by the NASBA method is shown diagrammatically in the figure (pages 226–227).

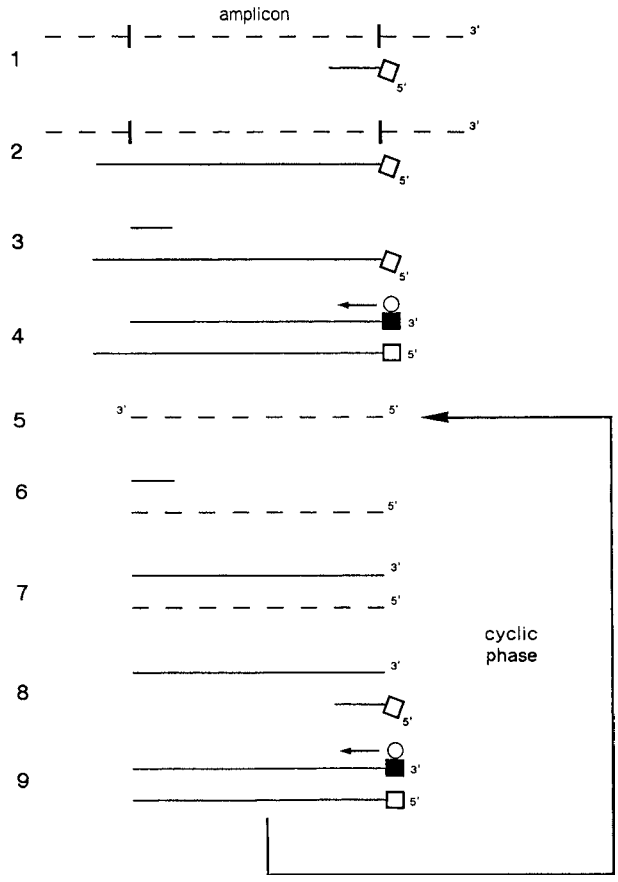
(For isothermal amplification of DNA see e.g. MOLECULAR ZIPPER and SDA.)

NASBA and TMA both reflect an earlier technique used for the *in vitro* amplification of nucleic acids: see entry SELF-SUSTAINED SEQUENCE REPLICATION.

The use of an RNA target sequence (as compared to DNA targets) is an advantage in some circumstances. For example, in a diagnostic assay of cellular pathogens, the abundance of rRNA molecules in every cell, as compared e.g. to (single- or double-copy) genes, increases the chances of detecting small numbers of pathogens in a clinical sample. Moreover, certain short-lived bacterial mRNAs are useful targets for detecting *viable* cells.

In the context of virology, NASBA has been useful e.g. for assaying genomic RNA of the retrovirus HIV-1 in a range of clinical specimens – including plasma, CSF (cerebrospinal fluid) and brain tissue.

NASBA is also useful for monitoring the appearance of transcriptional activity in latent DNA viruses. For example, transcripts of the late genes of cytomegalovirus (CMV) have been used as targets for monitoring those patients at risk from CMV-related conditions. Recipients of transplants have been



NASBA (nucleic acid sequence-based amplification): a method used for the amplification of an RNA sequence (diagrammatic, to show principle). The dashed lines are strands of RNA. The solid lines are strands of DNA. The following is an outline of the stages involved.

1. A strand of RNA showing the target sequence (the amplicon) delimited by the two, short vertical bars. A primer (primer 1) has bound to the 3′ end of the amplicon. The 5′ end of the primer has an extension (tag) which includes a short sequence (□) corresponding to the promoter of an RNA polymerase; the tag itself does not bind to the target RNA.
2. The enzyme reverse transcriptase has extended the primer to form a strand of cDNA.
3. The enzyme RNase H has degraded (removed) the strand of RNA, and a different primer (primer 2) has bound to the amplicon sequence in cDNA.
4. Reverse transcriptase (which can also synthesize DNA on a DNA template) has extended primer 2 to form double-stranded cDNA; note that primer 2 has been fully extended to form the complementary strand (■) of the promoter – so that a functional (double-stranded) promoter is now present. An RNA polymerase (○) has bound to the promoter and will synthesize a strand of RNA in the direction of the arrow, i.e. 5′-to-3′.
5. The newly synthesized strand of RNA. Note the polarity of the strand: it is *complementary* to the amplicon in the sample RNA (compare 5 with 1).
6. Primer 2 has bound to the strand of RNA.
7. Reverse transcriptase has synthesized cDNA on the strand of RNA.

monitored by assaying transcripts of the CMV gene *UL65*, which encodes protein pp67.

The products of a NASBA assay can be detected in various ways. In one approach, products are examined by agarose gel electrophoresis, and blotting, followed by hybridization of a biotinylated probe and detection by an alkaline phosphatase–streptavidin conjugate and a chemiluminescent substrate.

A system involving electrochemiluminescence is available commercially (NucliSens^{TM}, bioMérieux, Boxtel, Holland).

Real-time detection of NASBA products can be achieved e.g. with MOLECULAR BEACON PROBES.

In a different approach to the detection of NASBA products the reaction mixture included aminoallyl-tagged molecules of UTP – so that the products contained aminoallyl-UMP. After amplification, the products were labeled by adding an amine-reactive fluorophore which binds covalently to the aminoallyl tag. The fluorophore-labeled products were then examined on a MICROARRAY [BMC Biotechnol (2009) 9:45].

Quantitation of NASBA products has been achieved e.g. by using several internal 'calibrators', i.e. RNA molecules which are co-amplified and co-extracted with target RNA; because the initial concentration of each of the calibrators is known accurately, the initial concentration of the target sequence can be calculated.

Examples of the use of NASBA include measurement of plasma levels of RNA from human immunodeficiency virus (HIV-1) [Clin Vaccine Immunol (2006) 13(4):511–519], and detection of 16S rRNA from *Chlamydophila pneumoniae* in respiratory specimens [J Clin Microbiol (2006) 44(4):1241–1244]. NASBA was also reported to be useful for assaying noroviruses (common causal agents of viral gastroenteritis) in environmental waters – factors causing inhibition of reverse transcriptase PCR had little or no inhibitory effect on a real-time NASBA assay [Appl Environ Microbiol (2006) 72(8): 5349–5358]. This method has also been used to evaluate the persistence of noroviruses on food surfaces [Appl Environ Microbiol (2008) 74(11):3349–3355].

Quantitative NASBA (QT-NASBA) has been used to study the efficacy of antimalaria drugs by monitoring the clearance of asexual forms of *Plasmodium falciparum* with a real-time 18S rRNA QT-NASBA assay of nucleic acid extracted from blood [Malaria J (2008) 7:237].

NASBA was used for the assessment of *Plasmodium* spp in malaria-asymptomatic individuals [Malaria J (2009) 8:12].

NASBA has been adapted for the detection of genetically modified products in food or feedstuffs – see NAIMA in the entry GMO.

NASBA implemented microarray analysis See entry GMO.

National Bioforensics Analysis Center See NBFAC.

NBFAC National Bioforensics Analysis Center: the microbial forensics laboratory which forms part of the Department of Homeland Security in the USA; it works closely with the FBI (Federal Bureau of Investigation) and is a center specializing in the examination and analysis of material evidence in cases of actual or suspected criminal/terrorist activity.

(See also FORENSIC APPLICATIONS.)

ncRNAs See NON-CODING RNAS.

negative control (in operons) See OPERON.

negative selection (1) Selection that is based on the absence of a given feature, function or reaction.

(2) Any kind of selection which is manifested in the failure of cells or organisms to grow on the basis of a specific criterion or specific criteria.

(cf. POSITIVE SELECTION.)

negative-sense virus *Syn*. NEGATIVE-STRAND VIRUS.

negative-strand virus (negative-sense virus) Any virus whose genome (i.e. that present in the virion) is single-stranded and is complementary to – rather than homologous to – the viral mRNA. (cf. POSITIVE-STRAND VIRUS.)

(See also SUBGENOMIC MRNA.)

neighbor exclusion principle See INTERCALATING AGENT.

nelfinavir See PROTEASE INHIBITORS.

neomycin See AMINOGLYCOSIDE ANTIBIOTICS.

neoplasia Progressive, uncontrolled division of cells, resulting e.g. in a solid tumor or in leukemia.

(See also ONCOGENESIS.)

neoplasm A tumor.

neoschizomer See the entry ISOSCHIZOMER.

NASBA (*continued*)

8. RNase H has removed the strand of RNA, and primer 1 has bound to the amplicon sequence in cDNA.
9. Reverse transcriptase has synthesized a complementary strand of cDNA; note that the 3′ end of the template strand of cDNA has also been extended to form a functional (double-stranded) promoter for the RNA polymerase. RNA polymerase has bound to this promoter and will (repeatedly) synthesize strands of RNA identical to the one shown at stage 5; these strands can participate in the cyclic phase, leading to high-level amplification of the target.

The double-stranded molecules of cDNA in stage 9 are permanent products which are continually transcribed by the RNA polymerase. Operation of the cyclic phase produces many copies of the amplicon in the form of (i) *complementary* RNA and (ii) cDNA.

Reproduced from *Bacteria in Biology, Biotechnology and Medicine*, 6th edition, Figure 8.25, pages 264–265, Paul Singleton (2004) John Wiley & Sons Ltd, UK [ISBN 0-470-09027-8] with permission from the publisher.

neovascularization *Syn*. ANGIOGENESIS (q.v.).

nested deletions Any process in which segments are progressively removed from the end of a DNA molecule.

Unidirectional nested deletions in a cloned fragment within a circular vector can be prepared by first cutting the *vector*, with appropriate RESTRICTION ENDONUCLEASES, *at two sites close to one end of the insert*: (i) one enzyme to produce a 5′ overhang adjacent to the insert, and (ii) another enzyme to produce a 3′ overhang at the other end of the cleaved vector. The digested vector is then incubated with EXONUCLEASE III, and aliquots of the mixture are removed at intervals of time. Exonuclease III digests the 3′ end recessed under the 5′ overhang (but not the 3′ overhang at the other end of the cleaved vector). Increasing lengths of incubation with exonuclease III result in progressively greater digestion of the recessed 3′ end i.e. increasing the extent of deletion of the *insert* – so that aliquots removed later from the incubation will contain molecules in which more of the insert sequence has been deleted. The 5′ single-stranded sequence (resulting from digestion by exonuclease III) – and the 3′ overhang at the other end of the molecule – are removed by incubation with e.g. mung bean nuclease or ENDONUCLEASE S1. The resulting blunt-ended products can be ligated e.g. with T4 DNA ligase to re-form vector molecules containing inserts with a range of deletions.

The technique of using an enzyme with 3′→5′ exonuclease activity (to remove nucleotides from the 3′ end of a strand) is also used in various other forms of methodology in which the enzyme is required to create sticky ends:

In the IN-FUSION method, dsDNA target molecules (which are to be joined) are chosen such that appropriate sticky ends can be created when the 3′ ends are enzymically degraded to form 5′ overhangs. Exonucleolytic action is also involved in the STICKY RICE procedure.

(See also DNA CUTTING AND ASSEMBLY.)

nested PCR (nPCR) A form of PCR carried out in two phases, the first phase being a standard form of PCR. The amplicons produced in the first phase are then used as templates for the second phase; in this phase, the primers are complementary to *subterminal* (i.e. *internal*) sequences in the amplicon. (The primers used in the first phase are often referred to as 'outer' primers, while those used in the second phase may be called 'inner' primers.)

The conditions for the reaction may differ in the two phases of nested PCR because e.g. the optimum annealing temperatures of the outer and inner primers may differ.

nPCR has been used to improve both the sensitivity and the specificity of a PCR assay, and is regarded as being a particularly useful approach when the quantity of target material is limited or in poor condition.

One concern with nPCR was susceptibility to contamination of the reaction mixture when setting up the second phase of amplification. This was addressed in an early study [J Clin Microbiol (1997) 35:132–138] in which a 'one-tube protocol' was used for detecting *Trichomonas vaginalis* in samples of vaginal discharge. In this procedure, both the outer and inner primers were included in a single reaction mixture; initially, the primers were separated from the rest of the reaction mixture by a layer of wax (a so-called 'hot-start' approach) which melted at the start of temperature cycling. In the first phase of amplification, an annealing temperature of 62°C was used; at 62°C only the *outer* primers could bind. After 30 cycles, the annealing temperature was lowered to 45°C for a further 20 cycles; at this temperature (45°C) the inner primers were able to anneal and generated the second-phase amplicons (290 bp) by using the first-phase amplicons as templates. Subsequent detection of the 290-bp amplicons was facilitated by the use of inner primers pre-labeled with BIOTIN and DIGOXIGENIN.

Nested PCR has been used e.g. for diagnosis of taeniasis by targeting the *Tso31* gene of *Taenia solium* [J Clin Microbiol (2008) 46:286–289]; for a study on hydroxycinnamoyl transferase [Plant Physiol (2009) 150:1866–1879]; and studies on recombination and genetic diversity in human rhinoviruses [PLoS ONE (2009) 4(7):e6355].

neu In the rat: an ONCOGENE whose product is related to members of the EPIDERMAL GROWTH FACTOR RECEPTOR FAMILY; the product has TYROSINE KINASE activity that is associated with the intracellular domain.

Neu In the rat: the product of the *neu* gene (a homolog of the human HER2 protein).

Neulasta® See BIOPHARMACEUTICAL (table).

Neupogen® See BIOPHARMACEUTICAL (table).

neuraminidase (acylneuraminyl hydrolase, or sialidase) An enzyme (EC 3.2.1.18) which is found e.g. at the surface of some viruses (e.g. influenza viruses A and B, some members of the family Paramyxoviridae). It degrades mucus, thus helping to expose cell-surface receptor sites for the virus. It also hydrolyzes the *N*-acetylneuraminic acid (NANA) which is present in the virion-binding sites at the surface of red blood cells (RBCs) (see HEMAGGLUTININ); the neuraminidase can therefore cause elution of RBC-bound virions.

Antigenically distinct forms of viral neuraminidase are designated N1, N2, N3 etc. – see the entry INFLUENZAVIRUS.

nevirapine A NON-NUCLEOSIDE REVERSE TRANSCRIPTASE INHIBITOR (q.v.).

new-generation sequencing system Any procedure described in the entry MASSIVELY PARALLEL DNA SEQUENCING; these methods are also referred to as *second-generation* sequencing systems.

Newcastle disease virus A virus of the genus *Paramyxovirus*, family PARAMYXOVIRIDAE (q.v.).

Newcombe experiment An early (1949) classic experiment in which Newcombe demonstrated that *pre-existing* mutants are selected when a population of bacteria appears to adapt to a change in the environment.

A number of plates, containing nutrient agar, are each inoculated from a culture of phage-sensitive bacteria and are then incubated for a few hours. During the incubation period, each viable cell on the surface of the agar gives rise to a number of progeny cells (a clone), forming a colony.

At this stage, the plates are divided into two groups: group

A and group B.

In group A plates the bacterial growth is re-distributed over the surface of each plate by means of a sterile glass spreader. As a result, the cells of each colony are dispersed over the surface of the agar.

In group B plates the colonies are left undisturbed.

Every plate (in both groups, A and B) is then sprayed with a suspension of virulent phage and re-incubated.

During the period of re-incubation all phage-sensitive cells are lysed – thus any colonies that develop on the plates necessarily arise from phage-resistant cells.

If resistance to phage developed adaptively (as a response to the presence of phage) it would be observed only *after* the cells had been exposed to phage. Were this the case, then re-distribution of growth on group A plates should be irrelevant because it was done before the cells were exposed to phage. Hence – if resistance to phage developed adaptively – similar numbers of phage-resistant colonies should develop on all the plates (in groups A and B).

If, however, phage-resistant cells had been present *before* exposure to phage (on both the A and B plates), then each phage-resistant cell (mutant) would form a clone of cells (i.e. a single colony) during the initial incubation, i.e. before the plates were exposed to phage. In this case, the re-distribution of growth on group A plates would spread the cells of each (phage-resistant) colony – so that each cell in a colony would itself give rise to a separate colony during the period of re-incubation. By contrast, any colony of (phage-resistant) cells on group B plates would remain as a single colony because growth on these plates had not been re-distributed.

In practice, many more (phage-resistant) colonies are found on the group A plates, indicating the presence of pre-existing phage-resistant mutant cells in the original culture.

(cf. FLUCTUATION TEST.)

NF-κB Nuclear factor κB: a major transcription factor found in mammalian cells. An (inactive) form of NF-κB, complexed with the inhibitory factor IκB, is found in the cytoplasm. On activation, NF-κB (no longer complexed with IκB) can enter the nucleus and regulate the expression of a range of genes.

(See also BORTEZOMIB.)

An exonuclease-mediated ELISA-like assay has been used for detecting NF-κB–DNA binding activity [BioTechniques (2006) 41(1):79–90].

The DNA-binding activity of NF-κB has also been assayed with a double-stranded DNA probe that includes (i) a binding site for NF-κB and (ii) the recognition site for a RESTRICTION ENDONUCLEASE – ApoI – located *within* the binding site of NF-κB. One end of the probe has a pair of FRET-compatible fluorophores (FAM and TAMRA), one on each strand, so that the (intact) probe can produce a high-level FRET signal on appropriate excitation. When NF-κB is bound to the probe it blocks the recognition site of ApoI; hence, exposure to ApoI does not result in cleavage of the probe – so that a high-level FRET signal can be obtained on suitable excitation. Such a high-level signal is not obtained in the absence of binding between the probe and NF-κB because the probe can undergo restriction by ApoI and the resulting fragments do not remain double-stranded at the temperature of the reaction mixture. [Method: BioTechniques (2007) 43(1):93–98.]

NHEJ See NON-HOMOLOGOUS DNA END-JOINING.

NHGRI National Human Genome Research Institute.

nick In dsDNA: a break, in one strand of the duplex, at a particular phosphodiester bond.

(cf. GAP.)

nick-closing enzyme See TOPOISOMERASE (type I).

nick translation A technique for labeling DNA (for example, cloned fragments) by replacing the existing nucleotides with nucleotides carrying a radioactive or other label. Essentially, the (ds) DNA is exposed to an endonuclease which makes a number of (random) nicks. Using each of the nicks as a starting point, DNA polymerase I extends the 3′ terminus – the newly synthesized strand incorporating labeled dNTPs (that are abundant in the reaction mixture); during synthesis, the polymerase exerts 5′-to-3′ exonucleolytic activity, i.e. it progressively removes nucleotides from the nick's 5′ terminus.

nickel-charged affinity resin A resin charged with nickel ions (Ni^{2+}) that is able to bind proteins which have an (accessible) sequence of six consecutive histidine residues; it is used e.g. for the purification of 6×His-tagged recombinant proteins – including those produced from expression vectors such as the CHAMPION PET SUMO VECTOR and the VOYAGER VECTORS. (See also SIX-HISTIDINE TAG.)

Elution of proteins bound to the resin can be achieved e.g. by using a low-pH buffer or by histidine competition.

(See also PROBOND NICKEL CHELATING RESIN.)

nicking enzyme See RESTRICTION ENDONUCLEASE.

nicotine adenine dinucleotide See the entry NAD.

Niemann–Pick disease See GENETIC DISEASE (table).

nisin See LANTIBIOTIC.

nitrocefin An agent used to test for β-LACTAMASES; nitrocefin is a CEPHALOSPORIN which changes from red to yellow when cleaved by a β-lactamase.

(See also PADAC.)

nitrous acid (as a mutagen) Nitrous acid (HNO_2) brings about oxidative deamination of guanine (to xanthine), cytosine (to uracil) and adenine (to hypoxanthine). Replication of nitrous acid-treated DNA can lead to GC→AT mutation (deaminated cytosine) and AT→GC mutation (deaminated adenine).

(See also HYDROXYLAMINE.)

NLS See NUCLEAR LOCALIZATION SIGNAL.

NM23-H1 A gene encoding suppression of cancer metastasis in mice: see e.g. GENE-DELIVERY SYSTEM.

NMD See NONSENSE-MEDIATED MRNA DECAY.

NMR Nuclear magnetic resonance, a phenomenon exploited e.g. for monitoring gene expression *in vitro* and *in vivo*: see the entry MRI.

NNRTIs NON-NUCLEOSIDE REVERSE TRANSCRIPTASE INHIBITORS (q.v.).

nodaviruses A family of non-enveloped isometric viruses (∼30 nm diam.) in which the (bipartite) genome is positive-sense

ssRNA; the type virus, Nodamura virus, was first isolated in Japan from mosquitoes (*Culex tritaeniorhynchus*). Viruses of the genus *Alphanodavirus* (e.g. Flock House virus) primarily infect insects; viruses of the only other genus, *Betanodavirus*, infect mainly fish.

non-coding RNAs (ncRNAs) A broad category of RNAs which are transcribed from DNA but which do not encode proteins; they include the small, non-coding RNAs (see entry SRNAS) and the LONG NON-CODING RNAS.

(See also piRNAs (in the entry PIWI PROTEIN), SNRNAS, and Xist (in the entry X-INACTIVATION).)

non-coding strand (of a gene) See the entry CODING STRAND.

non-homologous DNA end-joining (NHEJ) A particular type of DNA REPAIR mechanism that is used e.g. for dealing with double-strand breaks in genomic DNA. The repair process is carried out by a number of components of the NHEJ system (see e.g. KU70), the ARTEMIS NUCLEASE having an early role in preparing the broken ends for ligation.

(See also X FAMILY and ZINC-FINGER NUCLEASE.)

non-invasive prenatal diagnosis See e.g. DOWN'S SYNDROME.

non-isotopic labeling (of probes) See PROBE LABELING.

non-Mendelian inheritance CYTOPLASMIC INHERITANCE.

non-nucleoside reverse transcriptase inhibitors (NNRTIs) A structurally diverse group of ANTIRETROVIRAL AGENTS that inhibit the viral reverse transcriptase.

The NNRTIs include delavirdine, efavirenz, etravirine and nevirapine. (See also GW678248 and SURAMIN.)

(See also the entry HAART.)

non-ribosomal peptide synthetase (NRPS) Any of a range of multisubunit enzymes which mediate ribosome-independent synthesis of peptides [e.g. NRPS from the bacterium *Pseudomonas aeruginosa*: J Bacteriol (2003) 185(9):2848–2855].

NRPSs have been genetically modified in order to generate new types of product; these and other advances have allowed the synthesis of e.g. various peptide antibiotics and integrin receptors.

[Chemoenzymatic and template-directed synthesis of bioactive macrocyclic peptides: Microbiol Mol Biol Rev (2006) 70(1):121–146.]

To synthesize their products, the NRPSs are organized in a modular fashion – in which each unit is involved in adding a part of the growing chain of the product; hence, the product develops sequentially by cooperative action of the individual modules. [Study of the *Escherichia coli* enterobactin synthetase EntF NRPS subunit: Nature (2008) 454(7206):903–906.]

nonsense-associated disease Any of various diseases (such as Duchenne muscular dystrophy) which may develop through a NONSENSE MUTATION that affects a specific protein.

nonsense codon (chain-terminating codon, stop codon, termination codon) Any of the three codons – UAA (*ochre* codon), UAG (*amber* codon) and UGA (*opal* codon or *umber* codon) – which usually specify termination of polypeptide synthesis during translation (see entry GENETIC CODE for exceptions).

In some cases the UGA codon can specify selenocysteine – see the entry SECIS.

nonsense-mediated mRNA decay (NMD) A process in which those mRNA molecules that have a so-called *premature* stop codon exhibit susceptibility to degradation by ribonucleases.

Recognition of a stop codon as 'premature' depends, at least to some extent, on the location of the stop codon within the mRNA; this can be determined, for example, during splicing of the pre-mRNA. Alternative splice forms which are down-regulated by NMD have been investigated in a genome-wide study in *Drosophila* [PLoS Genet (2009) 5(6):1000525].

Studies (in a mammalian context) suggested that the basis of NMD activity may depend on whether the target molecule contains recently evolved exons or ancient exons [BMC Biol (2009) 7:23].

nonsense mutation Any mutation that produces a NONSENSE CODON from a codon previously specifying a product.

nonsense suppressor See SUPPRESSOR MUTATION.

nonstop mRNA (nonSTOP mRNA etc.) mRNA in which there is no in-frame stop codon. In bacterial cells such mRNAs are degraded in the so-called *trans* translation process (see entry TRANS TRANSLATION).

In mammalian cells, nonstop mRNAs were reported to be repressed post-initiation [EMBO J (2007) 26 (9):2327–2338].

nopaline See CROWN GALL.

norfloxacin See QUINOLONE ANTIBIOTICS.

noroviruses RNA viruses which are common causal agents of viral gastroenteritis.

Noroviruses in samples of environmental water have been assayed e.g. by NASBA (q.v.).

Noroviruses of genogroups I and II have been detected and quantitated in clinical samples by real-time PCR carried out with *ad hoc* LUX PRIMERS (q.v.) [J Clin Microbiol (2008) 46 (1):164–170].

[The evaluation, and policy implications, of norovirus test performance for detection of outbreaks: J Transl Med (2009) 7:23.]

Northern blotting BLOTTING in which RNA is transferred from gel to matrix. (cf. SOUTHERN BLOTTING.) A Northern blot may be probed e.g. by labeled DNA.

Northern blotting can be used e.g. to reveal the occurrence of ALTERNATIVE SPLICING.

(See also WESTERN BLOTTING.)

NotI A RARE-CUTTING RESTRICTION ENDONUCLEASE (q.v.).

A LINKER or ADAPTOR containing a site for NotI cleavage can be used e.g. to bracket a fragment of DNA inserted into a (circular) vector; this could permit subsequent excision of the fragment by digestion with NotI, the fragment itself generally being safe from cleavage because recognition sites for NotI are generally very uncommon, and should not occur in either the vector or the insert.

(See also GC% (figure, part B).)

Novex® gels A range of gels (Invitrogen, Carlsbad CA, USA) used in gel electrophoresis of nucleic acids and proteins.

Novex® TBE (Tris, boric acid, EDTA) gels are used for the separation of fragments of DNA of a wide range of sizes.

Novex® TBE-urea gels have been optimized for fragments

of DNA in the size range ~20–800 bp.

novobiocin A coumarin derivative that binds to the B subunit of bacterial gyrase, blocking the enzyme's ATPase activity and, hence, blocking DNA synthesis.

(See also QUINOLONE ANTIBIOTICS.)

Novobiocin is not so effective against topoisomerase IV – (apparently) because of a one-residue difference in the target sequence.

A mutation in the B subunit of gyrase, causing *hypersensitivity* to novobiocin, was found to be partly suppressed by a further mutation within the A subunit.

NovoRapid® See BIOPHARMACEUTICAL (table).

NovoSeven® See BIOPHARMACEUTICAL (table).

NP-2 cell lines Cell lines that express most of the (known) cell-surface alternative co-receptors for HIV-1. NP-2 cells are used e.g. for studies on the infection process of HIV-1.

nPCR NESTED PCR.

NPVs NUCLEAR POLYHEDROSIS VIRUSES.

***nrdA* gene** (*dnaF* gene) In *Escherichia coli*: the gene encoding ribonucleotide reductase – which converts ribonucleotides to deoxyribonucleotides.

NRPS NON-RIBOSOMAL PEPTIDE SYNTHETASE.

NRTIs NUCLEOSIDE REVERSE TRANSCRIPTASE INHIBITORS.

nuage An electrodense body in the germ cells of e.g. mammals and invertebrates (including *Drosophila*). It is associated with e.g. RNA INTERFERENCE, and, in mice, silencing transposable elements in germline cells [Dev Cell (2008) 15(2):285–297].

nuclear egress (viral) See e.g. NUCLEAR LAMINA.

nuclear lamina In interphase cells: a protein layer that lines the nuclear membrane; it contains three types of protein: *lamins* A, B and C. During progression through the cell cycle, the nuclear membrane (including lamina) breaks down to permit mitosis and cell division. Such breakdown is promoted by the Cdc2/cyclin-dependent kinase 1 (lamin kinase) that depolymerizes the lamina by phosphorylating the lamin proteins. The nuclear lamina re-forms as the cell moves into interphase.

In human cells infected with the cytomegalovirus (CMV), a CMV-encoded protein, UL97, mimics Cdc2 by phosphorylating the lamin proteins – thus promoting the dissolution of the lamina and facilitating the egress of nascent virions from the nucleus [PLoS Pathogens (2009) 5(1):e1000275].

nuclear localization signal (NLS) A short peptide that is able to promote uptake (of e.g. a protein) into the nucleus.

A nuclear localization signal is incorporated e.g. in the REV PROTEIN of HIV-1.

(See also MONOMERIC RED FLUORESCENT PROTEIN.)

nuclear magnetic resonance (NMR) See the entry MRI.

nuclear polyhedrosis viruses (NPVs) A category of baculoviruses that are distinguished from GRANULOSIS VIRUSES e.g. on the basis of the characteristics of their occlusion bodies: see BACULOVIRIDAE. Those NPVs in which a single nucleocapsid is enveloped have been designated *S*NPV, while those in which multiple nucleocapsids are enveloped are referred to as *M*NPV.

Unlike granulosis viruses, NPVs occur in various orders of arthropods, including Coleoptera, Diptera and Hymenoptera as well as Lepidoptera.

The type species of NPVs is *Autographa californica* NPV (Ac*M*NPV) (genome 135 kb). Other NPVs: e.g. the silkworm pathogen *Bombyx mori* NPV (Bm*S*NPV) [genome: Nucleic Acids Res (2009) doi: 10.1093/nar/gkp.801] and Douglas Fir tussock moth virus *Orygia pseudotsugata* NPV (Op*M*NPV).

5′-nuclease PCR Any of various forms of PCR in which the 5′-exonuclease activity of the polymerase is used to degrade a labeled probe which is bound to an amplicon; the signal (for example, fluorescence) generated by degradation of the probe is monitored.

nuclease protection assay See entry RIBONUCLEASE PROTECTION ASSAY.

nuclease S1 *Syn*. ENDONUCLEASE S1 (q.v.).

nucleic acid A (typically) single- or double-stranded (hetero)-polymer of NUCLEOTIDES in which the adjacent nucleotides in a strand are linked by a phosphodiester bond between the 5′ and the 3′ positions of their pentose residues.

DNA is a nucleic acid consisting of *deoxy*ribonucleotides (in which the pentose residues carry –H, rather than –OH, at the 2′-position).

RNA is a nucleic acid consisting of ribonucleotides.

In a double-stranded nucleic acid the two strands are held together by base-pairing between *complementary* bases (i.e. between A in one strand and T (U in RNA) in the other, and between G in one strand and C in the other).

(See also entries TRIPLEX DNA and QUADRUPLEX DNA.)

A nucleic acid molecule may be linear (with free 5′ and 3′ ends) or covalently closed circular (ccc) with no free 5′ or 3′ ends. (A 'nicked' molecule is a double-stranded molecule that contains one or more nicks – see the entry NICK.)

Information is encoded in the *sequence* of the nucleotides in the polymer. In all organisms, the genome is composed of one or more molecules of nucleic acid. (See also CHROMATIN and CHROMOSOME.)

Inhibition of nucleic acid synthesis in microorganisms can be effected by certain types of ANTIBIOTIC (q.v.).

nucleic acid amplification (*in vitro*) Any of various methods used for the amplification (i.e. repeated copying) of (usually) specific sequences of nucleotides in DNA or RNA.

The table lists examples of methods for amplifying nucleic acids (see pages 232 and 233).

(See also DNA AMPLIFICATION and RNA AMPLIFICATION.)

nucleic acid enzyme Any molecule of nucleic acid, either RNA or DNA, which has enzymic capability – e.g. as a nuclease or as a ligase. See e.g. RIBOZYME and DEOXYRIBOZYME.

nucleic acid extraction See NUCLEIC ACID ISOLATION.

nucleic acid isolation (nucleic acid extraction) Any procedure which is used for removing and concentrating DNA or RNA from cells or tissues – see e.g. DNA ISOLATION, RNAQUEOUS TECHNOLOGY and PAXGENE BLOOD RNA KIT.

Reagents commonly used in the isolation of nucleic acids are listed in the table. Note that some of these reagents may inhibit amplification in processes such as PCR – indicating

NUCLEIC ACID AMPLIFICATION (summary of *in vitro* methods)[a]

Method	Targets in[b]	Temperature	Basis
cHDA[c]	DNA	Isothermal	Amplification of *circular* dsDNA templates by helicase-dependent strand separation followed by a rolling circle mechanism to produce concatemeric products; the helicase, DNA polymerase and single-strand binding proteins from bacteriophage T7 are key components
HDA[d]	DNA	Isothermal	Strands of (double-stranded) sample DNA are separated by helicase activity rather than by heat or recombinase activity (as in other methods)
LCR[e]	DNA	Cycling	Use of a binary probe (two oligonuleotides) binding specifically to two closely adjacent regions in the target sequence; closure of the gap between the two bound oligonucleotides by a DNA polymerase is followed by ligation and then temperature-based release of the newly formed amplicon – the newly formed and pre-existing amplicons then acting as templates for the next round of amplification. Thermostable enzymes (i.e. polymerase and ligase) are required. *Product detection*. One approach uses gel electrophoresis. A commercial system uses a microparticle enzyme immunoassay in which immobilized ligated probes are detected by an antibody–enzyme conjugate
MDA[f]	DNA	Isothermal	A form of WGA (whole-genome amplification) in which multiple sites are primed by random, exonuclease-resistant hexamers, and DNA synthesis is mediated by the highly processive DNA polymerase derived from bacteriophage φ29
Molecular zipper[g]	DNA	Isothermal	The target sequence is initially in the form of an ssDNA circle. A single primer is extended (by the rolling circle mechanism), yielding an ssDNA multimer of the target. The multimer binds many molecular zippers (each a composite, labeled probe) which, when extended via their 3′ ends, displace the downstream (bound and extended) zippers. Primers bind to the displaced zippers and are extended to form dsDNA products; extension of these primers displaces a fluorophore-labeled oligonucleotide hybridized to the original zipper, permitting detection of products via the fluorophore's label
NASBA[h]	RNA	Isothermal	Initially, primer 1 is extended, by reverse transcription on the RNA target sequence, forming a strand of cDNA as part of an RNA/DNA hybrid duplex; primer 1 carries a 5′ tag containing the promoter sequence of an RNA polymerase. The RNA strand of the duplex is degraded, and primer 2 binds to the first strand of cDNA; primer 2 is extended to form the second strand of duplex cDNA – thus creating a functional promoter. Transcription of the duplex cDNA forms single-stranded RNA products (complementary to the target sequence). Amplification of these products, first by primer 2 then by primer 1, yields further copies of the double-stranded cDNA which, as before, are transcribed into RNA products; this process constitutes a cyclic phase in which the products are (i) ssRNA amplicons (complementary to the target sequence) and (ii) dsDNA amplicons. *Product detection*. Gel electrophoresis (e.g. to compare size of amplicons with target sequence). Also, real-time monitoring e.g. with molecular beacon probes. In a commercial system: probe-based monitoring involving generation of an electrochemiluminescent signal
PAP[i]	DNA	Cycling	Pyrophosphorolysis-activated polymerization: a form of amplification used for high-level specificity (owing to minimal non-specific priming); it is employed e.g. for allele-specific amplification. Initially, primer extension is blocked because the 3′ terminal nucleotide of each primer is a dideoxy-ribonucleotide monophosphate (ddNMP); given correct hybridization to the target sequence, and the presence of pyrophosphate, the ddNMP is excised – so that primer extension can occur
PCR[j]	DNA	Cycling	Extension of primers on opposite strands of the target sequence following heat denaturation of double-stranded sample DNA, the two primer-binding sites delimiting the target sequence to be amplified. Temperature-based release of newly formed amplicons from the templates and subsequent binding of primers to newly formed and pre-existing target sequences for the next round of amplification. (See entry PCR for a detailed descripiton of the process.) *Product detection*. Gel electrophoresis (e.g. to compare the size of amplicons with expected size – as may be required in diagnostic tests). Monitoring with a dsDNA-binding fluorophore (such as SYBR® Green I). Quantitation of target sequences in the original sample by use of TaqMan® probes or by the LightCycler™ system. Monitoring by a combined probe–dye approach: see REAL-TIME PCR for details

(*continued*)

NUCLEIC ACID AMPLIFICATION (*continued*)

Method	Targets in[b]	Temperature	Basis
Rolling circle	DNA	Isothermal	Ongoing synthesis on a *circular* template, with concurrent displacement of the strand bound to the template strand (in a circular dsDNA target molecule). The target molecule may also be single-stranded; in some cases an RNA polymerase may be able to initiate synthesis on an ssDNA strand with GTP (see e.g. entry DNA NANOCIRCLE). (See also entry for ROLLING CIRCLE.) The product may be a concatemeric form of the target molecule; however, a double-stranded product can be produced if primers bind to this strand and undergo extension and ligation
RPA[k]	DNA	Isothermal	In this method, the binding of primers does not involve initial separation of the strands in sample dsDNA by heat or helicase activity (as in other methods). Instead, primer binding is mediated by a recombinase following the formation of primer–recombinase complexes. Extension of primers by a strand-displacing DNA polymerase produces a growing loop of ssDNA (from the target duplex) which is stabilized by single-strand binding proteins. Various factors within the reaction mixture are required to bring about a balance between development and disassembly of the primer–recombinase complexes. *Product detection*. Probe-based detection in which a bound probe with an abasic-site mimic is cleaved by an endonuclease, releasing a fluorophore from the influence of a quencher
SDA[l]	DNA	Isothermal	Two phases: target generation and amplification. In target generation, primers – 5′-tagged with the recognition sequence of HincII RESTRICTION ENDONUCLEASE – are extended, and the newly formed strands are displaced through extension of bumper primers. As the reaction mixture contains α-thiophosphoryl dATP (dATPαS) in place of normal dATP, the ongoing reaction sequence leads to the formation of double-stranded intermediates having two terminal HincII recognition sites that are hemi-modified asymmetrically – i.e. both recognition sites contain one modified strand and one unmodified strand but the modified/unmodified nucleotides are in different strands at the two sites (see figure in SDA). Nicking of a recognition site (in the unmodified strand), and extension of the 3′ end of the nick, displaces a strand with one terminal modified site; a primer extended on this strand yields a double-stranded product that feeds into the amplification phase (see figure in SDA). Amplification is based essentially on the nicking of a hemi-modified site (in the unmodified strand) and the displacement of a sense or antisense strand by extension of the 3′ end of the nick – such extension regenerating a new amplicon in readiness for the next round of nicking. *Product detection*. Various ways, including gel electrophoresis and probe-based approaches.
TMA[m]	RNA	Isothermal	Similar to NASBA. Commercial TMA-based diagnostic tests have been used for detecting pathogens such as *Chlamydia trachomatis* and *Mycobacterium tuberculosis* in clinical specimens *Product detection*. In a commercial system: probe-based monitoring involving generation of a chemiluminescent signal
WGA[n]	DNA	Isothermal	Whole-genome amplification: any of several techniques in which multiple regions of a sample of genomic DNA are amplified simultaneously by either random-sequence primers or e.g. bioinformatically optimized primers

[a]For further information see separate entries.
[b]Some methods can be adapted to amplify either DNA or RNA targets.
[c]Circular helicase-dependent amplification: see entry CHDA.
[d]Helicase-dependent amplification: see entry HELICASE-DEPENDENT AMPLIFICATION.
[e]Ligase chain reaction: see entry LCR.
[f]Multiple displacement amplification: see entry MULTIPLE DISPLACEMENT AMPLIFICATION.
[g]See entry MOLECULAR ZIPPER.
[h]Nucleic acid sequence-based amplification: see entry NASBA.
[i]See entry PYROPHOSPHOROLYSIS-ACTIVATED POLYMERIZATION.
[j]Polymerase chain reaction: see entry PCR.
[k]Recombinase polymerase amplification: see entry RPA.
[l]Strand displacement amplification: see entry SDA.
[m]Transcription-mediated amplification: see entry TMA.
[n]Whole-genome amplification: see entry WHOLE-GENOME AMPLIFICATION.

NUCLEIC ACID ISOLATION: some of the reagents commonly used in the isolation of nucleic acids from cells and tissues

Reagent	Function (e.g.)
EDTA (ethylenediamine-tetra-acetic acid)	Destabilizes the cell envelope (outer membrane) of Gram-negative bacteria by chelating those ions (e.g. Mg^{2+}) which confer stability
Guanidinium isothiocyanate	Cell lysis and inactivation of nucleases
Lysostaphin	Cleaves peptide cross-links in the cell-wall polymer peptidoglycan in *Staphylococcus aureus* (a Gram-positive bacterium), weakening the wall and thus promoting cell lysis
Lysozyme	Cleaves the glycan backbone chains in the bacterial cell-wall polymer peptidoglycan, promoting cell lysis by weakening the cell wall
Phenol–chloroform–isoamyl alcohol	Promotes partitioning of DNA in the upper layer; the DNA can be precipitated e.g. by ethanol or by isopropanol
Proteinase K	Non-specific digestion of proteins. It is also useful e.g. as an inactivator of nucleases
RNase A	Elimination of RNA from isolated DNA (used e.g. in plasmid preparations)
RNase inhibitors	Protection of RNA from RNA-degrading enzymes (see e.g. RNA*secure*™ in the entry RIBONUCLEASE)
Sodium hydroxide–sodium dodecyl sulfate	Disrupts cell membranes, allowing leakage
Tris (tris(hydroxymethyl)-aminomethane	As a hydrochloride: a buffering system which is used (e.g.) with EDTA
TRIzol®	Constituents: phenol and guanidine isothiocyanate. Used e.g. for the isolation of total RNA from cells. The integrity of the RNA is maintained during homogenization/lysis of the cells

the need for a pre-amplification clean-up. For example, lysozyme and proteinase K have both been reported as inhibitors of PCR.

nucleic acid sequence-based amplification See NASBA.

nucleic acid staining See e.g. DNA STAINING and RIBOGREEN. (See also ULTRAVIOLET ABSORBANCE.)

nucleocapsid A composite structure, consisting of nucleic acid and protein, that forms part, or all, of certain virions.

nucleofection A GENE-DELIVERY SYSTEM used for the TRANSFECTION (sense 1) of DNA or e.g. mRNA into cells by means of the Amaxa Nucleofector 10 (Amaxa, Köln, Germany); the method, which involves a combination of electrical and cell-type-specific factors, delivers nucleic acid directly into the *nucleus*, and is designed to be particularly useful for transfecting e.g. primary cells and 'hard-to-transfect' cells.

A study on the performance of nucleofection in *dendritic cells* reported that transfection of mRNA was more successful than transfection of DNA in terms of the observed intracellular *expression* of each type of nucleic acid [Clin Vaccine Immunol (2008) 15(9):1337–1344].

[Examples of use: Proc Natl Acad Sci USA (2009) 106(9): 3420–3425; PLoS Pathog (2009) 5(2):e1000315; Proc Natl Acad Sci USA (2009) 106(15):6321–6326; Neoplasia (2009) 11(7):683–691.]

nucleoid In a prokaryotic cell: a body which consists mainly of genomic DNA (one or more chromosomes) and NUCLEOID-ASSOCIATED PROTEINS; the DNA is highly compacted.

(See also A-TRACT, CHROMOSOME and DNA TOROID.)

Segregation of genetic material to daughter cells during cell division has been the subject of enquiry for a number of years – and various hypotheses were suggested. One supposed that segregation occurs in an entropy-driven, spontaneous manner due to the tendency of the two nucleoids (post-replication) to exhibit mutual repulsion, so maximizing the conformational entropy [Proc Natl Acad Sci USA (2006) 103(33):12388–12393]. An alternative proposal was that segregation needs a force-generating system involving the actin-like cytoskeletal protein MreB. MreB is an ortholog of eukaryotic actin [polymerization properties: Biochemistry (2008) 47(2):826–835] which is involved e.g. in the maintenance of cell shape; it has also been reported to be duplicated and segregated before cell division, leading to equipartition between the daughter cells [Proc Natl Acad Sci USA (2007) 104(45):17795–17800].

Studies on segregation of the paired chromosomes in *Vibrio cholerae* suggested that coordination may occur between the two chromosomes to ensure correct transmission to daughter cells [J Bacteriol (2006) 188(3):1060–1070].

Note. The term *nucleoid* is also used to refer to the DNA–protein complex in a mitochondrion and to the electron-dense core of a peroxisome.

nucleoid-associated proteins Certain small proteins, such as H-NS PROTEIN and HU PROTEIN, which are generally found to be abundant in the (prokaryotic) nucleoid.

nucleolar phosphoprotein B23 *Syn*. NUCLEOPHOSMIN.

nucleolus See NUCLEUS.

nucleophosmin (nucleolar phosphoprotein B23; numatrin) A multifunctional nuclear phosphoprotein which has roles e.g. in ribosome synthesis and cell proliferation (see also histone ACETYLATION). Nucleophosmin apparently regulates and/or stabilizes the tumor suppressors P53 and $p19^{Arf}$. One report indicated that it is formed in response to chemotherapy of lung cancer [Exp Cell Res (2007) 313(1):65–76].

nucleoporin See NUCLEUS.

nucleoprotein filament See HOMOLOGOUS RECOMBINATION.

nucleoside A compound consisting of the residue of a purine or pyrimidine base linked, covalently, to a *pentose* (i.e. 5-carbon sugar) – commonly ribose (in *ribo*nucleosides) or 2′-deoxyribose in *deoxyribo*nucleosides. A pyrimidine is linked via its 1-position to the sugar, a purine is linked via its 9-position.

A ribonucleoside which incorporates a residue of adenine, guanine, cytosine, uracil, thymine or hypoxanthine is called, respectively, adenosine, guanosine, cytidine, uridine, thymidine (or *ribothymidine*) or inosine. The corresponding *deoxyribo*nucleosides are deoxyadenosine, deoxyguanosine etc. In general, 'thymidine' is used to refer to deoxythymidine.

The letters A, C, G and T are commonly used to refer to the corresponding base, nucleoside or nucleotide (e.g. A is used to refer to adenine, adenosine and also adenosine 5′-monophosphate etc.).

It is common practice to omit 'deoxy' when referring to nucleosides or nucleotides (or their components) in DNA.

(cf. NUCLEOTIDE.)

nucleoside reverse transcriptase inhibitors (NRTIs) A class of synthetic nucleoside analogs used in antiretroviral therapy, e.g. anti-AIDS therapy; in cells, an NRTI is phosphorylated (to the triphosphate) and, when incorporated into DNA by the (viral) reverse transcriptase (RT), it blocks chain elongation by blocking formation of the next phosphodiester bond. All NRTIs have a similar mode of action.

The NRTIs include drugs such as ABACAVIR, adefovir, didanosine, LAMIVUDINE, stavudine, tenofovir, zalcitabine and ZIDOVUDINE.

The anti-HIV activity of an NRTI *in vivo* may be raised by raising the ratio of NRTI to the cell's own dNTPs; this can be done e.g. by making use of hydroxyurea to inhibit the host's enzyme ribonucleotide reductase (thereby causing a decrease in endogenous dNTPs). This approach gave clinical benefit e.g. when used in combination with didanosine.

Resistance to NRTIs can be due to a mutant RT gene. In one *in vitro* study of the activity of various NRTIs against a mutant (K65R) strain of HIV-1, all NRTIs except those with a 3′-azido moiety were less active against the mutant virus – indicating the value of including AZT in drug combinations [Antimicrob Agents Chemother (2005) 49(3):1139–1144].

Some NRTIs are also useful against DNA viruses: see e.g. LAMIVUDINE.

nucleosome See the entry CHROMATIN.

nucleotide A NUCLEOSIDE that carries one or more phosphate groups at the 5′ position of the sugar residue.

Nucleotides are subunits of the NUCLEIC ACIDS DNA and RNA.

Certain nucleotides, e.g. adenosine triphosphate (ATP) and guanosine triphosphate (GTP), have an important function in energy-transfer reactions as well.

Biosynthesis of nucleotides in microbial and/or tumor cells may be inhibited by various agents such as AZASERINE, DON, HADACIDIN, MYCOPHENOLIC ACID and PSICOFURANINE.

(See also UNIVERSAL NUCLEOTIDE.)

nucleotide excision repair *Syn*. UVRABC-MEDIATED REPAIR.

nucleotide flipping Enzymatic rotation of a nucleotide, within a duplex, during catalysis. [Flipping by restriction enzymes: Nucleic Acids Res (2009) doi: 10.1093/nar/gkp688.]

nucleus In a *eukaryotic* cell: a body comprising, minimally, of a (porous) nuclear membrane enclosing the CHROMOSOMES; the nuclear membrane consists of two bilayers – the inner nuclear membrane being lined by the NUCLEAR LAMINA and the nuclear *matrix*.

In the (mammalian) nucleus, the double membrane contains thousands of pores, each pore being composed of proteins (nucleoporins) of various types; these pores have essential roles in transporting molecules between the nucleus and cytoplasm, and they may also have other functions, such as gene regulation, which are independent of their roles in transport [review of nuclear-pore complexes: EMBO Rep (2009) 10(7):697–705].

The nuclear membrane can be seen during interphase but it is not present for part of the cell cycle.

In most cases (but not e.g. in a ciliate MICRONUCLEUS) the nucleus contains a *nucleolus*, a body which is associated e.g. with the synthesis of ribosomal RNA (rRNA).

(Some relevant entries: MATRIX-ATTACHMENT REGION and NUCLEAR LOCALIZATION SIGNAL.)

(cf. NUCLEOID.)

NucliSens™ See NASBA.

numatrin *Syn*. NUCLEOPHOSMIN.

numts Nuclear mitochondria-like sequences: those mtDNA-like sequences ('mitochondrial pseudogenes') which occur in nuclear (chromosomal) DNA. Numts are a possible source of contamination when mtDNA *specifically* is being targeted for amplification by PCR (e.g. in certain forensic investigations).

A total of 46 fragments of nuclear DNA, representing the entire (human) mitochondrial genome, have been sequenced [BMC Genomics (2006) 7:185].

(See also AEDES AEGYPTI and DNA BARCODING.)

nutrient agar A MEDIUM containing peptone, sodium chloride, beef extract and 1.5–2% agar. It supports the growth of e.g. *Escherichia coli*; many species will not grow on this (*basal*) medium unless serum or other enrichment is added.

O

***O*^{6}-alkylguanine-DNA alkyltransferase** (AGT) See the entry GENE FUSION (section: *Uses of gene fusion*).

O157:H7 A (serologically defined) strain of enterohemorrhagic *Escherichia coli*: see the entry EHEC.

(Note the upper-case 'O' – not zero (0) – in the designation of this strain.)

OB Occlusion body: see BACULOVIRIDAE.

***ob* gene** (*OB* gene, OB gene etc.) A gene encoding the signaling molecule *leptin*; mutation in the *ob* gene has been associated with obesity.

DNA sequence homology in this gene, in various species, is shown in the entry DNA SKYLINE (figure).

(See also BARDET–BIEDL SYNDROME.)

obesity (genetic aspects) See e.g. BARDET–BIEDL SYNDROME and OB GENE.

***o*^{c} mutant** See LAC OPERON.

occlusion body (OB) See BACULOVIRIDAE.

occlusion-derived virions (ODVs) See BACULOVIRIDAE.

ochre codon See NONSENSE MUTATION.

ochre suppressor See SUPPRESSOR MUTATION.

OCT plasmid A large (∼500 kb) plasmid, found in strains of *Pseudomonas*, which confers the ability to metabolize octane and decane.

Oct4 (transcription factor) See STEM CELL.

octopine See CROWN GALL.

ODC ORNITHINE DECARBOXYLASE (q.v.).

ODVs Occlusion-derived virions: see BACULOVIRIDAE.

OFAGE Orthogonal-field-alternation gel electrophoresis: one of a number of variant forms of PFGE which is used for the separation of large pieces of nucleic acid.

ofloxacin See QUINOLONE ANTIBIOTICS.

Okayama–Berg method An early method used for obtaining a full-length cDNA – in a known orientation within a vector – from an mRNA molecule with a poly(A) tail. Essentially, the vector was built around the mRNA in sequential steps.

The initial step involved base-pairing between the poly(A) of the mRNA and a poly(T) overhang present on an (added) fragment of dsDNA. The (3′) poly(T) overhang was extended to form the FIRST STRAND (q.v.).

The first strand was subjected to 3′ homopolymer tailing to enable base-pairing with (another) segment of dsDNA having a complementary 3′ overhang; the other end of this (second) segment was cut with a restriction endonuclease, enabling its ligation to the free end of the first fragment.

Okazaki fragment One of a large number of short, discontinuous sequences of deoxyribonucleotides synthesized during the formation of the lagging strand in DNA replication; each fragment is about 100–200 nt in eukaryotes (but reported to be ∼1000–2000 nt in *Escherichia coli*). These fragments are joined together to form the lagging strand.

In one model, the formation of an Okazaki fragment begins (in eukaryotes) with a hybrid RNA/DNA primer that is synthesized by the complex primase–DNA polymerase α. This primer consists of ∼10 nt of RNA and ∼20 nt of DNA (and has been referred to as iRNA/DNA); it is extended with deoxyribonucleotides – by DNA polymerase δ (pol δ) – until it reaches the 5′ terminus of the last Okazaki fragment. Pol δ continues to synthesize DNA, its strand-displacement activity displacing the iRNA/DNA of the (downstream) fragment as a 5′ 'flap'; this flap is excised before the (newly synthesized) Okazaki fragment is joined to the previous one. In the yeast *Saccharomyces cerevisiae*, excision of this flap may involve mainly *flap endonuclease 1* (FEN1) [J Biol Chem (2004) 279 (15):15014–15024]. In a (supplementary) mechanism for flap removal (in *S. cerevisiae*), the flap is coated with *replication protein A* (RPA); this promotes flap cleavage by the Dna2p helicase/nuclease protein.

(See also IDLING.)

oligo Abbreviation for 'oligonucleotide'. This abbreviation can be used only if the meaning is clear (otherwise there could be confusion with 'oligopeptide' etc.).

oligo(dT)-cellulose separation A procedure used e.g. to isolate (poly(A)-tailed) molecules of mRNA from a cell extract that contains total RNA. The extract is run through a column that contains immobilized oligo(dT) – oligonucleotides consisting solely of deoxythymidine monophosphate – and this selectively binds the poly(A) tails of mRNA molecules; the other species of RNA (e.g. ribosomal RNA) are therefore eliminated. The bound mRNAs are subsequently eluted from the column.

Oligofectamine™ A reagent (Invitrogen, Carlsbad CA) which is used e.g. for transfecting cells with oligonucleotides – see LIPOFECTION.

oligomer Any entity consisting of a relatively small number of units or subunits which may be e.g. nucleotides or amino acid residues; thus, an oligomer may be a strand of DNA or RNA that consists of (say) less than ∼10–50 nucleotides.

oligonucleotide A term that commonly refers to a short strand of DNA or RNA – e.g. a strand of 5–50 nt in length.

oligonucleotide-based therapy See SPLICE-SWITCHING OLIGONUCLEOTIDE.

oligonucleotide-directed mutagenesis See (the figure in) SITE-DIRECTED MUTAGENESIS.

oligonucleotide–peptide interactions See SOLVATOCHROMIC FLUORESCENT DYE.

(See also DIELS–ALDER REACTION.)

OliGreen® A fluorescent dye (Molecular Probes Inc., Eugene OR/Invitrogen, Carlsbad CA) used e.g. for the quantitation of ssDNA in solution. The sensitivity of the method is reported to be ∼10000-fold greater than that obtainable with measurements made by the ULTRAVIOLET ABSORBANCE method.

[Use of Oligreen® for quantitation of an (ss) cDNA library: BMC Genomics (2009) 10:219.]

(See also DNA STAINING and PICOGREEN ASSAY.)

omega (ω) protein See TOPOISOMERASE (type I).

OMIA Online Mendelian inheritance in animals: a database of

genes and inherited disorders/traits in animals *other than* man and mouse. The database can be reached via: http://www.ncbi.nlm.nih.gov/Literature/

OMIM® See entry ONLINE MENDELIAN INHERITANCE IN MAN.

***ompT* gene** In *Escherichia coli*: a gene encoding the protease OmpT, which is part of the (wild-type) cell's *outer membrane* (OM) – i.e. the outermost structure of the cell wall.

Some proteins secreted by *E. coli* comprise two parts, one of which forms a channel in the OM through which the other part is translocated to the cell's surface. This type of protein secretion is called an *autotransporter* system (it is also called type IV secretion by some authors, and type V secretion by others). The surface-exposed portion of the protein may be cleaved by the OmpT protease; this portion of the protein may then be released into the environment.

[*E. coli* autotransporters: Appl Environ Microbiol (2007) 73(5):1553–1562; EMBO J (2007) 26:1942–1952.]

In *Pseudomonas aeruginosa* an autotransporter is reported to be needed e.g. for cell motility and biofilm formation [J Bacteriol (2007) 189(18):6695–6703].

on-chip PCR See SOLID-PHASE PCR.

on-chip synthesis See MICROARRAY.

oncogene A gene which, if aberrantly expressed, may cause neoplastic transformation ('cancer'). Oncogenes were initially found in acutely oncogenic retroviruses, and subsequently, highly conserved homologs were found in a wide range of normal eukaryotic cells; these included human, fruitfly (*Drosophila*) and yeast (*Saccharomyces cerevisiae*) cells. Retroviral oncogenes appear to have derived from cellular genes.

An oncogene is usually given a three-letter designation that is based on the name of the retrovirus in which it was first identified. Two examples of a *viral* oncogene are: v-*fes* from the feline sarcoma virus, and v-*myc* from myelocytomatosis virus; the corresponding *cellular* homologs of these genes are c-*fes* (or proto-*fes*) and c-*myc* (or proto-*myc*) respectively.

At least some oncogenes have role(s) in the regulation and/or development of normal (i.e. non-neoplastic) cells (see e.g. MYB). Oncogenesis may arise as a result of dysregulation of these genes – e.g. through mutation, insertional activation, or chromosome rearrangements (e.g. ABL).

Viral oncogenes can cause cancer in various ways. In some cases the product of a viral gene may interfere with the cell's growth-related reactions which normally involve factors such as epidermal growth factor (EGF) or platelet-derived growth factor (PDGF). The v-*sis* product is almost identical to the B chain of PDGF. Some products may be active in the nucleus.

(See also entries for some individual oncogenes: ABL, FES, FPS, KIT, MYB, MYC, RAS, SIS, SRC.)

Cancers resulting from dysregulated tyrosine kinase activity have been treated with specific drugs, although resistance to these agents can arise through mutation in the oncogene. The problem of drug-resistance, similar to that in pathogenic bacteria, requires the ongoing development of new drugs.

Oncogenes occur in some DNA viruses – e.g. mastadenoviruses, polyomaviruses – although these oncogenes appear to have no cellular homologs.

(See also ZNF217.)

oncogenesis The development of a neoplastic condition such as a tumor or leukemia.

(See also NEOPLASIA.)

oncolytic virus Any virus with an innate, or engineered, ability to infect, and kill, cancer cells. Oncolytic viruses include e.g. strains of ADENOVIRUS and herpesvirus.

Engineered oncolytic viruses include the 'cancer terminator virus', a recombinant adenovirus with activity against cells of MELANOMA (q.v.).

Herpes simplex virus HSV1716 was modified by engineering its genome so that the virion's envelope includes an anti-CD38 scFV (SINGLE-CHAIN VARIABLE FRAGMENT); this gives the virus greater infectivity for $CD38^+$ human leukemic cells. HSV1716, displaying an anti-EGFR (epidermal growth factor receptor) scFV, exhibited enhanced destruction of ($EGFR^+$) tumors in mice on intravenous/intraperitoneal injection [Gene Ther (2008) 15:1579–1592].

An oncolytic adenovirus expressing a SHORT HAIRPIN RNA that targeted the mRNA of Ki-67 (see KI-67) has been used in murine renal cancer cells [Cancer Gene Therapy (2009) 16: 20–32].

(See also CANCER THERAPY and MDA-7/IL-24.)

oncomir (oncomiR) Any microRNA (see MICRORNAS) with a role in CANCER.

In certain cases, dysregulation of an miRNA – e.g. through mutation in its coding sequence – has been found to correlate with human cancer.

One study found that the expression of various miRNAs of the ONCOMIR-1 family was highest in cases of Wilms' tumor, as compared with other types of kidney tumor, and that these miRNAs were overexpressed relative to the levels found in normal kidney tissue [Cancer Res (2008) 68(11):4034–4038].

Oncomir-1 A family of oncogenic MICRORNAS usually understood to include the cluster miR-17 to miR-92; many or all of these miRNAs are found on chromosome 13.

[The essential and overlapping functions of the miR-17∼92 family: Cell (2008) 132(5):875–886.]

oncornaviruses The former name of viruses of the subfamily ONCOVIRINAE.

Oncovirinae A subfamily of viruses (family Retroviridae – see RETROVIRUSES); it includes all the oncogenic viruses within the Retroviridae. Some apparently non-oncogenic but related viruses have also been included in the subfamily. Members of the Oncovirinae were subdivided into types according to e.g. their morphology and mode of development.

one-tube protocol (in nested PCR) See NESTED PCR.

Online Mendelian Inheritance in Man (OMIM®) A database of human genes and genetic disorders [see Nucleic Acids Res (2009) 37(Database Issue):D793–D796] accessible at: http://www.ncbi.nlm.nih.gov/omim

ONPG The compound *o*-nitrophenyl-β-D-galactopyranoside; it is hydrolyzed by the enzyme β-galactosidase to galactose and

the (yellow) compound *o*-nitrophenol.

ONPG is used e.g. as an indicator for the presence of β-galactosidase in the SOS CHROMOTEST and in other tests. In some species (e.g. *Escherichia coli*) ONPG can pass through the cell envelope (i.e. it can enter cells) without the need for a specific permease.

***opa* genes** (of *Neisseria*) See OPA PROTEINS.

(See also PHASE VARIATION.)

Opa proteins Cell-surface antigens encoded by the *opa* genes in pathogenic strains of *Neisseria*; they affect the opacity and color of the bacterial colonies and bind to epithelial cells and polymorphonuclear leukocytes (e.g. via eukaryotic CD66a, c, d or e adhesins).

opal codon See NONSENSE MUTATION.

open complex See PROMOTER.

open reading frame (ORF) A sequence of nucleotides which starts with an initiator codon and ends with a stop codon – and which encodes an actual or potential polypeptide/protein or an RNA product. A reading frame is 'blocked' if it contains a stop codon close to the initiator codon.

Adjacent ORFs (which may be separated by a sequence of nucleotides) may *jointly* encode a polypeptide; in such cases, translation of the complete polypeptide requires a mechanism for coupling the ORFs [see: EMBO J (2000) 19:2671–2680].

operon Two or more contiguous genes coordinately expressed from a common promoter and transcribed as a polycistronic mRNA (see e.g. LAC OPERON).

The operon is sometimes regarded as a regulatory system specific to prokaryotic gene organization but a similar type of organization has been found in trypanosomes and (in some instances) in genes of the nematode *Caenorhabditis elegans* (see also TRANS SPLICING).

The following refers specifically to operons in *bacteria*.

The control mechanism varies according to operon. Some operons are controlled at the level of transcription, and some are controlled at the level of translation.

Some operons are regulated by CATABOLITE REPRESSION.

Each operon is subject to one or more modes of control:

Promoter control

This involves the synthesis of a regulator protein, which may be produced constitutively. One example of promoter control is the LAC OPERON of *Escherichia coli*. This operon is under *negative control*; this means that the regulator protein inhibits (or blocks) expression of the operon. In the *lac* operon, the regulator (repressor) protein binds to the so-called operator region immediately downstream of the binding site of RNA polymerase, thus inhibiting transcription of the operon.

An example of *positive control* is the *araBAD* operon of *E. coli* which encodes enzymes involved in the metabolism of L-arabinose to D-xylulose 5′-phosphate. If arabinose is present, the regulator protein, AraC (encoded by gene *araC*), acts as an *activator* of the *araBAD* operon – i.e. positive control. In the absence of arabinose (metabolic enzymes not needed) the regulator protein acts as a repressor (negative control).

Attenuator control

Operons under attenuator control are typically involved in the synthesis of amino acids. In the mRNA transcript, the (upstream) sequence of nucleotides (called the *leader* sequence) encodes a short *leader peptide* that is rich in the given amino acid whose synthesis is regulated by that operon. The leader region in the transcript also includes a so-called *attenuator*: a rho-dependent terminator of transcription located between the first gene of the operon and the sequence encoding the leader peptide.

Given adequate levels of the relevant amino acid (synthesis not required), the small leader peptide is synthesized by the ribosome (which is following behind the RNA polymerase); when this happens, the transcript adopts a conformation that stops transcription at the attenuator. However, if the level of amino acid is not adequate (synthesis needed), the ribosome stalls where the transcript contains codons of the given amino acid; this permits downstream parts of the transcript to base-pair – thus preventing formation of the attenuator and, hence, permitting ongoing transcription and translation of the genes.

Attenuator control is found e.g. in the *his* operon of *E. coli* – which regulates synthesis of histidine. This operon contains nine structural genes. The mRNA sequence that encodes the leader peptide includes seven successive histidine codons.

The *trp* operon of *E. coli* (synthesis of tryptophan) is under both attenuator and (negative) promoter control. This operon may be more stable than those operons with a less complex regulatory system [Microbiol Mol Biol Rev (2003) 67:303–342]. (See also the item *Optimization of transcription* in the entry OVEREXPRESSION.)

Translational control

In *E. coli*, genes *infC–rpmI–rplT* form an operon encoding, respectively, the translation initiation factor IF3 and the two ribosomal proteins L35 and L20. High levels of L20 repress the translation of both L35 and L20. L20 appears to bind to the mRNA upstream of the *rpmI* region, and it has been suggested that it stabilizes an RNA pseudoknot that includes the initiator codon – thus inhibiting translation of both *rpmI* and *rplT*. (cf. TRANSLATIONAL ATTENUATION (sense 1).)

operon fusion The engineered fusion of two operons such that genes of both operons are under the control of the regulatory regions (promoter, operator etc.) of one of the operons, or of a separate control system.

opine See CROWN GALL.

Op*M*NPV See NUCLEAR POLYHEDROSIS VIRUSES.

optical tweezers (also called: single-beam gradient force optical trap) A system in which the energy in a laser beam is used to *move* a cell or a micrometer-sized dielectric particle, or (e.g.) to *hold it* in a fixed (three-dimensional) position, either in air (or other gas(es)) or within a liquid; such 'optical trapping' is used in physical and biological procedures. [Optical tweezers for single cells: J R Soc Interface (2008) 5(24):671–690.] In this system, a particle – in the focus of a laser beam from an objective lens of high NA (numerical aperture) – scatters the incident photons and, in doing so, experiences forces which combine to move the particle or to maintain its fixed location.

(cf. MAGNETIC TWEEZERS.)

In DNA technology, optical tweezers have been used e.g. to fix the position of a DNA-bound bead while other forces are used to extend ('stretch') the DNA molecule: see SURFACE-INDEPENDENT TETHERING.

optimization (of gene expression) See e.g. the entries CODON HARMONIZATION, CODON OPTIMIZATION and OVEREXPRESSION.

ORC Origin recognition complex: see DNAA GENE.

Orc1–6 proteins See DNAA GENE.

ORF OPEN READING FRAME (q.v.).

ORFeome A set of open reading frames – see OPEN READING FRAME – representing many, most or all ORFs from a given species, often stored as inserts in a collection of plasmids. An ORFeome is useful as a source of material for functional analysis; thus, for example, a given ORF may be inserted into an expression vector in order to study an encoded protein. To facilitate use, ORFs are commonly inserted into a construct from the GATEWAY SITE-SPECIFIC RECOMBINATION SYSTEM – such as PDONR (q.v.) – to form entry clones; an ORFs in an entry clone is then inserted into a destination vector. This system was used with a *Staphylococcus aureus* ORFeome (2562 entry clones – 95% coverage) [BMC Genomics (2008) 9:321], and a *Caulobacter* ORFeome [Proc Natl Acad Sci USA (2009) 106(19):7858–7863]. [Use of RACE for defining an ORFeome of *Caenorhabditis elegans*: Genome Res (2009) doi: 10.1101/gr.098640.109.]

ORFmer sets Commercially available (Sigma-Genosys) sets of primer pairs designed to enable the PCR-based amplification of sequences from most or all ORFs (open reading frames) in the genomic DNA of a particular organism. The primers contain ORF-specific sequences.

The 5′ terminus of each primer consists of an *adaptamer*: a sequence that incorporates the recognition site of a particular restriction endonuclease; this facilitates insertion of a given PCR-amplified product into an appropriate vector.

[Use (e.g.): J Bacteriol (2009) 191(7):2296–2306.]

orientation (*DNA technol*.) The *direction* in which a given sequence is found, or inserted, in relation to other sequence(s) or in relation to the other parts of a sequence; for example, when using FLP recombinase, a target fragment which is bracketed by two recognition sequences (FRT sites) that are in the same orientation will be excised – but the target will be inverted if the two FRT sites are in opposite orientation.

origin licensing See RE-REPLICATION.

origin recognition complex (ORC) See DNAA GENE.

oriT The origin of transfer (in a conjugative plasmid): see e.g. CONJUGATION.

ornithine decarboxylase (ODC) See e.g. ANTIZYME and PEST DOMAIN.

orthologous genes Functionally analogous genes found within different species.

orthomere See MACRONUCLEUS.

Orthomyxoviridae A family of pleomorphic enveloped viruses in which the (segmented) genome consists of negative-sense ssRNA. (cf. PARAMYXOVIRIDAE.) These viruses are typically pathogens of the upper respiratory tract. Important pathogens in this family include influenza viruses A and B – which are classified in the genus INFLUENZAVIRUS (q.v.).

***Orygia pseudotsugata* NPV** A baculovirus within the category NUCLEAR POLYHEDROSIS VIRUSES (q.v.).

oseltamivir An inhibitor of viral NEURAMINIDASE used for the treatment of influenza caused by influenza virus A or B.

[Oseltamivir-resistant influenza virus A (H1N1) in Norway (2007–2008): Emerg Infect Dis (2009) 15(2):155–162.]

(See also *swine flu* in the entry INFLUENZAVIRUS.)

outer primer See NESTED PCR.

overexpression (*syn*. overproduction) (*DNA technol*.) The synthesis of high concentrations of a particular protein, usually a heterologous ('foreign') protein, in a recombinant organism – either a prokaryotic or eukaryotic organism; this procedure is used e.g. for manufacturing various products, especially those for therapeutic/diagnostic use, but also for research purposes. (See also BIOPHARMACEUTICAL.)

(***Note***. In *nature*, overexpression of proteins can occur e.g. when certain genes undergo GENE AMPLIFICATION (q.v.); this entry is concerned specifically with the *in vitro* processes in DNA technology.)

Proteins are synthesized in bacteria (particularly the Gram-negative species ESCHERICHIA COLI), yeasts (such as *Pichia pastoris*, *Saccharomyces cerevisiae*) and in cells from insects (such as *Bombyx mori*, *Spodoptera frugiperda*) and mammals – such as hamster (CHO: Chinese hamster ovary) and BHK (baby hamster kidney).

(See also BACULOVIRUS EXPRESSION SYSTEMS.)

Choice of cells

The choice of cells is influenced by factors such as the need for specific type(s) of post-translational modification in the recombinant product – and/or the preferred secretion of the recombinant protein by the producing cells (rather than intracellular accumulation) in order to facilitate the recovery and purification of the product.

For proteins produced on an industrial scale, the important factors include the need to maximize the yield and the ease of downstream processing of the product. Downstream processing may be simplified by choosing cells that secrete the given protein (see below).

Another consideration is the safety of the product; thus, for example, if a recombinant protein is synthesized in *E. coli*, or in other Gram-negative bacteria, there is a need for rigorous exclusion of endotoxin (see entry PYROGEN). Moreover, any agent used for the induction of transcription in the producing cells should not be associated with a risk of toxicity when the recombinant product is intended for therapeutic use.

Optimization of transcription

An obvious requirement is a strong PROMOTER to control the expression of the target gene. Moreover, transcription should be *tightly regulated* – meaning that the target gene should not be expressed prior to the time of its intended induction or de-repression. One reason for this is that the cells are commonly

grown to high density before expression of the target gene in order to maximize the yield of product. Were the gene to be expressed prior to the appropriate time then there may be a depression in the growth rate with a possible decrease in the final yield of recombinant protein.

Another reason for tight regulation is found in those cases where the recombinant protein is toxic to the producing cells; in these cases, an early expression of the recombinant protein could be expected to decrease final cell density – and, hence, yield.

Transcription of a given gene can be controlled (switched on or off) in various ways. For one method see the entry LAC OPERON. One alternative approach makes use of a promoter whose function is blocked if the concentration of a particular metabolite is above a certain level. Promoters like this are found in some of the operons involved in synthesis of amino acids; for example, the promoter of the *trp* operon – involved in tryptophan synthesis – switches off when the concentration of exogenous tryptophan exceeds a certain level (i.e. when the cell does not require to synthesize any more tryptophan). Thus, transcription via the *trp* promoter can be switched on (when required) by reducing the concentration of tryptophan in the medium to an appropriate level (which is still sufficient to permit protein synthesis).

Optimization of translation

In designing a recombinant gene there is a need to ensure that the characteristics of the *transcript* are optimal for the given translation system. Thus for example, the efficient translation of mRNA in *Escherichia coli* demands attention to factors that include the precise composition of the ribosome-binding Shine–Dalgarno (SD) sequence – and the number of nucleotides separating the SD sequence from the start codon. Moreover, optimal translation may depend on the provision of a translational enhancer ('downstream box') located close to the start codon. Failure to optimize these features may result in a marked reduction in the yield of recombinant protein.

Codon bias

The synthesis of a heterologous protein (for example, when a mammalian gene is expressed in *E. coli*) may be inefficient owing to CODON BIAS (q.v.).

(See also CODON OPTIMIZATION.)

Inclusion bodies

Inclusion bodies are insoluble aggregates of the unfolded, or incorrectly folded, heterologous protein which may develop within the cytoplasm of overexpressing cells. In *E. coli*, the formation of inclusion bodies can sometimes be inhibited e.g. by using a lower growth temperature (say, 30°C instead of 37°C) or by using the 'thiofusion' procedure (see below).

Another approach to the problem of inclusion bodies is to co-express *chaperone* proteins in overexpressing cells. These are normal constituents of cells that promote correct folding of nascent proteins. (See also CODON HARMONIZATION.)

[The use of 'folding modulators' to improve the production of heterologous proteins in *E. coli*: Microbial Cell Factories (2009) 8:9.]

If the inclusion bodies can be correctly folded *in vitro* (i.e. after isolation from the cells) then the formation of inclusion bodies can be an advantage in that they may facilitate purification of the product, i.e. by simplifying separation of product from cell debris etc.

The efficiency with which inclusion bodies can be released from disrupted cells may be assessed by a fluorescence-based method, and this information can be useful for developing the system to maximize yields of recombinant protein [BioTechniques (2007) 43(6):777–782].

The problem of proteolysis

Proteolysis of recombinant proteins in the cytoplasm of *E. coli* may be minimized in several ways. Thus, for example, the given protein may be targeted to the cell's *periplasm* (i.e. the region between the cytoplasmic membrane and the cell wall) – in which there are fewer proteolytic enzymes. This is possible e.g. by fusing the gene of interest with the gene of a periplasmic protein (such as DsbA).

Alternatively, the target protein may be made secretable by fusing its gene to the gene of a secreted protein.

A further possibility is re-coding the gene to eliminate particular proteolytic sites in the protein.

Additionally, use may be made of *rpoH* mutant cells of *E. coli*. RpoH can promote synthesis of the Lon protease (which degrades abnormal proteins), and these mutants have been found to give increased yields of recombinant proteins in *E. coli*.

Post-translational modification

Many proteins (particularly those used therapeutically) are non-functional unless they have appropriate post-translational modification, such as the glycosylation of specific amino acid residues; hence, only cells which are capable of such activity are used for the synthesis of these proteins. For this reason, a number of the commercial therapeutic agents are produced in e.g. Chinese hamster ovary cells or baby hamster kidney cells – i.e. mammalian cells which can carry out the correct forms of post-translational modification.

In some cases the absence of glycosylation has little or no effect on normal biological activity; for example, unglycosylated recombinant interleukin-2 (IL-2) has essentially normal biological activity compared with the natural protein, and this agent can be produced satisfactorily in *E. coli*. However, if a specific type of post-translational modification is essential, it may be possible to use a transgenic producer organism: see e.g. 'glycoengineered' *Pichia pastoris* in GLYCOSYLATION.

Secreted/intracellular proteins

The *secretion* of a recombinant protein can be advantageous in a production process because it tends to simplify isolation and purification of the product. Thus, if a protein remains intracellular it is necessary to lyse the cells in order to harvest the product; this entails (i) extra work to ensure efficient lysis of cells, and (ii) the need to separate the (one) target protein from the cell's own natural proteins (as well as from the other types of cell debris).

The production of recombinant proteins intracellularly can

be a problem particularly in *E. coli* and insect cells.

One approach to the problem of intracellular recombinant proteins is to design the expression vector in such a way that the recombinant gene is fused to another gene, or sequence, which promotes secretion/release of the fusion protein. When isolated *in vitro*, the product can be cleaved (with a suitable protease) to release the protein of interest.

In *E. coli*, the so-called thiofusion approach can yield extracellular recombinant products. In this process, the expression vector contains the gene of interest fused with the gene of the protein THIOREDOXIN. This protein is found normally in the cell envelope, and the fusion protein can be released from the cells simply by subjecting the cells to osmotic shock (which disrupts the outer membrane).

(See also SECRETION (of proteins).)

Detection/isolation/purification of recombinant proteins

In vitro detection, isolation and purification of recombinant proteins can be facilitated by using expression vectors which encode certain types of peptide/protein tag that form a fusion product with the protein of interest. For example, the LUMIO tag provides fluorescent detection of the recombinant protein, while the calmodulin-binding peptide (CBP) of the Affinity® system (Stratagene, La Jolla CA, USA) helps in the isolation of a protein by binding that protein to a CALMODULIN-based adsorption resin (see also the entry SIX-HISTIDINE TAG). The KEMPTIDE SEQUENCE permits robust labeling of the protein product.

Separation of the protein of interest from its tag may be achieved e.g. by using an endopeptidase which has a known cutting site (see e.g. ENTEROKINASE; see also TEV protease in the entry TEV).

(See also RAINBOW TAG.)

Quantitation of recombinant proteins

See e.g. Q-TAG (sense 1).

overhang See STICKY ENDS.

overlapping genes Two or more genes that share sequence(s) of nucleotides, part (or all) of a given gene being coextensive with part of another. This arrangement serves e.g. to maximize the amount of information carried by a genome. Such genes may differ e.g. in being translated in different reading frames, or they may be translated in the same reading frame but with different start/stop signals and/or a different pattern of splicing.

Overlapping genes are present in viruses, prokaryotes and eukaryotes (including mammals).

The region of overlap between genes may be on the same strand or on the complementary ('opposite') strand.

[(Same-strand) overlapping genes in bacteria: Biol Direct (2008) 3:36.]

[Overlapping genes in human and mouse genomes: BMC Genomics (2008) 9:169.]

[Database for evolutionary analysis of overlapping genes: Nucleic Acids Res (2009) 37(Database issue):D698–D702.]

overproduction (of proteins) *Syn*. OVEREXPRESSION (q.v.).

oxazolidinone antibiotics A group of synthetic antibiotics that inhibit protein synthesis (primarily in Gram-positive bacteria) by binding to 23S rRNA in the 50S ribosomal subunit – a site similar to the target site of chloramphenicol; these antibiotics, which include eperezolid and linezolid, are effective against e.g. sensitive strains of *Streptococcus pneumoniae*, staphylococci and enterococci.

In *S. pneumoniae*, resistance to linezolid has been reported to involve mutation in the gene encoding a methyltransferase which, in linezolid-sensitive strains, normally methylates the G2445 site in the ribosomal 23S rRNA [Genome Res (2009) 19:1214–1223].

8-oxoG 7,8-dihydro-8-oxoguanine: a mutagenic base produced in DNA e.g. by reactive oxygen species; if unrepaired (by the BASE EXCISION REPAIR system), it may cause a GC-to-TA transversion mutation as 8-oxoG (in the template strand) may pair with adenine during DNA replication.

During DNA replication, adenine misincorporated opposite a template-strand 8-oxoG can be excised by the *Escherichia coli* MutY DNA glycosylase (or by the human homolog of MutY, hMUTYH). The product of a putative MutY-encoding gene from *Helicobacter pylori* apparently complements *mutY* strains of *E. coli* [J Bacteriol (2006) 188 (21):7464–7469].

In duplex DNA, 8-oxoG can be excised from 8-oxoG:C pairs by the *E. coli* DNA glycosylase MutM (= Fpg protein).

The *E. coli* MutT protein prevents the incorporation of free 8-oxoG into DNA by eliminating it from the pool of oxidized guanines.

Structural studies of the human DNA polymerase κ (kappa) inserting adenine opposite an 8-oxoG in DNA are reported to show why polymerase κ is more efficient at inserting adenine (rather than cytosine) opposite 8-oxoG, and may also help to explain why pol κ is more efficient than other Y-family DNA polymerases in this role [PLoS ONE (2009) 4(6):e5766].

Yeast DNA polymerases δ (delta) and η (eta) both insert an adenine opposite 8-oxoG, and the efficiency and fidelity with which such 'bypass' occurs has been investigated [Nucleic Acids Res (2009) 37(9):2830–2840].

oxolinic acid See QUINOLONE ANTIBIOTICS.

OxyS A bacterial sRNA: see the entry SRNAS.

P

p (1) A prefix often used to indicate a recombinant PLASMID (e.g. pBR322 or any of a large number of commercial vectors such as pBluescript®).

(2) An indicator of the *short* arm of a CHROMOSOME; it is used when giving a particular location on a specific chromosome – e.g. 4p16, which refers to a location (16) on the short arm of chromosome 4.

(3) A prefix used to denote a particular polypeptide product of a retroviral gene.

P Proline (alternative to Pro).

P element Any of a range of transposable elements (up to ~3 kb) found in certain strains of *Drosophila*. Excision occurs at the flanking 31-bp inverted repeats, and an 8-bp duplication is created at the insertion site. Transposition occurs in germ line cells.

P elements are used e.g. for mutagenesis.

P-gel A system used for CELL-FREE PROTEIN SYNTHESIS (q.v.).

P site (of a ribosome) The 'peptidyl' site at which the initiator tRNA binds during translation.

(cf. A SITE.)

P1 plasmid The circular (extrachromosomal) form of the prophage of PHAGE P1.

pIII protein See PHAGE DISPLAY.

p21 protein See RAS.

p27 protein A eukaryotic protein which inhibits cyclin-dependent kinases and which can affect transition between phases of the cell cycle.

(See also IRES.)

p53 A gene (in human chromosome 17p) whose product arrests the cell cycle and induces APOPTOSIS following damage to DNA. Loss of *p53* activity leads to genetic instability and the survival of DNA-damaged cells.

p53 is a major tumor-suppressor gene which can e.g. inhibit oncogene-mediated transformation in cultured cells.

(See also NUCLEOPHOSMIN.)

p53 is frequently mutated in human cancers, and details of mutations in *p53* are saved in several databases (for example: www-p53.iarc.fr).

p53 is also involved in the regulation of specific groups of microRNAs (miRNAs) [see e.g. Mol Syst Biol (2008) 4:231].

The transcriptional control of human p53-regulated genes has been reviewed [Nature Rev Mol Cell Biol (2008) 9:402–412]. (See also METHYLATION.)

p53 protein has a short half-life – and is often undetectable immunohistochemically. Mutation may extend the half-life – so that detection of p53 may indicate a mutant allele.

A scintillation proximity assay was used to assay binding between p53 and DNA [BioTechniques (2006) 41(3):303–308].

Few, if any, of the SNPs in *p53* seem to be associated with perturbation of the function of p53; those polymorphisms that might have a negative role – as well as other polymorphic genes in the p53 pathway – have been reviewed [Nature Rev Cancer (2009) 9:95–107]. No association was found between *p53* polymorphisms and primary open-angle glaucoma in a Japanese population [Mol Vision (2009) 15:1045–1049].

P210 protein See the entry ABL.

PA-457 *Syn*. BEVIRIMAT.

PACE PCR-assisted contig extension. Following SHOTGUN SEQUENCING and the construction of clone contigs, PACE is used for closing *gaps* (i.e. regions in the genome for which no cloned fragments of DNA are available). Essentially, PCR is used (with genomic DNA) to amplify a sequence that extends from near the end of a contig into the (unknown) gap region. The outward-facing primers bind specifically to the known region near the end of the contig. The inward-facing primers have arbitrary sequences. In some cases an arbitrary primer will bind (with mis-matches) to a sequence within the unknown region at an amplifiable distance from the specific primer; the products formed can be sequenced, thus extending the known region outwards from the end of the contig.

PACE 2C A PROBE-based test (Gen-Probe, San Diego CA) for detecting the bacterial pathogens *Chlamydia trachomatis* and *Neisseria gonorrhoeae* in clinical samples in a single assay; the probe includes sequences specific to each of these organisms – the target sequences being in rRNA. Hybridization of the probe to target(s) is indicated by a hybridization protection assay (see the entry ACCUPROBE for details of the assay).

A positive PACE 2C screening test can be followed by two separate tests for *C. trachomatis* and *N. gonorrhoeae*. This is a particularly useful approach because co-infection with both pathogens is quite common.

This is one example of a *combination probe test*.

packaging (of phage DNA) Insertion of (phage) genomic DNA into the phage capsid during formation of an infective phage virion: a process that differs in different phages and which, in some cases, is not fully understood.

In PHAGE LAMBDA (λ), genome-size pieces of DNA are cut from a concatemer produced by ROLLING CIRCLE replication. Cuts occur at regularly repeated *cos* sites in the concatemer; a terminal *cos* site locates in the phage capsid, and DNA is fed into the capsid until the next *cos* site – when cleavage occurs. Thus, a capsid contains the DNA between two consecutive *cos* sites, and the packaged genome has terminal sticky ends.

Commercial products are available for packaging the phage λ genome: see e.g. GIGAPACK.

In phage φ29, packaging is dependent on energy from ATP hydrolysis. Moreover, the mechanism of insertion involves a 'motor' which contains various phage-encoded products: (i) a *connector* (consisting of 12 subunits) with a central channel, (ii) six molecules of RNA (which are designated pRNA), and (iii) a protein (gp16) whose copy number is not known. The motor may wind DNA into the capsid rather like a bolt drawn through a rotating nut. Two models for the packaging system in phage φ29 have been suggested. In one model, the pRNA molecules are attached to the capsid – and form part of the

stator; in another model a pRNA–connector complex forms the *rotor*. Studies using photoaffinity cross-linking indicated that pRNA molecules bind to the N-terminal amino acids of proteins in the connector.

packaging cell ('producer cell') Any cell in which replication-deficient virions of recombinant viral vectors are assembled with help from PACKAGING PLASMIDS: see e.g. AAV HELPER-FREE SYSTEM, VIRAPORT RETROVIRAL EXPRESSION SYSTEM, VIRAPOWER LENTIVIRAL EXPRESSION SYSTEM and also the vector pLenti6/V5-DEST™ (in the table) in entry GATEWAY SITE-SPECIFIC RECOMBINATION SYSTEM. Virions produced in these cells are used to infect the *target* cells, within which transfected DNA can be expressed.

In certain cases, expression of the gene of interest within packaging cells can reduce the final titer of virions produced; this can occur, e.g. if the gene of interest encodes a cytotoxic protein. Various approaches have been used to deal with this problem. In one method, the titer of recombinant adenovirus was improved 25-fold by using a *trans*-encoded suppressor which reduced transcription of the transgene [BioTechniques (2008) 44(1):85–89].

packaging plasmid A plasmid encoding certain component(s) of a virion (e.g. capsid proteins) which is co-transfected into cells of a 'producer cell line' (see PACKAGING CELL) together with a plasmid that encodes (i) a replication-deficient part of a viral genome and (ii) the gene/fragment of interest. The (replication-deficient) viral particles that are produced in the packaging cells are used to infect *target* cells – within which the gene of interest can be expressed.

Packaging plasmids are used e.g. in certain vector systems – see e.g. the lentiviral vector pLenti6/V5-DEST™ in entry GATEWAY SITE-SPECIFIC RECOMBINATION SYSTEM (table).

(See also AAV HELPER-FREE SYSTEM, VIRAPORT RETROVIRAL GENE EXPRESSION SYSTEM and VIRAPOWER LENTIVIRAL EXPRESSION SYSTEM.)

paclitaxel A diterpenoid antitumor agent used e.g. in the treatment of breast and ovarian cancers. Paclitaxel promotes polymerization of tubulin to microtubles, and inhibits depolymerization, thereby arresting cells in the G2/M phase of the cell cycle.

(See also the note in the entry EPSTEIN–BARR VIRUS.)

pactamycin An antibiotic which inhibits the initiation stage of translation in protein synthesis.

PADAC A violet-colored reagent (a cephalosporin derivative) which forms a yellow product when it is hydrolyzed e.g. by a β-lactamase.

(See also NITROCEFIN.)

pAd/CMV/V5-DEST™ vector A destination vector in the GATEWAY SITE-SPECIFIC RECOMBINATION SYSTEM (see the table in that entry).

padlock probe A linear ssDNA PROBE in which the two *end* sections are complementary to *adjacent* regions of the target sequence; when bound to the target, the two end sections of the probe are therefore juxtaposed and can be ligated – thus circularizing the probe. Ligation is possible only if there is appropriate probe–target complementarity. Non-circularized probes can be eliminated by exonucleases.

The central region of each padlock probe may include a unique identification sequence (the *ZipCode* or *zipcode*); all probes with the same *target specificity* have the same zipcode – and different zipcodes are found on probes with different target specificities. Universal primer-binding sites (which are the same on all the probes, regardless of zipcode) permit the (circularized) probes to be amplified; target-specific products can then be detected by a MICROARRAY of probes consisting of sequences complementary to the zipcode sequences.

This method was used e.g. for the simultaneous detection of a range of plant pathogens [Nucleic Acids Res (2005) 33(8): e70].

Padlock probes have also been used for detecting the SARS virus [J Clin Microbiol (2005) 43(5):2339–2344] and detecting miRNAs [RNA (2006) 12(9):1747–1752], and were used in targeted molecular analyses [see Nucleic Acids Res (2009) 37(1):e7].

Optimized padlock probe ligation, coupled with microarray detection, has been used to test for multiple (non-authorized) GMOs (genetically modified organisms) by means of a single reaction [BMC Genomics (2008) 9:584].

(See also ZIPCODE ARRAY.)

painting (chromosome) See CHROMOSOME PAINTING.

palifermin See Kepivance® in BIOPHARMACEUTICAL (table).

palindromic sequence In a double-stranded nucleic acid molecule, a pair of INVERTED REPEAT sequences such as:

5′. . ...CCATCGATGG. . ...3′
3′. . ...GGTAGCTACC. 5

This region of the double-stranded nucleic acid molecule has twofold rotational symmetry. If the two inverted repeats were separated by a sequence of nucleotides, the region would be said to have *hyphenated dyad symmetry* – for example:

5′. . ...CCATCNNNNNGATGG. . ...3′
3′. . ...GGTAGNNNNNCTACC. . ...5′

An example of a palindromic sequence with hyphenated dyad symmetry is the REP SEQUENCE.

A palindromic sequence may occur as a linear molecule (as shown above) or it may be present as a (cross-like) *cruciform* in which each of the strands forms a *hairpin* by *intrastrand* base-pairing.

Palindromic sequences are commonly present in various sites associated with protein–DNA binding – including e.g. operators, and recognition sequences of certain RESTRICTION ENDONUCLEASES. (See also CRE–LOXP SYSTEM.)

Palindromic sequences of length greater than about 40 bp were reported to be significant target sites for the integration of recombinant AAV vectors (see the entry AAVS).

palindromic unit *Syn*. REP SEQUENCE.

PAMAM Polyamidoamine: a polymer used e.g. in the preparation of certain types of DENDRIMER.

Pan B Dynabeads See DYNABEADS.

Pan T Dynabeads See DYNABEADS.

panning (biopanning) See PHAGE DISPLAY.

Panton–Valentine leucocidin (PVL) A product of some strains of STAPHYLOCOCCUS which lyses the macrophages and the polymorphonuclear leucocytes of certain species (including human and rabbit).

PVL consists of two proteins – protein S (encoded by gene *lukS*) and protein F (encoded by gene *lukF*).

In some reports, the production of PVL has been associated more with community-acquired MRSA than with hospital-acquired MRSA (see MRSA); for this reason, some diagnostic tests have used the *lukS* gene as indicative of CA-MRSA.

pantropize To increase the range of types of (host) cell that can be infected by a given type of virus by PSEUDOTYPING that virus – e.g. by engineering the virion to express the surface proteins of vesiculostomatitis virus (see the entry VSV).

PAP PYROPHOSPHOROLYSIS-ACTIVATED POLYMERIZATION.

PAP-LMPRC See LIGATION-MEDIATED PCR.

papillomaviruses A category of small (∼55 nm), icosahedral, non-enveloped viruses (genome: ccc dsDNA) that infect epithelial cells. More than 100 types of human papillomavirus (HPV) have been identified; of these, some (e.g. HPV6) are considered low-risk (as they give rise e.g. to warts) while others, such as HPV16 and HPV18, are considered high-risk because they can cause e.g. cervical cancer.

Until recently it was believed that the papillomaviruses are refractory to propagation in cell cultures. However, they can be propagated, in high titers, in primary human keratinocytes [Genes Dev (2009) 23(2):181–194].

Vaccines against HPVs have been made by using virus-like particles (see VLP) that consist of the major (immunogenic) capsid protein.

Detection of papillomaviruses in gynecologic samples: see e.g. DIGENE HYBRID CAPTURE 2 ASSAY.

***par* loci** (in plasmid partition) Loci, found within certain low-copy-number PLASMIDS, concerned with the segregation of plasmids to daughter cells during cell division; such plasmids require an active form of partition to ensure the maintenance of a given plasmid within a bacterial population.

Members of the ParA protein family include e.g. the SopA protein (encoded by the F PLASMID) and the (chromosomally encoded) Soj protein in *Bacillus subtilis*, which is apparently concerned with both plasmid and chromosomal segregation. All ParA proteins appear to share certain characteristics: (i) they have ATPase activity, (ii) they bind to DNA, and (iii) in at least some cases, they have been reported to move within cells (such movement being demonstrated e.g. by GFP-fusion techniques).

Early studies found that the SopA protein polymerizes into filaments, *in vitro*, and that the rate at which these filaments elongate was similar to the rate at which plasmids separate in dividing cells. It appears to be generally understood that partition, in at least many low-copy-number plasmids, involves a mechanical contribution by *par*-encoded proteins to plasmid movement and separation.

parA See F PLASMID and PLASMID (partition).

parainfluenza viruses Various (avian, bovine, canine, human, murine and ovine) viruses in the genus *Paramyxovirus* (family PARAMYXOVIRIDAE (q.v.)) which are associated e.g. with respiratory tract infections.

paramere See MACRONUCLEUS.

Paramyxoviridae A family of enveloped viruses in which the genome consists of non-segmented, negative-sense ssRNA. (cf. ORTHOMYXOVIRIDAE.) Paramyxoviruses include various important pathogens, such as the measles and mumps viruses, Newcastle disease virus (the causal agent of avian pneumoencephalitis), the canine distemper virus, and rinderpest virus (the causal agent of bovine typhus, or cattle plague).

The viruses can be propagated in cell cultures.

The virions (∼200–300 nm diam.) contain enzymes – e.g. a transcriptase, polyadenylate transferase and mRNA methyltransferase. In the members of one genus, *Paramyxovirus* – including e.g. the mumps and Newcastle disease viruses, the Sendai virus, various PARAINFLUENZA VIRUSES – the virions have both HEMAGGLUTININ and NEURAMINIDASE activities; neither of these activities are found in the virions of another genus (*Pneumovirus*) which includes the respiratory syncytial virus (RSV). Virions of the genus *Morbillivirus* (that includes both the measles and rinderpest viruses) have hemagglutinin but no neuraminidase activity.

Paramyxoviruses can be inactivated e.g. by lipid solvents, non-ionic detergents, oxidizing agents and formaldehyde.

(See also SUBGENOMIC MRNA.)

Paramyxovirus A genus within the family PARAMYXOVIRIDAE (q.v.).

paramyxoviruses A term which may refer either to viruses of the family PARAMYXOVIRIDAE (q.v.) or viruses of the genus *Paramyxovirus*.

paranemic Refers to an unstable juxtaposition of two strands of DNA in which the strands are not wound around each other.

(cf. PLECTONEMIC.)

parB See F PLASMID and PLASMID (partition).

parC See R1 PLASMID.

***parD* system** In the R1 PLASMID: a system promoting stability of COPY NUMBER. The Kid protein is a nuclease which was believed to act solely by cleaving host-encoded mRNAs in plasmid-free cells, the plasmid-encoded Kis protein acting as an antidote to Kid. Kid was subsequently reported to cleave host mRNAs *and* a plasmid mRNA (encoding a repressor of plasmid replication) when the copy number falls.

parent-of-origin-specific gene expression An alternative (and less frequently used) name for GENETIC IMPRINTING.

parenteral administration Administered by a non-oral route.

PARN Poly(A)-specific ribonuclease: a eukaryotic exonuclease which specifically degrades the poly(A) tail of mRNA molecules. In at least some cases, deadenylation of the poly(A) tail is reported to be enhanced by interaction between PARN and the 5′ CAP region of the mRNA. PARN is apparently

involved in the CYTOPLASMIC POLYADENYLATION mechanism. (See also CAF1.)

ParR See R1 PLASMID.

partial digest (by restriction enzymes) See LIBRARY (sense 1).

partial fill-in See e.g. the entry LAMBDA FIX II VECTOR.

particle gun *Syn*. GENE GUN.

partition (of plasmids) (*syn*. segregation) See PLASMID.

partitioning complex See R1 PLASMID.

partner gene See GENE FUSION and GENE FUSION VECTOR.

passenger strand (in ds siRNA) The sense strand. (cf. GUIDE STRAND.)

Passport See RECONSTRUCTED TRANSPOSABLE ELEMENT.

PathDetect® systems Reporter systems (Stratagene, La Jolla CA) used for investigating the regulatory function of a given, intracellularly expressed gene product, or of an extracellular stimulus (e.g. a cytokine), on SIGNAL TRANSDUCTION PATHWAY(S) within mammalian cells.

Types of PathDetect® system

There are two forms of the PathDetect® system. (i) The *cis*-reporting systems can be used e.g. to determine the ability of a given (intracellularly expressed) gene product to activate a signal transduction pathway, causing a factor (such as NF-κB or p53) to bind to a sequence upstream of the promoter of a reporter gene – thus promoting transcription of the reporter. (ii) The *trans*-reporting systems are used e.g. for determining the ability of a given (intracellularly expressed) gene product to activate one of certain specific transcription factors (such as CREB or ELK-1).

The cis-reporting system

In the *cis*-reporting system, a plasmid that contains the gene of interest is co-transfected into mammalian cells with a plasmid containing the *cis*-reporter gene with a regulatory region. The *cis*-reporter plasmid contains: (i) a reporter gene – either a gene encoding LUCIFERASE or one that encodes the (low-toxicity) variant of green fluorescent protein: the humanized *Renilla* GFP (see HRGFP); (ii) a promoter sequence for the reporter gene; (iii) upstream of the promoter region, tandem repeats of a particular binding sequence (which is referred to as the *binding element*, the *enhancer element* or the *response element*). Various *cis*–reporter plasmids are available, each plasmid with a different response element; thus, for example, one type of *cis*-reporter plasmid contains a response element consisting of 5 tandem repeats of the binding site of NF-κB, while another has a response element consisting of 15 tandem repeats of the binding site of p53. Other response elements, present in different *cis*-reporter plasmids, include the tandem repeats of binding sites of CRE, ISRE, NFAT, SRE, SRF and TARE. (A response element of choice can be inserted into a plasmid supplied with a POLYLINKER.)

If the expressed protein of interest activates a signal transduction pathway which then gives rise (e.g.) to an active form of NF-κB, the NF-κB can bind to the response element and promote expression of the reporter gene.

The trans-reporting system

The *trans*-reporting system includes a reporter plasmid and a *trans*-activator plasmid, both of which are co-transfected into cells together with the expression vector encoding the protein of interest. (Positive and negative control plasmids are used.)

The reporter plasmid of the *trans*-reporting system includes (i) a reporter gene encoding e.g. luciferase or β-galactosidase; (ii) a promoter region; and (iii) upstream of the promoter, a fivefold tandem repeat of a GAL4-binding element.

The *trans*-activator plasmid of the *trans*-reporting system includes a sequence encoding a *fusion protein*: the DNA-binding domain of GAL4 fused to a transcription factor (e.g. CREB or Elk-1); other fusion partners for GAL4 – used in different *trans*-activator plasmids – include c-Jun and c-FOS. The fusion partner of GAL4 forms the so-called *activation domain* of the fusion protein. Transcription of this plasmid is mediated by a CMV promoter.

If the expressed protein of interest causes phosphorylation, directly or indirectly, of the *activation domain* of the fusion protein (i.e. the fusion partner of GAL4), then the activated GAL4 fusion protein binds to the tandem repeat binding site for GAL4 on the reporter plasmid, thus promoting expression of the reporter gene.

[Examples of use of PathDetect®: J Neurochem (2006) 96 (1):65–77; and (*trans*-reporting system) Mol Biol Cell (2009) 20(3):1020–1029.]

pathogenesis-related proteins (PR proteins) (*plant pathology*) Low-molecular-weight proteins produced in the tissues of certain plants as a response to infection by a pathogen; some are reported to have antimicrobial properties. PR proteins are formed e.g. in tobacco cultivars exposed to infection with the tobacco mosaic virus, and are also produced in other plants infected with viruses, viroids, fungi or bacteria. (cf. PHYTOALEXINS; see also GENE-FOR-GENE CONCEPT.)

pathogenicity island (PAI) In a microbial (usually bacterial) genome: a cluster of several to many genes encoding specific determinants of pathogenicity/virulence.

Commonly, the GC% of a PAI is different from that of the organism's chromosome; this has been taken to indicate that, in general, PAIs were acquired by horizontal gene transfer, i.e. by a process such as conjugation. The results of one study were reported to show that both conjugation and homologous recombination play a major role in the horizontal transfer of a PAI between strains of *Escherichia coli* [PLoS Pathog (2009) 5(1):e1000257].

Many PAIs are flanked by DIRECT REPEATS; the LEE PAI (see below) is one example of a PAI which is not flanked by direct repeats. In *Moraxella bovis* (hemolytic strains) the *mbx* operon is flanked by ∼700-bp imperfect repeats; this, together with e.g. the GC% of *mbx* genes, suggested that this region of the genome is a PAI [J Med Microbiol (2006) 55:443–449].

PAIs are found in both Gram-negative and Gram-positive bacteria.

The PAI designated LEE (*locus of enterocyte effacement*) is a 35-kb element found e.g. in all strains of enteropathogenic *Escherichia coli* (EPEC). Genes in the LEE PAI encode a so-called type III secretory system which apparently enables the

strains of EPEC to translocate effector proteins directly into (eukaryotic) target cells. [Type III systems and PAIs: J Med Microbiol (2001) 50:116–126.] Proteins encoded by the LEE PAI are responsible for characteristic 'attaching and effacing' lesions produced by this pathogen on intestinal epithelium. Transcription of at least some of the LEE genes is controlled from a regulatory region in the EAF plasmid (EAF = EPEC adherence factor) which occurs in all strains of EPEC.

The chromosomal site occupied by LEE is occupied by a different PAI in strains of UPEC (uropathogenic *E. coli*); the products of this 70-kb PAI include a hemolysin.

A number of distinct PAIs occur in *Salmonella* spp. One of them, SPI-1, contains the *inv–spa* genes which encode a type III secretory system involved in the invasion of intestinal epithelial cells; the PAI SPI-3 encodes the means of survival in macrophages.

In the (gastric) pathogen *Helicobacter pylori*, expression of the *cagA* gene (in the *cag* PAI), and of the (unlinked) gene *vacA*, is linked to severe gastrointestinal disease. It appears that CagA is injected into gastroepithelial cells by *H. pylori* and causes e.g. proliferation of these cells. VacA is a secreted cytotoxin (95 kDa) which, *in vitro*, blocks the proliferation/activity of T lymphocytes (T cells).

The VPI pathogenicity island in *Vibrio cholerae* encodes e.g. cell-surface appendages (the so-called 'toxin co-regulated pili' or TCP) which function as receptors for CTXΦ – the phage whose genome encodes cholera toxin.

In certain strains of *Staphylococcus aureus* a PAI encodes the toxic shock syndrome toxin.

[PAIs in evolution: BMC Evol Biol (2007) 7(suppl 1) S8; in *Francisella*: BMC Microbiol (2007) 7:1.]

paucibacillary specimen Any specimen which contains a *small* number of target bacteria per unit volume – e.g. a specimen that contains fewer than $\sim 10^4$ colony-forming units (cfu) per mL. Such specimens usually permit detection of the target organism by culture, but nucleic-acid-based tests – which use a much smaller volume of the sample – may fail to detect the organism. Improved detection by a nucleic-acid-based test may be achieved if the test is preceded by a technique which effectively concentrates the pathogen (e.g. immunomagnetic separation).

PAX6 See AXENFELD–RIEGER SYNDROME.

PAXgene™ blood RNA kit A product used for the isolation of RNA from whole blood. (Sources of PAXgene cited in the literature include: PreAnalytiX, Hombrechtikon, Switzerland, Becton Dickinson, and Qiagen.) The system was designed to avoid the problem of instability of RNA *in vitro* (i.e. following collection of blood samples); during collection, storage and transport of blood – at ambient temperatures – the copy number of a given RNA species may change significantly owing to (i) degradation of the RNA, and/or (ii) the ongoing expression of genes following collection of the sample from the patient.

This approach involves collecting the blood in specialized tubes containing a mixture of reagents that stabilize the RNA profile by (i) protecting RNA from degradation by RNases and (ii) inhibiting gene expression.

The collecting tube (containing sample) is centrifuged. The washed pellet is re-suspended in buffers, with proteinase K, and incubated. After centrifugation, supernatant is transferred to a fresh tube and ethanol is added. The supernatant is added subsequently to the PAXgene spin column for abstraction of RNA. Following several washing steps with buffers the RNA is eluted.

The procedure generally removes most of the DNA. For a more rigorous elimination of DNA it is possible to include DNase treatment in the protocol.

Uses (e.g.): studies on the time course of proinflammatory/immunomodulatory cytokine mRNAs [Infect Immun (2006) 74(7):4172–4179]; studies on expression of genes encoding the glucocorticoid receptor [Proc Natl Acad Sci USA (2006) 103(14):5496–5501]; a study on Parkinson's disease [Proc Natl Acad Sci USA (2007) 104(3):955–960]; transcriptional profiling in simian immunodeficiency virus (SIV) infections [PLoS Pathog (2009) 5(2):e1000296]; study of gene expression in primary human leukocytes [BMC Med Genom (2009) 2:1].

(See also NUCLEIC ACID ISOLATION.)

pBAD-DEST49 vector A destination vector in the GATEWAY SITE-SPECIFIC RECOMBINATION SYSTEM (see table in entry).

[Use (e.g.): PLoS ONE (2009) 4(2):e4509.]

pBLOCK-iT™3-DEST vector A destination vector (see entry GATEWAY SITE-SPECIFIC RECOMBINATION SYSTEM (in the table)) in which an insert, encoding an shRNA, can be expressed on a long-term basis.

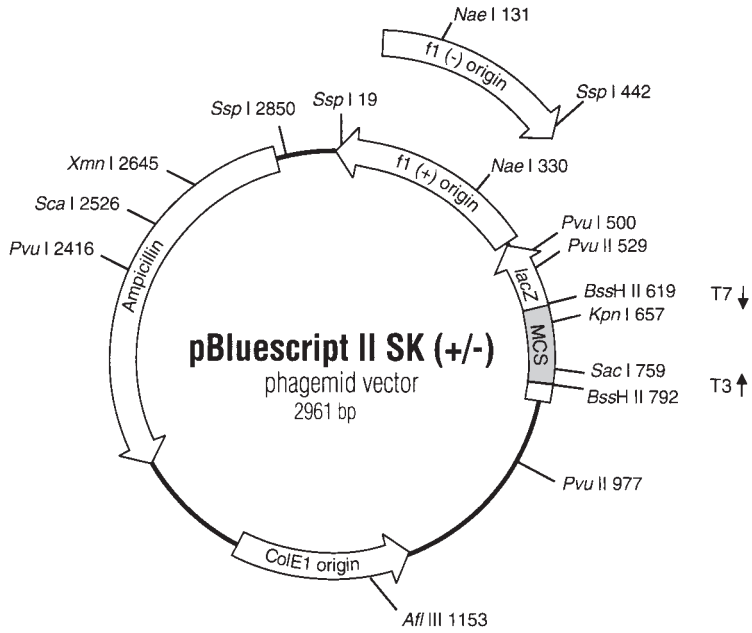
[Use (e.g.): PLoS ONE (2009) 4(6):e5877.]

pBluescript® Any of a set of ~3-kb PHAGEMIDS (marketed by Stratagene, La Jolla CA). Derived from plasmid pUC19, the phagemids contain: (i) the replication origin from plasmid ColE1, (ii) the replication origin from bacteriophage f1, (iii) an ampicillin-resistance gene, (iv) a sequence that encodes α-PEPTIDE (part of the *lacZ* gene), (v) the *lac* promoter, and (vi) a POLYLINKER, situated between the *lac* promoter and the α-peptide sequence. The polylinker is flanked by promoter sites for T3 and T7 RNA polymerases (allowing transcription of an insert in either direction).

The f1 origin can be inserted in either orientation – so that either the sense or antisense strand can be amplified if the host cell contains a helper phage.

pBluescript II SK (+/–), a phagemid of 2961 bp, is shown in the diagram (on page 248).

These vectors can be replicated in strains of *Escherichia coli* (e.g. JM105 or JM109) which contain a chromosomal *lacZΔM15* mutation that deletes part of the N-terminal region of the enzyme β-galactosidase. When these strains of *E. coli* are grown on lactose-containing media (or supplied with IPTG) the LAC OPERON (which includes *lacZ*) will be induced – giving rise to the *inactive* (mutant) form of β-galactosidase. However, if the phagemid-encoded α-peptide is synthesized it complements the mutant enzyme (by non-covalent binding)



pBluescript® II SK (+/–) PHAGEMID VECTOR One of a range of phagemid vectors in the pBluescript® series. SK indicates that the orientation of the polylinker is such that transcription of *lacZ* proceeds in the direction SacI to KpnI.

A vector containing the f1(+) origin (the origin of replication of the filamentous phage f1) permits recovery of the sense strand of the *lacZ* gene when a strain containing the pBluescript® vector is co-infected with a helper phage. The f1(–) origin permits recovery of the antisense strand of the *lacZ* gene.

The ColE1 origin of replication can be used in the absence of a helper phage.

The inclusion of a portion of the *lacZ* gene permits the identification of recombinant phagemids by the blue–white selection procedure (see entry for details).

Courtesy of Stratagene, La Jolla CA, USA.

to form an *active* (functional) β-galactosidase. The presence of the functional enzyme can be determined e.g. by growing the bacteria on plates containing X-GAL – bacteria containing a functional enzyme will hydrolyze this compound to form a blue product and will form *blue* (or blue-green) colonies.

If the phagemid contains a gene/fragment (inserted in the polylinker), the α-peptide is usually not formed because the presence of the insert prevents transcription of a functional α-peptide; thus, without a functional α-peptide, these cells lack a functional β-galactosidase – and they therefore give rise to colorless (*white*) colonies on X-gal media as they are unable to hydrolyze the X-gal.

Hence, a colony's appearance on media containing X-gal generally indicates whether or not the (intracellular) phagemids are *recombinant*, that is, whether they contain an insert (white colonies) or lack an insert (blue colonies); this is *blue–white screening*. (See also BLACK–WHITE SCREENING.)

Exceptionally, a *small* insert that does not affect the reading frame of the α-peptide sequence *may* allow the formation of a functional, or weakly functional, α-peptide – leading to the formation of blue colonies (despite the presence of an insert). Moreover, certain events, such as a phase-shift mutation, may prevent synthesis of the α-peptide – leading to the formation of a white colony (despite the absence of an insert).

[Uses of pBluescript® (e.g.): BMC Plant Biol (2009) 9:36; (SK+) J Biol Chem (2009) 284(15):10056–10066, and PLoS ONE (2009) 4(3):e4861; (SK–) J Biomed Biotechnol (2009) 985140.]

PBMCs Peripheral blood mononuclear cells.

PBP 2a (PBP 2′) See the entry MECA GENE.

PBPs Penicillin-binding proteins: see β-LACTAM ANTIBIOTICS.

pBR322 A recombinant PLASMID (~4360 bp) constructed from several naturally occurring plasmids. Small plasmids like this are useful as VECTORS in genetic engineering; they are taken up by TRANSFORMATION more efficiently (as compared with larger plasmids) and they are also less susceptible to damage during manipulation.

pCAL vectors A set of vectors used in the AFFINITY PROTEIN EXPRESSION AND PURIFICATION system (q.v.).

pcDNA™6.2/GeneBLAzer™-DEST™ A fusion vector, used for the expression of proteins in mammalian cells, in which reporter activity involves the FRET principle – see the table in GATEWAY SITE-SPECIFIC RECOMBINATION SYSTEM.

pcDNA™6.2/nLumio™-DEST™ A destination vector in the GATEWAY SITE-SPECIFIC RECOMBINATION SYSTEM (see the table in that entry).

PCNA Proliferating cell nuclear antigen: a nuclear factor which is responsible for processivity during DNA replication; thus, for example, processivity of DNA polymerase δ (associated with synthesis of the lagging strand) is reported to be more than 100-fold enhanced in the presence of PCNA. (cf. THIOREDOXIN.)

PCNA is sometimes called 'the clamp'.

***pcnB* mutant** (in *Escherichia coli*) See MULTICOPY PLASMID.

PCR Polymerase chain reaction: a method used primarily for 'amplifying' (= copying) specific sequences of nucleotides in DNA (or in RNA, by using REVERSE TRANSCRIPTASE PCR); sequences up to kilobases in length can be copied – millions of copies of the given sequence being produced within a few hours.

The principle of PCR is shown – diagrammatically – in the figure on page 250.

Essentially, PCR depends on the ability of a *thermostable* DNA polymerase to extend primers, bound to each strand of the target sequence (AMPLICON), and to do this, repeatedly, in a process which involves temperature cycling; temperature cycling is needed to allow separation of the completed copies from their templates – so that fresh primers can bind to the templates.

PCR has a wide spectrum of uses in biomedical science – see later under *Some modified protocols* for examples. (See also FORENSIC APPLICATIONS.)

[The development of miniaturized on-chip PCR (a review): Nucleic Acids Res (2007) 35(13):4223–4237.]

[PCR's next frontier (technology feature): Nature Methods (2007) 4:869–875.]

(For alternative methods of amplifying nucleic acids see the section *Other methods for amplifying nucleic acids* at the end of the entry.)

Basic requirements for successful PCR

Attention to detail is essential for successful PCR. Not least is the requirement to exclude extraneous DNA, i.e. non-target DNA (see *Contamination by extraneous DNA*).

The choice of primers is important to ensure specificity and also e.g. to avoid the formation of PRIMER–DIMERS. (The use of ZIP NUCLEIC ACID primers may be an advantage e.g. when amplifying AT-rich target sequences.)

The use of a proof-reading DNA polymerase is required for high-fidelity copying.

The correct concentration of magnesium ions (Mg^{2+}) in the reaction mixture is critical for optimization of the reaction.

Accurate and consistent temperature cycling is necessary in order to obtain satisfactory and reproducible results.

Quantity of sample DNA required

Samples containing less than 100–200 picograms DNA were reported to be involved in the LCN DNA TEST (q.v.), although the amount normally available is appreciably greater.

Volume of reaction mixture*; *choice of enzyme

PCR is often carried out in a total volume of approx. 25–50 μL, although smaller volumes (e.g. 12.5 μL) have been used.

The choice of enzyme (i.e. DNA polymerase) depends on specific requirements. A wide range of thermostable enzymes is available from commercial suppliers. For example, some DNA polymerases are designed to amplify 'difficult' target sequences (e.g. GC-rich sequences: see e.g. ACCUPRIME GC-RICH DNA POLYMERASE); others are designed specifically to introduce errors into the products (in studies on mutagenesis: see e.g. GENEMORPH). A recombinant polymerase, which is activated during the initial high-temperature stage of PCR, is used in one form of HOT-START PCR.

Note that some DNA polymerases (for example RTTH DNA POLYMERASE XL) have PROOF READING ability while others (e.g. *Taq* DNA polymerase) do not. Proof reading ability is essential when high-fidelity copying is required.

The TAQ DNA POLYMERASE is used when the products are required for cloning with TOPO TA CLONING KITS (q.v.).

Two DNA polymerases are used (simultaneously) in LONG PCR.

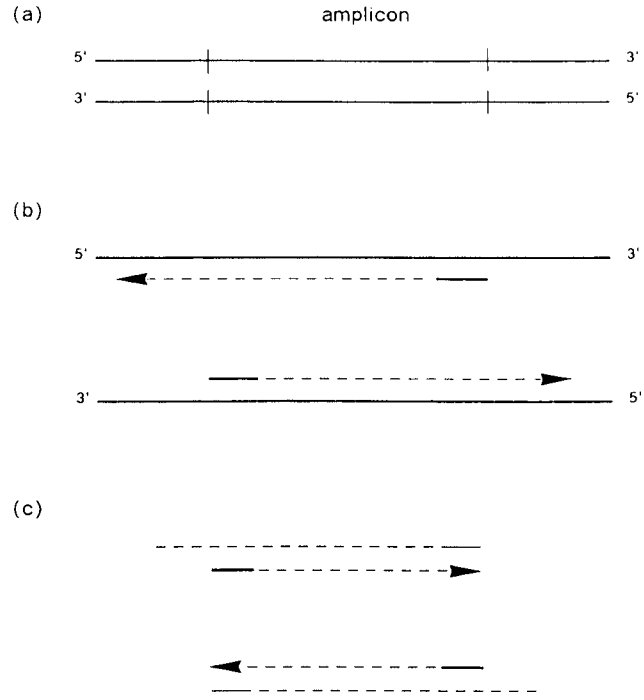
Temperature cycling

As indicated in the figure (legend), temperature cycling often follows the pattern:

$$95°C \rightarrow 45–65°C \rightarrow 72°C \rightarrow 95°C \rightarrow$$

These temperatures are designed for the cyclical repetition of (i) strand separation (denaturation of the sample dsDNA), (ii) binding (*annealing*) of primers, on opposite strands, at sites which bracket the amplicon, and (iii) extension (synthesis of DNA from the 3′ terminus of each primer) to form *products* – i.e. amplicons. Following the synthesis of amplicons, a rise in temperature separates the amplicons from their templates.

Although the temperatures for denaturation and extension are usually ~94/95°C and ~72°C, respectively, the annealing temperatures reported in the literature vary significantly. One reason for this is that GC%-rich primers need higher annealing temperatures in order to avoid 'mispriming' – i.e. binding to inappropriate (i.e. non-complementary) sequences within the template strand. (Mispriming yields spurious products.)



PCR (polymerase chain reaction): the basic method (principle, diagrammatic).

The reaction mixture includes: (i) The sample of (double-stranded) DNA – which may be, for example, genomic DNA or fragments of genomic DNA; (ii) millions of copies of the two types of primer, each primer being about 20–30 nucleotides in length; (iii) a thermostable DNA polymerase (e.g. the *Taq* DNA polymerase); (iv) deoxyribonuleoside triphosphates of all four types (ATP, CTP, GTP, TTP); and (v) appropriate buffer.

For PCR, the reaction mixture undergoes a cyclical series of changes in temperature, such changes being controlled automatically in an instrument called a *thermocycler*. The cycle is repeated about 20–40 times, amplification of the target sequence occurring at each cycle.

(a) Part of the sample double-stranded DNA: two antiparallel strands held together by base-pairing. The sequence to be amplified – the *amplicon* – is shown between the short vertical bars.
(b) Initial heating to about 95°C has separated the two strands of DNA. Subsequently, transient cooling to about 45–65°C has allowed the primers – shown as short, solid lines – to bind (*anneal*) to their complementary sites on the two template strands. Importantly, notice the positions of the bound primers in relation to the two ends of the amplicon. The temperature is then increased to e.g. 72°C, and DNA synthesis (mediated by the thermostable DNA polymerase, and shown here as a dashed line with an arrowhead) proceeds from the 3′ end of each primer. Each of these newly synthesized strands is longer than the amplicon.
(c) The second round of replication begins with initial heating to ~95°C – separating the new daughter strands (synthesized in the first round of replication) from their templates. Each of the new daughter strands (as well as the two parent strands) can act as a template in the next round of replication. In the diagram, a primer (short, solid line) has bound to each new daughter strand and has been extended (dashed line with arrowhead). Note that the strand synthesized on a new daughter strand is the same length as the amplicon; such strands rapidly become dominant in subsequent rounds of replication.

Reproduced from Figure 4.1 on page 57 of *DNA Methods in Clinical Microbiology* [ISBN 07923-6307-8], Paul Singleton (2000), with the kind permission of Springer Science and Business Media.

PCR (polymerase chain reaction): some of the many variant forms and their uses (see separate entries for each variant form)

Variant form	Uses (e.g.)
Allele-specific PCR	Amplification of a *particular* allele of a given gene (in the presence of other alleles of that gene)
Anchored PCR	Amplification of an unknown sequence flanking a known sequence
Asymmetric PCR	Preparation of probes and the preparation of ssDNA e.g. for sequencing
Bubble-linker PCR	See Vectorette PCR (below)
Competitive RT-PCR	Estimation of the number of RNA target sequences in a given sample
Headloop suppression PCR	Protocol arranged to promote suppression of amplification of sequences *related to* the given target sequence (without affecting the amplification of the intended target sequence)
Hot-start PCR	Cycling delayed until the temperature of the reaction mixture is sufficient to inhibit non-specific binding of the primers. The object is to increase the specificity and the sensitivity of the assay
IAN-PCR	Inverse affinity nested PCR: a modification of inverse PCR used for the recovery of flanking sequences from low-copy-number fragments: see entry METAGENOME
Immuno-PCR	Ultrasensitive detection of antigen; the antigen-binding antibody is linked to a DNA fragment containing a sequence which is amplified by PCR; the amplicons (visualized by gel electrophoresis) provide an indication of the presence of specific antigen
Inverse PCR	Amplification of a sequence that *flanks* a known sequence
Island rescue PCR	Amplification of CpG islands in cloned fragments
LMPCR	Amplification of fragments produced in an *in vivo* footprinting assay (see entry LIGATION-MEDIATED PCR)
Methylation-specific PCR	Monitoring methylation status of CpG islands in genomic DNA
Quantitative PCR (QPCR)	Absolute or relative quantitation of DNA or RNA target sequences (see entry QUANTITATIVE PCR)
Real-time PCR	Determination of the number of specific target sequences in the reaction mixture prior to cycling
Relative RT-PCR	Comparison of the numbers of RNA targets, e.g. transcripts of a given gene, in samples from different cells/tissues etc.
Sexual PCR	See entry DNA SHUFFLING.
Splice overlap extension PCR	Fusion of two sequences of DNA (present either on the same molecule or on separate molecules) without the use of restriction enzymes
Terminal transferase-dependent PCR	Detection of the presence of any obstruction or barrier to DNA polymerase during replication
Touchdown PCR	A protocol designed to optimize stringency and specificity of the assay by avoiding mispriming at the annealing stage
Vectorette (bubble-linker) PCR	Amplification of an unknown sequence adjacent to a known sequence

Optimization of the annealing temperature may be achieved e.g. by TOUCHDOWN PCR.

A different approach to avoiding mispriming is HOT-START PCR (q.v.) which includes a range of techniques such as heat-activatable primers, and a (meltable) wax barrier to separate components of the reaction mixture.

In some cases, depending e.g. on the primers involved, a *two-temperature protocol* has been used – e.g. 94°C for the denaturation stage and a combined annealing/extension temperature of 68°C.

Theoretically, temperature cycling could be carried out by manual control, but in routine laboratory conditions it is conducted in an instrument called a THERMOCYCLER (q.v.).

Rapid-cycle PCR

This refers to a type of PCR protocol which involves e.g. the optimization of temperature control and which can achieve a total run time of less than 30 minutes: see the entry THERMOCYCLER.

Stringency

The physicochemical conditions (e.g. temperature, electrolyte concentrations) under which the primers anneal to their target sites on the template strands affect the accuracy (specificity)

of the polymerase chain reaction. For example, if a primer is allowed to anneal to a template strand at a temperature which is too low, binding may occur aberrantly at a site which is not complementary to the primer's sequence (and which may be outside the amplicon); in this case, extension of the 3′ end of the primer will give rise to a product which is not the one required. For this reason, such *low-stringency* conditions are not generally used – although they are used e.g. for the initial phase of AP-PCR (q.v.).

Normally, PCR uses *high-stringency* conditions (which are dictated primarily by temperature). Under such conditions the binding (annealing) of primers to the target DNA occurs only when they are exactly complementary, or very nearly so, to their appropriate binding sites; thus, the higher the stringency the greater will be the degree of matching required between primer and target sequence for successful binding and, hence, extension, to occur.

Contamination by extraneous DNA

Extraneous DNA that contaminates a reaction mixture may contain (i) sequence(s) to which the primers can bind, and/or (ii) sequence(s) that are able to prime irrelevant sites in the sample DNA. In either case, unwanted sequences may be amplified, resulting in a troublesome mixture of products; in addition, the efficiency of the legitimate amplification may be reduced owing to sequestration of polymerase and nucleotides.

Successful PCR needs rigorous exclusion of contaminating DNA – which may be derived from the environment and/or from the apparatus and reagents.

An important source of contaminating DNA consists of the amplicons from previous reactions that had been carried out within the same laboratory.

The two basic approaches to the problem of contamination are AMPLICON CONTAINMENT and AMPLICON INACTIVATION (q.v.).

Inhibitors of PCR

In some cases the sample DNA is derived from a source that may include substance(s) which are inhibitory to PCR; this may give rise to a false-negative result from a sample which, in fact, contains the sequence of interest.

Inhibitors of PCR are found e.g. in various types of clinical specimen – including bone marrow aspirates, CSF (cerebrospinal fluid), feces and urine; specific inhibitory substances include hemoglobin (in red blood cells), lactoferrin (in white blood cells) and various polysaccharides (in feces).

Other inhibitors include AGAR, hemin, and the blood anticoagulants heparin and sodium polyanetholesulfonate (SPS); see also SYBR GREEN I for the effects of certain dyes used as indicators in real-time PCR.

Mutant versions of *Taq* DNA polymerase (an enzyme often used in PCR) have been reported to be resistant to the effects of inhibitors – allowing DNA amplification from samples of whole blood or crude soil samples [Nucleic Acids Res (2009) 37(5):e40].

Detection of inhibitors can be approached in several ways. For example, if an aliquot of the sample is 'spiked' with the target DNA, prior to assay, then failure to amplify the target in this aliquot may suggest the presence of inhibitor(s). This approach, however, increases the risk of contamination.

Alternatively, an internal control can consist of molecules of nucleic acid which (i) include binding sites for the target-specific primers, but which (ii) can be distinguished from the actual target by a unique PROBE-binding sequence. Given the absence of inhibitors, the control and target sequences should both be amplified; failure to obtain any amplification could indicate the presence of inhibitor(s), while amplification of the control sequence (only) would suggest the absence of the target sequence in the sample DNA.

The effects of an inhibitor may be reduced by diluting the sample, but this concomitantly reduces the number of copies of the target sequence (if present) and may therefore lower the sensitivity of the assay. Alternatively, use can be made of a magnetic separation technique (using e.g. DYNABEADS); in this approach, copies of the target sequence are captured, and held, on magnetic beads while any inhibitory substances are removed by washing.

Facilitators of PCR

Certain agents may enhance the efficiency of a PCR assay in the presence of inhibitory substances. These agents include bovine serum albumin (BSA) and betaine; for example, BSA has been reported to improve the performance of PCR in assays on fecal specimens (which characteristically contain a range of inhibitory substances).

Some facilitators are effective against a range of inhibitors, while others are more selective in their activity.

It was reported that the accuracy of PCR may be promoted by the presence of the RecA protein.

It has been reported that CARBON NANOTUBES enhance the efficiency and/or specificity of LONG PCR.

PCR with damaged/ancient DNA

The amplification of poor-quality DNA, for forensic or other purposes, may be facilitated by the use of DNA polymerases of the Y FAMILY (q.v.).

(cf. SPEX.)

Long PCR

This form of PCR involves a special protocol which is used for amplifying very long target sequences (e.g. >40 kb) – see the entry LONG PCR.

Detection of PCR products

Given that the products of PCR (amplicons) are of a *precise* length, i.e. an exact number of nucleotides delimited by the primer pair, they can be readily detected and identified by gel electrophoresis. Thus, a given product forms a single band at a precise location in the gel; the size of the product forming such a band can be verified by using a so-called DNA ladder (see entry GEL ELECTROPHORESIS). Bands of fragments in a gel may be further examined e.g. by SOUTHERN BLOTTING followed by hybridization of a target-specific, labeled PROBE.

To verify that an amplicon of a specific size and sequence has been amplified, the amplicons can be subjected to a part-

icular RESTRICTION ENDONULEASE directed at a known cutting site in the amplicon; the resulting products of restriction can then be checked, by gel electrophoresis, to see whether fragments of the *expected* sizes were produced.

Gel electrophoresis is a slow process. Hence, in some cases use is made of instrumentation that permits monitoring of the amplification (i.e. increase in amplicons) *during* the reaction: so-called REAL-TIME PCR.

The products of PCR amplification can also be detected, independently of the actual sequence which has been amplified, by using the *universal template* approach – see, for example, the use of this approach in the entry DIGITAL PCR.

Production of modified PCR products

By using specially designed primers in the PCR reaction, it is possible to produce amplicons that are modified in various ways (for various purposes): see the entry TAG (sense 2).

Selecting low-copy-number targets from dominant targets

See e.g. PCR CLAMPING.

Real-time PCR

In this form of PCR, in which amplification of the target can be monitored as it occurs, the formation of products can be followed by a dye-based approach, a PROBE-based approach, or a combined probe-and-dye-based approach: see the entry REAL-TIME PCR.

Quantitative PCR

PCR can be used to give an estimate of the number of copies of the target sequence which were present in the sample *prior to* amplification: see QUANTITATIVE PCR.

Errors in PCR products

If PCR is carried out with a DNA polymerase (such as *Taq* DNA polymerase) which lacks PROOF READING ability, one error (incorrect nucleotide) is reported to occur, on average, in approximately 10^4 nucleotides. The smaller the amplicon, the less likely it is for an error to occur in any given amplicon (and conversely). The presence of an error within a particular amplicon will be important, for example, if, by chance, that amplicon is selected for cloning; in such a case, all the cloned amplicons will contain the error. Cloned PCR amplicons can be checked for accuracy by sequencing.

Inserting PCR products into cloning vectors or expression vectors

When PCR is carried out with the *Taq* DNA polymerase the amplicons characteristically have an 'extra' adenosine residue at the 3′ ends (see EXTENDASE ACTIVITY). This can be useful because the 3′ overhang facilitates insertion of the amplicons into linearized vector molecules that have 3′ thymidine overhangs (see e.g. TA CLONING KITS).

In a different approach, the PCR reaction mixture employs primers which have a 5′ tag (see TAG, sense 2) that includes a recognition site for a specific RESTRICTION ENDONUCLEASE; the resulting (double-stranded) amplicons can thus be cleaved (at each end) by the given restriction enzyme to form STICKY ENDS, facilitating insertion into a vector with complementary sticky ends.

In yet another approach, the primers' 5′ tag contains an *att*B site to enable insertion into the pDONR™ vector for working in the Gateway® system (see entry GATEWAY SITE-SPECIFIC RECOMBINATION SYSTEM).

Multiplex PCR

With appropriate experimental arrangements, more than one target sequence can be amplified in a single reaction: see the entry MULTIPLEX PCR.

Some modified protocols

The flexibility of PCR can be illustrated by some examples:

- ALLELE-SPECIFIC PCR: use of an allele-specific primer (and a gene-specific primer) to amplify a particular allele among a mixture of alleles.
- ASYMMETRIC PCR: a procedure in which the concentration of one primer is much lower than that of the other; it is used e.g. for copying a particular strand of the template dsDNA.
- DIGITAL PCR: a procedure which has been used e.g. for the calibration of sequencing libraries.
- DNA SHUFFLING: a primer-less form of PCR in which fragments of DNA, in the reaction mixture, prime each other; this technique is used to promote *in vitro* recombination and the development of a product with new properties.
- Emulsion PCR (emPCR): PCR carried out within aqueous microdroplets in an emulsion – used e.g. in some of the new-generation methods for DNA sequencing: see e.g. GS FLX and SOLID TECHNOLOGY.
- EXSITE SITE-DIRECTED MUTAGENESIS KIT: a commercial system in which mutations (including point mutations, additions or deletions) can be introduced into plasmids by using appropriately designed primers in PCR.
- GENEMORPH: a commercial system in which an error-prone DNA polymerase is used for promoting mutagenesis in the amplicons.
- HEADLOOP SUPPRESSION PCR: use of modified primers that selectively suppress amplification of unwanted sequence(s) without affecting amplification of the wanted sequence.
- IMMUNO-PCR: a technique used for ultrasensitive detection of e.g. antigens.
- Inverse affinity nested PCR (IAN-PCR): allows access to rare genes in a metagenome (see the entry METAGENOME).
- INVERSE PCR: used e.g. for amplifying DNA which *flanks* a known region, and for the preparation of deletion mutations. (See also the entry TAIL-PCR.)
- ISLAND RESCUE PCR: used e.g. for amplifying CpG islands.
- LCN DNA TEST: a form of PCR used for the amplification of *minute* quantities of DNA.
- LIGATION-MEDIATED PCR: a procedure used for analyzing products obtained e.g. from an *in vivo* footprinting assay.
- MEGAPRIMER MUTAGENESIS: a technique for site-directed mutagenesis.
- MULTIPLEX PCR: amplification (simultaneously, in a single assay) of more than one target sequence.
- NESTED PCR: used to increase specificity and sensitivity.
- PCR CLAMPING: use of a PNA probe to favor amplification of a given sequence by blocking amplification of unwanted sequence(s).

- SELECTOR-BASED MULTIPLEX PCR: a procedure in which a number of different target sequences can be amplified with a single pair of primers (cf. conventional MULTIPLEX PCR).
- SOLID-PHASE PCR (on-chip PCR): the extension of primers on immobilized target sequences.
- SPLICE: an *in vitro* method for obtaining a complete coding sequence from a split gene in genomic DNA.
- SPLICE OVERLAP EXTENSION PCR: a procedure which can be used to fuse two sequences of DNA (either on the same molecule or on different molecules) without using restriction endonuleases.
- SUICIDE POLYMERASE ENDONUCLEASE RESTRICTION: a procedure for enhancing amplification of a low-copy-number template (in the presence of so-called dominant templates) – an alternative to PCR clamping.
- TAIL-PCR: like INVERSE PCR, a technique which is used for copying an unknown sequence adjacent to a known sequence.
- TERMINAL TRANSFERASE-DEPENDENT PCR: a method used for detecting an obstruction or barrier, in a dsDNA sample, which may inhibit the DNA polymerase during amplification.
- UNIVERSAL TEMPLATE PCR ASSAY: a method that enables detection of amplified products independently of sequence.
- VECTORETTE PCR: a modified form of ANCHORED PCR – involving a dsDNA linker with a central mismatched region – used for copying an unknown sequence adjacent to a known sequence.

Other methods for amplifying nucleic acids

PCR is used widely throughout the biomedical sciences, but there are various other methods which can be used to amplify nucleic acid sequences.

Alternative methods include the LIGASE CHAIN REACTION, which, like PCR, involves temperature cycling.

Among the other methods, some – e.g. NASBA and SDA – are conducted *isothermally*, i.e. they are carried out at a fixed temperature.

For details of various methods used to amplify DNA (and RNA) see the table in NUCLEIC ACID AMPLIFICATION.

PCR-assisted contig extension See the entry PACE.

PCR clamping In a PCR reaction: an arrangement by which the amplification of one or more specific sequences is facilitated by concomitant blocking of amplification of other (specific) sequence(s). PCR clamping has a range of uses, one of which is the detection of rare (low-copy-number) mutant templates in the presence of a great excess of the corresponding wild-type templates; in such cases, without clamping, a dominant template may preclude detection, or reduce the possibility of detecting, minority template(s) e.g. by the early depletion of reagents.

There are two main types of PCR clamping (see later), each of which involves the use of an oligomer of peptide nucleic acid (see entry PNA) as a clamp. The properties of PNA that make it particularly useful for this role are:

- duplexes formed from complementary strands of PNA and DNA have greater thermal stability than the corresponding DNA:DNA duplexes (the T_m of a PNA:DNA duplex is higher than that of the corresponding DNA:DNA duplex);
- a PNA:DNA duplex is destabilized (T_m significantly lower) by a single mismatch;
- a PNA oligomer, bound to its complementary sequence, is not extendable by any DNA polymerase and therefore cannot act as a primer.

Elongation arrest

This type of PCR clamping uses a PNA oligomer which is complementary to an internal (non-terminal) sequence in the amplicon. An oligomer with a wild-type sequence will bind strongly to its target site in *wild-type* templates; accordingly, the wild-type templates will not be amplified because primer extension is blocked by the oligomer. *Mutant* templates – containing one or more mutant bases in the oligomer's target site – will be amplified normally as the absence of a bound PNA oligomer will allow primer extension.

***Primer exclusion* (= *competitive clamping*)**

In this type of PCR clamping one of the primers is targeted to a binding site that includes a particular (required) mutant sequence. If this sequence occurs in a given template then that template will be amplified (through extension of primers). If the PNA oligomer is complementary to the *wild-type version* of the primer's binding site it will block amplification of the wild-type templates but will not interfere with amplification of the mutant templates. With this scheme, point mutations at different sites in the amplicon sequence can be investigated by using primers with appropriate complementarity.

In one study, rare K-*ras* DNA was detected by a clamping procedure in which a fluorophore-labeled PNA oligomer acted as both the PCR clamp and sensor probe, the latter enabling detection of specific mutations by melting curve analyses [Nucleic Acids Res (2006) 34(2):e12].

An alternative approach to handling dominant templates is SUICIDE POLYMERASE ENDONUCLEASE RESTRICTION.

In an analogous context, dominant *RNA* templates may be dealt with e.g by the use of COMPETIMERS.

PCR cloning A phrase which generally refers to a procedure in which amplicons, produced by PCR, are inserted into vector molecules and replicated within cells.

PCR-RFLP analysis A form of RFLP ANALYSIS in which the sample DNA is copied from genomic DNA (or from another source) by PCR. One advantage of this approach is that the amplicons (copies of the target sequence) differ from source DNA in being non-methylated; this absence of methylation permits cutting by those restriction enzymes that are inhibited by the methylation of bases in DNA isolated from cells. The PCR-RFLP procedure is widely used e.g. for TYPING bacteria and for detecting mutations/polymorphisms.

PCR-RFLP has been used e.g. for studies on polymorphism in the cyclo-oxygenase-2 (COX-2) gene [BMC Gastroenterol (2009) 9:7], and for studying polymorphisms in the immuno-regulatory *FCRL3* gene [Mol Vision (2009) 15:955–961].

PCR-ribotyping An alternative approach to the standard form of RIBOTYPING in which PCR is used to amplify the intergenic spacer region between the 16S rRNA gene and the 23S

rRNA gene. One of the PCR primers is complementary to a conserved sequence in the 16S rRNA gene while the other is complementary to a conserved sequence in the gene for 23S rRNA. The amplicons are subjected to gel electrophoresis in order to generate a fingerprint which is based on variation in the length of the intergenic spacer region in different strains; strains are then compared on the basis of their fingerprints.

[Use of PCR-ribotyping (e.g.): characterization of strains of *Clostridium difficile*: BMC Microbiol (2009) 9:6.]

PCR-SSCP See SSCP ANALYSIS.

PD-(D/E)XK endonucleases See HOMING ENDONUCLEASE.

pDNA Plasmid DNA.

(cf. pRNA: see PACKAGING (of phage DNA).)

pDONR™ A vector, containing a pair of *att*P sites, that is used in the GATEWAY SITE-SPECIFIC RECOMBINATION SYSTEM for the initial incorporation of a target sequence which is flanked by *att*B sites. Insertion of the target sequence into a pDONR vector, mediated by BP Clonase, yields an *entry clone*, i.e. a vector-borne *att*L-flanked target fragment that can be inserted into any desired destination vector.

[Uses of pDONR™ (e.g.): BMC Genomics (2008) 9:321; Biochem J (2009) 419(3):577–584.]

PDR Polymorphism discovery resource: see entry DNA POLYMORPHISM DISCOVERY RESOURCE.

PEG See POLYETHYLENE GLYCOL.

PEGfilgrastim See Neulasta® in entry BIOPHARMACEUTICAL (table).

pegylated A term used to refer to an *in vitro* construct prepared with POLYETHYLENE GLYCOL (PEG).

A molecule (e.g. an enzyme) may be conjugated to PEG in order e.g. to stabilize it.

Pegylation of a DENDRIMER surface was used e.g. facilitate uptake by tumor cells.

PEI Polyethyleneimine: an agent used to facilitate transfection. The role of PEI in transfection was reported to be augmented by ultrasound [J Control Release (2007) 118(1):126–135].

[PEI in gene delivery (e.g.): Cancer Gene Therapy (2009) doi: 10.1038/cgt.2009.11.]

(See also MELITTIN.)

Peltier system See THERMOCYCLER.

penicillin Any of a class of β-LACTAM ANTIBIOTICS characterized by a thiazolidine ring fused to the β-lactam ring. As in all β-lactam antibiotics, the penicillins (which act only on *growing* cells) kill or inhibit bacteria by interfering with the function of the so-called *penicillin-binding proteins* (PBPs) – proteins which have enzymic roles in the synthesis of the cell wall polymer peptidoglycan (which is present in both Gram-positive and Gram-negative species); cells lyse when they are killed by penicillins (cf. STREPTOZOTOCIN).

Some penicillins, e.g. penicillin G (= benzylpenicillin), are active primarily, or solely, against Gram-positive bacteria as they cannot penetrate the outer membrane of Gram-negative species; others (e.g. ampicillin, amoxycillin and carbenicillin) have activity against both Gram-negative and Gram-positive species.

Penicillins, like cephalosporins, are produced by fungi; they include various semi-synthetic antibiotics – such as AMPICILLIN, cloxacillin, flucloxacillin, METHICILLIN and oxacillin.

Some of the other β-lactams (e.g. carbapenems and monobactams) are produced by bacteria.

In some cases, bacteria are resistant to penicillins because they produce enzymes (β-lactamases) capable of cleaving the β-lactam ring.

Some bacteria contain the *mecA* gene which encodes a PBP that is resistant to the action of various β-lactam antibiotics.

penicillin-binding protein 2a See MECA GENE.

penicillin-binding proteins See β-LACTAM ANTIBIOTICS.

PEP Phosphoenolpyruvate.

(See also the entry PTS.)

peplos (*virol.*) See the entry CAPSID.

α-peptide The N-terminal 146 amino acid residues of β-galactosidase (an ezyme encoded by the *lacZ* gene in *Escherichia coli*). This peptide can complement a mutant (inactive) form of the enzyme that lacks part of the N-terminal region; that is, the α-peptide can associate (*non*-covalently) with the mutant enzyme to form an active enzyme. Such association is called *α-complementation* and is exploited in a method for selecting recombinant vectors (see e.g. PBLUESCRIPT).

(See also BIMOLECULAR FLUORESCENCE COMPLEMENTATION and ENZYME FRAGMENT COMPLEMENTATION.)

peptide nucleic acid See PNA.

perfloxacin See QUINOLONE ANTIBIOTICS.

permeaplast A cell-wall-deficient cell obtained by treating a cyanobacterial cell with LYSOZYME and EDTA. Permeaplasts have been used e.g. for TRANSFORMATION in certain species.

permease See e.g. the entry PTS.

pertussis toxin An exotoxin, formed by the Gram-negative bacterial pathogen *Bordetella pertussis*, which is a virulence factor in whooping cough (pertussis); the toxin has an AB_5 arrangement of subunits, i.e. a central A subunit surrounded by a pentameric ring of (heterogeneous) B subunits – the A subunit being to one side of the plane of the ring.

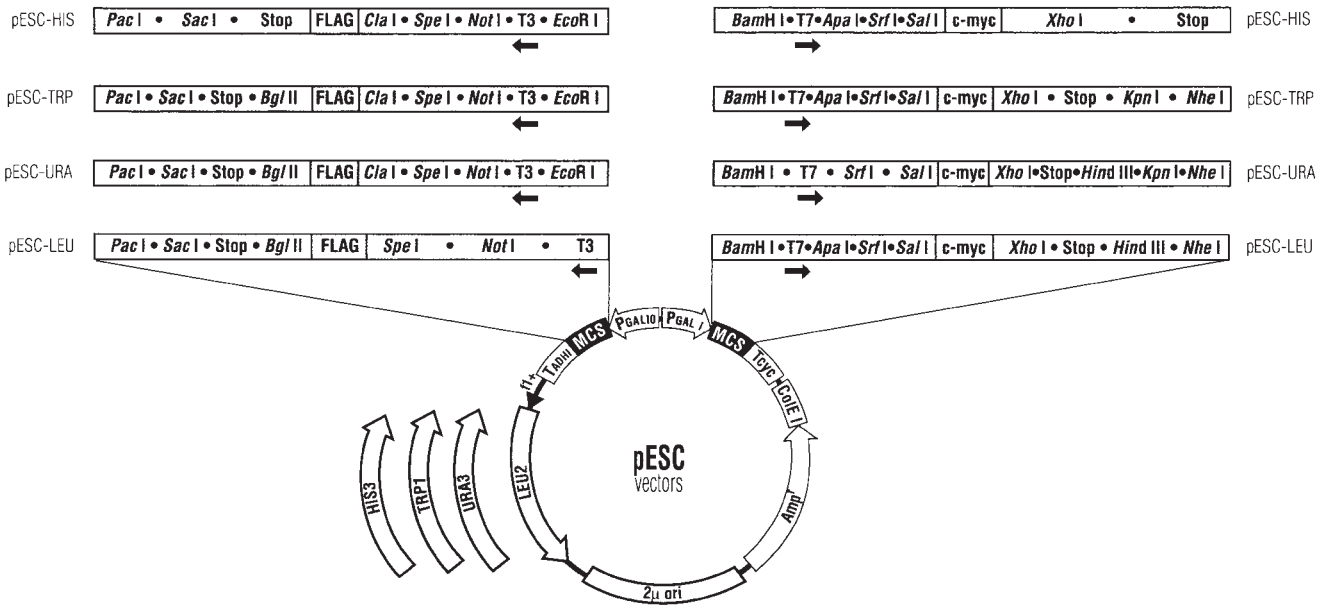
The A subunit (~26 kDa) has ADP-ribosyltransferase activity (see entry ADP-RIBOSYLATION); in target cells, it ADP-ribosylates the G_1 protein – thereby stimulating ADENYLATE CYCLASE, which leads to an increase in the concentration of intracellular cyclic AMP (cAMP).

pESC vectors Plasmid vectors used e.g. for expressing eukaryotic genes in yeast (*Saccharomyces cerevisiae*); these vectors include sequences for EPITOPE TAGGING which facilitate the detection/isolation of the expressed proteins.

A series of pESC vectors marketed by Stratagene (La Jolla CA) contain sequences encoding FLAG and c-*myc* epitopes (used for tagging target proteins). Each vector also includes GAL promoters, and one of four selectable markers – *HIS3*, *TRP1*, *URA3* and *LEU2* (see figure on page 256).

Two different genes can be expressed in each vector (transcription occurring in opposite directions).

pESC vectors were used e.g. for investigations on gene reg-



GAL1 polylinker

```
                                  Srf I                  Start                                                                  Stop      Kpn I        Nhe I
   BamH I                         Apa I        Sal I          c-myc epitope sequence                              Xho I       Hind III
5' GGATCCGTAATACGACTCACTATAGGGCCCGGGCGTCGAC ATG GAA CAG AAG TTG ATT TCC GAA GAA GAC CTC GAGTAAGCTTGGTACCGCGGCTAGC 3'
3' CCTAGGCATTATGCTGAGTGATATCCCGGGCCCGCAGCTG TAC CTT GTC TTC AAC TAA AGG CTT CTT CTG GAG CTCATTCGAACCATGGCGCCGATCG 5'
```

GAL10 polylinker

```
                                                Start                                                    Stop
   EcoR I                    Not I      Spe I   Cla I      FLAG epitope sequence                 Bgl II   Sac I   Pac I
5' GAATTCAACCCTCACTAAAGGGCGGCCGCACTAGTATCG ATG GAT TAC AAG GAT GAC GAC GAT AAG ATCTGAGCTCTTAATTAA 3'
3' CTTAAGTTGGGAGTGATTTCCCGCCGGCGTGATCATAGC TAC CTA ATG TTC CTA CTG CTG CTA TTC TAGACTGCAGAATTAATT 5'
```

pESC VECTORS A versatile system of vectors for expression of proteins in the yeast *Saccharomyces cerevisiae* (see entry for details). The polylinker (MCS) sequence downstream of each of the GAL promoters is shown at the bottom of the figure.

Courtesy of Stratagene, La Jolla CA, USA.

ulation in yeast [Genetics (2006) 173(1):75–85]; for studying the suppression of viral RNA recombination by cellular exoribonuclease [J Virol (2006) 80(6):2631–2640]; and for the synthesis of cyanophycin in transgenic *Saccharomyces cerevisiae* [Appl Environ Microbiol (2008) 74(11):3410–3418].

PEST domain A proline–glutamate–serine–threonine-rich sequence in ornithine decarboxylase. If the nucleotide sequence encoding the PEST domain is used as a tag for a fusion gene it promotes rapid degradation of the resulting fusion protein in a PROTEASOME. (For an example of use see entry GREEN FLUORESCENT PROTEIN.)

Petri dish A round, shallow, flat-bottomed and vertically sided dish, made of plastic or glass, which has a flat, loosely-fitting lid.

A Petri dish is commonly about 10 cm in diameter.

(See also PLATE.)

Peutz–Jeghers syndrome (PJS) A rare autosomal dominant condition characterized by intestinal hamartomatous polyposis and pigmentation in the oro-facial region; complications include intestinal obstruction, and there is an increased risk of gastrointestinal malignancy.

Mutation within the gene for serine-threonine kinase (*STK11*/*LKB1*, chromosomal location 19p13.3) seems to account for at least the majority of cases, and new mutations in the *STK11* gene continue to be reported [BMC Med Genet (2008) 9:44].

pEXP-DEST vectors See CELL-FREE PROTEIN SYNTHESIS.

pEXP3-DEST vector A destination vector used in the GATEWAY SITE-SPECIFIC RECOMBINATION SYSTEM (see table).

[Use (e.g.): PLoS ONE (2009) 4(2):e4509.]

(See also CELL-FREE PROTEIN SYNTHESIS.)

pEXP4-DEST vector A destination vector used in the GATEWAY SITE-SPECIFIC RECOMBINATION SYSTEM (see table).

pFastBac^{TM} dual vector kit A kit (Invitrogen, Carlsbad CA) containing a vector (pFastBac^{TM} Dual – 5.2 kb) for use in the BAC-TO-BAC expression system; the pFastBac vector contains two promoters (polyhedrin and p10), in opposite orientations (and with appropriate polylinkers), allowing the simultaneous expression of two proteins (for details of this vector see the diagram in the entry BAC-TO-BAC).

The pFastBac dual vector was adapted for the expression of *Manduca sexta* proteins [Insect Biochem Mol Biol (2008) 38 (1):89–98].

PFGE Pulsed-field gel electrophoresis: a specialized form of GEL ELECTROPHORESIS which can be used to separate large pieces of nucleic acid of up to ~10–15 million base pairs. The fragments are exposed, alternately (in pulses), to two separate electric fields that are arranged at different angles to the gel.

The ability to separate fragments of linear DNA depends on several factors that include pulse time, total running time (of the experiment), the applied field strength and the positions of the electrodes. The pulse *length* affects the size range of the fragments separated: short pulses can be used to separate the relatively smaller fragments, while long pulses are needed to separate the larger fragments.

Disadvantages of PFGE include the time factor (the process may take 2–3 days) and the susceptibility of the isolated genomic DNA to endogenous nucleases.

Variant forms of PFGE differ e.g. in the arrangement of the electrodes; they include:

CHEF (*contour-clamped homogeneous electric field electrophoresis*). The electrodes are arranged on sides 2, 3, 5 and 6 of a hexagonal chamber.

Double inhomogeneous field electrophoresis. Two, long negative electrodes (cathodes) are on one side of the gel, and two short positive electrodes (anodes) are on the other side.

Field inversion electrophoresis. Two electrodes of equal size are present, one on each side of the gel; one acts as an anode, the other as a cathode – but periodically (depending on the pulse length) the polarity is switched.

PFGE (on samples of fragmented genomic DNA) has been used e.g. for TYPING an extensive range of bacteria (such as methicillin-resistant *Staphylococcus aureus*, MRSA), and is regarded by some as the 'gold standard' in molecular typing methods for bacteria.

pfu PLAQUE-FORMING UNIT (q.v.).

***PfuTurbo*®** The trade designation (Stratagene, La Jolla CA) for certain types of DNA polymerase which are formulated for high-fidelity replication in PCR.

(See also ARCHAEMAXX.)

The enzyme *PfuTurbo*® C_x Hotstart DNA polymerase has been found useful for replicating DNA on uracil-containing templates [Nucleic Acids Res (2006) 34(18):e122].

PGD Preimplantation genetic diagnosis: a procedure which can be carried out in association with IVF and which is used for detecting specific genes linked to certain inherited disorders. Essentially, a cell, taken by micromanipulation from an 8-cell embryo, is subjected to genomic analysis by *sequencing* the relevant gene(s) (see entry DNA SEQUENCING); it seems that the embryo is not harmed by the loss of a cell.

(See also ALLELE DROP-OUT.)

During *pregnancy*, non-invasive diagnosis of fetal chromosomal aneuploidy can be made from fetal DNA in samples of maternal plasma: see e.g. DOWN'S SYNDROME.

pGEX-2T plasmid See TEV.

Ph^{1} The Philadelphia chromosome: see the entry ABL.

phage A common abbreviation for BACTERIOPHAGE (q.v.); see also the entries below for details of specific phages.

phage BPP-1 A temperate BACTERIOPHAGE which infects the Gram-negative bacteriun *Bordetella*; it exhibits a host-range switching mechanism involving a diversity-generating retroelement (see DGR).

phage Bxb1 A BACTERIOPHAGE that infects a rapidly growing species of *Mycobacterium*: *M. smegmatis*. An integrase from phage Bxb1 (a serine recombinase) was reported to mediate efficient SITE-SPECIFIC RECOMBINATION in mammalian cells [BioTechniques (2006) 40(4):460–464].

[Bxb1-mediated excision: Proc Natl Acad Sci USA (2008) 105(9):3238–3243.]

phage conversion See LYSOGENY.

phage CTXΦ A filamentous BACTERIOPHAGE whose ssDNA genome encodes e.g. cholera toxin; CTXΦ infects the Gram-negative bacterium *Vibrio cholerae*. The manner of insertion of (double-stranded) versions of the phage genome into the host cell chromosome was reported to affect the ability of the prophage to give rise to phage virions [J Bacteriol (2000) 182 (24):6992–6998].

(See also (specialized) TRANSDUCTION.)

phage display A method in which phages are used for selecting a ligand that binds to a given target molecule with high-level affinity/specificity.

Initially, molecules of the given target are immobilized and exposed to a large population of recombinant phages. These phages have been genetically modified, *in vitro*, so that their coat proteins (i.e. surface proteins) differ from wild-type coat proteins; genetic modification of the phages is carried out on a random basis so that, collectively, the population of phages exhibits an enormous diversity of recombinant coat proteins, each having a potential binding site for a given type of target molecule. Such a genetically modified population of phages is called a *phage library* (see later).

After the target molecules have been exposed to the phage library, any phages which do *not* bind to the target molecules are removed by washing (and discarded). The phages which *do* bind are eluted (removed from the target molecules) by physical means (e.g. adjusting the pH); these phages are used to infect their host bacteria – within which they are replicated to produce vast numbers of progeny phages. These progeny phages are therefore an expanded population of those phages which were selected from the library owing to their ability to bind to the given target; they usually consist of a number of different strains with various forms of modified coat protein.

Purification of these phages can be carried out by a cesium chloride equilibrium density gradient centrifugation method, or by chromatography using Sephacryl^{TM} or hydroxyapatite.

Target molecules are exposed to the progeny phages under more stringent conditions – i.e. conditions under which only those phages with the highest level of affinity/specificity will be able to bind to the target molecules. Unbound phages are, as before, removed by washing.

The entire selective process is repeated.

The technique described above, i.e. selecting phages from a phage library by exposing the library, and successive batches of progeny phages, to the target ligand, is called *biopanning* (or simply *panning*). (The basic principle involved is similar to that used in the selection of aptamers by SELEX (see entry APTAMER).)

The final batch of phages, characterized by strong, selective binding to the target, are examined to determine the nature of the phage's binding site – and also the coding sequence of the binding site in the phage genome.

Phage display technology makes use of filamentous phages (e.g. M13) which are typically very long and thin and which (compared with e.g. phage T4) have a relatively high ratio of surface protein to nucleic acid.

Preparation of phage libraries

A phage library may be prepared by inserting a gene library – consisting of unrelated genes/fragments – into the DNA of unmodified phages. Many of the genes/fragments will fuse with a phage gene that specifies a coat protein (e.g. protein pIII), and the resulting fusion protein will replace the phage's normal coat protein. Each hybrid gene will encode a different type of fusion protein – so that the coat proteins of the phage population, as a whole, will display a wide range of potential binding sites for any given target molecule.

Another procedure for making a phage library involves *in vitro* randomization of a given sequence of nucleotides in the gene of a phage coat protein. The result is an immense range of different combinations of nucleotides, and the coat protein gene in a given phage particle will contain only one of these combinations; consequently, the coat proteins of the phage population collectively display an immense range of potential binding sites.

The target ligands for phage display technology include e.g. those involved in gene therapy techniques and various other mammalian cell-surface receptors (such as receptors for cytokines, and ligands relating to receptor-mediated endocytosis). The need to optimize receptor–ligand binding in the context of gene therapy is clearly important as this can help to avoid wasteful delivery of genes to untargeted cells and, therefore, help to reduce the required dosage for the patient.

Applications of phage display

Examples of use of phage display include: studies on *in vitro* protein evolution to optimize binding of an IL-13-neutralizing antibody [Proc Natl Acad Sci USA (2006) 103(20):7619–7624]; generating Epstein–Barr virus antigen mimotopes to assist diagnosis [J Clin Microbiol (2006) 44(3):764–771]; use of scFv antibody phage display libraries in studies on a tumor antigen [Proc Natl Acad Sci USA (2006) 103(4):1041–1046]; studies on epitope characterization of a monoclonal antibody [PLoS ONE (2009) 4(3):e4924]; studying transcription factor interactions with DNA [PLoS Genet (2009) 5(4):e1000449]; identifying peptides which can block *in vitro* infection by the Andes virus [J Virol (2009) 83(17):8965–8969].

phage display-mediated immuno-PCR See IMMUNO-PCR.

phage f1 A filamentous ssDNA BACTERIOPHAGE that is related to PHAGE M13.

phage fd A filamentous ssDNA BACTERIOPHAGE that is related to PHAGE M13.

phage L54a A BACTERIOPHAGE which infects *Staphylococcus aureus*, causing phage conversion (see LYSOGENY).

phage lambda (λ) A temperate BACTERIOPHAGE of the family Styloviridae; host: *Escherichia coli*. The virion has an icosahedral head (55 nm diameter) and a non-contractile tail (∼150 × 10 nm).

The genome (within the virion) is linear dsDNA (∼48.5 kb) with single-stranded 12-base 5′ STICKY ENDS; circularization

occurs intracellularly, the sticky ends hybridizing to form the so-called *cos* site.

Lysogeny/lysis decision after phage infection of bacteria

During transcription of the phage early genes, the result of infection may be either LYSOGENY or cell lysis. The outcome will depend e.g. on the activity of two mutually antagonistic products of the early phage genes: the cI protein (a repressor of transcription from certain operators) whose activity favors lysogeny, and the Cro protein – which inhibits transcription of the gene encoding cI and promotes the lytic phase. Factors that affect the lysogeny/lysis decision also include starvation (favoring lysogeny) and the multiplicity of infection (MOI) – a high MOI (i.e. each bacterium infected with many phages) favoring lysogeny.

Lysogeny

Lysogeny involves integration of the (circularized) prophage into the host cell's chromosome. This occurs by *site-specific recombination* between an 'attachment site' (designated *att*P) in the prophage and another attachment site (designated *att*B) in the bacterial chromosome (located between the *gal* and *bio* genes in the chromosome).

Integration of the prophage into the chromosome involves a phage-encoded integrase (*int* gene product) and the bacterial INTEGRATION HOST FACTOR (IHF).

The bacterial and phage attachment sites are not identical, although there is a central 'core' sequence (O) that is common to both sites. The bacterial (*att*B) site contains two 'arms' (B and B′) either side of the central O sequence, i.e. the structure of *att*B is BOB′. Analogously, the structure of *att*P is POP′.

The recombination event involves strand breakage in the O region of each *att* site, strand exchange, and ligation. Hence, the integrated prophage is flanked by hybrid *att* regions BOP′ (= *att*L) and POB′ (= *att*R).

The *att* sites are exploited in commercial systems: see e.g. GATEWAY SITE-SPECIFIC RECOMBINATION SYSTEM.

Phage λ also encodes another recombination system which is exploited as the LAMBDA (λ) RED RECOMBINATION system.

Lytic cycle

The lytic cycle of phage λ may begin on phage infection of the cell and circularization of the genome. (Alternatively, in a lysogenic cell, the prophage may leave the bacterial chromosome to form a (circularized) phage genome; this needs the product of phage gene *xis*: an excisionase that recognizes the *att*L and *att*R sites.) Initially, the circularized phage genome is replicated by the CAIRNS MECHANISM (= theta replication) and later by the ROLLING CIRCLE mechanism. Rolling circle replication results in the formation of a continuous chain of double-stranded phage genomes, joined end-to-end, in which the c*os* site is repeated at regular intervals.

A *cos* site is cleaved (by an enzyme) and a *cos* sticky end locates in a phage head; the DNA is then fed into the phage head until the next *cos* site is reached – and cleaved. Hence, the phage head contains linear DNA with terminal *cos* sticky ends. The tail is then added to the phage head, and the phage is released by cell lysis. Lysis involves the phage *S* gene (see below).

Lambda-based vectors and constructs

Modified forms of the phage λ genome are used as cloning or expression vectors (see e.g. LAMBDA FIX II VECTOR and ZAP EXPRESS VECTOR).

Some lambda-based vectors encode a temperature-sensitive (i.e. mutant) form of the cI protein (cI_{857} or *cI857* mutation). With such vectors, lysogenic growth occurs at, say, 32°C (as cI is active at this temperature), but a switch to the lytic phase can be achieved simply by raising the temperature to 42°C (which inactivates cI). [Use of this mutant protein in a study on the induction of lysogenic phage λ: Mol Microbiol (2008) 68(1):29–36.]

In some lambda-based vectors there is a mutant form of the *S* gene that inhibits lysis of the host cell; such inhibition tends to increase the yield of progeny phages because delayed lysis gives more time for the maturation of phage particles.

The phage λ *cos* site is used in COSMIDS and PHASMIDS.

[The impact of phage lambda (from restriction to recombineering): Biochem Soc Trans (2006) 34(2):203–207.]

[*In vivo* gene delivery and expression by phage λ vectors: J Appl Microbiol (2007) 102(5):1337–1349.]

phage library See PHAGE DISPLAY.

phage M13 A filamentous, non-lytic BACTERIOPHAGE able to infect strains of *Escherichia coli* containing an F PLASMID (or a related plasmid). When infecting the cells of *E. coli*, phage M13 initially binds to the *tips* of (plasmid-encoded) F pili or F-like pili (see PILUS); later, phage DNA enters the cell. Note that this phage cannot infect cells which lack the appropriate pili.

The genome of M13 is covalently closed ssDNA.

Phage M13 has been used in SITE-DIRECTED MUTAGENESIS (q.v.).

Replication of the phage M13 genome can be divided into three stages:

In *stage I*, a complementary strand (*c* strand, *minus* strand) is synthesized on the (circular) viral ssDNA (*v* strand, or *plus* strand) to form the so-called (dsDNA) *replicative form* (RF). When supercoiled (by the cell's gyrase), RF is referred to as RFI. Phage genes are transcribed from the *c* strand of RFI. One of the early phage-encoded products is an endonuclease which starts stage II by nicking the *v* strand at a specific site.

In *stage II*, the 3′ end of the nicked *v* strand is elongated by a ROLLING CIRCLE process, displacing the 5′ *v* strand. Genome lengths of *v* strand are cut and circularized. Initially, the new, circularized *v* strands are converted to (double-stranded) RFs. Subsequently, a phage-encoded protein accumulates in the cell and coats the newly formed *v* strands; this prevents their conversion to RFs. *Stage III* of the replication process in M13 is characterized by asymmetric DNA synthesis in which only the circularized *v* strands are formed.

The M13 coat proteins are synthesized with an N-terminal *signal sequence* that targets them to the bacterial cytoplasmic

membrane. Assembly of virions – a THIOREDOXIN-dependent process in M13 and related phages – occurs in the bacterial cell envelope, and phage particles are extruded from (living) host cells.

The above account refers also to phages fd and f1; the other phages in this group appear to be generally similar.

Phage M13 and derivatives are used e.g. for making single-stranded copies of sample DNA and (e.g. strain M13K07) as a HELPER PHAGE for single-strand rescue from PHAGEMIDS.

phage Mu A temperate BACTERIOPHAGE which infects various enterobacteria (including *Escherichia coli*). The virion has an icosahedral head (∼60 nm diam.) and a contractile tail (∼150 nm long, extended) with a base plate, spikes and fibers; the head contains linear dsDNA (∼39 kb) consisting of the phage genome (∼37.5 kb) flanked, at both ends, by a sequence of bacterial DNA acquired during the lytic/packaging process – packaging occurs by the 'headful' mechanism, as in phage T4.

Mu DNA includes: the α region, which contains most of the phage genes – including the *c* (repressor) gene that is needed for LYSOGENY; a small central region, G, that encodes some phage tail components, and which determines the host range of the phage; and the β region, encoding e.g. a RECOMBINASE which can invert the G segment (and, hence, change the host range of the phage) and a DNA METHYLTRANSFERASE which methylates (i.e. 'modifies') adenine residues in phage DNA in the sequence:

$$5' \ldots . (C/G)A(C/G)NPy \ldots . 3'$$

Modification of the phage DNA protects it from degradation by the host cell's RESTRICTION ENDONUCLEASES; moreover, modified Mu DNA is resistant to a range of REs *in vitro*.

Locations in the Mu DNA have been referred to as *left end* (meaning the terminal, outer/distal, part of the α region) and *right end* (or *S end*) (meaning the terminal, outer/distal, part of the β region).

On infection of a host cell, the phage DNA integrates into the host's chromosome at a more or less random location, and integration gives rise to a 5-bp duplication of the target site. Integration commonly inactivates a gene in which it occurs; in a population of bacteria lysogenized by Mu, up to ∼3% of the cells contain a detectable mutation. (The name 'Mu' refers to the mutagenic activity of this phage.)

Integration of phage Mu involves the Mu *A* gene product, a TRANSPOSASE which recognizes terminal (outer) sequences in Mu DNA and mediates the insertion process. Lysogeny is promoted by the product of phage gene *c*.

During the lytic cycle, replication of Mu DNA occurs by repeated *replicative* transposition of the integrated prophage (see TRANSPOSABLE ELEMENT), a process which involves the transposase (product of phage gene *A*); other factors involved include the host's HU PROTEIN and the product of phage gene *B*. Acting as a giant transposon, phage DNA is duplicated at random sites in the bacterial genome; this kills the cell, and about 100 progeny phages are released after about 60 minutes infection.

Mu-type transposition in vitro

While *in vivo* transposition is a complex process, involving a number of phage and host factors, a simplified system can be operated *in vitro* using transposase and incorporating transposase-binding sites (Mu-end sequences) in DNA substrates. The method was described some years ago [see e.g. Nucleic Acids Res (1999) 27(13):2777–2784], and has been exploited in a number of contexts – e.g. for identifying SNPs [Nucleic Acids Res (2007) 35(6):e44], and for making gene-targeting vectors for conditional knock-out mice [PLoS ONE (2009) 4(2):e4341].

An analogous procedure is based on transposon Tn*5* – see the entry TN5-TYPE TRANSPOSITION.

(See also G LOOP and MINI-MU.)

phage N15 A BACTERIOPHAGE with a linear dsDNA genome which has covalently closed (hairpin) terminal regions (*telomeres*); hosts include *Escherichia coli* and strains of *Yersinia*.

The prophage of N15 does not integrate with the bacterial chromosome.

The mode of DNA replication of this phage has features in common with that of chromosomal replication in BORRELIA BURGDORFERI (q.v.). (See also PROTELOMERASE.)

A construct derived from phage N15 is used as a (linear) vector in the BIGEASY LINEAR CLONING SYSTEM.

DNA replication in PHAGE φ29, another phage with a linear dsDNA genome, involves a different mechanism.

phage P1 A temperate BACTERIOPHAGE of the family Myoviridae; host bacteria include *Escherichia coli*. The virion has an icosahedral head, a long contractile tail and six tail fibers.

The genome is linear dsDNA, 93.6 kb, which is circularly permuted and terminally redundant. Within the host cell, the prophage (genome) circularizes, with loss of redundancy, and usually remains as an extracellular low-copy-number plasmid (one or two copies per chromosome) during lysogenic infection; the plasmid is referred to as the 'P1 plasmid'.

Within a lysogenic cell, copies of the plasmid may dimerize by homologous recombination. Nevertheless, at cell division, plasmid *monomers* are segregated to daughter cells; monomerization is mediated by the (phage-encoded) site-specific recombination system, the CRE–LOXP SYSTEM, that resolves plasmid dimers into monomers by a recombination event at the *loxP* site.

[P1 genome: J Bacteriol (2004) 186(21):7032–7068.]

phage φ29 A BACTERIOPHAGE (host: *Bacillus* species) with a linear dsDNA genome. Each of the 5′ ends of the DNA is covalently linked to a phage-encoded protein (designated TP).

DNA replication starts at both ends of the genome. DNA polymerase (encoded by the phage) forms a complex with a *free* molecule of TP and localizes at the 3′ end of each strand; a molecule of dAMP is then covalently linked to this TP – and acts as the initial nucleotide for the synthesis of the new strand. After insertion of nucleotide 10 in the new strand the polymerase dissociates from TP and continues to synthesize the remainder of the strand. TP remains bound to the 5′ end of the new strand. Such *protein priming* is one example of

the strategies used for replication of linear dsDNA genomes. A different strategy is used for replicating the linear chromosomal DNA of the bacterium BORRELIA BURGDORFERI, and for replication in PHAGE N15.

phage φKO2 A BACTERIOPHAGE with a linear dsDNA genome that has closed (hairpin) terminal regions (*telomeres*). During DNA replication the chromosomal dimer is resolved by the enzyme PROTELOMERASE (q.v.).

phage Rac A LAMBDOID PHAGE that lysogenizes certain strains of *Escherichia coli* K-12; the prophage encodes the RecET recombination functions (genes *recE*, *recT*) that are exploited (as ET RECOMBINATION) in RECOMBINEERING (q.v.).

phage T3 A BACTERIOPHAGE (hosts: enterobacteria) which is closely related to PHAGE T7 (q.v.).

phage T4 A virulent (lytic) BACTERIOPHAGE (hosts: enterobacteria) of the family Myoviridae. The T4 genome is linear, circularly permuted and terminally redundant dsDNA (∼170 kb) [T4 genome (a review): Microbiol Mol Biol Rev (2003) 67:86–156]; the presence of hydroxymethylcytosine (instead of cytosine) involves the activity of phage-encoded enzymes. The virion is structurally complex; it has an elongated head (∼110 × 85 nm) and a long, contractile tail.

Following infection of a bacterial cell, the phage genes are transcribed by the host cell's RNA polymerase. Contrary to the situation in PHAGE T7, the host's polymerase appears to be used throughout – although modified (e.g. ADP-ribosylated) in a series of phage-induced alterations.

The genome contains more than one origin of replication; replication occurs bidirectionally from a given origin. During replication, the 3′ end of a lagging strand template remains single-stranded, and this single-stranded region can displace a homologous sequence in a double-stranded region in another copy of the replicating genome. Such 3′ ends act as primers, resulting in the formation of a complex, branching network of phage genomes.

DNA packaging occurs by the *headful mechanism*: one end of a multi-genome molecule is fed into a head until the head is full; the DNA is then cleaved.

Phage-encoded lysozyme (and a *lysis protein*) weaken the cell envelope, and the virions are released by osmotic lysis.

phage T7 A virulent (lytic) BACTERIOPHAGE (hosts: enterobacteria) of the family Podoviridae. The genome is a linear ds DNA molecule (∼40 kb) that has terminally redundant ends. Each virion has of an isometric head (diameter ∼60 nm) and a tail with six short fibers.

Products of the early phage genes (transcribed by the host's RNA polymerase) include a protein kinase (which inactivates the host's RNA polymerase) and an RNA polymerase which transcribes the intermediate and late genes of the phage; the intermediate genes are concerned mainly with DNA replication, and late genes encode functions such as phage assembly and host cell lysis.

During DNA replication the T7 genome remains as a linear molecule. Initiation occurs at a single site near one end of the genome; the primers are apparently synthesized by the phage RNA polymerase, and elongation occurs bidirectionally, with leading and lagging strands, from the replication origin. The newly formed duplexes recombine to form concatemers that are further replicated. Concatemers are cleaved into genome-length pieces; a maturation process includes the formation of 160-bp terminally redundant sequences. Mature virions are released by host cell lysis.

In the initial infection of a bacterium, the phage apparently ejects channel-forming proteins which permit the transfer of DNA through the cell envelope; it was speculated that some of the ejected proteins may form components of a 'motor' that actively transports DNA into the host cell.

(See also BACTERIOPHAGE (table).)

phage typing A method for TYPING bacteria in which a given strain is exposed to a range of different phages and examined for sensitivity or resistance to each of the set of phages used. The strain is then defined in terms of the pattern of sensitivity or resistance to each of the given set of phages.

Essentially, a FLOOD PLATE of the test strain is prepared, and excess surface moisture allowed to dry. A grid is drawn on the base of the plate, and the agar surface is inoculated with one drop of each phage suspension – each drop placed in a separate square of the grid. The plate is then incubated at the growth temperature of the given strain. After incubation, the plate is examined for PLAQUES. Any phage which lyses the test bacterium will have formed macroscopic plaques in the opaque surface layer of bacterial growth on the plate; any phage to which the strain is resistant will not form plaques.

Certain species of bacteria, e.g. *Staphylococcus aureus* and *Yersinia enterocolitica*, are commonly typed by this method.

phagemid A circular, dsDNA hybrid VECTOR (constructed *in vitro*) which includes (i) a plasmid's origin of replication; (ii) a POLYLINKER; (iii) a marker, such as an antibiotic-resistance gene; and (iv) the origin of replication of an ssDNA bacteriophage (such as phage f1).

As a CLONING VECTOR, a phagemid can be used to make double-stranded copies of a gene or fragment by replication in bacteria. The gene/fragment is inserted into the phagemid, via the polylinker, and copies of the recombinant phagemid are inserted into bacteria (e.g. by transformation); within the bacteria, phagemids replicate as plasmids – using the plasmid origin of replication.

A phagemid can make *single-stranded* copies of the target DNA if the phagemid-carrying bacteria are infected with a so-called HELPER PHAGE – for example, PHAGE M13 – which provides functions necessary for replication from the *phage* origin in the phagemid; copies of ssDNA, including the gene/fragment of interest, are packaged in phage coat proteins that are encoded by the helper phage, and the phagemid particles are exported to the external medium from (living) bacteria. After removing the coat proteins, the ssDNA can be used e.g. for sequencing.

A phagemid can also be used as an EXPRESSION VECTOR by inserting the target DNA and also inserting an upstream, in-frame promoter of a type that functions in the host cell.

A phagemid can be used for *in vitro* transcription by including a promoter for T3 or T7 RNA polymerase (enzymes derived from phages) upstream of (and in phase with) target DNA.

Phagemid particles, with modified coat protein, have been prepared as vectors for gene delivery to eukaryotic cells (see the entry GENE THERAPY).

One example of a phagemid is PBLUESCRIPT (q.v.).

pharmacogenomics A study of the interactions between drugs and the genome, including modification of gene expression by drugs, and the effect of mutation(s) in those genes whose products are involved in the response to specific drugs.

phase-shift mutation *Syn*. FRAME-SHIFT MUTATION.

phase variation ANTIGENIC VARIATION which includes on–off switching of antigen-encoding genes, and changing from one antigen-encoding sequence to another *pre-existing* sequence.

For example, in many strains of *Salmonella* the flagellar filament is composed of one or other of two distinct types of protein (H1 and H2). Expression of the H2 gene (filament made with the H2 protein) is accompanied by production of a repressor of the H1 gene. The sequence that contains the H2 promoter is flanked by two sites which are recognized by a recombinase; SITE-SPECIFIC RECOMBINATION, mediated by the recombinase, inverts this sequence – thus inactivating the H2 promoter and stopping the production of the H1 repressor. Transcription of the H1 gene results in a filament made with H1 protein.

A different mechanism regulates *opa* genes in *Neisseria*. In these genes the leader region contains tandem repeats of the (pentameric) sequence CTCTT, and the *number* of repeats in a given gene may be changed e.g. through the mechanism of SLIPPED-STRAND MISPAIRING; such changes may affect the reading frame in an *opa* gene and, hence, change the resulting OPA PROTEIN.

(See also VSG.)

phasmid A hybrid (engineered) construct formed from a phage genome and a PLASMID.

For example, a construct can be prepared by recombination between a phage λ genome and a (modified) plasmid which contains (i) an *att* site (permitting site-specific recombination with the λ phage genome) and (ii) a *cos* site. Recombination between phage and plasmid can occur within a bacterial cell and requires e.g. the (phage-encoded) integrase. Phasmids are encapsidated in phage heads during replication of the phage, and any target DNA (carried by the plasmid moiety) can be inserted into other bacteria by infecting them with phasmid-containing phage particles.

A phasmid was used during the construction of a recombinant mycobacteriophage; the recombinant phage – carrying a fluorescent reporter gene – was used for detecting antibiotic resistance in *Mycobacterium tuberculosis* [PLoS ONE (2009) 4(3):e4870].

(cf. PHAGEMID.)

Note. The term 'phasmid' is also used for a chemoreceptor in the caudal region of certain types of nematode worm.

phenocopy (1) (*noun*) Any cell or organism that exhibits a new phenotype which mimics a phenotype that would result from a genetic change – i.e. the cell or organism *appears* to have experienced a genetic change.

One example of a phenocopy is observed when bacterial F^+ cells are grown to high density. These cells lose their donor characteristics and mimic (transfer-deficient) F^- cells; however, when these F^- phenocopies are grown in fresh medium they revert to the F^+ donor condition.

(2) (*verb*) To produce an effect similar to that produced by another event. For example, a null mutation in protein X may give rise to a phenotype identical to that produced by a null mutation in protein Y, one mutation being said to phenocopy the other.

phenol–chloroform–isoamyl alcohol A mixture of reagents used e.g. for isolating DNA. The DNA tends to partition in the top layer and may then be precipitated by isopropanol or ethanol.

phenotype Those characteristic(s) of a given organism which can be observed/detected/measured – manifestations of the organism's genotype; in genetics, phenotype is often used to refer to specific, designated characteristic(s).

(cf. GENOTYPE.)

phenotypic lag A delay in the expression of an allele which has been newly acquired e.g. by mutation or by gene transfer. In bacteria, such delay may be due e.g. to SEGREGATION LAG.

phenotypic mixing A phenomenon which may occur if a cell is infected simultaneously with two genetically distinct viruses: e.g. incorporation of the genome of one virus into the capsid of the other (*transcapsidation*, *genetic masking*), or inclusion of structural component(s) from one virus in virions of the other.

(See also REASSORTANT VIRUS.)

phenotypic screen Any screening process in which individuals in a given population are identified according to one or more phenotypic characteristics.

4-phenylbutyric acid An inhibitor of histone deacetylases (see the entry HDAC).

phenylketonuria An autosomal recessive disease, the classical form of which is characterized by phenylpyruvate in the urine and high levels of phenylalanine in blood and other tissues; it is caused by inadequacy of phenylalanine mono-oxygenase. The condition leads to mental retardation. Treatment includes limiting dietary phenylalanine.

(See also GENETIC DISEASE (table).)

pheromone (1) (in quorum sensing) See QUORUM SENSING.

(2) (in conjugation) See CONJUGATION.

φ29 phage (phi 29 phage) See PHAGE φ29.

Philadelphia chromosome See the entry ABL.

phleomycin D1 *Syn*. ZEOCIN.

phorbol ester (TPA) See ZEBRA.

phosphodiester bond In DNA and RNA: the bond linking the 5′-phosphate group of one sugar residue to the 3′-OH group of the next (upstream, 5′) sugar residue.

(See also DNA LIGASE.)

phosphoenolpyruvate-dependent phosphotransferase system See the entry PTS.

phosphoramidite Any of a class of compounds used e.g. in the chemical synthesis of oligonucleotides, and also for adding a group to the terminal 5′-phosphate of an existing oligonucleotide.

phosphorodiamidate morpholino oligomer See MORPHOLINO ANTISENSE OLIGOMER.

photo-cross-linking assay See CROSS-LINKING ASSAY.

photolyase (DNA photolyase) In *Escherichia coli* (and in some other bacteria): a flavoenzyme (product of gene *phr*) which can cleave the bonds in THYMINE DIMERS – produced when DNA is damaged by ULTRAVIOLET RADIATION; this reaction (*photoreactivation*, *photorestoration*), which uses the energy in blue light, can restore a normal base sequence.

[Electrochemistry of *E. coli* photolyase action: Proc Natl Acad Sci USA (2005) 102(31):10788–10792.]

Photolyase is reported to be absent in some bacteria (e.g. *Bacillus subtilis*). It is absent in *Ehrlichia* spp, but occurs in the (related) bacterium *Neorickettsia*; however, this enzyme is related more closely to the photolyase of *Coxiella burnetii* than to the *E. coli* enzyme [PLoS Genetics (2006) 2(2):e21].

Photolyases also occur in some eukaryotes. A photolyase in the rice plant (*Oryza*) appears to differ, structurally, from the *E. coli* enzyme [Plant Physiol (2008) 146(4):1941–1951].

photoreactivation See PHOTOLYASE.

photorestoration See PHOTOLYASE.

***phr* gene** The gene encoding PHOTOLYASE in *Escherichia coli*.

phred A program [Genome Res (1998) 8(3):186–194] designed to assist in estimating a probability of error in the context of parameters from trace data from high-throughput sequencing systems.

phylotype (*microbiol.*) Any hitherto uncultured species which has been defined solely on the basis of data from nucleic acid sequence(s).

physical map (of a genome) The complete sequence of nucleotides in the genome.

When applied e.g. to a plasmid vector molecule, the term 'physical map' may refer simply to a diagram which shows the locations of e.g. specific gene(s) and restrictions sites.

phytoalexins (*plant pathol.*) Low-molecular-weight chemically diverse antimicrobial compounds which are formed by higher plants when exposed to certain types of microorganism or to some metals (such as copper or mercury), detergents or other chemicals, or to physical stimuli such as freeze/thaw damage.

Phytoalexins appear to be synthesized locally in the plant, in response to the given stimulus. In some cases, production of a phytoalexin enables a plant to resist the development of disease on infection with a given pathogen, but in other cases phytoalexins appear not to be the sole factor that determines resistance.

(cf. PATHOGENESIS-RELATED PROTEINS.)

(See also GENE-FOR-GENE CONCEPT.)

phytokinins *Syn.* CYTOKININS.

Pichia A genus of budding yeasts, related to *Saccharomyces*; species are widespread. The sexually derived spores (= ascospores) are characteristically shaped in some species.

[Sequence of the (9.43 Mbp) genome of *P. pastoris*: Nature Biotechnol (2009) 27:561–566.]

In DNA technology, *P. pastoris* has been used extensively for the production of a wide variety of heterologous proteins. Reasons for choosing this organism include: (i) similarity to *Saccharomyces cerevisiae* (with well characterized genetic/cultural properties); (ii) high-level synthesis of heterologous proteins; (iii) the ability to synthesize intra- and extracellular proteins; and (iv) an ability to carry out various types of post-translational modification – e.g. glycosylation. Nevertheless, strains of *Pichia* need to be modified for some purposes: see the entry GLYCOSYLATION.

Regulated, *inducible* protein synthesis in *P. pastoris* can be achieved by using expression vectors that include the strong promoter of the *Pichia AOX1* gene.

For *constitutive* expression, one can use the promoter of the gene encoding glyceraldehyde 3-phosphate dehydrogenase.

(See also β-LACTAMASES (*Detection of β-lactamases*) and PPICZ VECTOR.)

pico- A prefix meaning 10^{-12}. For other prefixes see the section *Ready reference* at the front of the dictionary.

PicoGreen® assay An assay (Molecular Probes, Eugene OR/Invitrogen, Carlsbad CA) for quantitating dsDNA in solution. A reagent is added and incubated for 5 minutes; when bound to double-stranded DNA this reagent fluoresces on excitation at 480 nm. (Fluorescence is reported to be linear over at least four orders of magnitude.)

The assay is reported to be >10000-fold more sensitive than quantitation with ULTRAVIOLET ABSORBANCE at 260 nm.

[Uses (e.g.): PLoS ONE (2009) 4(2):e4584, and PLoS ONE (2009) 4(4):e5304.]

A PicoGreen approach has been used for detecting dsRNA [BMC Biotechnol (2009) 9:38].

(See also OLIGREEN and RIBOGREEN.)

picornaviruses Icosahedral, ether-resistant viruses (~22–30 nm in diam.) with a genome of linear, positive-sense ssRNA that is polyadenylated at the 3′ end and linked (covalently) at the 5′ end to a virus-encoded protein.

Most picornaviruses can be grown in cell culture.

Some picornaviruses cause disease in humans and animals – see e.g. APHTHOVIRUS and ENTEROVIRUSES.

(See also POLYPROTEIN.)

PicoTiterPlate™ See GS FLX.

pif In the F PLASMID: a locus (*pif* = phage-inhibiting function) associated with the inability of particular phages (e.g. T7) to replicate in F-plasmid-containing cells.

(See also FEMALE-SPECIFIC PHAGE.)

piggyBac A TRANSPOSABLE ELEMENT isolated initially from a baculovirus-infected species of lepidopteran: *Trichoplusia ni* (the cabbage looper). This 2.4 kb TE includes 13-bp terminal INVERTED REPEATS. The *piggyBac* transposase (C-terminal) includes a nuclear targeting signal [BMC Mol Biol (2008) 9:72]. The *piggyBac* target site is reported to be TTAA.

piggyBac-based vectors are used for manipulating genomes in invertebrates (including *Drosophila* and other insects) and vertebrates (including mammalian cells).

piggyBac-based transposition was used e.g. for mutagenesis in *Drosophila* [BMC Dev Biol (2009) 9:14], and conversion of mammalian fibroblasts to 'induced pluripotent stem cells' [Nature (2009) doi: 10.1038/nature07863].

piggyBac-based insertional mutagenesis has been used for preparing a genome-wide library of mutated Blm^{-} murine ES (embryonic stem) cells [Genome Res (2009) 19:667–673].

pilus (*bacteriol.*) (plural: pili) An appendage that projects from the surface of those Gram-negative bacteria which contain a conjugative PLASMID and in which the transfer operon of the plasmid is being expressed; the genes which encode the pilus are found in the plasmid.

These appendages have also been called *conjugative pili* or *sex pili*.

Some pili are known to be essential for plasmid-mediated CONJUGATION, and it is assumed that all pili, *as defined*, play an essential role in this process.

There are morphologically (and serologically) distinct types of pilus; different types of pilus are encoded by genetically distinct plasmids. The *type* of pilus on a donor cell influences the physicochemical conditions under which the conjugal process occurs – see the entry CONJUGATION for details.

Pili encoded by the F PLASMID are thin, flexible, protein filaments of up to several micrometers in length and ~8 nm in diameter; this type of pilus has an axial channel of ~2 nm in diameter. Other types of pilus include e.g. the short, rigid, thorn-like or nail-like appendages encoded by plasmids of the IncN incompatibility group.

Some types of BACTERIOPHAGE infect bacteria by initially binding to specific sites on pili; infection does not occur in the absence of pili. Examples of phages that infect via F pili (encoded by the F plasmid) include e.g. phages f1, fd, M13 (and derivatives), MS2 and Qβ. (See also HELPER PHAGE.)

F-like pili (encoded e.g. by plasmids ColV and ColI-K94) resemble F pili and they bind similar types of phage.

I-like pili (encoded e.g. by plasmid ColIb-P9) are similar, morphologically, to F pili but they differ serologically; they bind a distinct range of phages – including e.g. If1 and If2.

Note on the use of the terms pilus and fimbria

The two terms *fimbria* (plural: fimbriae) and pilus have been commonly used synonymously. This lack of discrimination is unfortunate because it is logical to distinguish between those appendages that have a specific role in conjugation (pili), and those appendages (fimbriae) which have other roles – e.g. in adhesion – and which are generally distinct morphologically. In this dictionary, and in the *Dictionary of Microbiology and Molecular Biology* (revised 3rd edition, 2006), 'fimbria' and 'pilus' are not used synonymously.

pipemidic acid See QUINOLONE ANTIBIOTICS.

P_iPer^{TM} pyrophosphate assay kit A product from Molecular Probes Inc. (Eugene OR, USA) used for the detection of free PYROPHOSPHATE in solution.

In the assay, inorganic pyrophosphatase first cleaves pyrophosphate to inorganic phosphate. Maltose (in the reaction mixture) is converted to a mixture of glucose and glucose 1-phosphate (catalyzed by maltose phosphorylase). Glucose is converted, by glucose oxidase, to gluconolactone and hydrogen peroxide. Hydrogen peroxide – in the presence of horseradish peroxidase – reacts with an added reagent to generate *resorufin*, whose fluorescence (maxima at approximately 563 nm and 587 nm) is monitored.

(See also ENZCHEK PYROPHOSPHATE ASSAY.)

piRNA See ARGONAUTE and PIWI PROTEIN.

***PITX2* gene** See AXENFELD–RIEGER SYNDROME.

Piwi-interacting RNAs (piRNAs) See PIWI PROTEIN.

Piwi protein Any member of the ARGONAUTE subfamily of proteins which appear to occur primarily in germ line cells. The Piwi proteins are reported to be involved (jointly, with Piwi-interacting RNAs – the so-called piRNAs) e.g. in silencing transposable elements. (See also NUAGE.)

Two categories of piRNAs have been described: (i) repeat-associated piRNAs (also called rasiRNAs), which appear to target transposable elements, and (ii) a second group whose target(s) are unknown.

(See also AUBERGINE.)

pJAZZ-KA See BIGEASY LINEAR CLONING SYSTEM.

pJET1/blunt See CLONING VECTOR.

pKM101 See AMES TEST.

PKR kinase See e.g. the entry RNA INTERFERENCE.

PLA PROXIMITY LIGATION ASSAY (q.v.).

plant pathogen forensics See FORENSIC APPLICATIONS.

plaque (phage) A usually 'circular' region in a confluent layer of bacterial growth (or in the upper layer of a PLAQUE ASSAY plate) in which some or all of the cells are lysed, or growth-impeded, owing to phage activity.

In the PLAQUE ASSAY, on a plate which has been prepared from a high dilution of the sample, any given plaque is likely to have developed as a result of the activity of a single lytic phage; progeny phages derived from lysis of the first-infected cell diffuse outwards and progressively infect, and lyse, the neighboring cells until a visible and often 'circular' area (the plaque) is formed.

The general appearance of a plaque may indicate the nature of the infecting phage. For example, a clear plaque, in which all the cells have been lysed, is characteristic of a virulent phage, but a turbid plaque, in which only a proportion of the cells are lysed, is characteristic of a temperate phage. Some non-lytic phages – e.g. FILAMENTOUS PHAGES such as f1 and M13 – may also form turbid plaques by reducing the growth rate of infected cells within the plaque.

plaque assay (of phage) A procedure used for determining the PLAQUE TITER of a sample containing a given type of phage.

Initially, log dilutions of the sample are prepared. From a given dilution, a measured volume is added to molten, semi-solid agar (~45°C) containing an excess of phage-susceptible bacteria; the whole is well mixed and is then poured onto a plate of solid nutrient agar – forming a layer of ~1–2 mm in

thickness. A similar procedure is used for each of the log dilutions. The plates are incubated at the growth temperature of the bacteria. After incubation, the plates are examined for plaques (see PLAQUE). Assuming that each observed plaque on a given plate was caused by a single phage, the plaque titer can be calculated from the volumes and dilutions used.

An analogous procedure may be used for viruses other than phages provided that a suitable monolayer culture of susceptible cells is available. In one method, each of a set of monolayer cultures is drained of growth medium and inoculated with one of the log dilutions of the sample; after gentle rocking (to enable viral adsorption to the cells), each monolayer culture is overlaid with a buffered growth medium in semi-solid agar and is then incubated and examined after a specific time interval.

plaque-forming unit (pfu) (of phage) Generally, a single phage particle which, under appropriate conditions, can give rise to a single PLAQUE in a confluent layer of susceptible bacteria or in a PLAQUE ASSAY plate.

plaque titer (of phage) A measure of the number of PLAQUE-FORMING UNITS in a given volume of sample.

plasmid A covalently closed circular, or linear, DNA molecule – distinct from chromosomal, mitochondrial, chloroplast or kinetoplast DNA – which can replicate autonomously (i.e. independently of other replicons) within a prokaryotic and/or eukaryotic cell. Plasmids are commonly dispensable, i.e. not essential to their host cells, and not all cells contain plasmids; for example, plasmids are commonly absent in bacteria of the genera *Anaplasma*, *Bartonella*, *Brucella* and *Rickettsia*.

The following account refers mainly to *bacterial* plasmids.

Features encoded by plasmids

Collectively, the plasmids encode a vast range of functions. Some encode resistance to antibiotic(s) and/or to other agents such as mercury (or other heavy metals). In some pathogenic

PLASMID: a diversity of molecules (some examples)

Plasmid	Notes, context
CAM	A natural plasmid (~500 kb) present in certain strains of the Gram-negative bacterium *Pseudomonas*. It encodes the enzymic capacity that enables these organisms to metabolize the terpenoid camphor
Cit	A natural plasmid found e.g. in strains of *Escherichia coli*. It encodes a citrate transport system which confers on the host cell the ability to use citrate as the sole source of carbon and energy. Cells of *E. coli* which lack the plasmid are citrate-negative in biochemical tests
DFRS	Dual-fluorscence reporter/sensor plasmid: a type of construct which has been used e.g. to study the dynamics of microRNAs (miRNAs) in living cells. Each plasmid contains sequences that encode two different fluorophores (which have distinct emission spectra), one of the fluorophores acting as the reporter (i.e. signaling intracellular expression of the plasmid). The mRNA of the other fluorophore includes a sequence complementary to the intracellular target; if present, the intracellular target binds to this mRNA and inhibits translation – indicated by a loss of fluorescence from this (sensor) fluorophore
F	A natural, low-copy-number plasmid (~95 kb) found e.g. in strains of the Gram-negative bacterium *Escherichia coli*. The plasmid may exist as an integral part of the bacterial chromosome or as an extrachromosomal element. The presence of the plasmid confers on the host cell the ability to act as a donor of DNA (either plasmid DNA or plasmid *and* host DNA) during conjugation; for this purpose the plasmid encodes various products required for the mobilization and transfer of DNA, including a bacterial cell-surface appendage (see PILUS) required for establishment of contact between the donor and recipient cells. In DNA technology, cells containing the F plasmid are used e.g. to permit infection by certain types of bacteriophage (e.g. phage M13) which are used e.g. for the preparation of single-stranded DNA; these phages infect cells by initially binding to the pilus (cells lacking a pilus cannot be infected). The so-called *ccd* mechanism (see entry), encoded by the F plasmid, has been used to promote stable maintenance of the plasmid in populations of bacteria under experimental conditions
OCT	A natural plasmid (~500 kb), found in some strains of *Pseudomonas*, which enables these strains to metabolize octane and decane
pBR322	A small (~4.3 kb) engineered, multicopy plasmid encoding resistance to ampicillin and tetracycline. It has been widely used e.g. as a cloning vector. Many plasmids have been derived from pBR322 or from components of this plasmid
pCAL	Any of a range of engineered plasmid vectors used in the AFFINITY PROTEIN EXPRESSION AND PURIFICATION system (q.v.); each of the plasmids includes a calmodulin-binding peptide (CBP) tag to facilitate isolation/purification of the protein of interest

(*continued*)

PLASMID (*continued*)

Plasmid	Notes, context
pCT	A natural plasmid found in most strains of the Gram-negative pathogen *Chlamydia trachomatis*. A sequence of nucleotides in this plasmid has been used as a target for a diagnostic test involving the ligase chain reaction (LCR). Rare isolates of *C. trachomatis* have been reported to lack the plasmid
pESC	Any of a range of engineered commercial plasmid vectors used e.g. for gene expression in the yeast *Saccharomyces cerevisiae*. The plasmids include sequences (e.g. FLAG) for epitope tagging; this facilitates detection/isolation of the target proteins (see entry PESC VECTORS)
pET SUMO	A small (5.6 kb) engineered plasmid used as an expression vector for high-level expression of recombinant proteins in *Escherichia coli* hosts. The plasmid encodes an 11-kDa solubility-enhancing protein that forms a fusion product with the protein of interest. This plasmid is part of a commercial protein expression system (see entry CHAMPION PET SUMO VECTOR)
pJAZZ-KA	A linear, engineered plasmid derived from the genome of a bacteriophage (coliphage N15). It replicates intracellularly as an extrachromosomal element. This plasmid is a commercial cloning vector used in the BigEasy^{TM} linear cloning system (see entry)
pJET1/blunt	An engineered plasmid designed for cloning blunt-ended PCR products (see entry CLONING VECTOR)
pKM101	A plasmid encoding e.g. certain error-prone repair enzymes; it is used in the AMES TEST
pMG7	A plasmid (found in some strains of the Gram-negative bacterium *Pseudomonas aeruginosa*) which inhibits replication of certain phages (e.g. F116, G101)
pOG44	An engineered plasmid encoding the Flp recombinase (see FLP). It is used in conjunction with FLP-IN CELL LINES
pSM19035	A natural plasmid found e.g. in strains of the Gram-positive bacterium *Streptococcus pyogenes*. It exhibits a high level of segregational stability in host cells – a feature due to the combined activity of a number of distinct mechanisms for controlling copy number
pXO1	A natural plasmid found in virulent (anthrax-causing) strains of the Gram-positive bacterium *Bacillus anthracis*. It encodes the anthrax toxin
pXO2	A natural plasmid found in virulent (anthrax-causing) strains of the Gram-positive bacterium *Bacillus anthracis*. It encodes a poly-D-glutamic acid capsule which enables the organism to evade phagocytosis in the host
R1	A natural, low-copy-number plasmid found in strains of *Escherichia coli* and related bacteria. It encodes resistance to certain antibiotics (including ampicillin and streptomycin), and also confers on the host cell the ability to act as a donor of DNA in conjugation. During cell division, segregation of the plasmids to daughter cells appears to involve a mitosis-like process in which a plasmid-encoded protein polymerizes to form force-generating filaments that are involved in mechanical separation of the daughter plasmids
R6K	A natural multicopy plasmid (~38 kb) found in *Escherichia coli* and related bacteria. It encodes e.g. a β-lactamase (and, hence, confers resistance to certain β-lactam antibiotics) and confers on the host cell the ability to act as a DNA donor in conjugation
R100	A natural, low-copy-number plasmid found in strains of *Escherichia coli* and related bacteria. It encodes resistance to e.g. chloramphenicol, sulfonamides and mercury; a gene for tetracycline resistance is carried on transposon Tn*10*. R100 confers on the host cell the ability to act as a DNA donor in conjugation
Ti	A large plasmid found in (virulent) strains of the bacterium *Agrobacterium tumefaciens*. It encodes the means of transferring DNA into plant cells, and this is exploited in a procedure called AGROINFECTION which is used for the experimental infection of plants by viruses. In nature, DNA injected into plants by *A. tumefaciens* causes tumors (galls – see entry CROWN GALL)
TOL	A natural plasmid (~117 kb) found in strains of the Gram-negative bacterium *Pseudomonas*. The plasmid genes are arranged in two operons which jointly confer on the bacterial host the ability to metabolize toluene and xylene; these substrates are first oxidized to carboxylic acids which are then converted to pyruvate and acetaldehyde (which enter the tricarboxylic acid cycle). The availability of iron may be a factor in this metabolic activity. One isolate of a TOL plasmid was found to encode a homolog of DNA polymerase V which was reported to promote survival in DNA-damaged cells

bacteria (e.g. *Bacillus anthracis*, causal agent of anthrax) the toxin(s) and/or other virulence factors are plasmid-encoded. (In some pathogens the virulence factors are *phage*-encoded.) Some plasmids encode products (such as particular enzymes) which enhance the metabolic potential of a cell; for example, the Cit plasmid in certain strains of *Escherichia coli* enables those strains to use citrate as the sole source of carbon and energy (an ability which is lacking in wild-type strains of *E. coli*). Again, the TOL plasmid confers on certain strains of *Pseudomonas* the capacity to metabolize toluene and xylene. Certain plasmids (e.g. ColE1) encode a COLICIN or other type of BACTERIOCIN. In the archaean *Halobacterium*, structural components of gas vacuoles are plasmid-encoded.

Engineered plasmids

Plasmids have been engineered in order to carry out a great diversity of functions. For example, plasmids are widely used as cloning vectors and as expression vectors. They are also used e.g. in studies on the mechanisms of intracellular signaling. [Plasmid vector design and production (a review): Gene Therapy (2009) 16:165–171.]

A plasmid encoding a transposase may be used as a 'helper plasmid' in cells within which a specific type of transposition is required.

A plasmid may contain e.g. genes of the LAMBDA (λ) RED RECOMBINATION system in order to carry out an intracellular replacement of DNA (such as a gene knock-out).

A plasmid may act as a PACKAGING PLASMID.

A DFRS PLASMID can be used for detecting an intracellular target sequence.

Of the enormous range of natural and engineered plasmids, only a very limited selection is shown in the table.

Size, transformability

In size, plasmids typically range from about several kilobases to several hundred kilobases. In DNA technology, the small circular plasmids can be inserted into cells by transformation more readily than large plasmids; moreover, small plasmids are less susceptible to damage by shearing forces. In general, linear plasmids are poorly transformable, e.g. in normal cells of *E. coli* (although linear DNA is used with methods such as the lambda (λ) red recombination system).

Circular/linear plasmids

In the literature, plasmids have been described *ad nauseam* as 'small circles of DNA'; as *linear* plasmids have been known for many years, this is clearly misleading. While many, even most, plasmids *are* ccc dsDNA molecules, linear plasmids (and chromosomes) occur e.g. in a number of bacteria; thus, e.g. a bacterium with one of the largest genomes (~9.7 Mbp), *Rhodococcus* sp RHA1, has three linear plasmids as well as a linear chromosome [see Proc Natl Acad Sci USA (2006) 103: 15582–15587], and linear plasmids also occur e.g. in species of *Borrelia* and *Streptomyces*.

Linear plasmids are exploited in DNA technology: see e.g. BIGEASY LINEAR CLONING SYSTEM and USER CLONING.

The presence of transposable elements in plasmids

Some plasmids contain one or more transposons or insertion sequences (see TRANSPOSABLE ELEMENT). For example, the transposon Tn*3* occurs in R1*drd*-19 while Tn*10* and Tn*21* are found e.g. in R100. The insertion sequence IS*3* is found e.g. in the F plasmid (see FINOP SYSTEM).

Copy number

For a given plasmid, the copy number is characteristic, under given conditions, in a particular type of host cell.

Plasmids are referred to as 'low-copy-number' if they occur singly, or in a few copies, in each cell; the F PLASMID is one example of a low-copy-number plasmid.

The multicopy plasmids are usually present in appreciably higher numbers – e.g. >10 copies per cell; ColE1 plasmid (often 10–30 copies) is one example.

Factors which influence copy number include the plasmid's specific type of replication control system. In one mechanism for maintaining copy number, encoded by plasmid R1, the Kid protein acts as a nuclease: if copy number falls it cleaves host-cell mRNAs and also cleaves the mRNA of a (plasmid-encoded) replication repressor; hence, under these conditions, bacterial growth rate decreases and the plasmid's replication rate rises, tending to restore normal copy number and avoiding loss of plasmids from the population [EMBO J (2005) 24(19):3459–3469].

Replication of bacterial plasmids

To maintain their presence within a (growing) population of bacteria, plasmids must replicate; moreover, plasmid progeny must be distributed efficiently to daughter cells (see Partition, below).

Control of plasmid replication is encoded in the plasmid itself, although the cell's biosynthetic potential is an essential requirement.

The frequency with which a plasmid replicates in the host cell varies with plasmid. It is determined by the frequency at which replication is *initiated*, and this feature is regulated in a plasmid-dependent way. (Early reports indicated that, in at least some plasmids (e.g. pSC101), initiation of replication involves the cell's DnaA protein (needed for *chromosomal* replication); however, it now appears that DnaA is specific for the initiation of replication of the host cell's DNA and that plasmid-encoded factors regulate the replication of plasmid DNA.)

Control of plasmid replication involves different plasmid-dependent mechanisms. For example, in the ColE1 plasmid, control of replication occurs at the primer level and involves competitive interaction between two (plasmid-encoded) RNA molecules, one of which can prime replication (for details see COLE1 PLASMID).

A different mechanism operates in various small, circular plasmids found in Gram-positive bacteria. In these plasmids, replication – involving a ROLLING CIRCLE mechanism – is initiated by a plasmid-encoded Rep protein. The Rep protein nicks a specific strand of the plasmid at the origin, thus forming a free 3′ end suitable for rolling circle synthesis. A single round of replication yields a complete (circular) ssDNA copy of the plasmid that is subsequently replicated to the double-

stranded state from an RNA primer. Initiation of replication may be controlled by regulating the synthesis or activity of the Rep protein. For example, in some cases, translation of the Rep-encoding gene may be regulated by a small plasmid-encoded antisense RNA molecule that binds to the mRNA. Alternatively, Rep may be inactivated after only a single use by the binding of a small plasmid-encoded oligonucleotide.

Some plasmids contain *iterons*: sequences of direct-repeats associated with their origin of replication; these regions are central to replication control. Thus, initiator proteins can bind to these regions and give rise to a nucleoprotein complex as a prerequisite to the initiation of replication. The mechanism of control may be complex, and in at least some cases initiator proteins are autoregulated at the level of transcription. Moreover, in some cases a mechanism known as HANDCUFFING appears to be involved in initiation control.

Stringent and relaxed control of replication

Plasmids whose replication is dependent on protein synthesis (in the host cell) fail to replicate if the host cell is treated with certain agents (such as the antibiotic chloramphenicol) which inhibit protein synthesis; in these cases, replication is said to be under *stringent control*.

Other plasmids, such as ColE1, can continue to replicate in the presence of agents that inhibit protein synthesis; for these plasmids, replication is sometimes said to be under *relaxed control*. Plasmids in this category can exploit pre-formed host proteins (such as polymerases). (Note that, in plasmid ColE1, the regulation of initiation of replication involves synthesis of oligonucleotides rather than the synthesis of proteins.) Given continued replication in the presence of bacteriostatic agents, some plasmids – such as ColE1 – achieve higher-than-normal copy numbers within their host cells.

Partition (segregation) of bacterial plasmids

In certain – high-copy-number – plasmids there seems to be no special mechanism for active segregation of the plasmid progeny to daughter cells during cell division. It is generally believed that, in such cases, partition may involve a random distribution of plasmid progeny into daughter cells – the high copy number of the plasmid in the parent cell ensuring that both daughter cells will receive an appropriate complement of plasmids.

In low-copy-number plasmids there is a need for an active mechanism for partition in order to ensure that each daughter cell receives at least one copy of the plasmid – in this way achieving a stable inheritance of the plasmid in the bacterial population. Hence, these plasmids contain specific loci designated e.g. *par*, *sop*, *sta* (for example, *parA* and *parB* in the F PLASMID – see also PAR LOCI) which are associated with the partition function.

In some plasmids the mechanism of partition appears to be more complex than previously understood. For example, in the R1 PLASMID (q.v.), segregation of plasmids to daughter cells apparently involves a mitosis-like process that includes a form of mechanical propulsion. Proteins encoded by other plasmids are reported to polymerize, *in vitro*, into filaments that may be involved in partition – e.g. the SopA protein of the F plasmid and the ParF protein of plasmid TP228.

Compatibility*: *Inc groups

Different types of plasmid can occur in the same cell: those plasmids which have different modes of replication and partition are said to exhibit *compatibility*; such plasmids are able to co-exist – stably – within the same cell. On the other hand, plasmids with similar or identical systems of replication/partition are incompatible: they are not able to co-exist, stably, in the same cell. This is the basis of the *incompatibility groups* (the so-called Inc groups): a given Inc group consists of those plasmids which have similar or identical replication/partition systems and which cannot co-exist, stably, in the same cell.

Among enterobacteria the Inc groups include:

IncFI (e.g. the F plasmid, ColV-K94, R386)
IncFII (e.g. R1, R6)
IncIα (e.g. ColIb-P9, R64)
IncN (e.g. R46, R269N-1)
IncX (e.g. R6K)

Among *Pseudomonas* species, the Inc groups have been designated IncP-1, IncP-2 etc.; for example:

IncP-2 (e.g. CAM plasmid, OCT plasmid)
IncP-9 (e.g. TOL plasmid)

Some of the enterobacterial and pseudomonad Inc groups are shared (equivalent).

Conjugative plasmids

Some bacterial plasmids encode the means for transferring a copy of the plasmid DNA to another host cell by the process of CONJUGATION – involving cell–cell contact during which donor DNA is transferred to the recipient cell; one example of such plasmids is the F PLASMID. The conjugative plasmids contain a so-called transfer operon that includes those genes required for the synthesis of products (such as the PILUS and certain enzymes) necessary for conjugative transfer of DNA.

When integrated into the chromosome, these plasmids may be able to mobilize transfer of *chromosomal* DNA (as well as the plasmid DNA) during conjugation.

Plasmid persistence in a bacterial population

In some cases, one or more mechanisms militate against the loss of a plasmid from a bacterial population – see e.g. CCD MECHANISM. (The R1 mechanism, involving the Kid protein (see above: *Copy number*), also helps to prevent plasmid loss from a bacterial population.)

Experimental elimination of plasmids

In some cases there is a need for plasmid-free cells. Various approaches have been used to eliminate the plasmids from a bacterial population: see e.g. CURING and SUICIDE VECTOR.

Isolation of plasmids from bacteria in the laboratory

Various commercial methods can be used for the isolation of plasmids from bacteria. Using a QIAGEN® plasmid mini kit,

the bacterial sample is subjected to controlled alkaline lysis with NaOH (sodium hydroxide) and SDS (sodium dodecyl sulfate); this kind of treatment is designed to disrupt the cells' cytoplasmic membrane and to release e.g. RNA, proteins and plasmids. (If the procedure is carried out correctly, chromosomal DNA is trapped within the leaky cells.) Contaminating RNA is digested with RNase. The lysis period (i.e. the period of exposure to the alkaline lysis regime) is optimized in order to promote maximum release of plasmids and to avoid their irreversible denaturation. After the lysate is neutralized, the whole is centrifuged to obtain a cleared lysate (containing the plasmids) – which is exposed to a special QIAGEN® anion-exchange resin in a controlled electrolyte environment; after adsorption to the resin, plasmids can be eluted by adjustment of the electrolyte/pH.

A different procedure is used in the PURELINK PLASMID PURIFICATION SYSTEMS.

plasmid artificial modification See TRANSFORMATION (under subheading *Enhanced efficiency of bacterial transformation*).

plasmid profiling See REAP.

Plasmodium A genus of protozoa (phylum: Apicomplexa) that includes causal agents of human malaria – e.g. *P. falciparum* [genome: Nature (2002) 419:498–511], *P. malariae* (earlier name: *P. rodhaini*), *P. ovale* and *P. vivax*. Some other species (such as *P. cynomolgi*) are parasites in non-human primates, but they occasionally cause malaria in humans and have been used as models in research on human malaria.

[Malaria parasites (a genomic database): Nucleic Acids Res (2009) 37(Database issue):D539–D543.]

Infection of erythrocytes (red blood cells) by *Plasmodium* has been detected with the DNA-staining dye YOYO-1 [Proc Natl Acad Sci USA (2009) 106(2):361–368].

(See also REFRACTORY GENE.)

plate (1) A PETRI DISH containing a (sterile) AGAR- or gelatin-based medium (in either case, a *gel*). A plate is prepared by pouring the (molten) medium into a Petri dish and allowing it to set. A sterile plate may be *inoculated* with a given microorganism, or material containing the microorganism, and then *incubated* at an appropriate temperature for a suitable period of time; if the organism grows, the result is a *plate culture*. (Note that a plate culture may also be referred to as a 'plate'.) (2) (*verb*) To inoculate the surface of a plate (sense 1) with an *inoculum* containing (or expected to contain) a particular type (or particular types) of microorganism – with the object of initiating a plate culture.

plate culture See PLATE (sense 1).

Platinum® *Taq* DNA polymerase A recombinant form of TAQ DNA POLYMERASE (Invitrogen, Carlsbad CA, USA) which is designed for improved specificity in hot-start PCR; the polymerase incorporates a thermolabile inhibitor (containing an anti-*Taq* monoclonal antibody) which is degraded during the first (high-temperature) denaturation step, releasing the active *Taq* DNA polymerase.

plectonemic Refers to a (stable) arrangement of two strands of DNA in which the strands are wound around each other, each strand being base-paired to the other.

(cf. PARANEMIC.)

pleiotropic Refers (e.g.) to a gene whose expression can affect multiple – apparently unrelated – phenotypic characteristics.

pleiotropic mutation A mutation affecting multiple, apparently unrelated, targets/systems.

pLenti6/V5-DEST™ vector A destination vector in the GATEWAY SITE-SPECIFIC RECOMBINATION SYSTEM (see the table in that entry).

plenum ventilation See SAFETY CABINET (class I).

PLG Poly(lactide-co-glycolide) – a co-polymer of lactide and glycolide used as a matrix for the encapsulation of complexes composed of cationic lipid and nucleic acid; this system was used for transfecting cells with DNA encoding neurotrophic factors [Biomaterials (2006) 27(11):2477–2486].

ploidy The number of sets of chromosomes per cell – e.g. one set (haploid), two sets (diploid) etc. (cf. ANEUPLOID).

In some organisms (e.g. certain algae, fungi and protozoa) the life cycle exhibits a so-called *alternation of generations* in which generation(s) of mature haploid individuals alternate with generation(s) of mature diploid individuals.

pluripotent stem cell See STEM CELL.

plus strand (of a gene) See CODING STRAND.

PML Progressive multifocal leukoencephalopathy: a fatal demyelinating condition caused by the JC VIRUS.

pML31 A 9-kb plasmid which was derived from the F PLASMID by digestion with restriction enzyme EcoRI and the insertion of an antibiotic-resistance gene in a recircularized segment.

PMN (granulocyte) Polymorphonuclear leukocyte: a type of white blood cell with lobed nucleus and granular cytoplasm; PMNs include neutrophils, basophils and eosinophils.

The term *polymorph* is sometimes used to refer specifically to neutrophils.

PMO (phosphorodiamidate morpholino oligomer) See MORPHOLINO ANTISENSE OLIGOMER.

pMT-DEST48 vector A destination vector in the GATEWAY SITE-SPECIFIC RECOMBINATION SYSTEM (see the table in that entry).

pMT/V5-His vector See the entry DES.

pMT/V5-His-TOPO® vector See the entry DES.

pMyr vector See CYTOTRAP TWO-HYBRID SYSTEM.

PNA Peptide nucleic acid: a DNA analog in which the backbone chain is a modified polypeptide – rather than a chain of sugar–phosphate units.

Two strands of PNA can form a duplex that resembles the helical duplex of DNA. Moreover, a strand of PNA can bind to a strand of DNA more stably than can the complementary strand of DNA (cf. ZIP NUCLEIC ACID).

Unlike DNA, PNA is not sensitive to the concentration of electrolyte.

Chemically modified oligomers of PNA have been used to form high-affinity TRIPLEX DNA (q.v.).

PNA has been used e.g. in LIGHTUP PROBES, and an *antisense* PNA, specific for an essential gene in *Escherichia coli*, was reported to exert potent antibacterial activity.

Owing to inherent properties, PNA oligomers are especially useful as clamps in PCR CLAMPING. [Use of PNA as both a PCR clamp and sensor probe for melting curve analyses for detecting rare mutant K-*ras* DNA: Nucleic Acids Res (2006) 34(2):e12.]

PNA–peptide conjugates were reported to have been used for intracellular gene control – e.g. correcting the splicing in HeLa cells – without the use of a transfection agent [Nucleic Acids Symp Series (2008) 52(1):31–32].

PNA-mediated PCR clamping See PCR CLAMPING.

***pncA* gene** In *Mycobacterium tuberculosis*: a gene encoding the enzyme pyrazinamidase; *in vivo*, pyrazinamidase converts the anti-tuberculosis prodrug pyrazinamide to the active form of the drug: pyrazinoic acid.

Clinical isolates of *M. tuberculosis* which are resistant to pyrazinamide are commonly found to have a mutant form of the *pncA* gene.

Pneumovirus A genus of viruses within the family PARAMYXOVIRIDAE (q.v.).

PNPase POLYNUCLEOTIDE PHOSPHORYLASE (q.v.).

pOG44 A plasmid encoding the Flp recombinase (see FLP). It is used in conjunction with FLP-IN CELL LINES for inserting the gene of interest into the host cell's chromosome: pOG44 is co-transfected into (engineered) host cells with the Flp-In™ vector (carrying the gene of interest) – the gene of interest then integrating into the chromosome by (Flp-mediated) site-specific recombination.

point mutation (1) Any mutation in which a single nucleotide is replaced by another type of nucleotide (see TRANSITION MUTATION and TRANSVERSION MUTATION).

In certain cases a point mutation can make the difference between a normal gene and a disease-causing gene (see e.g. RAS).

In pathogenic microorganisms, some point mutations can confer resistance to antibiotics. For example, mutations in the *rpoB* gene in *Mycobacterium tuberculosis* are responsible for resistance to RIFAMYCINS; these mutations include serine→leucine in codon 456, and aspartic acid→valine in codon 441.

Point mutations can be detected by various procedures such as e.g. ARMS, CLEAVASE FRAGMENT LENGTH POLYMORPHISM ANALYSIS, SINGLE-BASE EXTENSION and SSCP ANALYSIS.

[A highly specific microarray method for the detection of point mutations: BioTechniques (2008) 44(1):119–126.]

Any kind of point mutation in a sample of nucleic acid can potentially interfere with sequence-based diagnostic tests or with other procedures used for examining/amplifying specific sequences of nucleotides.

A point mutation can be inserted into a target sequence by various techniques: see, for example, RECOMBINEERING and SITE-DIRECTED MUTAGENESIS. (See also CHIMERAPLAST.)

(2) The phrase is sometimes used to include – in addition – those cases in which a single nucleotide is inserted or deleted, leading to a FRAME-SHIFT MUTATION.

poky mutant (of *Neurospora*) See MATERNAL INHERITANCE.

pol See RETROVIRUSES.

pol α (pol alpha) See B FAMILY.

pol β (pol beta) See X FAMILY.

pol δ (pol delta) See B FAMILY.

pol ϵ (pol epsilon) See B FAMILY.

pol ζ (pol zeta) See DNA POLYMERASE ζ.

pol β (pol eta) See Y FAMILY.

pol θ (pol theta) See A FAMILY.

pol ι (pol iota) See Y FAMILY.

pol κ (pol kappa) See Y FAMILY and 8-OXOG.

pol λ (pol lambda) See X FAMILY.

pol μ (pol mu) See X FAMILY.

pol Q See A FAMILY.

polar body (*embryol.*) A small haploid cell which is produced during meiosis in an oocyte. A polar body disintegrates when the oocyte is fertilized.

(See also SCNT.)

polar mutation Any mutation which affects not only the gene in which it occurs but also other gene(s) located downstream on the same operon. For example, in the LAC OPERON, a mutation in *lacY* may affect the synthesis of permease and transacetylase. If promoter-distal genes are strongly inhibited the mutation is said to be strongly polar.

polar trophectoderm See BLASTOCYST.

polarity (of strands in DNA) See the entry DNA.

polioviruses See ENTEROVIRUSES.

polishing A non-specific term that has been used e.g. to refer to the removal of 3′ overhangs, using an exonuclease, to prepare fragments for blunt-ended ligation. [Example of use of term (polishing genomic DNA prior to cloning): Genetics (2009) 181(3):1129–1145.]

A 'polishing kit' (marketed by Stratagene, La Jolla CA) employs the 3′ exonuclease function of *Pfu* DNA polymerase to create blunt-ended products from PCR-generated amplicons formed by *Taq* DNA polymerase; the *Taq* products have 3′ overhangs owing to EXTENDASE ACTIVITY.

(See also END-IT DNA END-REPAIR KIT.)

polony A 'PCR colony' formed when a target sequence of DNA is amplified by PCR within a thin film of polyacrylamide gel on a glass microscope slide [early description: Nucleic Acids Res (1999) 27(24):e34]. The purpose of the polyacrylamide is to retard diffusion of the PCR products – which therefore remain localized and accumulate to form a 'colony'. A large number of such colonies may be formed on a single slide.

DNA in polonies was used for *in situ* transcription, translation and protein folding [Nucleic Acids Res (2005) 33(17): e145].

POLQ See A FAMILY.

poly-MBD proteins See CPG ISLAND.

poly(A) polymerase An enzyme which adds adenosine residues to the 3′ end of an mRNA (forming the 'poly(A) tail'). See the entry MRNA.

(See also COLE1 PLASMID, MULTICOPY PLASMID, POLY(A) TAILING.)

poly(A)-specific ribonuclease See PARN.

poly(A) tail See MRNA.

poly(A) tailing (of RNA) (*DNA technol.*) The process in which a poly(A) tail is added to the 3′ end of an RNA molecule, e.g. by using a poly(A) polymerase from *Escherichia coli*.

polyacrylamide A polymer of ACRYLAMIDE prepared e.g. with a chemical catalyst such as METHYLENE-BIS-ACRYLAMIDE – which promotes cross-linking (polymerization) of the acrylamide.

Polyacrylamide gels are used for separating small/medium-sized fragments of DNA or RNA by GEL ELECTROPHORESIS (cf. AGAROSE); they are also used e.g. for separating proteins.

In polyacrylamide gels, *pore size* depends on the (i) concentration of acrylamide, and (ii) proportion of cross-linking agent in the mixture. In the electrophoresis of nucleic acids, the ratio acrylamide:cross-linker (Bis) is often ∼19:1. For the electrophoresis of proteins, higher ratios (up to e.g. 40:1) may be used.

(See also SDS–PAGE.)

polyadenylation The sequential addition (*in vivo*) of adenosine residues to the 3′ end of a pre-mRNA molecule, forming a chain of adenosines (a 'poly(A) tail'). The typical (*eukaryotic*) pre-mRNA is polyadenylated after cleavage at a site downstream of a polyadenylation signal (see AAUAAA). Some prokaryotic transcripts also are polyadenylated.

(See also CYTOPLASMIC POLYADENYLATION.)

In some mammalian genes, the site at which the pre-mRNA is cleaved varies according to conditions; hence, under differing conditions, a given gene can give rise to mRNAs which differ in nucleotide sequence at their 3′ end. Such *alternative polyadenylation* can affect the results of gene expression profiling, particularly when mRNAs are examined by a MICROARRAY in which the probes are biased towards a target region at the 3′ end of the mRNAs [Appl Bioinformatics (2006) 5 (4):249–253].

(See also THREE-PRIME ARRAY.)

polybrene (hexadimethrine bromide) A polycationic compound used e.g. for enhancing viral infection of cells *in vitro*; polybrene is reported to enhance receptor-independent adsorption of at least some virions to the membranes of target cells.

Polybrene® is a registered trade mark of Sigma-Aldrich, St Louis MO, USA.

[Uses of polybrene: Hum Mol Genet (2008) 17:3487–3501; J Biomed Biotechnol (2009) doi: 10.1155/2009/901079; and J Clin Invest (2009) 119:1727–1740.]

polycistronic mRNA A molecule of mRNA that encodes two or more gene products. Polycistronic mRNAs are produced e.g. during transcription of a bacterial OPERON. For example, in the LAC OPERON, the polycistronic mRNA encodes LacZ, LacY and LacA – and each gene (*lacZ*, *lacY* and *lacA*) has its own initiator codon and stop codon; these genes are therefore translated as separate proteins.

The *genome* in positive-sense ssRNA viruses such as PICORNAVIRUSES acts as a polycistronic mRNA. However, in this case, translation gives rise to a POLYPROTEIN (q.v.).

polycomb-group genes Genes, found in diverse species, whose products have an epigenetic function; they are involved e.g. in preventing inappropriate expression of HOMEOTIC GENES.

In the fruit-fly, *Drosophila*, polycomb-group genes encode transcriptional repressors which function via DNA elements designated polycomb-group response elements. (Trithorax-group genes of *Drosophila* encode transcriptional activators which function via trithorax-group response elements.)

In *Drosophila*, polycomb-group products have established roles in regulating homeotic genes, but they are also reported to be involved in regulating the expression of cyclin A; this would give the polycomb-group products a role in regulation of the cell cycle [Genes Dev (2006) 20:501–513].

Deregulation of the human polycomb-group genes may be linked to various hematological and epithelial cancers [Hum Mol Genet (2005) 14(suppl 1):R93–R100].

The target genes of polycomb-group proteins were mapped in a genome-wide study of (human) embryonic fibroblasts [Genes Dev (2006) 20:1123–1136].

polyethylene glycol (PEG) Any of a range of polymers having the general formula: $H(OCH_2.CH_2)_nOH$; those polymers with a molecular weight above ∼1000 are white solids – the others are liquids.

PEG has a variety of uses. For example, it is used for CELL FUSION and as a precipitating agent in clinical chemistry. An aqueous solution of PEG is used for concentrating various solutions and suspensions (including serum); the solution or suspension is placed inside a tube of semipermeable material which is partly submerged in the PEG solution – water being withdrawn from the tube by osmosis.

A brief incubation in an alkaline (pH 13.3–13.5) solution of 60% PEG 200 has been reported to allow bacteria, eukaryotic tissue, whole blood etc. to be used directly for PCR without further preparation [BioTechniques (2006) 40(4):454–458].

polyethyleneimine See the entry PEI.

polygenic mRNA *Syn*. POLYCISTRONIC MRNA.

polyhedrin A protein encoded by the NUCLEAR POLYHEDROSIS VIRUSES that forms the matrix of the intracellular *occlusion bodies* of these viruses: see BACULOVIRIDAE for details.

(cf. GRANULIN.)

The promoter of the polyhedrin-encoding gene is used e.g. in BACULOVIRUS EXPRESSION SYSTEMS because it acts as a strong promoter within insect cells.

poly(lactide-co-glycolide) See the entry PLG.

polylinker (MCS, multiple cloning site) A DNA segment that contains a number of close or overlapping recognition sites for *different* types of RESTRICTION ENDONUCLEASE. A polylinker is included in a VECTOR for flexibility in manipulation (i.e. the choice of various restriction enzymes) when inserting target DNA into the vector.

When inserting a gene/fragment into a vector molecule, the *orientation* of the insert in the vector can be pre-determined e.g. by using a vector that includes a polylinker containing an asymmetric array of cloning sites, and using fragments which are flanked by dissimilar STICKY ENDS; this permits insertion in only one orientation.

poly-L-lysine A reagent used e.g. to bind DNA to glass slides

(non-covalently). (cf. ATMS.)

polymerase See the entries DNA POLYMERASE and RNA POLYMERASE.

polymerase chain reaction (PCR) See the entry PCR.

polymorph See the entry PMN.

polymorphic (*adj*.) Refers to a locus, in a given sequence of nucleic acid, at which the composition varies among different samples of that sequence.

polymorphonuclear leukocyte See the entry PMN.

polynucleotide kinase Any ATP-dependent enzyme of EC subclass 2.7 which catalyzes the addition of phosphate to the 5′-OH of a (terminal) nucleotide. The enzyme has a role e.g. in DNA repair [mechanism of DNA substrate recognition by the mammalian polynucleotide kinase: Nucleic Acids Res (2009) doi: 10.1093/nar/gkp597].

This enzyme is used e.g. for 5′ END LABELING with labeled phosphate (see e.g. PROBE LABELING).

(See also END-IT DNA END-REPAIR KIT.)

polynucleotide phosphorylase In (e.g.) *Escherichia* coli: an enzyme that degrades RNA exonucleolytically and which can also polymerize ribonucleoside diphosphates. Another role is the addition of nucleotides to poly(A) tails [Proc Natl Acad Sci USA (2000) 97(22):11966–11971] – a role for which this enzyme appears to be relatively unimportant in the bacterium *Bacillus subtilis* [J Bacteriol (2005) 187(14):4698–4706].

This enzyme is a component of the DEGRADOSOME.

polyomaviruses Icosahedral, non-enveloped viruses (∼45 nm; genome: ccc dsDNA) which include BK virus, JC VIRUS and SV40; they appear to be common in mammalian (e.g. human) hosts.

Polymaviruses can be propogated readily in cell cultures, and are resistant to lipid solvents and low pH.

[Primate polyomaviruses (gene expression): J Virol (2009) 83:10846–10856.]

polyprotein A protein produced by uninterrupted translation of two or more genes specified by a *polycistronic* mRNA; the polyprotein is cleaved during or after translation into separate proteins.

Polyproteins are encoded e.g. by some ssRNA viruses that have positive-sense genomes (e.g. PICORNAVIRUSES).

polyribosome *Syn*. POLYSOME.

polysome (polyribosome) A molecule of mRNA which carries a number of ribosomes distributed along its length, the stage of translation of the encoded polypeptide being different at each ribosome.

Polysome analysis in CELL-FREE PROTEIN SYNTHESIS systems has identified certain rate-limiting steps. For example, dilution of translation factors in the lysate is one limitation, and it was shown that addition of purified elongation factors can increase the rate of protein synthesis [Biotechnol Bioeng (2005) 91(4):425–435].

mRNA isolated from the polysomes of solid tissues was reported to be usable for reverse-transcriptase PCR (rt-PCR) and for profiling experiments [RNA (2007) 13(3):414–421].

polytene chromosome See POLYTENIZATION.

polytenization Repetitive replication of chromosomes in which daughter chromosomes do not separate – producing a ribbon-like *polytene* chromosome (a 'giant chromosome') consisting of many chromosomes arranged parallel to one another.

Polytenization occurs in certain ciliates – and also in some of the cells in plants and insects.

polyvinylidene fluoride See PVDF.

population replacement In genetic engineering: replacement of a given population with a genetically distinct population e.g. by promoting the spread of transgene(s) throughout an existing, wild-type population so that progeny which have the required genetic modification replace the original population – see e.g. the entry GMMS.

positive control (in operons) See OPERON.

positive selection (direct selection) Selection which is based on the presence of a given feature, or the occurrence of a given reaction.

(Compare NEGATIVE SELECTION.)

positive-sense virus *Syn*. POSITIVE-STRAND VIRUS.

positive-strand virus (positive-sense virus) Any virus whose genome (i.e. that present in the virion) is single-stranded and is homologous to – rather than complementary to – the viral mRNA. (cf. NEGATIVE-STRAND VIRUS.)

(See also SUBGENOMIC MRNA.)

post-genomics A term which has been used to refer to studies on the genome *as a whole* – including characteriztion of the TRANSCRIPTOME and analysis of the (complete) complement of expressed proteins in cells/organisms at a given time under given conditions.

post-segregational killing Any plasmid-encoded mechanism in which, following division of the host cell, any daughter cell which *lacks* a plasmid is killed by a plasmid-encoded toxin; such a mechanism promotes the presence of the particular plasmid in the bacterial population by eliminating plasmid-free cells.

One example is the CCD MECHANISM.

post-transcriptional gene silencing See PTGS.

potyviruses Viruses of the potato virus Y group. These viruses include potato virus Y, soybean mosaic virus and the tobacco etch virus (see TEV).

PP_i See PYROPHOSPHATE.

PPI Protein–protein interaction.

pPICZ vector A 3.3 kb plasmid vector (Invitrogen, Carlsbad CA) designed for high-level, inducible, intracellular expression of recombinant proteins in the yeast *Pichia pastoris* (see also PICHIA).

The plasmid, carrying the gene of interest, integrates stably in the *Pichia* genome.

Using the strong promoter of the *AOX1* gene, control of expression is achieved with a *lac*-based system. Subsequent purification of the isolated recombinant protein can be facilitated e.g. via a SIX-HISTIDINE TAG (encoded as a C-terminal sequence in the plasmid).

The pPICZ plasmid includes a gene encoding resistance to ZEOCIN. By using increasing concentrations of Zeocin™ it is

possible to select for cells in which multiple copies of the construct have integrated in the genome; in some cases it has been found that a higher copy number of the vector results in increased levels of expression of the recombinant protein.

A related plasmid, pPICZα (3.6 kb), is designed for use in the production of a *secreted* recombinant protein; it contains an N-terminal sequence (α-factor) which acts as a secretion signal.

PR proteins (*plant pathol.*) See entry PATHOGENESIS-RELATED PROTEINS.

Prader–Willi syndrome See the entry GENETIC IMPRINTING.

PRD (PTS regulation domain) See CATABOLITE REPRESSION.

pre-miRNA See MICRORNAS and PRI-MIRNA.

pre-mRNA A PRIMARY TRANSCRIPT formed by transcription of a SPLIT GENE. To be translated, the intron(s) are removed by SPLICING, and exons are ligated together, serially, to form a complete coding sequence; further modifications – e.g. the addition of a CAP – are required in order to form the mature mRNA (see MRNA).

(See also QUADRUPLEX DNA.)

pre-natal diagnosis (of inherited disorders) See e.g. DOWN'S SYNDROME and PGD.

(See also EPIGENETIC ALLELIC RATIO ASSAY.)

pre-RC (pre-replication complex) See RE-REPLICATION.

pre-replication complex See RE-REPLICATION.

precursor miRNA See PRI-MIRNA.

preimplantation genetic diagnosis (PGD) See the entry PGD.

prey protein See e.g. YEAST TWO-HYBRID SYSTEM.

pri-miRNA (primary miRNA) The transcript within which an miRNA is embedded prior to its subsequent development. In general, the maturation process requires cleavage of the pri-miRNA (involving the MICROPROCESSOR) to form a ~70 bp hairpin-containing intermediate called the *precursor miRNA* (pre-miRNA). However, in some cases the Microprocessor is bypassed: see MIRTRON. Later, cleavage of the pre-miRNA, or mirtron, (involving DICER) releases a duplex form of the miRNA, one strand of which is used as the effector molecule.

(See also MICRORNAS.)

Pribnow box The PROMOTER consensus sequence TATAAT in e.g. *Escherichia coli*.

primary miRNA See PRI-MIRNA (under 'pri-', above).

primary mutation See BACK MUTATION.

(See also SUPPRESSOR MUTATION.)

primary transcript The RNA transcribed directly from a DNA template strand. If a SPLIT GENE were transcribed the primary transcript is called a *pre-mRNA* because it is a precursor of mRNA; conversion of pre-mRNA to mRNA involves several types of processing, including capping, POLYADENYLATION and SPLICING.

(See also PRI-MIRNA.)

primase See RNA POLYMERASE.

prime plasmid A (circular) plasmid which has excised from a chromosome aberrantly. During the excision it may have left behind a part of its sequence – but taken some chromosomal DNA; this is a type I prime plasmid. A type II prime plasmid is an *entire* plasmid which contains *additional* chromosomal DNA.

primer An oligonucleotide which (usually) must be base-paired to the template strand before a DNA polymerase (or a reverse transcriptase) can synthesize a complementary strand on that template. The DNA polymerase, or the reverse transcriptase, *extends* a primer by adding a sequence of specific nucleotides to the 3′ end of the primer – as dictated by the sequence of nucleotides in the template strand; nucleotides are thus added in the 5′-to-3′ direction.

In vivo, a primer is generally a short oligonucleotide (RNA) synthesized by an RNA polymerase as an initial step in DNA synthesis – the primer being extended by a DNA polymerase and the primer itself being subsequently replaced by DNA.

In at least some eukaryotes, when DNA is synthesized discontinuously, as a series of Okazaki fragments (as in lagging strand synthesis), a primer consisting of a mixture of deoxyribonucleotides and ribonucleotides is apparently synthesized to initiate each Okazaki fragment (see OKAZAKI FRAGMENT).

In some instances, an oligonucleotide primer is *not* a prerequisite for DNA synthesis – see e.g. *protein priming* in the entry PHAGE φ29.

In vitro, primers are normally designed for the synthesis of *specific sequences* of nucleotides – a given primer serving to fix the location at which synthesis begins in a template strand – i.e. primers determine that part of a template strand which is to be copied. An exception to this is the primer of arbitrary sequence: see 'Arbitrary-sequence primer' in the table.

Factors which govern the binding ('annealing') of a primer to its target sequence are the same as those factors governing probe–target binding (see PROBE). The annealing of primers in PCR-based diagnostic tests is carried out under conditions of the highest stringency in order to achieve a suitable level of specificity. The high-affinity binding of ZIP NUCLEIC ACID primers may be advantageous e.g. when amplifying AT-rich sequences by PCR.

The normal functioning of a primer requires that its 3′-end terminal nucleotide base-pairs with the correct (complementary) base in the template strand – even if mismatched base(s) occur elsewhere in the primer; extension from the 3′ end of a primer is either significantly inhibited or blocked by a mismatch at this site. (See also MISMATCH EXTENSION.)

Primer design (in terms of base sequence) reflects various requirements. In diagnostic tests, for example, primers must anneal to a highly specific target sequence. By contrast, so-called broad-range bacterial primers are designed to anneal to conserved sequences found in all bacterial 16S rRNA genes, and such primers can be used to detect *any* species of bacteria in a sample. Computer programs can assist in the design of primers [e.g. for bisulfite-treated DNA: Nucleic Acids Res (2005) 33(1):e9]. [PCR PrimerBank for human/mouse DNA: Nucleic Acids Res (2009) doi: 10.1093/nar/gkp1005.]

Modification of primers for extra functions

In addition to their primary role, primers can be modified in various ways in order to fulfill extra functions. For example,

PRIMER: some examples

Primer	Use (e.g.)
Anchor primer	A primer which binds to a specific *known* sequence – used when the other primer in a primer pair binds to an unknown or variable sequence (see e.g. entry ANCHORED PCR)
Arbitrary-sequence primer	A primer with a non-specific or random sequence used in certain forms of TYPING; some of these primers bind to 'best-fit' sequences in genomic DNA and may then be extended (see entry AP-PCR)
Broad-range primer	A primer which can be used to amplify DNA of a given sequence from each of a wide range of species, all the species sharing a common target sequence; broad-range primers can be used, for example, to detect the presence of *any* bacterial species by amplifying a conserved region of the 16S rRNA gene common to all bacteria
Bumper primer	A primer whose extension is used to displace a strand located downstream on the same template (see entries STRAND DISPLACEMENT and SDA)
Competimer	A *non*-extendable primer used in certain types of quantitative PCR (see entry COMPETIMER)
Degenerate primer	One of a set of primers whose binding site has been predicted e.g. on the basis of the amino acid sequence in a polypeptide (see entry DEGENERATE PRIMER)
Forward primer	A primer which binds to a non-coding strand; when extended it replicates the coding strand (see entry FORWARD PRIMER)
Inner primer	One of a pair of primers used in the second phase of nested PCR; it binds to a subterminal region of the initial amplicon
Outer primer	One of a pair of primers used in the first phase of nested PCR
PCR primer	Any primer used in PCR; PCR primers are sometimes referred to as *amplimers*
Reverse primer	A primer which binds to a coding strand; when extended it replicates the non-coding strand (see entry FORWARD PRIMER)
Self-reporting primer	A PCR primer carrying a quenched fluorophore which is activated during the formation of double-stranded amplicons (see entry SELF-REPORTING PRIMER)
T_m-shift primer	A type of primer involved in the formation of (dsDNA) products which can be identified by their melting curve characteristics (see e.g. SNP GENOTYPING)
Universal primer	A broad-range primer or, in general, any primer which anneals to a common sequence on a range of different molecules

primers may be labeled with a dye to facilitate detection of amplified products, or 5′-end-labeled with BIOTIN in order to immobilize a particular strand of amplified DNA (allowing it to be captured by a STREPTAVIDIN-based system).

A primer for PCR may carry a 5′ tag (i.e. an extension at the 5′ end) which incorporates a promoter sequence – so that the (dsDNA) products can be transcribed; note that the *tag* itself does not bind to the target sequence.

PCR-generated fragments examined by DGGE (q.v.) may incorporate a 5′ GC-rich tag (again, not bound by the primer-binding site) which serves to prevent unwanted *total* melting of dsDNA fragments at the concentrated end of the gradient.

Primers with a short, AT-rich 5′ flap (tag) were reported to increase the fluorescent signal in real-time PCR, although the reason for this was not given [BioTechniques (2007) 43(6): 770–774].

A *non-extendable* primer is used in some types of quantitative PCR: see e.g. COMPETIMER.

(See also SINGLE-BASE EXTENSION.)

Heat-activatable primers have been reported to be effective in HOT-START PCR (q.v.).

A few examples of primers are listed in the table. (See also e.g. LUX PRIMER.)

primer–dimer In PCR: an artefact formed when the 3′ ends of two primers are mutually complementary; during extension, each primer uses the other as a template, forming two double-stranded primers in tandem.

Owing to the use of multiple primer pairs, primer–dimers may be particularly problematic in MULTIPLEX PCR.

primer exclusion See PCR CLAMPING.

primer extension (from mismatched base-pairs) See the entry MISMATCH EXTENSION.

(cf. PRIMER-EXTENSION INHIBITION.)

primer extension assay A procedure which has been used e.g. for identifying the nucleotides that form the 5′ end of a given molecule of mRNA.

The initial requirement is for a (labeled) primer that binds to the particular mRNA of interest, with high specificity, at a site near the 5′ end; thus, on extension of this primer (with a reverse transcriptase), the last nucleotide to be added will be paired with the 5′ terminal nucleotide of the mRNA.

Copies of this primer are added to a preparation of mRNAs isolated from cells or tissue expressing the particular gene of interest. Primers bind to the corresponding mRNA molecules, and each primer is extended by reverse transcriptase to form a (labeled) strand of cDNA.

A segment of genomic DNA containing the gene of interest is then sequenced (see DNA SEQUENCING) – *using the same primer*. Sequencing thus begins at the same primer-binding site and continues through the gene's START SITE. Products of the sequencing reaction will therefore include one fragment that terminates at the gene's start site; this particular product will thus be the same length as the molecules of cDNA which were produced by reverse transcription of the given mRNA.

All the products from the sequencing reaction are run in a (4-lane) electrophoresis gel – an extra (5th) lane being used for the cDNAs. The results of sequencing will therefore give the identity of each nucleotide between the 3′ terminus of the primer and the start site – the start site being identified by the product that forms a band in the same relative position as that formed by the cDNA.

primer-extension inhibition The principle underlying several techniques – see e.g. TERMINAL TRANSFERASE-DEPENDENT PCR and TOEPRINTING.

primer sequestration A technique (USB Corporation, Cleveland OH) used in HOT-START PCR; it involves sequestration of the primers at the lower temperatures prior to the initial stage of denaturation. In this process, the primers bind to a recombinant protein (present in the reaction mixture) which is inactivated at the initial denaturation step, following which the primers are released to function in the normal way.

primer walking The sequential use of a number of different primers which bind at successive sites along the length of a template strand, the binding sequence of a given primer being within the 3′ end of the sequence covered by extension of the previous primer.

Primer walking can be used e.g. for solid-phase sequencing of DNA (by the DIDEOXY METHOD) in which strands of 5′-BIOTINylated template DNA (up to about 10 kb) are initially immobilized on Dynabeads® coated with streptavidin. After each sequencing reaction (with a given primer) the newly synthesized strands are eluted by formamide and examined in a sequencing gel. The immobilized template strand is then reconditioned by treatment with sodium hydroxide, washed, and used again for the next round of sequencing.

(See also DNA SEQUENCING.)

pri-miRNA See the entry above (under 'pri-').

primordial germ cells The cells, in a developing embryo, from which the gametes will be formed.

printing (*biotechnol.*) The process of adding probes to a glass or other surface – e.g. by *in situ* synthesis.

pRNA See PACKAGING (of phage DNA).

probe (*mol. biol.*) (1) Any short (single) strand of DNA, RNA or PNA which is *complementary* to a specific target sequence of nucleotides, and which is used to detect, identify, amplify or monitor that target sequence by binding to it, permanently or transiently; binding, by the probe, signals the presence of the given target sequence (qualitatively or quantitatively) and is registered in various ways – usually involving the direct or indirect *labeling* of the probe (see PROBE LABELING and see also UNLABELED PROBE).

(cf. BINARY PROBE.)

(See sense (2), below, for a note on *double-stranded* DNA probes.)

Often, the DNA in which the *target* sequences may occur is immobilized and then exposed to many copies of one or more types of *free* probe (i.e. probes in solution) under appropriate conditions; after removal of unbound probes, by washing, the presence of (any) bound probes may be detected e.g. by their labels.

The term *probe* is also used to refer to immobilized oligonucleotides in a MICROARRAY. (See also LINE PROBE ASSAY and REVERSE HYBRIDIZATION.)

Probe–target binding is strongly influenced by factors such as electrolyte concentration and temperature. The conditions can be arranged so that probes will bind to their targets *only* when probe and target sequences are *exactly* complementary to one another; such conditions are said to be *high-stringency* conditions and are used to promote high-level specificity in probe–target binding – necessary, for example, in diagnostic tests. Under these conditions, mismatches in base-pairing are not tolerated.

Low-stringency, in which the effect of mismatched base(s) is minimized by an appropriate choice of conditions, may be used for particular purposes (see e.g. AP-PCR and TAIL-PCR).

Probe–target binding may occur (e.g.) ~25°C below the T_m (see entry THERMAL MELTING PROFILE) of the corresponding (duplexed) probe–target sequence. The temperature at which probe–target hybridization occurs can be manipulated by the use of agents (such as formamide) which decrease the T_m by a specified amount (depending on the concentration of agent used).

The probing of *RNA* target sequences (e.g. in Northern blot analysis) may be facilitated by using probes that consist of 2′-*O*-methyl oligoribonucleotides (rather than 2′-deoxy oligoribonucleotides) [Nucleic Acids Res (1998) 26:2224–2229]. The stability of binding of these probes may be raised by the use of LOCKED NUCLEIC ACIDS [Nucleic Acids Res (2005) 33(16):5082–5093].

PNA is used e.g. in the LIGHTUP PROBE.

[Fluorescent probes for super-resolution imaging in living cells: Nature Rev Mol Cell Biol (2008) 9:929–943.]

Some uses of probes

The uses of probes include:

- Detection of single-base mutations/SNPs (e.g. MOLECULAR INVERSION PROBE GENOTYPING)

PROBE: some examples

Probe	Use (e.g.)
Abasic-site mimic	A (quenched) fluorescent probe used for detecting products in an isothermal method for DNA amplification (see RPA, sense 1)
bDNA probe	Quantitation of pathogens by a signal amplification process (see entry BDNA ASSAY)
Binary probe	A composite probe used e.g. in FRET-based systems and in the PYRENE BINARY PROBE approach
Capture probe	A biotinylated probe used e.g. for capturing the target sequence and linking it to streptavidin-coated DYNABEADS in a CROSS-LINKING ASSAY (also used in other procedures)
Degenerate probe	One of a set of probes used e.g. when the target sequence is not completely known but has been predicted (see the entry DEGENERATE PROBE)
End probe	See the entry CHROMOSOME WALKING
Hybridization protection probe	Detection of pathogens in clinical samples (see the entry ACCUPROBE)
LightCycler™ probe	Monitoring REAL-TIME PCR; quantitation of target sequences
LightUp® probe	Monitoring real-time PCR (see the entry LIGHTUP PROBE)
2′-*O*-methyl oligo probe	Probing RNA, e.g. in Northern blot analysis
Molecular beacon probe	Monitoring real-time PCR and NASBA; quantitation of target sequences (see the entry MOLECULAR BEACON PROBE)
Molecular inversion probe	SNP genotyping (see the entry MOLECULAR INVERSION PROBE GENOTYPING)
Molecular zipper	Isothermal, real-time amplification of DNA (see the entry MOLECULAR ZIPPER)
PACE 2C probe	Combination probe diagnostic test for detecting the bacterial pathogens *Chlamydia trachomatis* and *Neisseria gonorrhoeae* in clinical samples in a single assay (see the entry PACE 2C)
Padlock probe	Confirmation of sequences. Identification of variation in sequences. Detection of SARS virus in clinical samples (see the entry PADLOCK PROBE)
Pyrene binary probe	Detection of specific mRNAs (see the entry PYRENE BINARY PROBE)
Quantum dot probe	Multiplex approach in Western blot analysis (see the entry QUANTUM DOT PROBE)
Reporter probe	A probe involved e.g. in generating a fluorescent signal in a CROSS-LINKING ASSAY (q.v.)
RPA probe	Detection of products in a method for isothermal amplification of DNA: see abasic-site mimic, above
Scorpion probe	Monitoring PCR products (see the entry SCORPION PROBE)
Selector probe	Amplification of a number of specific target sequences in a single assay (see the entry SELECTOR-BASED MULTIPLEX PCR)
Smart probe	Detection of single-base mismatches, SNPs. Essentially similar in principle to the molecular beacon probe, but quenching of the fluorophore (in the unbound probe) is due to guanosine residues in the probe itself (see the entry SMART PROBE)
TaqMan® probe	Monitoring real-time PCR; quantitation of target sequences (see the entry TAQMAN PROBE)
Triple-strand probe	Detecting SNPs or e.g. single-base mismatches (see the entry TRIPLE-STRAND PROBE)

- Confirmation of sequence variation (e.g. PADLOCK PROBE)
- Monitoring progress in REAL-TIME PCR (q.v.)
- Quantitation of target sequences in real-time PCR
- Isothermal amplification of DNA (MOLECULAR ZIPPER)
- Amplification of specific target sequences (LIGASE CHAIN REACTION; SELECTOR-BASED MULTIPLEX PCR)
- Detection of mRNAs (e.g. PYRENE BINARY PROBE)
- Detection of pathogens in clinical samples (e.g. PACE 2C)
- Detection of toxin (or other) genes in DNA samples
- Localization of specific intracellular target sequences by *in situ* hybridization
- Identification of specific cloned sequences (e.g. COLONY

HYBRIDIZATION)

Preparation of probes

Probes can be made in various ways. For example, they can be synthesized, directly, according to any required sequence of nucleotides; this option is available commercially. They can also be prepared e.g. by asymmetric PCR, the appropriate primer being used in excess, or by cloning.

(2) Any ligand used for the detection of a specific target – for example, a labeled antibody of the type used in Western blot analysis (e.g. QUANTUM DOT PROBE).

Note that a *double-stranded* DNA probe has been used to detect certain DNA-binding proteins (transcription factors) in the EMEA procedure; see also the entry for NFκB.

(3) An agent or system used for obtaining information about a specific structure, process or reaction – for example, photoreactive tRNA derivatives used to probe the peptidyltransferase center of the ribosome [RNA (2007) 13(5):793–800].

probe labeling Any procedure in which a PROBE is modified – either before or after binding to the target – in order to enable it to indicate the presence, quantity or location of the target.

There are many ways of labeling probes, and the following refers primarily to the labeling of *oligonucleotide* probes.

Labeling may be direct or indirect.

In *direct* labeling the label is an intrinsic part of the probe.

In *indirect* labeling the label is bound to the probe – often non-covalently – via an intermediate molecule.

Direct labeling methods

A radioactive label such as ^{32}P can be added to the terminal 5′-OH group of DNA (or RNA) by the enzyme polynucleotide kinase using [$\gamma^{32}P$]ATP, i.e. ATP labeled with ^{32}P at the terminal (γ) phosphate group; in this (end-labeling) reaction, the radioactive phosphate group is transferred to the 5′-OH terminus of the oligonucleotide.

Alternatively, a probe may be synthesized e.g. as a segment within a circular double-stranded cloning vector (such as a plasmid) and labeled by NICK TRANSLATION in a reaction mixture which includes radioactively labeled dNTPs. In this process, labeled nucleotides are incorporated randomly in the plasmid. The relevant sequence is subsequently excised and denatured to produce the single-stranded probes.

Radioactive probes can be detected by autoradiography.

Radioactive (= *isotopic*) labeling has been widely used e.g. for *in situ* hybridization. Radioactive labels are robust and are associated with good sensitivity; however, they have limited resolution. Other forms of labeling (*non-isotopic labeling*) are often preferred because of the hazards and costs that are associated with handling radioactive material.

Fluorescent labels are now popular because they are easy to handle and because there is currently an extensive range of fluorophores with diverse emission spectra, permitting multiplex working, if required.

Existing oligonucleotide probes can be labeled e.g. by a commercial process (ULYSIS®) in which a platinum–label complex forms a stable adduct at the N7 position of guanine; any of a variety of fluorescent dyes can be used as the label. This is a rapid and convenient method of labeling. Examples of use: studies on chromosomal instability [Neoplasia (2009) 11(3):260–268]; studies on the design of microarray probes [BMC Genomics (2008) 9:496].

As an alternative, ChromaTide™ nucleotides (marketed by Molecular Probes Inc, Eugene OR) can be incorporated into DNA or RNA during synthesis. These products consist of a nucleotide (e.g. UTP) covalently bound to a fluorescent label via a spacer link; again, any of a variety of dyes can be used as label. Depending on the particular dye–nucleotide product, incorporation may be carried out during e.g. PCR, or during *in vitro* transcription and/or reverse transcription; many of the products can be used for 3′-end labeling using the enzyme terminal deoxynucleotidyl transferase.

Another procedure for fluorescent labeling (ARES™, Molecular Probes) is a two-step method. Amine-modified nucleotides – 5-(3-aminoallyl)-dUTP – are first incorporated into DNA during synthesis, e.g. during reverse transcription; subsequently, an amine-reactive fluorophore binds covalently to the available amine groups in the DNA. Any of a number of amine-reactive dyes can be used – e.g. various Alexa Fluor® dyes.

When labeling DNA probes during synthesis, the efficiency of the process can vary according e.g. to the type of DNA polymerase and to the site of labeling.

(See also LIGHTUP PROBE.)

Indirect labeling methods

One common procedure involves the initial insertion of a molecule of BIOTIN into the probe. This is often achieved by incorporating covalently complexed biotin–nucleotide molecules into DNA or RNA during synthesis; this can be carried out e.g. during PCR. The probe can then be detected by using molecules of STREPTAVIDIN which are themselves linked to either a fluorophore (detected by fluorescence) or an enzyme – e.g. ALKALINE PHOSPHATASE (AP).

With an AP label the probe can be detected in several ways, depending on the experimental arrangements. In some cases it may be appropriate to identify the location of a bound AP-labeled probe by a chromogenic (color-generating) reaction; for example, AP forms a dark bluish/purple precipitate with 5-bromo-4-chloro-3-indolyl phosphate (BCIP) and nitroblue tetrazolium (NBT) – both reagents being used in the reaction.

Alternatively, AP can produce chemiluminescence (chemically generated light) with e.g. 1,2-dioxetane substrates such as AMPPD® and CSPD®.

Although the streptavidin–label complex is usually added *after* probe–target hybridization, it was reported earlier that short biotinylated probes which are *already* labeled with the streptavidin–alkaline phosphatase complex before use permit satisfactory probing [Nucleic Acids Res (1999) 27:703–705].

Indirect labeling can be carried out in a similar way using e.g. a DIGOXIGENIN (rather than a biotin) tag.

ProBond™ nickel chelating resin A (pre-charged) NICKEL-CHARGED AFFINITY RESIN (marketed by Invitrogen, Carlsbad CA) which is used e.g. for the purification of 6×His-tagged

proteins – e.g. those produced by expression vectors such as CHAMPION PET SUMO VECTOR or VOYAGER VECTORS.

[Uses (e.g.): PLoS Genet (2008) 4(9):e1000194; Mol Cell Biol (2008) 28(15):4851–4861; Plant Physiol (2009) 150(1): 521–530; J Exp Bot (2009) 60(6):1743–1757.]

processive enzyme Any enzyme that remains bound to the substrate as it (repeatedly) carries out its function.

(See also PCNA.)

procyclin A member of a family of glycophosphatidylinositol-associated proteins which occur in a surface layer on certain protozoa, e.g. on some *Trypanosoma* spp within the gut of an insect vector (cf. VSG); procyclins may protect trypanosomes from enzymic attack within the insect's gut.

Post-transcriptional regulation of procyclin genes involves a protein – *Tb*ZFP3 – which was found to associate with the mRNAs of some procyclin isoforms [PLoS Pathog (2009) 5 (2):1000317].

producer cell line See e.g. PACKAGING CELL and PACKAGING PLASMID.

proenzyme *Syn*. ZYMOGEN.

profile See e.g. DNA PROFILE.

programable endonuclease A RESTRICTION ENDONUCLEASE that is linked covalently to a molecule that binds to a specific sequence of nucleotides in double-stranded DNA; sequence-specific binding of the latter molecule directs the activity of the restriction endonuclease to a *particular* cutting site in the region of binding.

The molecule linked to the restriction endonuclease may be e.g. a DNA-binding protein or a (synthetic) triplex-forming oligonucleotide (see TRIPLEX DNA). A conjugate that formed between a TFO and the restriction enzyme PvuII achieved a high level of cutting specificty at a given PvuII restriction site [Nucleic Acids Res (2005) 33(22):7039–7047].

(See also ZINC-FINGER NUCLEASE.)

programed ribosomal frameshifting See FRAMESHIFTING.

progressive multifocal leukoencephalopathy See JC VIRUS.

prohead The precursor of a phage head (i.e. capsid) prior to the insertion of the genomic nucleic acid.

proinsulin (human, recombinant) See HUMAN INSULIN PRB.

prokaryote Any organism, limited to species of the ARCHAEA or BACTERIA, whose cellular structure is characterized by the absence of a nuclear membrane and the absence of organelles such as mitochondria. (cf. EUKARYOTE.) There are various other basic differences between prokaryotes and eukaryotes – e.g. differences in ribosomes (and in translation factors etc.); in the types of enzyme which act on nucleic acids; in proteasomes; in membrane lipids; in various aspects of metabolism; and in the general organization of the genome.

The *typical* prokaryotic genome is a circular dsDNA molecule (chromosome) which may be present in more than one copy, depending e.g. on growth temperature and species. Although prokaryotes usually contain only one *type* of chromosome, the cells of e.g. *Vibrio cholerae* contain two dissimilar chromosomes which jointly constitute the bacterial genome [Nature (2000) 406:477–483]. (See also NUCLEOID.)

Some prokaryotic chromosomes are not ccc dsDNA. For example, the chromosomes are molecules of *linear* dsDNA in the bacterium BORRELIA BURGDORFERI and in species of e.g. *Rhodococcus* and *Streptomyces*.

proliferating cell nuclear antigen See PCNA.

proline–glutamate–serine–threonine-rich sequence See PEST DOMAIN.

promiscuous plasmid Any conjugative plasmid which is able to promote self-transmission among bacteria of widely differing species/genera.

promoter A sequence of nucleotides, at the beginning (5′ end) of a gene, which is necessary for transcription of the gene; it is the (direct or indirect) binding site of various components of the transcription machinery, including the RNA POLYMERASE and other essential factors. (See also SIGMA FACTOR.)

Following assembly of functional transcription machinery, the strands of dsDNA separate in the region of the 'start site' to form the *open complex* – i.e. exposing the single-stranded DNA template.

The START SITE (q.v.) is the precise nucleotide at which transcription begins, i.e. it determines the first (5′) nucleotide of the nascent transcript.

Strong promoters (e.g. the phage T7 promoter, the *AOX1* promoter of the yeast *Pichia pastoris*, the *GAL1* promoter of *Saccharomyces cerevisiae*, the SV40 viral promoter, and the baculovirus polyhedrin gene promoter) are efficient ones, i.e. they support high-level transcription – while *weak* promoters are relatively inefficient. (See also UP MUTATION.) However, the strength of a given promoter may differ in different types of cell, i.e. it cannot be assumed that a promoter that is strong in one type of cell will be strong in other types of cell.

Some promoters are 'leaky' – for example, the *lac* promoter (see entry LAC OPERON) permits a low level of transcription, even in the absence of inducer; for this reason, when the *lac* system is employed for transcriptional control, use may be made of a plasmid-borne *lacI*^{q} gene in order to form higher-than-normal levels of the *lac* repressor protein. By contrast, some promoters (for example, the phage T7 promoter) offer tighter control of transcription.

Some regions within eukaryotic promoters are mentioned in CORE PROMOTER.

In the prokaryote *Escherichia coli*, the promoter includes: (i) the CONSENSUS SEQUENCE TATAAT, centered on or near the –10 nucleotide (see START SITE); (ii) another consensus sequence, TTGACA, centered on/near the –35 nucleotide; (iii) an AT-rich *UP element* found between nucleotides –40 and –60.

Promoter activity/function may be investigated as part of a study on gene expression: see relevant item in the entry GENE FUSION. (See also SPLIT-CRE.)

promoter control (in operons) See OPERON.

promoter core *Syn*. CORE PROMOTER.

promoter-distal *Downstream* of a PROMOTER, i.e. in the 3′ direction from the promoter.

promoter-probe vector A construct, containing a PROMOTER-

less REPORTER GENE, which has various uses. For example, it can be used for studying, quantitatively, the role of a specific promoter by inserting that promoter upstream of the reporter gene and measuring gene expression under given conditions.

It is also used e.g. to search for those promoters, in a given genome, that are functional under particular conditions. For this purpose, fragments of the genome can be cloned in the promoter-probe vector, and the resulting clones screened to detect those in which the reporter gene is expressed under the specified conditions.

A promoter-probe vector is also useful for defining the promoter of a given gene. Fragments of DNA that include (i) the gene's *start site* (determined e.g. from a PRIMER EXTENSION ASSAY), and (ii) regions upstream of the start site, are cloned in a promoter-probe vector and the resulting clones screened for reporter activity; reporter activity will be detected only if an active promoter occupies the cloning site of the promoter-probe vector. Any clones exhibiting reporter activity can be examined further by sequencing.

[Example of use in a quantitative study: J Bacteriol (2008) 190(10):3456–3466.]

(See also GENE TRAPPING.)

promoter-proximal *Upstream* of a promoter, that is, in the 5′ direction.

promoter trapping See GENE TRAPPING.

Pronase (proprietary name) An enzyme mixture which includes both exopeptidases and endopeptidases; it is obtained from a Gram-positive bacterium, *Streptomyces griseus*. This product is used for extensive, non-specific degradation of proteins.

proof reading (editing) During DNA replication: removal of an incorrectly incorporated nucleotide and its replacement with the correct nucleotide; proof reading is carried out by certain types of DNA-DEPENDENT DNA POLYMERASE – those which have a 3′-to-5′ exonuclease capability.

prophage See BACTERIOPHAGE and LYSOGENY.

propidium iodide A fluorescent phenanthridinium dye which binds to DNA and RNA; it is generally excluded from living cells and is used e.g. as a stain for dead cells (see e.g. LIVE/DEAD BACLIGHT BACTERIAL VIABILITY KIT). The dye fluoresces red on appropriate excitation.

Compared with the related dye, ETHIDIUM BROMIDE, this dye is more water-soluble and less membrane-permeant.

(See also DNA STAINING.)

2-propylpentanoic acid Valproic acid: see the entry HDAC.

prostate cancer (detection) See CANCER DETECTION.

prostate-specific antigen (PSA) See CANCER DETECTION.

protamine sulfate An agent used e.g. for the enhancement of virus-mediated transfer of genes into cells (by both retroviral and non-retroviral vectors).

[Use (retroviral transfection: 4 mg/mL): PLoS ONE (2009) 4(3):e4779.]

(See also ENDO-PORTER and MELITTIN.)

protease inhibitors (antiretroviral) A category of ANTIRETROVIRAL AGENTS which inhibit *retroviral* protease: an enzyme (encoded in the *pol* region of the viral genome: see RETROVIRUSES) which is required for post-translational processing of viral proteins; the protease inhibitors include amprenavir, indinavir, nelfinavir, ritonavir and saquinavir.

Viral resistance to protease inhibitors can arise as a result of mutation in the protease gene – such mutation reducing the binding of protease inhibitors. It was reported that mutations in the *gag* region of the HIV-1 genome can also contribute to HIV-1 resistance to protease inhibitors [PLoS Pathog (2009) 5(3):e1000345].

The agent BEVIRIMAT inhibits protease activity, but it does not bind to the enzyme.

proteasome A hollow, cylindrical, intracellular structure with internal active enzymic sites involved e.g. in the degradation of earmarked proteins and processing of exogenous antigens.

Prokaryotic proteasomes are of the 20S type while those of eukaryotes are of the 26S type. The cellular degradative pathway also differs in prokaryotes and eukaryotes in that e.g. the process is UBIQUITIN-dependent in eukaryotic cells. (See also BORTEZOMIB.)

For degradation, a given protein must initially be unfolded in order to enter the proteasome; this is an energy-dependent process, and AAA ATPASES are known to form a part of the multi-protein complex of at least some proteasomes.

Degradation of a eukaryotic protein is preceded by the sequential (cascade) binding of a number of ubiquitin molecules to the given protein; this involves three types of enzyme: an activating enzyme (E1), a ubiquitin-conjugating enzyme (E2) and a ubiquitin–protein ligase (E3). (Such ubiquitinylation is involved not only in degradative action but e.g. in regulating protein activities and influencing subcellular location.)

(See also SUMOYLATION.)

The ubiquitin-binding cascade seems to proceed as follows. In an ATP-dependent reaction, the active-site cysteine of E1 forms a thioester with ubiquitin; ubiquitin is then transferred to E2. The ubiquitin ligase (E3) then mediates attachment of ubiquitin to (usually) the ϵ-amino group of a lysine residue in the target protein. Some E3 enzymes appear to form an intermediate thioester with ubiquitin (before transferring it to the target protein); however, most do not form such a bond – but bind to E2 and the target protein to facilitate ubiquitinylation of the protein. The way in which ubiquitin passes from E2 to the protein is apparently not understood. Some data suggest that most types of ubiquitin ligase activate E2 enzymes allosterically.

[Proteasomes (minireview): Cell (2007) 129:659–662.]

[Chemical genetics: exploring the role of the proteasome in cell biology using natural products and other small-molecule proteasome inhibitors: J Med Chem (2008) 51:2600–2605.]

[Structure of human 26S proteasomes: J Biol Chem (2008) 283(34):23305–23314.]

(See also the entries CHIP, CULLIN-RING COMPLEXES and N-END RULE.)

protein A A high-molecular-weight protein found in the cell wall in most strains of *Staphylococcus aureus*; protein A can be released, intact, from its covalent attachment to the (cell

wall) polymer peptidoglycan by the enzyme LYSOSTAPHIN.

One of the domains of protein A binds with high specificity to the Fc portion of most types of IgG antibody; the protein also binds to the Fab region of immunoglobulins of other isotypes. Such binding is exploited in a number of techniques – e.g. a scintillation proximity assay (see the entry SPA).

Protein A has also been used in affinity chromatography (for purification of IgG antibodies), and (in derivative form) in a QUANTUM DOT PROBE. (See also IMMUNO-PCR.)

(cf. PROTEIN G and SORTASE.)

protein analysis (*in vivo*) See e.g. TEV.

protein–biotin ligase See IN VIVO BIOTINYLATION.

protein chip *Syn*. PROTEIN MACROARRAY.

protein degradation See e.g. N-END RULE and PEST DOMAIN.

protein design The computer-aided design of synthetic (novel) proteins with the required characteristics or properties.

protein detection See e.g. PROXIMITY LIGATION ASSAY.

(See also IMMUNO-PCR and IMMUNOBLOTTING.)

protein–DNA interactions See DNA–PROTEIN INTERACTIONS.

protein engineering The creation of modified forms of a given protein (with required characteristics or properties) e.g. by recoding the protein's gene.

protein G A protein, derived from group G streptococci, which apparently binds IgG of all subclasses; unlike PROTEIN A, it does not bind immunoglobulins of other isotypes.

protein kinase Any of various ATP-dependent enzymes which phosphorylate particular residues in proteins – e.g. TYROSINE KINASE.

protein labeling Labeling proteins *in vitro*, and at cell-surface locations in living cells, can be achieved e.g. by site-specific conjugation between the target protein and a given label, catalyzed by transglutaminase [J Am Chem Soc (2006) 128(14): 4542–4543]. In this procedure, the target protein is initially fused to a short peptide tag (a so-called Q-tag) which acts as a recognition sequence for the enzyme transglutaminase. The recombinant (Q-tagged) protein can be prepared by fusing the protein-encoding gene to a sequence encoding the Q-tag (see GENE FUSION and FUSION PRODUCT). Once the target protein has been Q-tagged it can be (covalently) linked to any of a wide range of labels in a reaction that is mediated by transglutaminase.

Labeling of proteins is also carried out with other enzymes, each of which recognizes its own (enzyme-specific) peptide tag – which is fused to the target protein. For example, labeling can be carried out with phosphopantetheinyl transferase and also with biotin ligase. The latter enzyme is used in DNA technology e.g. for IN VIVO BIOTINYLATION.

protein macroarray A membrane on which specific proteins are fixed at known sites. Such a membrane may be exposed to proteins in a given sample and examined to see whether any binding has occurred between proteins in the sample and those on the membrane. If the sample proteins had been antibodies which were expected to bind to one or more protein(s) on the membrane, then, following exposure to the sample, the membrane can be treated with labeled *secondary* antibodies, i.e. antibodies that will bind to any of the antibodies from the sample which had bound to membrane proteins.

In one scheme, target proteins, at specific locations on the membrane, were synthesized *in situ* by recombinant bacteria. In this scheme, the bacteria were initially transformed with a cDNA library. Individual bacterial clones were then grown, separately, in microtiter wells. Each bacterial clone was then transferred, robotically, from its microtiter well to a specific location on a nitrocellulose membrane – and there induced to express the cDNA (i.e. to synthesize the protein encoded by the cDNA) *in situ* (i.e. at the site on the membrane occupied by that clone) [RNA (2007) 13(5):704–712].

A protein macroarray was used e.g. to study autoantigens in COPD patients [Respir Res (2009) 10(1):20].

protein overexpression (protein overproduction) See OVEREXPRESSION.

protein pIII See PHAGE DISPLAY.

protein priming (of DNA replication) See e.g. PHAGE φ29.

protein–protein interactions See e.g. BACTERIOMATCH TWO-HYBRID SYSTEM, BIMOLECULAR FLUORESCENCE COMPLEMENTATION, CO-IMMUNOPRECIPITATION, the CYTOTRAP™ TWO-HYBRID SYSTEM, YEAST TWO-HYBRID SYSTEM and YEAST THREE-HYBRID SYSTEM.

(See also SOLVATOCHROMIC FLUORESCENT DYE.)

protein splicing See INTEIN.

protein synthesis (*in vitro*) See the entry CELL-FREE PROTEIN SYNTHESIS.

protein *trans* splicing See *split intein* in the entry INTEIN.

proteinase K A non-specific serine protease which is effective over a wide range of pH.

[Uses of proteinase K (e.g.): World J Gastroenterol (2009) 15(21):2679–2683; (in studies on prions) PLoS ONE (2009) 4(6):e5829; PLoS ONE (2009) 4(6):e5916.]

protelomerase An enzyme, encoded e.g. by certain bacteria (such as BORRELIA BURGDORFERI) and bacteriophages (e.g. PHAGE N15 and phage φKO2), which is required during DNA replication to resolve the chromosome dimers by catalyzing the formation of closed (hairpin) terminal regions of genomic DNA (see TELOMERE sense 2). Target sites for protelomerase activity were reported to involve inverted repeat sequences, recognition involving both structure-specific and sequence-specific interaction.

Studies were carried out on the crystal structure of TelK (a protelomerase encoded by phage φKO2) in complex with its palindromic target DNA. The studies indicated that an interlocked TelK dimer distorts DNA structure during formation of the hairpin telomeres. It was proposed that protein-induced distortion of DNA made the hairpinning reaction effectively irreversible by preventing re-ligation of strand-cleavage sites [Mol Cell (2007) doi: 10.1016/j.molcell.2007.07.026.]

proto-*onc* A *cellular* (as opposed to viral) ONCOGENE.

proton PPase See PYROPHOSPHATE.

protoplast An osmotically sensitive structure produced *in vitro* by removal of the cell wall from a cell suspended in an isotonic or hypertonic medium; a protoplast therefore consists of

the cell membrane with all that it encloses.

A protoplast can metabolize, and some are able to revert to normal cells under suitable conditions.

Bacterial protoplasts resist infection by phages because the phage receptor sites are lost when the outer layer of the cell is removed.

(See also AUTOPLAST.)

prototroph See the entry AUXOTROPHIC MUTANT.

prototype (of a restriction enzyme) See ISOSCHIZOMER.

provirus Viral DNA integrated in the host's genome. (A retroviral provirus is a reverse-transcribed viral genome.)

(See also *prophage* in BACTERIOPHAGE and LYSOGENY.)

proximal box See RNASE III.

proximity ligation assay (PLA) A sensitive, DNA-based technique used e.g. for detecting specific proteins.

PLA involves the use of protein-binding entities (such as monoclonal antibodies, or APTAMERS) to which are attached oligonucleotides (of known sequence); these protein-binding entities are known as *proximity probes*. When two proximity probes bind to *site-specific* regions on the target protein, the distal ends of their oligonucleotides become close enough for enzymic ligation. Ligation of the two oligonucleotides may be assisted by adding so-called 'connector oligos' that base-pair with the distal ends of the proximity probe oligos – thus holding them in close proximity for ligation. The non-bound proximity probes are silent, but any ligated oligos, on bound probes, provide an amplifiable DNA sequence for detection and/or quantitation by REAL-TIME PCR.

(See also IMMUNO-PCR.)

Basic PLA was developed into the *triple-specific proximity ligation assay* (3PLA). In 3PLA, the binding of three distinct proximity probes (1, 2 and 3) to a given target is required in order to form amplifiable DNA. The distal end of the oligo in probes 1 and 2 is initially bound to a short, blocking oligo but the oligo in probe 3 is not blocked in this way. When all three of the probes bind to the target, the oligo of probe 3 displaces the blocking oligos from probes 1 and 2 and then binds to the oligos of probes 1 and 2 either side of a gap; this allows the oligos of probes 1 and 2 to be ligated on either side of a short (added) cassette oligo. This creates amplifiable DNA that can then be examined by PCR.

A PLA was used e.g. for detecting phosphorylated PDGFβ [Mol Cell Proteomics (2007) 6:1500–1509].

Preparation of proximity probes

One way to prepare proximity probes is to link (covalently) the proximal end of each oligo to STREPTAVIDIN; this conjugate can then bind strongly to biotinylated antibodies. An alternative, convenient approach is to bind biotinylated oligos to the biotinylated antibodies via streptavidin [BioTechniques (2007) 43(4):443–450].

proximity probe See PROXIMITY LIGATION ASSAY.

PS-341 *Syn*. BORTEZOMIB (q.v.).

PSA (prostate-specific antigen) See CANCER DETECTION.

pSC101 A ~9-kb low-copy-number mobilizable plasmid which encodes resistance to tetracycline; it has one EcoRI cleavage site. pSC101 has been used e.g. as a CLONING VECTOR.

pseudo-exons See SPLICING.

pseudo-mRNA A type of RNA molecule which resembles a normal mRNA but which has a disrupted reading frame so that it does not encode a full-length protein.

[Pseudo-mRNA: PLoS Genetics (2006) 2(4):e23.]

pseudo-wild phenotype See SUPPRESSOR MUTATION.

pseudogene (truncated gene) A sequence of nucleotides which resembles an existing (functional) gene but which does not yield the corresponding product; some pseudogenes resemble a DNA copy of an mRNA – having a short A–T tract, which corresponds to a poly(A) tail, and the absence of intron-like sequences. (cf. RETROGENE.)

Some pseudogenes exist in fragmented form in the genome.

Pseudogenes have been regarded as non-functional.

Pseudogenes occur in both eukaryotes and prokaryotes. For example, truncated copies of the 5′ end of functional genes, possibly formed through duplication events, were reported in the channel catfish (*Ictalurus punctatus*), and the genome of the prokaryote *Mycobacterium leprae* was reported to contain ~27% pseudogenes; pseudogenes were also recorded in K-12 strains of *Escherichia coli*.

Mitochondrial pseudogenes are reported to be prevalent in the mosquito *Aedes aegypti* [BMC Genet (2009) 10:11].

In recent years, some pseudogenes have been reported to be active within cells. One study on the *BRAF* pseudogene has concluded that its activity may be linked to the development of thyroid tumors [Neoplasia (2009) 11(1):57–65].

[Pseudogene families database: Nucleic Acids Res (2009) 37(Database issue):D738–D743.]

pseudoisocytosine (in PNA oligomers) See TRIPLEX DNA.

Pseudomonas aeruginosa A species of Gram-negative aerobic, oxidative (i.e. respiratory) bacteria which is catalase-positive, oxidase-positive and usually motile. It is a major opportunist pathogen in man and other animals; mucoid (i.e. alginate-producing) strains are particularly problematic e.g. in CYSTIC FIBROSIS. At least some of the virulence factors are expressed via QUORUM SENSING. *P. aeruginosa* is inherently insensitive to a number of common antibiotics.

The GC% of the genomic DNA is 67.

P. aeruginosa is the type species within the genus *Pseudomonas*.

pseudoparticle (of hepatitis C virus) An *in vitro* construct prepared by coating retroviral core particles with glycoproteins of hepatitis C virus (HCV); pseudoparticles may also include GREEN FLUORESCENT PROTEIN to facilitate the monitoring of infection.

Pseudoparticles have been used e.g. in studies on the entry stage during the infection of cells by HCV, e.g. an analysis of the RGE/RGD motif of the E2 protein [Virol J (2009) 6:12].

pseudopromoter (1) Any promoter that functions *in vitro* but not *in vivo*.

(2) A promoter, constructed *in vitro*, that is able to function *in vitro* and *in vivo* [use of term (e.g.): Proc Natl Acad Sci USA (2002) 99(1):54–59].

pseudorecombination (1) During cycling in a PCR reaction: the formation of hybrid products owing to interactions between various templates (which may prime one another); this may occur when the reaction mixture contains different target sequences which have an appropriate level of homology.

While this kind of effect is exploited in '*in vitro* evolution' (see e.g. DNA SHUFFLING), it may produce erroneous results e.g. when attempts are being made to quantitate a number of similar sequences in a multiplex PCR. One assay which was designed to estimate this effect reported recombination frequencies >20% – a higher rate of chimeric products apparently being formed from longer templates [BioTechniques (2006) 40(4):499–507].

(2) In plant virology: a form of genetic change that occurs in viruses with bipartite or multipartite genomes; it is due to an exchange of (whole) sections of the genome (rather than to crossing-over).

pseudotype (*virol.*) A virion, formed by PHENOTYPIC MIXING, which contains the genome of one virus and component(s) of a co-infecting virus.

pseudotyping (*DNA technol.*) The genetic modification of a given virus resulting in replacement of surface components of the virion with those of a different virus. This is done in order to modify the range of receptors, or types of cell, to which the given virus can attach.

[Viral envelope pseudotyping: PLoS Pathog (2009) 5(4): e1000377.]

[Production of SARS S-pseudotyped murine leukemia virus (MLV) particles: Proc Natl Acad Sci USA (2009) 106(14): 5871–5876.]

(See also VSV.)

Re-targeting of adenovirus type 5 vectors to $\alpha_v\beta_6$ INTEGRIN cell-surface receptors was reported to be associated with lower hepatotoxicity and greatly improved uptake by tumor cells on systemic delivery [J Virol (2009) 83(13):6416–6428].

The concept has been broadened by using *non*-viral surface components in order to target the virion to a particular ligand. For example, the surface of a herpesvirus was modified so as to display a SINGLE-CHAIN VARIABLE FRAGMENT – targeting the virion to $CD38^+$ leukemic cells, in culture, and to the cells of $EGFR^+$ tumors in mice [Gene Ther (2008) 15:1579–1592].

pseudovirion A particle that consists of a viral capsid which encloses nucleic acid from the host cell.

ψ (psi) (1) A sequence in a retroviral genome that promotes the packaging of viral RNA during the formation of virions.

(2) A symbol for pseudouridine (5-β-D-ribofuranosyluracil).

psicofuranine (9-β-D-psicofuranosyladenine) A compound that has antimicrobial and antitumor activity; it is synthesized by the (Gram-positive) bacterium *Streptomyces hygroscopicus*. This agent inhibits biosynthesis of guanosine monophosphate by inhibiting xanthosine 5′-monophosphate (XMP) aminase.

Decoyinine has a similar action.

PSK POST-SEGREGATIONAL KILLING.

pSM19035 A PLASMID, found e.g. in strains of *Streptococcus pyogenes*, that exhibits a high level of segregational stability resulting from the combined activity of several mechanisms for controlling COPY NUMBER – e.g. resolution of multimers (promoting segregation of monomers) and post-segregational killing in which plasmid-free cells are killed. A further, novel system has been reported for plasmid segregation which is encoded by genes δ and ω [J Bacteriol (2006) 188(12):4362–4372].

psoralens (furocoumarins) Various (tricyclic) INTERCALATING AGENTS (present e.g. in certain fungi, and in tropical fruits) which can insert into dsDNA and dsRNA. After intercalation, irradiation (e.g. ∼365 nm) causes the formation of covalent inter-strand cross-links between pyrimidine bases.

Targeted, site-specific cross-linking with a psoralen may be achieved e.g. by a triplex-directed method; in this approach, a psoralen is linked to a so-called triplex-forming oligonucleotide (TFO: see the entry TRIPLEX DNA) which is targeted to the required sequence.

*Iso*psoralens are used e.g. in AMPLICON INACTIVATION.

pSos vector See CYTOTRAP TWO-HYBRID SYSTEM.

PspGI A highly thermostable RESTRICTION ENDONUCLEASE from a strain of *Pyrococcus* (growth temperature 95°C); this (type IIP) enzyme has a recognition sequence 5′-CCWGG-3′ and was used e.g. in a method for signal generation in real-time PCR: fluorescent amplicon generation (see FLAG).

pT-REx^{TM}-DEST31 vector A destination vector in the GATEWAY SITE-SPECIFIC RECOMBINATION SYSTEM (see the table in that entry).

pTcINDEX See EXPRESSION VECTOR.

PTGS Post-transcriptional GENE SILENCING: in animals, plants, fungi and bacteria, any of various types of phenomenon in which the transcript of a gene is either blocked or degraded. (The term *co-suppression* has been used to refer to PTGS in plants.)

One of these PTGS phenomena involves endogenous, genomically encoded molecules of RNA: see MICRORNAS and SRNAS. In another type of process, effector molecules are formed e.g. in response to intracellular double-stranded RNA (dsRNA) of a certain minimum length: see the entry RNA INTERFERENCE. In eukaryotes, processing of the precursors of miRNAs and siRNAs involves similar types of enzyme (see e.g. DICER).

As well as involvement in PTGS, siRNAs are also involved in *transcriptional* gene silencing (TGS): see the entry SIRNA.

PTS Phosphoenolpyruvate-dependent phosphotransferase system: a transport system involved in transmembrane uptake of various substrates by bacteria; some of the components of the PTS are also involved in the regulation of (certain) catabolic OPERONS, including the LAC OPERON of *Escherichia coli* (see also CATABOLITE REPRESSION).

In all cases, the energy (for transport and phosphorylation) in PTS is derived from phosphoenolpyruvate (PEP).

The following descriptions of two examples of PTS-based transport in *E. coli* are used to illustrate various aspects of PTS involvement in the regulation of a number of catabolic operons (i.e. operons which specify enzymes etc. involved in

the catabolism (breakdown) of particular types of substrate).

In each case, energy is fed into the system by the sequential transfer of phosphate, from the PEP, to two soluble (cytoplasmic) energy-coupling proteins: *enzyme I* (designated I) and *HPr protein* (designated H). Thus, phosphate from PEP is initially transferred to I and then from I to H. From H, phosphate is transferred to a *permease* (= *enzyme II*, designated II) in order to provide the energy for uptake of a given substrate. A cell contains a number of different permeases – a particular substrate being taken up by a particular permease. Note that the molecules of I and H are not specific to any particular permease, i.e. they can be used for the phosphorylation of any of the cell's permeases.

The *mannitol* permease of *E. coli* is a membrane-associated complex consisting of three domains: IIA, IIB and IIC. The IIC domain seems to contribute to a transmembrane channel for uptake of the substrate. It appears that the IIA domain is phosphorylated by H and then transfers phosphate to the IIB domain, at which site the substrate (mannitol) is phosphorylated prior to entering the cytoplasm; phosphorylation of the substrate is an essential part of the uptake process in all types of PTS.

The *glucose* permease of *E. coli* consists of (i) a membrane-bound complex of domains IIC and IIB and (ii) domain IIA on a separate, *cytoplasmic* protein. In the *presence* of glucose IIA is phosphorylated by H, and the phosphate is transferred from IIA to IIB – and thence to the substrate (glucose). In the *absence* of glucose, IIA remains in a phosphorylated state and activates the enzyme adenylate cyclase, stimulating the synthesis of cyclic AMP (cAMP). cAMP binds to the cAMP-receptor protein (CRP) – and the resulting complex permits transcription of various catabolic operons in the presence of their respective inducers (see CATABOLITE REPRESSION); for example, in the absence of glucose and in the *presence* of lactose, the cAMP-bound CRP protein complex permits transcription of the *lac* operon (and, hence, utilization of lactose).

Components are arranged in different ways in some of the permeases. Thus, for example, in enteric bacteria the fructose permease consists of a membrane-bound protein that includes the IIC and IIB domains together with a cytoplasmic protein which incorporates the functions of both IIA and H. In the (non-enteric) species *Rhodobacter capsulatus*, the functions of IIA, H and I are combined in a single, separate protein.

Notice that, in catabolite repression, the IIA component of the glucose permease has a regulatory role in the typical PTS-mediated mechanism in the Gram-negative bacteria, whereas the IIB and HPr∼P components have a regulatory role in PRD-mediated mechanisms.

[Genetic dissection of specificity determinants in the interaction between HPr and enzymes II: J Bacteriol (2007) 189 (13):4603–4613.]

PTS regulation domain (PRD) See CATABOLITE REPRESSION.

pUC plasmids A family of small high-copy-number plasmids, derivatives of which are used e.g. as CLONING VECTORS; they contain multiple restriction sites (for inserting target DNA), a sequence from the *lacZ* gene (encoding the α-PEPTIDE), and an ampicillin-resistance gene. When these plasmids replicate in suitable host cells, the presence of target DNA (as an insert within the plasmid) can be detected by blue–white screening (see PBLUESCRIPT for the mechanism).

pBluescript® is a derivative of pUC19.

pulsed-field gel electrophoresis See PFGE.

PureLink™ plasmid purification systems Systems (Invitrogen, Carlsbad CA, USA) that are used for the rapid isolation of plasmids from bacteria (in about 30 minutes).

The apparatus consists of rows of cups in a 'filter plate' that are aligned, vertically, over cups in a 'receiver plate'; cups in the receiver plate contain isopropanol.

Bacterial (liquid) culture is added to the cups in the filter plate; lysis buffer is then added to each cup and incubation is carried out for 10 minutes. The whole is then centrifuged for 15 minutes at 3000 *g* – during which plasmids (released from the cells) pass through the filter plate into the corresponding cups in the receiver plate.

The filter plate is discarded, and DNA pellets, within cups in the receiver plate, are washed with 70% ethanol.

After discarding the supernatant the pellet is air-dried for 5 minutes and then suspended in buffer.

puromycin A nucleoside antibiotic (6-dimethyl-3′-deoxy-3′-*p*-methoxyphenylalanyl-amino adenosine) which blocks protein synthesis by entering the ribosomal A site and forming a covalent bond with peptidyl-tRNA in the P site, thereby causing premature chain termination.

Puromycin is an effective antibiotic against prokaryotic and eukaryotic cells.

[Examples of use of puromycin as a selective agent: PLoS Biol (2006) 4(10):e309; Retrovirology (2006) 3:51; Proc Natl Acad Sci USA (2009) 106(23):9373–9378; and J Clin Invest (2009) 119(7):2100–2112.]

pVAX™200-DEST vector A destination vector in the GATEWAY SITE-SPECIFIC RECOMBINATION SYSTEM (see the table in that entry).

pVAX vectors have been used e.g. to study DNA vaccines against tuberculosis [Genet Vaccines Ther (2009) 7:11, Clin Vaccine Immunol (2009) 16(1):122–126].

PVDF Polyvinylidene fluoride: a compound used for making membranes – an alternative to nitrocellulose. Membranes of PVDF have greater mechanical strength (as compared with nitrocellulose), and they are reported to have better protein-retention characteristics under chemically harsh conditions.

The FluoroTrans® PVDF membrane (Pall Life Sciences, Ann Arbor MI, USA) has been used e.g. in studies on gastric carcinoma [J Natl Cancer Inst (2009) 101(8):592–604], and on entomopoxvirus infection [J Virol (2008) 82(24):12406–12415].

PVL PANTON–VALENTINE LEUCOCIDIN (q.v.).

pXO1, pXO2 Plasmids found in virulent strains of BACILLUS ANTHRACIS (causal agent of anthrax); they encode anthrax toxin (pXO1) and the poly-D-glutamic acid bacterial cell cap-

sule (pXO2), both of which are required for virulence of the pathogen.

Py–MS See PYROLYSIS.

pYES-DEST52 vector A destination vector in the GATEWAY SITE-SPECIFIC RECOMBINATION SYSTEM (see the table in that entry).

This vector has been used for cloning [PLoS ONE (2009) 4(4):e5154] and for expression [BMC Microbiol (2009) 9:9].

pyknosis (*histopathol.*) The shrinkage of a eukaryotic nucleus, forming a densely staining body.

(See also KARYOLYSIS.)

Pyr(6–4)Pyo See PYRIMIDINE DIMER.

pyrazinamidase See PNCA GENE.

***pyrC* gene** (in *Escherichia coli*) See the entry TRANSLATIONAL ATTENUATION (sense 1).

pyrene binary probe A BINARY PROBE in which both of the oligonucleotides are labeled with a pyrene molecule such that these molecules are juxtaposed (and thus capable of forming an excited dimer – an *excimer*) when the probe is correctly hybridized to its target sequence. After a pulse of excitation, the lifetime of the fluorescence from the pyrene excimer is considerably longer than that of the background fluorescence which characterizes some cellular extracts; thus, this permits the detection of fluorescence specifically from the excimer by using a *time-resolved emission spectra* (TRES) approach – in which the measurement of fluorescence begins only *after* the short-term background fluorescence has decayed.

With TRES, pyrene binary probes have been useful e.g. for the detection of specific mRNAs [Nucleic Acids Res (2006) 34 (10):3161–3168].

Pyrene fluorescence has also been exploited for monitoring the folding of RNA molecules [Nucleic Acids Res (2006) 34 (1):152–166].

In one approach to probing nucleic acids, pyrene monomers were incorporated in the 5′ end of the stem in a MOLECULAR BEACON PROBE, and a quencher (DABCYL) was incorporated in the 3′ end; on binding of the molecular beacon probe to its target sequence the stem opened (separating pyrenes from the quencher) and, on excitation, generated an excimer emission signal [J Am Chem Soc (2008) 130(1):336–342].

***pyrF* gene** (in *Escherichia coli*) A homolog of the *URA3* gene in *Saccharomyces cerevisiae*.

pyrimidine dimer Two adjacent pyrimidine residues, in the same strand of DNA, which have become covalently cross-linked e.g. as a result of ULTRAVIOLET RADIATION; one example is the THYMINE DIMER (q.v.).

Cyclobutyl pyrimidine dimers (CPDs) may be repaired by PHOTOLYASE, but this enzyme does not repair certain other products of ultraviolet radiation: the so-called (6–4) photoproducts (also referred to as Pyr(6–4)Pyo). A (6–4) photoproduct may form between a thymine residue and its adjacent (3′) cytosine residue – or between adjacent cytosine residues – by the covalent linking of the 6-position and the 4-position of the two pyrimidines, respectively; (6–4) photoproducts can be corrected e.g. by the EXCISION REPAIR system.

Pyrimidine dimers can block DNA replication. They can be detected e.g. by TERMINAL TRANSFERASE-DEPENDENT PCR.

Pyrococcus A genus of thermophilic archaeans (see ARCHAEA). The species *P. furiosus* is the source of certain thermostable DNA polymerases used e.g. in PCR.

(See also FLAG.)

pyrogen Any of various, potentially lethal substances which, if introduced into the bloodstream, can affect body temperature; pyrogens typically induce fever if present at an appropriate concentration.

Recombinant proteins which are synthesized within Gram-negative bacteria (such as *Escherichia coli*) must be rendered free of the potent pyrogen LIPOPOLYSACCHARIDE (LPS), also called *endotoxin*, which forms a (normal) part of the Gram-negative cell envelope – and which can therefore be shed into the product.

LPS can be detected e.g. by the LAL TEST (q.v.).

During the manufacture of recombinant proteins in *E. coli*, LPS can be removed from the product by appropriate downstream processing involving an efficient form of chromatographic separation, such as ion-exchange or gel filtration.

pyrogram See PYROSEQUENCING.

pyrolysis Thermal degradation of a sample (*in vacuo* or in an inert gas) forming a range of (low-molecular-weight) compounds which can be analysed. Pyrolysis has been used e.g. in microbial taxonomy – the sample being a small quantity of microbial culture; the products of pyrolysis (the *pyrolysate*) are taken to be characteristic of the organism being studied – and they may be analyzed by pyrolysis–mass spectrometry (Py–MS).

A *filament* pyrolyser is one in which a platinum wire, coated with the sample, is heated by passing an electric current through it (cf. CURIE POINT PYROLYSIS). Pyrolysis can also be carried out in a laser-based instrument.

pyrophosphate Inorganic pyrophosphate (PP_i) or diphosphate; the ion $P_2O_7^{4-}$.

Pyrophosphate is released by nucleoside triphosphates e.g. when they are incorporated (by a DNA or RNA polymerase) into a growing strand of nucleic acid.

Pyrophosphate can be monitored by techniques such as the ENZCHEK PYROPHOSPHATE ASSAY and the PIPER PYROPHOSPHATE ASSAY KIT.

Pyrophosphate is synthesized e.g. within animal and yeast mitochondria, and in some photosynthetic bacteria; synthesis involves a proton PPase.

(See also PYROPHOSPHOROLYSIS-ACTIVATED POLYMERIZATION and PYROSEQUENCING.)

pyrophosphate assay See the entries ENZCHEK PYROPHOSPHATE ASSAY and PIPER PYROPHOSPHATE ASSAY KIT.

pyrophosphorolysis-activated polymerization (PAP) A form of nucleic acid amplification, involving temperature cycling, in which the extension of each primer is initially blocked by its 3′ terminal dideoxyribonucleotide (ddNMP); on annealing *correctly* to the template, in the presence of pyrophosphate, the ddNMP is excised (forming ddNTP) – so that the primer

can be extended by a DNA polymerase. Primers are used in pairs (as in PCR), and cycling involves temperatures suitable for denaturation, annealing and extension. PAP is reported to be characterized by high-level specificity.

PAP has been used e.g. for allele-specific amplification. A multiplex form of PAP was used to detect deletions in the human gene for factor IX [BioTechniques (2006) 40(5): 661–668].

PAP has been used in a modified form of ligation-mediated PCR in order to improve the analysis of so-called 'difficult' sequences by increasing specificity at primer extension stage [Nucleic Acids Res (2008) 36(3):e19].

PAP has also been employed in microarray-based re-sequencing studies [Anal Biochem (2008) 374(1):41–47].

Pyrosequencing™ A technique for rapidly sequencing short, single-stranded fragments of DNA produced, for example, by PCR (Pyrosequencing, Uppsala, Sweden).

Pyrosequencing is exploited e.g. in the 'massively parallel' sequencing method GS FLX.

The following is an account of the basic technique.

Target fragments are first immobilized on a solid surface.

Initially, primers bind to the immobilized target fragments. The reaction mixture includes (i) DNA polymerase, (ii) ATP sulfurylase, (iii) adenosine 5′-phosphosulfate (APS), (iv) firefly LUCIFERASE, (v) luciferin, the substrate for luciferase, and (vi) *apyrase*, an enzyme that degrades nucleotides.

The four types of deoxyribonucleoside triphosphate (dNTP) are added – separately and sequentially – to the reaction mixture at intervals of about 1 minute; this addition is repeated cyclically. Thus, e.g. the addition of:

[. . ..A. . ..T. . ..G. . ..C. . ..A. . ..T. . ..G. . ..C. . .A]

takes about 9 minutes.

If a given, added nucleotide is complementary to the next free (unpaired) base in the (single-stranded) template (target), it will extend the primer strand *and release pyrophosphate*.

Pyrophosphate is used by the enzyme ATP sulfurylase to convert adenosine 5′-phosphosulfate (APS) to adenosine 5′-triphosphate (ATP); this ATP then triggers the luciferase system, producing a burst of light (of intensity proportional to the amount of ATP). The light is recorded automatically as a spike on a time-based graph.

Note. If the *normal* nucleotide dATP (deoxyadenosine triphosphate) were used in this technique, the (ATP-dependent) luciferase system would generate light each time dATP was added to the reaction mixture – whether or not the nucleotide had been incorporated into the primer strand. Hence, instead of dATP, use is made of deoxyadenosine α-thiotriphosphate (dATPαS); this modified nucleotide is used as a substrate by DNA polymerase *but not* by luciferase.

If a particular (added) nucleotide is complementary to the next *two* free bases in the template strand, it will extend the primer by two nucleotides, releasing a correspondingly larger amount of pyrophosphate; hence, the number of consecutive sites into which a given nucleotide is incorporated is signaled by the amount of light generated (shown by the height of the spike on the graph).

Apyrase continually degrades nucleotides in the reaction mixture. This enzyme acts on (i) each nucleotide added to the mixture, and (ii) the ATP derived from pyrophosphate.

If a given, added nucleotide is *not* complementary to the next free base, its degradation (by apyrase) will be complete by the time the next dNTP is due to be added; in this case, no light is generated because the nucleotide does not extend the primer and (therefore) no pyrophosphate is formed.

Thus, generation of light, or an absence of light generation, at a given time, can be linked to a particular (added) dNTP – and, hence, to a particular nucleotide in the template strand.

The sequential addition of nucleotides to the primer strand is recorded as a series of spikes on the time-based graph (the *pyrogram*). Because a given spike of light (at a specific time) corresponds to the addition of a particular, known nucleotide, the final result (the series of spikes) indicates the sequence of nucleotides in the template.

A problem in the interpretation of results may arise if the target sequence contains one or more homopolymer stretches (such as TTTTT. . ...); this problem is due to the non-linear generation of light when more than about five nucleotides *of the same type* are incorporated consecutively.

With allele-specific primers, Pyrosequencing has been used in a method for determining the allele-specific methylation in samples of bisulfite-treated, PCR-amplified DNA [BioTechniques (2006) 41(6):734–739].

Pyrosequencing has been used e.g.: for subtyping *Helicobacter pylori* [FEMS Microbiol Lett (2001) 199:103–107]; for forensic analysis of mitochondria [BioTechniques (2002) 32:124–133]; for identifying bacterial contamination [FEMS Microbiol Lett (2003) 219:87–91]; for typing *Neisseria gonorrhoeae* (the *porB* gene) [J Clin Microbiol (2004) 42:2926–2934]; for detecting lamivudine resistance in hepatitis B virus (YMDD motif mutations) [J Clin Microbiol (2004) 42:4788–4795]; for studying polymorphism in a malaria vaccine antigen [PLoS Med (2007) 4(3):e93]; for studies on phylogeny of the Y chromosome PLoS ONE (2009) 4(6):e5792]; for the genomic analysis of *Nosema ceranae* [PLoS Pathog (2009) 5(6):e1000466]; for the analysis of wound bacteria using 16S rRNA gene-based pyrosequencing [PLoS ONE (2009) 4(7): e6462]; for studies on maize transcriptome [BMC Genomics (2009) 10:299]; and for analysis of splicing patterns [Nucleic Acids Res (2009) doi: 10.1093/nar/gkp626].

Q

q An indicator of the *long* arm of a CHROMOSOME; it is used when giving a particular location on a specific chromosome – e.g. 9q14, which refers to a location (14) on the long arm of chromosome 9.

Q L-Glutamine (alternative to Gln).

Q bases Certain modified forms of guanine found in molecules of tRNA. QUEUOSINE is one example.

(cf. Y BASES.)

Q-PNA Quencher-labeled PNA probe: see SELF-REPORTING PRIMER.

Q-tag (1) A 45-amino-acid part of the enzyme β-galactosidase. Q-tags are used for the quantitative estimation of a (soluble) recombinant protein expressed from a vector. In the vector, the sequence encoding the protein and that encoding the Q-tag form a fusion gene. When expressed, the protein – which is Q-tagged at the C-terminal or the N-terminal – can interact with a non-functional form of β-galactosidase; complementation between the Q-tag and non-functional β-galactosidase produces a functional enzyme which permits quantitation of the given protein with a chemiluminescence-generating substrate. This system is marketed by Stratagene (La Jolla, CA, USA) under the trade name VariFlex™.

(See also SOLUBILITY ENHANCEMENT TAG.)

(2) See PROTEIN LABELING.

QD QUANTUM DOT.

qdot QUANTUM DOT.

QEXT (quencher extension) A method, involving the FRET principle, which has been used e.g. for SNP detection; in the presence of an SNP, a probe carrying a 5′ reporter dye undergoes SINGLE-BASE EXTENSION with a TAMRA-labeled chain-terminating dideoxyribonucleotide – causing the reporter dye to either fluoresce or be quenched (according to the particular reporter used).

Multiplex QEXT uses different reporter dyes for detecting different SNPs [BioTechniques (2006) 40(3):323–329].

QIAamp® A system (QIAGEN GmbH, Hilden, Germany) that is used for rapid isolation and purification of nucleic acids (DNA or RNA) from various specimens.

Cell lysis is achieved by the use of optimized buffers, and the required nucleic acid (DNA or RNA) is removed from the lysate by adsorption to a specialized membrane in a spin column. After spin-washing, the given nucleic acid is eluted from the membrane into a collecting tube.

Separate kits are available for different types of sample. For example, one kit is designed for the extraction of DNA from blood, buffy coat or body fluids etc., while another kit is used for isolating pure viral RNA from samples of plasma containing hepatitis C virus.

QIAGEN® plasmid mini kit See DNA ISOLATION.

qiRNA See the entry ARNA (sense 2).

***qnr* genes** Plasmid-borne genes, present e.g. in some bacteria of the family Enterobacteriaceae, which encode resistance to the fluoroquinolone antibiotics (see QUINOLONE ANTIBIOTICS).

QPCR (qPCR) QUANTITATIVE PCR (q.v.).

QQI Quadruplex–quadruplex interaction – see QUADRUPLEX DNA.

Qrr1–Qrr4 Bacterial sRNAs: see SRNAS (*Prokaryotic sRNAs*); these sRNAs are involved in regulating QUORUM SENSING in *Vibrio cholerae*.

QSY dyes See FRET.

QT-NASBA Quantitative NASBA.

quadruplex DNA (G-quadruplex) Four, G-rich, single-stranded DNA sequences associated to form a structure whose core is a stack of parallel G-QUARTETS [see Figure 1, Nucleic Acids Res (2003) 31(8):2097–2107].

In vitro, quadruplex DNA may be formed by certain G-rich sequences such as those found in the (eukaryotic) TELOMERE region, and those involved in the immunoglobulin switching mechanism.

It was suggested that, when present in the promoter region of a certain oncogene, quadruplex DNA inhibits transcription and may exist in equilibrium with duplex DNA (which may favor transcription) [Nucleic Acids Res (2006) 34(9):2536–2549]. Quadruplex DNA has since been found in promoters throughout the human genome [Nucleic Acids Res (2007) 35 (2):406–413]. In telomeric DNA, isoflavones were reported to regulate competition between quadruplex DNA and duplex DNA [Nucleic Acids Res (2009) 37(8):2471–2482].

[The direct visualization of G-quadruplexes in DNA using atomic force microscopy: Nucleic Acids Res (2009) doi: 10.1093/nar/gkp679.]

Experimental evidence has been presented for interaction between two contiguous G-quadruplexes in the insulin-linked polymorphic region (ILPR) [Nucleic Acids Res (2009) doi: 10.1093/nar/gkp181].

Structures similar to those found in DNA may be formed by RNA. [Database of quadruplex-forming G-rich sequences in alternatively processed mammalian pre-mRNAs: Nucleic Acids Res (2006) 34(Database issue):D119–D124.]

Quantiplex™ assay See BDNA ASSAY.

quantitative PCR (QPCR, qPCR) Any form of PCR which is used to determine either the absolute or the relative number of copies of a given target sequence in a sample.

Absolute numbers of a given sequence may be assessed by REAL-TIME PCR (real-time RT-PCR for RNA targets) or e.g. by COMPETITIVE RT-PCR.

Relative numbers of a given target sequence in two or more samples may need to be assessed e.g. when comparing the expression of a given gene in different tissues or in different individuals; such quantitation may be achieved e.g. by using RELATIVE RT-PCR.

[Twentyfive years of quantitative PCR for gene expression

analysis (review of the use of qPCR for quantitation of RNA levels): BioTechniques (2008) 44(5):619–626.]

quantum dot (QD; qdot) A NANOPARTICLE composed of semiconductor materials that exhibits fluorescence on suitable excitation.

The wavelength of the fluorescence emission peak becomes longer as the size of the QD increases, but QDs of different size can be excited with a single source of radiation.

Fluorescence from quantum dots is characterized by a large STOKES SHIFT.

The properties of QDs, together with the narrow emission spectrum of a given QD, provide opportunities for a range of high-resolution *in vitro* and *in vivo* reporter systems (see e.g. QUANTUM DOT PROBE).

(See also IN VIVO FLUORESCENCE.)

[Quantum dots: characterization, preparation, usage in biological systems: Int J Mol Sci (2009) 10(2):656–673.]

quantum dot probe A PROBE labeled with a QUANTUM DOT (q.v.).

In one approach, a quantum dot probe was developed for use in Western blot analysis. The probe consisted essentially of a dimerized domain (based on the Fc-binding portion of PROTEIN A) carrying a BIOTINylated peptide sequence which bound a STREPTAVIDIN-coated QD. Such a probe permits a QD (with specified characteristics) to be conjugated (via the Fc-binding moiety of the probe) to the Fc portion of any of a range of IgG antibodies. Thus, a QD, with specified emission characteristics, can be conjugated to a particular monoclonal antibody which is highly specific for a given target antigen – while a different QD, with different emission characteristics, can be conjugated to a different antibody that has a different target antigen; on a Western blot, these QDs would bind to separate targets, and each target would be identified with a distinct label – even though both QDs were excited from a single source of radiation. This, therefore, permits detection of multiple proteins on a Western blot.

[Quantum dots for live cell and *in vivo* imaging (a review): Int J Mol Sci (2009) 10(2):441–491.]

quartet (G-quartet) See G-QUARTET.

quasispecies A population of (e.g.) a given RNA virus which, due to normal, ongoing replication errors, consists of genetically heterogeneous individuals; the range of genomes present in the population will depend e.g. on the effects of selection.

quelling A form of post-transcriptional gene silencing (PTGS) in fungi.

quencher extension See QEXT.

query sequence Any sequence (e.g. a sequence of nucleotides) which is to be compared with other sequence(s) on a database (such as BLAST) e.g. for identification purposes or to assist in the annotation of an unknown gene.

queuosine A Q BASE (7-[4,5-*cis*-dihydroxyl-1-cyclopentene-3-aminomethyl]-7-deazaguanosine) that is found in the wobble position (see WOBBLE HYPOTHESIS) of tRNAs which carry aspartic acid, asparagine, histidine and tyrosine; this modified base is found in both eukaryotic and prokaryotic tRNAs, but apparently does not occur e.g. in *Saccharomyces cerevisiae*.

Queuosine pairs with either cytosine or uracil residues.

QuikChange™ site-directed mutagenesis kit A kit (Stratagene, La Jolla CA, USA) which is used for SITE-DIRECTED MUTAGENESIS. An outline of the procedure is given below.

A small (<8 kb) plasmid, carrying the target DNA, is heat-denatured to expose primer-binding sites. Two primers, each containing the required mutation, are allowed to anneal – one on each strand – to target DNA; the primers bind in slightly staggered positions on the target sequence. These primers are then extended by a (thermostable) DNA polymerase, forming dsDNA copies of the plasmid with the required mutation and staggered nicks.

Ongoing thermal cycling forms many copies of the mutant plasmid (which remain unligated).

Parent template DNA is then degraded with the restriction enzyme DpnI. The parent DNA is susceptible to this (methylation-*dependent*) RESTRICTION ENDONUCLEASE (q.v.) because it was produced in *Escherichia coli* and will (therefore) have undergone *dam* methylation (see DAM GENE).

Because the mutant plasmids were formed *in vitro* they are non-methylated and are therefore resistant to DpnI.

The mutant strands of the plasmid hybridize to form mutant dsDNA plasmids. Each plasmid is nicked in both strands but, because the nicks are in staggered positions, stable plasmids are formed. The nicks are repaired *in vivo* when the mutant plasmids are inserted into *E. coli* for replication.

Other versions of QuikChange™ are used e.g. to introduce mutations at multiple sites.

[Uses (e.g.): PLoS ONE (2009) 4(1):e4259; PLoS Pathog (2009) doi: 10.1371/journal.ppat.1000273; and Retrovirology (2009) 6:11.]

quinacrine 6-Chloro-9-(4-diethylamino-1-methylbutylamino)-2-methoxyacridine: a yellow fluorescent compound that has been used e.g. for staining the DNA in chromosomes.

quinolone antibiotics Synthetic antibiotics (based on a substituted 4-quinolone ring) which target certain bacterial enzymes involved in the replication of DNA: the A subunit of gyrase, and topoisomerase IV. (cf. NOVOBIOCIN.)

Bacterial resistance to quinolone antibiotics often derives from mutation in the *gyrA* gene (gyrase subunit A), although resistance may also result e.g. from lowered permeability of the cell envelope and/or efflux mechanisms.

The early quinolone antibiotics, which were active mainly against Gram-negative bacteria (but not against e.g. *Pseudomonas aeruginosa*), included cinoxacin, nalidixic acid, oxolinic acid and pipemidic acid.

*Fluoro*quinolones such as ciprofloxacin, enoxacin, norfloxacin, ofloxacin and perfloxacin were developed later; their range of activity included *Pseudomonas aeruginosa* as well as certain Gram-positive bacteria.

A subsequent generation of the fluoroquinolones included e.g. gatifloxacin, gemifloxacin, grepafloxacin, levofloxacin, moxifloxacin and trovafloxacin.

quinomycins See QUINOXALINE ANTIBIOTICS.

quinoxaline antibiotics A category of bifunctional INTERCALATING AGENTS which include quinomycins and triostins; the molecule is a cyclic octapeptide dilactone linking two quinoxaline 2-carboxylic acid chromophores. Some of these antibiotics exhibit a preference for GC-rich sequences, while a synthetic member, TANDEM (des-*N*-tetramethyltriostin A), shows a preference for AT-rich regions.

Quinoxalines have antimicrobial and antitumor activity.

quorum sensing In certain microorganisms: the phenomenon in which a particular characteristic is expressed by cells only when their concentration (number of cells per unit volume) is higher than a certain minimum. Hence, cells in a high-density population may exhibit characteristics that are not exhibited when *cells of the same type* are in low-density populations.

The archetypal presentation of quorum sensing is the BIOLUMINESCENCE exhibited by high-density populations of the bacterium *Vibrio fischeri* (= *Photobacterium fischeri*) within the light-emitting organs of certain fishes; in the free-living state (sea water: low-density populations) cells of *V. fischeri* exhibit little or no bioluminescence.

Quorum sensing involves a gene-regulatory mechanism. It is mediated by certain *secreted* low-molecular-weight molecules; in a high-density population of the secreting cells these molecules reach a threshold concentration which is sufficient to trigger activity in particular gene(s) within the cells via a signal transduction mechanism. These low-molecular-weight signaling molecules are referred to as *autoinducer* molecules because they are synthesized by the cells themselves.

The autoinducer in many Gram-negative bacteria is an *N*-acyl-L-homoserine lactone (AHL). (In the case of *V. fischeri*, an AHL triggers activity in the *lux* operon.)

[Regulation of AHLs in *Agrobacterium vitis*: J Bacteriol (2006) 188(6):2173–2183.]

For Gram-positive bacteria, signaling molecules in quorum sensing are generally *peptides* (*pheromones*). The pheromone ComX, formed by *Bacillus subtilis*, promotes competence in TRANSFORMATION by activating a certain TWO-COMPONENT REGULATORY SYSTEM, which activates a transcription factor needed for competence.

In the fungus *Candida albicans*, farnesol, a sesquiterpene alcohol, was reported to act as a quorum sensing molecule and to inhibit yeast-to-mycelium transition in this organism. [Quorum sensing in fungi: Eukaryotic Cell (2006) 5(4):613–619; Appl Environ Microbiol (2006) 72:3805–3813.]

Quorum sensing is reported to regulate the expression of a wide range of characteristics in many organisms – including, e.g. the formation of biofilms in *Staphylococcus aureus*, the expression of type III protein secretion systems in EHEC and EPEC strains of *Escherichia coli*, swarming in *Burkholderia cepacia*, as well as development of competence in *Bacillus subtilis* (see above). (See also CROWN GALL.)

A mutant, autoinducer-negative strain of *Chromobacterium violaceum* (CV026) produces the purple pigment *violacein* (a derivative of tryptophan) when exposed to any of a variety of exogenous inducers, including all tested AHLs – as well as AHT (*N*-acyl-homocysteine thiolactone) having *N*-acyl side-chains in the range C_4–C_8. This strain is therefore useful as a biosensor for detecting the relevant range of autoinducers.

Quorum sensing has been studied by the RIVET procedure – see the entry IVET.

R

R (1) A specific indicator of ambiguity in the recognition site of a RESTRICTION ENDONUCLEASE (or in any other nucleic acid sequence); for example, in GTY↓RAC (enzyme HincII) the 'R' indicates A or G. In the example, 'Y' is C or T.
(2) L-Arginine (alternative to Arg).

R loop See D LOOP.

R plasmid A plasmid encoding resistance to antibiotics etc.

R–M system See MODIFICATION.

R1 plasmid A low-copy-number conjugative PLASMID, found in strains of *Escherichia coli* and other enterobacteria, which encodes resistance to e.g. ampicillin and streptomycin.

Replication of R1 DNA is controlled by the synthesis of the (plasmid-encoded) RepA protein which promotes replication of the plasmid from its origin.

RepA synthesis appears to be regulated at several different levels. When the plasmid COPY NUMBER is normal, the CopB protein (encoded by the plasmid) inhibits transcription from the *repA* gene promoter. Moreover, under these conditions, an ANTISENSE RNA molecule, CopA (also plasmid-encoded), inhibits translation of the mRNA of gene *repA*; the inhibitory action of CopA becomes less effective when copy number is low. When the copy number falls, it appears that the lower concentration of CopB exerts less inhibitory action on *repA* transcription, allowing the plasmid's copy number to recover to normal levels.

A further mechanism for maintaining a stable copy number of plasmid R1 involves the Kid protein, a component of the *parD* system. When the copy number falls, Kid, acting as a nuclease, cleaves both host-encoded mRNA (e.g. *dnaB*) and a plasmid-encoded mRNA (which encodes a *repressor* of R1 replication) at the specific sequence UUACU; the effects of such cleavage are to inhibit bacterial growth (thus delaying cell division) and to help restore the normal copy number of the plasmid [EMBO J (2005) 24(19):3459–3469].

Partition (that is, segregation of plasmids to daughter cells during cell division) appears to involve a remarkable mitosis-like process mediated by actin-like filaments. In this process, it appears that a pair of plasmids are linked by the (plasmid-encoded) ParR protein – which binds to the 'centromere-like' region *parC*, present in each of the plasmids, to form a so-called *partitioning complex*. (ATP-bound) ParM molecules then polymerize, forming actin-like filaments which – as they lengthen in an axial direction between the two plasmids – concurrently propel the plasmids to opposite poles of the cell. The filaments subsequently de-polymerize following hydrolysis of filament-bound ATP.

[Structural analysis of the ParR/*parC* complex: EMBO J (2007) 26(20):4413–4422.]

Proteins encoded by some other plasmids were reported to polymerize into filaments *in vitro* – e.g. the F plasmid SopA protein and a protein encoded by plasmid TP228.

R5 strains (of HIV-1) See the entry HIV-1.

R5X4 strains (of HIV-1) See the entry HIV-1.

rAAV Recombinant adeno-associated virus.

RACE Rapid amplification of cDNA ends: an approach used (i) to copy the 5′ end and/or 3′ end of an mRNA, and/or (ii) to prepare full-length copies of the corresponding cDNA.

In the 3′ RACE procedure, an oligo(d)T primer binds to the 3′ poly(A) tail of the mRNA, and this primer is extended by reverse transcriptase. After degradation of the RNA template (e.g. with RNase H), a gene-specific primer (complementary to a sequence *within* the FIRST STRAND) may be extended to form the second strand. Then, a nested PCR may be used to amplify smaller sections of the double-stranded cDNA. This approach may be improved e.g. by adding a 5′ tag (see entry TAG, sense 2) to the oligo(d)T primer – which can provide a more sequence-specific binding site for the primers during subsequent amplification. [Example of use of 3′ RACE: Mol Biol Cell (2009) 20(1):464–480.]

In the 5′ RACE procedure, a gene-specific primer (i.e. one complementary to a sequence within the mRNA) is extended (with reverse transcriptase) to the 5′ terminus of the mRNA template. The mRNA is then degraded (e.g. with the enzyme RNase H), and the 3′ end of the first strand is tailed with e.g. dCTP. This permits the design of a second-strand primer that includes a section: 5′. . . . . ..GGG-3′. The ds cDNA may then be amplified by PCR using the two primers.

(See also PRIMER EXTENSION ASSAY.)

Given that the complete 3′ and 5′ ends of the mRNA have been copied it is then possible to design primers which can be used to obtain full-length copies of the corresponding cDNA.

One problem with this procedure (particularly 5′ RACE) is the formation of non-specific products; for example, in the 5′ RACE reaction the second-strand primer may base-pair with any complementary sequence (. . .CCC. . .) which happens to be available in the mixture.

(See also CAPFINDER and CAPSELECT.)

rad (radiation absorbed dose) A unit of the amount of radiation absorbed: 100 ergs of energy per gram (0.01 joule per kilogram).

(cf. GRAY.)

Rad51 A type of eukaryotic protein – analogous to the bacterial RecA protein and archaeal RadA protein – which is involved e.g. in strand exchange during homologous recombination. As is the case with the (*Escherichia coli*) RecA protein, and yeast (*Saccharomyces cerevisiae*) Rad51 protein, the human Rad51 protein has been reported to show a preference for binding to sequences of ssDNA that are over-represented for guanine residues and under-represented for adenine and cytosine residues [Nucleic Acids Res (2006) 34(10):2847–2852].

[Direct imaging of (human) Rad51 nucleoprotein dynamics on DNA molecules: Proc Natl Acad Sci USA (2009) 106(2): 361–368.]

Rad51-mediated homologous pairing and strand exchange were reported to be inhibited by the reagent 4,4′-diisothiocyanostilbene-2-2′-disulfonic acid [Nucleic Acids Res (2009)

doi: 10.1093/nar/gkp200].

RadA A protein, found in members of the ARCHAEA, which is analogous to the bacterial RecA protein and to the eukaryotic Rad51 proteins; RadA is involved in strand exchange during homologous recombination.

rainbow tag A small fusion tag, expressed as a part of a recombinant protein, which facilitates purification and quantitation of the protein by imparting *color* – thus allowing tracking during the purification process and enabling quantitation via measurement of the specific absorbance signature of the tag.

raltegravir An INTEGRASE INHIBITOR (q.v.).

random access combinatorial chemistry See MICROARRAY.

random amplified polymorphic DNA (RAPD) See AP-PCR.

random insertional mutagenesis See the entry INSERTIONAL MUTAGENESIS.

random match probability See CODIS.

random mutagenesis See MUTAGENESIS.

RAPD analysis See AP-PCR.

rapid amplification of cDNA ends See RACE.

rapid-cycle PCR See THERMOCYCLER.

rare-cutting restriction endonuclease A RESTRICTION ENDONUCLEASE whose recognition sequence generally occurs only rarely in many or most samples of DNA.

One example is NotI. NotI has no recognition sites at all in some AT-rich prokaryotes – such as *Staphylococcus aureus* (GC% 32.8) and *Campylobacter jejuni* (GC% 30.5) – and it has none in bacteriophages λ and φX174, in plasmid pBR322 or in the genome of simian virus 40 (SV40). However, NotI has >350 sites in the genome of the archaean *Halobacterium* (GC% 67.9) and >250 sites in the bacterium *Mycobacterium tuberculosis* (GC% 65.6).

Other rare-cutters include e.g. FseI, SfiI and SwaI (see table in the entry RESTRICTION ENDONUCLEASE).

[Engineering a rare-cutting restriction enzyme (selection of NotI variants): Nucleic Acids Res (2006) 34(3):796–805.]

(See also part 'B' of the figure in the entry GC%.)

ras (*RAS*) A family of ONCOGENES, first identified in Harvey and Kirsten strains of murine sarcoma virus (H-*ras* or Ha-*ras*, K-*ras* or Ki-*ras*, respectively); the cellular homolog, c-*ras*, is widespread in eukaryotes, from humans to yeast.

The *ras* protein (generic designation p21) is located in the plasma membrane. This protein binds (and hydrolyzes) GTP, and it is involved in the transmission of signals which affect functions such as cell proliferation.

A transforming *ras* gene can differ by only a point mutation from the normal cellular *ras* – important sites for mutation being e.g. the 12th and 61st amino acids in the p21 protein; such mutations appear to leave the protein in a permanently active (i.e. transforming) state.

Mutated *ras* genes are common in human tumors.

rasiRNAs See the entry PIWI PROTEIN.

rATP ATP which contains a ribose (rather than a deoxyribose) residue.

RBS (1) Ribosome-binding site.
(2) Rep-binding site.

RCA (1) ROLLING CIRCLE amplification.
(2) RNA CHAPERONE ACTIVITY.

RCR vector Replication-competent retroviral vector.

RdDM RNA-DIRECTED DNA METHYLATION (q.v.).

RDM RIBOSOME DENSITY MAPPING (q.v.).

rDNA Genomic DNA encoding ribosomal RNA (rRNA).

RDO RNA/DNA oligonucleotide: a CHIMERAPLAST (q.v.).

RDR Replication-deficient retroviral vector.

RE RESTRICTION ENDONUCLEASE (q.v.).

re-animated transposable element An alternative name for a RECONSTRUCTED TRANSPOSABLE ELEMENT.

re-replication In (eukaryotic) cells: aberrant *repetition* of DNA replication, from a given origin of replication, during a given cell cycle.

Normally, a given origin in the genome 'fires' (is initiated) only *once* per cell cycle (during the S phase) – this being preceded by mitosis. This sequence of events is regulated by an 'origin licensing' system in which an origin cannot fire without being licensed. Licensing involves certain factors, such as the ORC proteins, which jointly act to enable assembly of the pre-replication complex (pre-RC) at replication origins in the G1 phase; hence, the (multiple) origins of replication present in a eukaryotic genome can fire only if they are preceded by mitosis and only if they are suitably licenced. Under normal conditions therefore, re-replication is prevented by inhibition of the premature assembly of the pre-RC.

In some test strains of the yeast *Saccharomyces cerevisiae*, multiple re-initiation events occur at some of the origins.

In certain cells, normal development includes consecutive S phases that are not separated by mitosis; in this case, repeated replication from a given origin is called *endoreduplication*.

(Note that the term 're-replication' has been used to refer to both re-replication, as defined above, and endoreduplication – even though endoreduplication is a normal process, in some cells, while re-replication refers to an aberrant process.)

re-sequencing Sequencing a given fragment of nucleic acid (or a genome), of which the wild-type (or reference) sequence is already known, in order (e.g.) to look for mutations or polymorphisms.

REA Restriction enzyme analysis: see DNA FINGERPRINTING.

read length (in massively parallel sequencing) See MASSIVELY PARALLEL SEQUENCING.

reading frame In a sequence of bases in a nucleic acid: a series of consecutive three-base groups, *with a specific base as the starting point*, in which each three-base group corresponds to a CODON during translation of that nucleic acid into a protein or polypeptide. For example, a sequence:

GTACCGTGAT. . . . .

could be read as:

|**G**TA|CCG|TGA|T. . . . .

or as:

G|**T**AC|CGT|GAT|. . . . .

or as:

GT|**A**CC|GTG|AT. . . . .

with the starting base as G, T or A, respectively. These three reading frames therefore show three possible ways in which a (single-strand) nucleic acid sequence could be read.

The above refers to the three reading frames delimited in a *single* strand of nucleic acid. Six possible reading frames can therefore be delimited in a double-stranded sample of nucleic acid.

readthrough (1) (transcription) The continuation of transcription through a stop signal.

(2) (translation) Continuation of translation through any stop codon – an amino acid being inserted at the stop codon; the product is a *readthrough protein*.

Occasional readthrough (of a UGA codon) in bacteriophage Qβ produces the A1 protein (not the normal coat protein) – translation continuing to the subsequent UAG stop codon.

readthrough protein See READTHROUGH.

real-time PCR A form of PCR in which it is possible to follow the progress of amplification – that is, the ongoing increase in numbers of specific amplicons in the reaction mixture – *while it is happening*; this approach also permits estimation of the number of specific target sequences that were present in the reaction mixture *before* the beginning of cycling (one form of QUANTITATIVE PCR).

The increase in numbers of amplicons in a given reaction can be monitored in two main ways: (i) by the use of target-specific probes (which give rise to a fluorescent signal in the presence of specific amplicons), and (ii) by the use of certain dyes which exhibit dsDNA-dependent fluorescence, i.e. dyes which fluoresce when they bind to double-stranded DNA (or which increase fluorescence, from a minimal to a significant level, when they bind to dsDNA). Each of these approaches (probe and dye) provides a type of information which is not available from the other (see later). In both of these methods, however, an increase in numbers of amplicons is monitored by recording a rising level of fluorescence while the reaction is in progress. (For a novel type of signal generation in real-time PCR see the entry FLAG.) (See also LUX PRIMER.)

During the reaction, fluorescence is recorded *cycle by cycle* – beginning at cycle 1 and continuing throughout the reaction to the last cycle (which may be e.g. cycle ∼25–35).

To estimate the number of target sequences present in the mixture prior to cycling it is necessary to determine the so-called *threshold cycle* (C_t, or C_T), i.e. the first cycle in which the probe-based or dye-based fluorescence from the reaction exceeds the background fluorescence by a specified amount.

The threshold cycle allows a quantitative estimation to be made of the initial number of target sequences in the reaction mixture as there is an *inverse linear* relationship between (i) threshold cycle and (ii) the logarithm of the number of target sequences prior to cycling. That is, a linear graph is obtained, over a range of values, if threshold cycle is plotted against the logarithm of the initial number of target sequences. We can understand this intuitively – the lower the number of target sequences present in the reaction mixture prior to cycling, the higher will be the number of cycles of amplification required to reach the level of fluorescence which corresponds to the threshold cycle (and vice versa).

Probes used in real-time PCR

The amplification of the target sequence can be monitored by several types of probe which differ mechanistically: see e.g. the LIGHTUP PROBE, MOLECULAR BEACON PROBE and TAQMAN PROBE.

Probes are included in the reaction mixture prior to cycling.

Note that, in a given cycle, the time at which fluorescence is monitored will depend on the type of probe used. Using a molecular beacon or LightUp® probe, fluorescence is monitored at the annealing stage of the PCR cycle; this is the stage at which these probes bind to specific amplicons (their target sequences) and fluoresce under appropriate excitation. With this kind of probe, the level of fluorescence from the reaction mixture increases step-wise at each annealing stage; fluorescence is not produced at the (high-temperature) stage of denaturation because these probes are not bound to their target sequences under these conditions.

By contrast, fluorescence from TaqMan® probes increases at the primer-extension stage; the level of fluorescence from the reaction mixture increases in a step-wise fashion with each cycle – although in this case the fluorescence is not lost between cycles because the molecules of dye released from probes in previous cycles remain free (and fluorescent) in the reaction mixture.

Another type of probe is associated with the LightCycler™ instrument (Roche Diagnostics, Basel, Switzerland) used for real-time PCR. It consists of two (separate) oligonucleotides that bind adjacently on the target sequence – the 3′ end of one oligo (probe 1) binding close to the 5′ end of the other (probe 2). The 3′ end of probe 1 carries a molecule of fluorescein, and the 5′ end of probe 2 carries a molecule of the fluorophore LightCycler™ Red 640. Suitable excitation produces fluorescence from the fluorescein, but (while the two probes remain unbound) no fluorescence is produced from the molecule of LightCycler™ Red 640. If both probes are correctly bound to the target sequence, activation of the fluorescein provides excitation for the LightCycler™ Red 640 (see FRET) and the latter dye emits red light (640 nm); this emission (i.e. 640 nm) signals probe–target binding and is monitored by the instrument. Fluorescence occurs during the *annealing* stage; subsequent extension of the primer displaces (but does not degrade) the two oligonucleotides – which can therefore take part in subsequent cycles. In the LightCycler™ system, therefore, fluorescence increases step-wise and is monitored in the

annealing stage.

Probes used in real-time PCR may be significantly shorter if they contain LOCKED NUCLEIC ACIDS, and LNA-containing probes may be more discriminatory than the standard form of probe [BioTechniques (2005) 38(1):29–32].

Primers which have short, 5′ AT-rich flaps were reported to increase the fluorescent signal in real-time PCR, although the reason for this was not stated [BioTechniques (2007) 43(6): 770–774].

Multiplex real-time PCR with probe-based monitoring

Simultaneous monitoring of two or more separate reactions (in the same reaction mixture) can be achieved by the use of several fluorescent dyes that have distinct emission spectra. In this mode, probes which bind to different target sequences are labeled with different fluorophores – so that separate and distinguishable signals are produced during the amplification of each different type of target sequence.

Various types of fluorophore have been used for labeling probes – see e.g. FAM and TAMRA.

dsDNA-binding dyes used in real-time PCR

One of the most commonly used dyes is SYBR GREEN I. Dyes which fluoresce when bound to dsDNA enable progress in real-time PCR to be monitored at the end of the extension phase – i.e. when the number of (double-stranded) amplicons is at a maximum. (See the entry SYBR GREEN I for notes on the dyes SYTO-13 and SYTO-82.)

(See also BOXTO.)

Combined probe-based and dye-based monitoring

While a target-specific probe detects the required product, it cannot of course detect any non-specific products or primer–dimers which may be formed during the reaction; however, spurious products such as these, if formed, may interfere with the amplification of the actual target – and they may reduce the yield of specific amplicons.

On the other hand, dsDNA-binding dyes signal *all* dsDNA, i.e. they do not distinguish between specific amplicons and any spurious products which may be present. However, these dyes are useful because, at the end of the reaction (i.e. after completion of cycling), it is possible to examine the products by melting curve analysis to detect any illegitimate products that may have been formed. (The basic rationale of melting curve analysis is that the temperature–fluorescence characteristics of a given dye-bound product depend on the product's length and composition; different dye-bound products exhibit different melting characteristics, and the presence of *mixed* products may be detected in this way if only a single product is expected.)

Hence, probe-based and dye-based monitoring offer different types of information in real-time PCR. To obtain both types of information, real-time PCR can be carried out with both types of monitoring, allowing detection of the required products and also allowing detection of any non-specific and interfering products that may have been formed in the reaction mixture [BioTechniques (2006) 40(3):315–319].

real-time protein synthesis (*in vitro*) See CELL-FREE PROTEIN SYNTHESIS.

reanimated transposable element An alternative name for a RECONSTRUCTED TRANSPOSABLE ELEMENT.

REAP Restriction endonuclease analysis of plasmid DNA (also called *plasmid profiling*).

One approach is to subject PCR-amplified plasmid DNA to restriction enzyme analysis. This approach has been used e.g. in studies on sequence variation among virulence plasmids in the equine pathogen *Rhodococcus equi* and was also used for evaluating these plasmids (as epidemiological markers) in a global surveillance program; geographic differences could be demonstrated in the distribution of the plasmids on the basis of REAP studies.

Some bacteria – e.g. species of *Brucella* and *Rickettsia* – normally lack plasmids, so that this method is not applicable to these organisms.

REase (RE) RESTRICTION ENDONUCLEASE.

reassortant virus A (hybrid) form of a SEGMENTED GENOME virus, i.e. one which contains segment(s) of the genome of a genetically distinct virus – e.g. segment(s) from another virus that is co-infecting the cell.

(See also GENETIC SHIFT.)

REBASE® An excellent source of information on restriction endonucleases and methyltransferases etc. at:
http://rebase.neb.com/rebase/rebase.html
[Nucleic Acids Res (2009) doi: 10.1093/nar/gkp874.]

RecA The bacterial counterpart of the eukaryotic RAD51 protein and the archaeal RADA protein; the RecA protein is involved e.g. in strand exchange in HOMOLOGOUS RECOMBINATION. (See also SOS SYSTEM.)

[DNA sequence specificity of the *Escherichia coli* RecA protein: Nucleic Acids Res (2006) 34(8):2463–2471.]

The RecA protein was reported to stimulate the relaxation activity of topoisomerase I, although it is apparently without effect on the supercoiling activity of gyrase [Nucleic Acids Res (2007) 35(1):79–86].

recA-deficient strains of bacteria are sometimes used for the stable cloning of plasmids because they tend to hinder the occurrence of homologous recombination between plasmids.

If there is a need to carry out homologous recombination in a *recA*-minus strain of bacteria, a *temporary* RecA function may be made available by inserting a *recA* gene on a suicide plasmid – see e.g. the entry SUICIDE VECTOR.

The RecA protein has various uses in DNA technology: see e.g. RECACTIVE.

RecActive™ A system (Active Motif) that was designed for rapidly isolating specific clone(s) from e.g. a cDNA library or a genomic library – avoiding the time-consuming methods often used for library screening.

To isolate a particular clone, the initial step is to prepare BIOTINylated, double-stranded probes (~200–600 bp) which represent a segment within the required clone. The probes are denatured to the single-stranded state and are then coated with RECA protein. When mixed with the library, the single-stranded, RecA-coated probes bind to homologous sequences

within the required clones, forming stable hybrids. The RecA proteins are then removed by treatment with sodium dodecyl sulfate (SDS), thus leaving the single-stranded probes bound to their target sequences.

The probe-labeled, biotinylated clones are then bound by STREPTAVIDIN-coated magnetic beads that are held in a magnetic field while the remaining clones are removed by washing.

After such enrichment, molecules of the required clone can be isolated and used as required.

RecBCD See HOMOLOGOUS RECOMBINATION.

***recE* gene** (of *Escherichia coli*) See PHAGE RAC.

RecF See HOMOLOGOUS RECOMBINATION.

***recG* gene** (of *Escherichia coli*) A gene encoding a monomeric HELICASE which can unwind DNA at replication forks and at Holliday junctions.

RecG is inhibited by certain hexameric peptides [Nucleic Acids Res (2008) 36(16):5319–5334].

recipient (conjugational) See F PLASMID and CONJUGATION.

recloning An approach in which fetal cells from an embryo – produced by a cloning procedure (see the entry SCNT) – are used as the nuclear donor cells in a further round of cloning.

[Example of recloning: J Reprod Dev (2008) 54(1):58–62.]

recognition site (of a restriction enzyme) See RESTRICTION ENDONUCLEASE.

recombinant (*adj*.) See RECOMBINATION.

recombinase Any enzyme which, alone or in combination with other factors, can mediate CROSSING OVER (breakage and reunion) between two regions of dsDNA.

Recombinases which mediate SITE-SPECIFIC RECOMBINATION are active only at their *specific* recognition sequences. Thus, for example, the Int recombinase of PHAGE LAMBDA (λ) acts on *att* sites, while the Cre enzyme of the CRE–LOXP SYSTEM acts on *loxP* sites. (See also FLP.)

A (chimeric) form of Cre is one example of a recombinase which is *inducible* (ligand-dependent) *in vivo*: see entry CRE–LOXP SYSTEM.

Other types of recombinase lack a *specific* recognition site and may mediate crossing over between any two *homologous* regions of dsDNA; this kind of enzyme can be useful e.g. for site-directed recombination between two suitably engineered molecules: see e.g. LAMBDA (λ) RED RECOMBINATION.

Resolvase-type and integrase-type recombinases

Resolvase-type recombinases cut all of the four strands concurrently – with 5′-phosphate bound to the enzyme in order to conserve energy for subsequent ligation. One example of this type of enzyme is a recombinase encoded by transposon Tn*3* (which is used to resolve the cointegrate formed during replicative transposition: see TRANSPOSABLE ELEMENT (figure legend)).

Integrase-type recombinases (e.g. the Int protein of phage λ, Cre protein of phage P1, and the Flp recombinase) cut and join strands in a pairwise fashion, forming an intermediate structure that is referred to as a HOLLIDAY JUNCTION (q.v.). (The λ integrase function involves the (bacterial) *integration host factor*.)

Serine and tyrosine recombinases

In *serine* recombinases (encoded e.g. by phages R4 and by transposon Tn*3*), the N-terminal catalytic site contains a serine residue.

In *tyrosine* recombinases (including Cre and Flp enzymes), the C-terminal catalytic site contains a tyrosine residue. (A tyrosine recombinase is also encoded by transposon Tn*4655* (q.v.).)

The VLF-1 (very late expression factor 1) of baculoviruses is reported to have homology with members of the tyrosine recombinase family.

recombinase-mediated cassette exchange (RMCE) Exchange of cassettes between two regions, or molecules, of DNA in which each cassette is flanked (on both sides) by the recognition site of a given recombinase.

In an example, each cassette is flanked by a wild-type *loxP* site (see CRE–LOXP SYSTEM) and a modified *loxP* site; Cre recombinase can mediate the exchange of these two cassettes – the modified *loxP* sites (one in each cassette) being used in order to maintain the cassettes' orientation after exchange.

[RMCE: tagging genes for cassette-exchange sites: Nucleic Acids Res (2005) 33(4):e44.]

RMCE was used e.g. to transfer a cassette from a plasmid to a *loxP*-augmented viral vector in human embryonic kidney cells which were engineered to express the Cre recombinase constitutively [Nucleic Acids Res (2005) 33(8):e76]. RMCE has also been used e.g. for carrying out genetic modification of stem cells [Nucleic Acids Res (2005) 33(4):e43; Proc Natl Acad Sci USA (2005) 102(18):6413–6418] and for studies on transcription [PLoS Genetics (2007) 3(2):e27].

recombinase polymerase amplification See the entry RPA.

recombination (*DNA technol*.) Any *in vivo* or *in vitro* process in which one or more molecules of nucleic acid undergo rearrangement, and/or interact with one another, in such a way that new sequence(s) of nucleotides are created; recombination thus includes events such as (e.g.) additions, inversions, replacements and amalgamations.

A molecule which has undergone recombination is referred to as a *recombinant* molecule. The term 'recombinant' is also used to describe cells and viruses in which recombination has occurred. Moreover, the term 'recombinant' is used to refer to a REASSORTANT VIRUS; although the development of such a virus does not require crossing over, the genome, as a whole, contains a new sequence of nucleotides.

Recombination includes e.g. HOMOLOGOUS RECOMBINATION, SITE-SPECIFIC RECOMBINATION and transpositional recombination.

Recombination is involved in events such as crossing over (in meiosis); activity at an INTEGRON; regulation of gene expression (e.g. of flagellar filament alleles in *Salmonella* – see PHASE VARIATION); and gene inactivation by insertion of transposable elements.

(See also INSERTION–DUPLICATION RECOMBINATION.)

recombination-based IVET (RIVET) See the entry IVET.

recombination machine See HOMOLOGOUS RECOMBINATION.

recombineering An approach with which site-specific changes can be made in chromosomes (or plasmids etc.), *within living cells*, by using phage-derived recombination systems (see e.g. ET RECOMBINATION and LAMBDA (λ) RED RECOMBINATION); with these systems, *linear* dsDNA fragments, whose terminal sequences are homologous to the target sequences, are inserted into cells and then undergo recombination (at the targeted sites) in the chromosome or plasmid etc. Terminal homologous sequences on the fragment can be as short as e.g. 50 bp. Thus, recombineering can be seen as *facilitated homologous recombination*, and is one of the approaches used for *in vivo* SITE-DIRECTED MUTAGENESIS.

The nucleotide sequences in the recombineering fragments can enable e.g. mutations to be inserted into specific genes; insertion of *loxP* sites etc.; deletion of specific genes; or gene replacement.

Fragments used in recombineering can be prepared e.g. by PCR using primers whose 5′ tags incorporate the homologous end sequences.

As well as dsDNA fragments, the recombineering process can use single-stranded oligonucleotides which have terminal homologies. This is simpler than working with fragments of dsDNA; moreover, it needs only the expression of the single-strand-annealing protein (Bet in the lambda Red system, or RecT in the RecET system).

The well-known recombineering systems (ET and Red) are both encoded by phages from Gram-negative bacteria. However, a number of functional recombineering systems are also found in some Gram-positive bacteria [the identification and analysis of recombineering functions in Gram-negative and Gram-positive bacteria: Proc Natl Acad Sci USA (2008) 105 (5):1626–1631].

Recombineering has been used e.g. for inserting point mutations or *loxP* sites etc. into bacterial artificial chromosomes [Nucleic Acids Res (2005) 33(4):e36], and making in-frame deletions/specific mutations in genomic DNA in *Salmonella enterica* [Appl Environ Microbiol (2009) 75(6):1575–1580].

reconstructed embryo (*embryol*.) See SCNT.

reconstructed transposable element (re-animated transposable element, engineered transposable element) Any TRANSPOSABLE ELEMENT from a *vertebrate* species that was originally discovered as a non-functional TE and engineered so that it can now act as a functional TE.

One reconstructed TE, known as *Sleeping Beauty*, was first discovered in a salmonid fish and was considered to contain a number of inactivating mutations that had accumulated over an evolutionary period of time; it was reconstructed – using molecular phylogenetic data (and eliminating the inactivating mutations) – and is now able to mediate precise cut-and-paste transposition in both mouse and human cells [Cell (1997) 91(4):501–510]. (The name *Sleeping Beauty* derives from the notion of an element that has re-awakened after evolutionary quiescence.) A hyperactive form of *Sleeping Beauty*, derived from a large-scale screen in mammalian cells, was reported to enable robust and stable gene transfer within vertebrates [Nature Genetics (2009) 41:753–761].

Frog Prince is a TE that was initially isolated from the frog species *Rana pipiens*. After engineering, it can now mediate cut-and-paste transposition in various types of cell (including mammalian) [Nucleic Acids Res (2003) 31(23):6873–6881].

More recently, a TE referred to as *Passport* – isolated from a fish, *Pleuronectes platessa* – was reported as the first *active* TE discovered in the native genome of a vertebrate [Nucleic Acids Res (2009) 37(4):1239–1247].

(Active TEs from *invertebrates* include PIGGYBAC.)

RecoverEase™ See DNA ISOLATION.

RecQ protein A member of a family of HELICASES which have an essential role in maintaining genome stability; helicases of this family are present in organisms ranging from bacteria to mammals (including humans).

Mutation (functional deficiency) in RecQ helicases is found in certain genetic diseases – e.g. in BLOOM'S SYNDROME and WERNER SYNDROME.

RecQ1 helicase (human) See HELICASE and RECQ PROTEIN.

***recT* gene** (of *Escherichia coli*) See PHAGE RAC.

Red (Red-mediated recombination) See entry LAMBDA (λ) RED RECOMBINATION.

RED REPEAT-EXPANSION DETECTION (q.v.).

red fluorescent proteins Fluorescent proteins whose emission spectra are in the longer wave-length band (i.e. red) (see e.g. MONOMERIC RED FLUORESCENT PROTEIN and MRUBY). One advantage of the red fluorescent proteins is that their longer wave-length emission is subjected to less scattering – making them particularly suitable for deep-tissue imaging.

(cf. GREEN FLUORESCENT PROTEIN, and see also YELLOW FLUORESCENT PROTEIN.)

Red-mediated recombination See the entry LAMBDA (λ) RED RECOMBINATION.

Red/ET recombination *Syn*. RECOMBINEERING (q.v.).

reduced-genome *Escherichia coli* A strain of *Escherichia coli* from which part(s) of the normal genome have been removed by DNA technology.

One advantage of a reduced-genome strain is that it may be more suitable (than a wild-type strain) for certain procedures. For example, a reduced-genome strain (MDS42) was found to be useful for cloning lentiviral vectors; attempts to clone lentiviral vectors in routine *E. coli* host cells usually result in instability – possibly due to the occurrence of homologous recombination between LTR regions in the vector [improved stability of lentiviral vectors in *E. coli* strain MDS42: BioTechniques (2007) 43(4):466–470].

Reduced-genome strains are also used to improve productivity in an industrial context [a reduced-genome strain of *E. coli* for L-threonine production: Microb Cell Factories (2009) 8:2].

reduced representation bisulfite sequencing An approach for comparing and analyzing methylation in different samples of genomic DNA. In the original description of RRBS [Nucleic Acids Res (2005) 33(18):5868–5877], two genomic samples

were compared: (i) the genomic DNA from (normal) mouse embryonic stem cells, and (ii) genomic DNA from embryonic stem cells that lacked DNA METHYLTRANSFERASES Dnmt3A and Dnmt3B (and expressed lower levels of Dnmt1). Both of these samples were initially digested with the RESTRICTION ENDONUCLEASE BglII, and restriction fragments of 500–600 bp were selected. Adaptors (each containing a primer-binding site for PCR) were ligated (to both ends) of all the fragments. This was followed by treatment with BISULFITE; the bisulfite-treated fragments were PCR-amplified and cloned.

Some 960 clones from the methylation-deficient DNA were sequenced, yielding 343 kb of non-redundant genomic DNA; this sequence included 66212 genomic cytosines – of which >99.9% had been converted to uracil by the bisulfite. Of the 38 cytosines which had *not* been converted to uracil, 35 were present in CpG and 3 in CpT dinucleotides; methylation at non-CpG sites was found to be >250-fold less than that found in wild-type embryonic stem cells, possibly reflecting role(s) for Dnmt3A and/or Dnmt3B in the methylation of CpA and CpT.

The RRBS approach has been used e.g. for the generation of genome-scale DNA-methylation maps [Nature (2008) 454: 766–770].

reference gene *Syn*. HOUSEKEEPING GENE.

Refludan® See BIOPHARMACEUTICAL (table).

refractory gene (in malaria) Any gene which, when introduced into, and expressed in, mosquitoes, inhibits their ability to act as vectors in the spread of the malaria parasite (*Plasmodium* spp).

(See also GMMS and PLASMODIUM.)

The mosquito *Anopheles stephensi* was made refractory to (i.e. unable to transmit) the causal agent of *rodent* malaria (*P. berghei*) by engineering it to produce certain proteins which interrupt the life cycle of *P. berghei*; these proteins bind to, and block, sites involved in the passage of plasmodial sporozoites through the mosquito's gut wall to the salivary glands.

regulation of genes *in vivo* (*DNA technol*.) See CONDITIONAL GENE ACTIVATION/REGULATION.

regulon A gene-regulatory system in which two or more non-contiguous genes/operons – each gene/operon having its own promoter – are controlled by a common regulatory molecule that is recognized by similar sequences in each of the genes/operons.

One example of a regulon is the so-called *araBAD* operon (see OPERON) which is involved in the uptake and transport of arabinose; this system includes a number of genes, several of which (*araE*, *araF*) occur at separate loci; all of the genes are controlled by the AraC protein.

relative molecular mass (M_r; 'molecular weight') Of a given molecule: the ratio of the mass of the molecule to one twelfth of the mass of a ^{12}C atom (a unitless number).

relative RT-PCR A form of QUANTITATIVE PCR employed for comparing the number of copies of a given target sequence in two or more samples – as, for example, when comparing the transcripts of a given gene from different cells or individuals, or from the same cells under different conditions.

Essentially, the target (in a given sample) is co-amplified with an internal control – each sequence (target and control) having its own pair of primers. The products of PCR can be separated by gel electrophoresis and scanned for quantitation.

The internal control may be a sequence of RNA encoded by a so-called housekeeping gene (= reference gene), such as the gene encoding 18S rRNA. This kind of control is expressed constitutively, but (e.g. in the case of 18S rRNA) its high level of expression means that amplification is likely to go beyond the exponential phase of PCR within a few cycles – thus failing to fulfill the requirement of the internal control; this problem may be addressed e.g. by using an appropriate COMPETIMER–primer ratio.

relaxed control (of plasmid replication) See PLASMID.

relaxed phenotype See *relA* in ESCHERICHIA COLI (table).

relaxing enzyme See TOPOISOMERASE (type I).

relaxosome See CONJUGATION.

release ratio In e.g. a trial of genetically modified mosquitoes: the ratio transgenic:wild type at the time of release.

remodeling (of chromatin) See CHROMATIN.

***Renilla reniformis* GFP** See HRGFP.

rep-PCR A PCR-based method used for TYPING those bacteria whose chromosomes contain a repetitive sequence of nucleotides.

Repetitive sequences exploited in this method include the so-called *repetitive extragenic palindromic sequence*, which is found e.g. in *Escherichia coli* and also in other members of the family Enterobacteriaceae (see REP SEQUENCE), the ERIC SEQUENCE and SERE. (See also MIRU in the entry W-BEIJING STRAIN.)

For rep-PCR with the REP sequence (REP-PCR – compare rep-PCR), a primer is designed which binds to a consensus sequence at one end of a REP sequence. Extension from each of the primers bound to REP sequences in the genomic DNA produces fragments which reflect REP-to-REP distances, i.e. primer extension from a given REP sequence continues until the newly synthesized strand reaches the next REP sequence in the genome; any further extension is blocked by the bound primer – all newly synthesized strands remaining as separate entities. As the REP-to-REP distances along the length of the chromosome may vary from strain to strain, analysis of PCR-generated amplicons by gel electrophoresis results in fingerprints that are characteristic of the given strains.

REP-PCR was used e.g. for typing *Escherichia coli* [Emerg Infect Dis (2009) 15(5):741–748].

SERE-PCR was used for typing *Streptococcus oralis* from dental plaque [Appl Environ Microbiol (2000) 66(8):3330–3336].

ERIC-PCR was used for studies on *Salmonella enteritidis* [World J Gastroenterol (2008) 14(7):1120–1125], and studies on rodent intestinal flora [World J Gastroenterol (2009) 15 (18):2220–2227].

REP-PCR See the above entry.

Rep protein See e.g. ROLLING CIRCLE.

REP sequence (palindromic unit) Repetitive extragenic palindromic sequence: a sequence of nucleotides – containing a PALINDROMIC SEQUENCE with hyphenated dyad symmetry – which is found in multiple copies (∼500–1000 per chromosome) in the genomic DNA of e.g. various enterobacteria (including *Escherichia coli* and *Salmonella typhimurium*). REP sequences are located in intergenic regions (so far they have not been detected in coding sequences) and they occur e.g. in promoter-distal regions (beyond genes) in many operons.

REP sequences have been used for TYPING various species of bacteria (see REP-PCR).

It has been suggested that the role of REP sequences *in vivo* may include a contribution to nucleoid structure.

REP sequences were reported to be targets for IS elements (insertion sequences) and 'hot spots' for transposition [BMC Genomics (2006) 7:62].

repeat-associated piRNAs See the entry PIWI PROTEIN.

repeat-expansion detection (RED) A method for detecting an abnormally expanded series of short nucleotide repeats (e.g. repeated trinucleotide units) in a sample of genomic DNA; an abnormal increase in the number of repeated units (produced, perhaps, by the mechanism of SLIPPED-STRAND MISPAIRING) has been associated with various types of inherited disease in humans (e.g. HUNTINGTON'S DISEASE and MYOTONIC DYSTROPHY).

(See also MICROSATELLITE DNA.)

In this approach, genomic DNA is initially heat-denatured, and *repeat-specific* oligonucleotides are allowed to anneal to the sample DNA. Those oligonucleotides that bind *adjacently* (i.e. contiguously) are ligated by a thermostable DNA LIGASE. Ligated oligonucleotides, which are released from the DNA by heating, consist of a population of multimers of different lengths, some containing two ligated oligonucleotides, others three, etc.

The cycle of annealing oligonucleotides, ligation, and heat-denaturation is repeated a number of times. The population of multimers is then subjected to gel electrophoresis and blotted onto a suitable membrane. This membrane is examined with labeled probes that are complementary to the given (repeat-specific) oligonucleotide. The sizes (i.e. lengths) of the multimers may indicate the presence/absence of expanded repeats.

(See also MLVA.)

repetitive extragenic palindromic sequence REP SEQUENCE.

replacement therapy In the present context: cell replacement therapy (see the entry STEM CELL).

replacement vector A CLONING VECTOR containing a dispensable (non-essential) sequence (the *stuffer*); the stuffer region is replaced by target DNA prior to cloning. One example of a replacement vector is the LAMBDA FIX II VECTOR.

(cf. INSERTION VECTOR.)

replica plating A method which may be used e.g. for isolating mutants which arise in a population of microorganisms under non-selective conditions; it can be used e.g. for the isolation of AUXOTROPHIC MUTANTS in a population of prototrophs.

For isolating auxotrophic mutants, a *master plate* is made by inoculating a complete medium with an inoculum from a population of prototrophs; when incubated, all the cells, i.e. prototrophs and any auxotrophic mutants that may be present in the inoculum, can grow and form individual colonies.

After incubation of the master plate, a disk of sterile velvet or similar material, fixed to one end of a cylindrical support, is lightly pressed onto the surface of the plate and then withdrawn. During contact between disk and plate a small amount of growth from each of the colonies will adhere to the velvet. The disk is then used to inoculate one (or several) plates of minimal medium (*replica plates*) – on which only the prototrophs can grow. During these procedures, it is necessary to keep a careful record of the orientation of the disk in relation to the master plate and in relation to the replica plate(s).

The replica plates are then incubated. The colonies which develop on these plates are entirely those of prototrophs – as no auxotroph is able to grow on minimal medium. Hence, by comparing the positions of colonies on the master and replica plate(s), it is possible to identify colonies on the master plate which are absent on the replica plate(s) and which are therefore those of presumptive auxotrophs.

replicase (1) Any RNA-DEPENDENT RNA POLYMERASE.

(2) An RNA-dependent RNA polymerase which specifically replicates a viral genome.

An RNA polymerase that synthesizes mRNA is sometimes called a TRANSCRIPTASE.

replication (of DNA) See DNA REPLICATION.

(See also DNA AMPLIFICATION.)

replication protein A See e.g. OKAZAKI FRAGMENT.

replicon Any molecule of nucleic acid capable of autonomous (independent) replication in an appropriate setting.

repliconation See the entry CONJUGATION.

reporter gene Any gene used in an experimental system for the purpose of signaling the occurrence (or otherwise) of specific activity, or event(s), by its own expression. Examples: genes for e.g. chloramphenicol acetyltransferase, β-galactosidase, GREEN FLUORESCENT PROTEIN and LUCIFERASE.

Increased activity of reporter gene products was achieved by inserting a short peptide-encoding sequence into the gene [Cancer Gene Therapy (2009) doi: 10.1038/cgt.2009.54].

A β-lactamase gene (*bla*) is used as a reporter gene in some commercial vectors: e.g. pcDNA^{TM}6.2/GeneBLAzer^{TM} (see GATWAY SITE-SPECIFIC RECOMBINATION SYSTEM (table)).

PCR-generated cassettes containing various reporter genes were inserted into bacterial genomes by the LAMBDA (λ) RED RECOMBINATION system [Appl Environ Microbiol (2007) 73 (13):4234–4242].

In some cases a (defective) reporter gene may be included in a transfected sequence in order to signal the integration of that sequence into genomic DNA; on integration, *cis*-acting sequence(s) in the genome may supply essential function(s) (e.g. a promoter) needed for expression of the reporter gene. (In this case, the reporter gene is expressed only if the transfected sequence integrates, appropriately, in the genome.)

(See also ENHANCER TRAPPING, IVET and PROMOTER-PROBE

VECTOR.)

repressor titration A technique for detecting the intracellular presence of a small, multicopy plasmid vector. The method uses an engineered strain of *Escherichia coli* that contains a kanamycin-resistance gene whose expression is *repressed* via a *lac* operator system (see LAC OPERON).

The multicopy plasmid – whose presence is to be detected – also incorporates the *lac* operator.

Cells which lack the vector will not grow on kanamycin-containing media: in the absence of lactose (or of IPTG), the *lac* promoter is blocked by the binding of LacI repressor to the *lac* operator – so that the kanamycin-resistance gene is not expressed.

Cells which contain the (multicopy) vector will have many copies of the (plasmid-borne) *lac* operator – and these extra copies of the *lac* operator compete with the chromosomal *lac* sequence for LacI repressor protein. Because so much LacI will be bound by the plasmid-borne *lac* operator sequences, insufficient LacI will be available to bind to the chromosomal *lac* sequence; hence, under these conditions, the kanamycin-resistance gene will be transcribed and the cells will be able to grow on kanamycin-containing media.

In addition to the simplicity of the method, one advantage is that the vector molecule can be *small* because there is no need for it to include an antibiotic-resistance marker gene (as is common in other methods). Moreover, small vector molecules are transformable with greater efficiency.

reptation In gel electrophoresis: the movement of a (linear) molecule of nucleic acid in an end-on orientation (parallel to the current).

rereplication See (earlier) entry under 're-'.

resequencing See (earlier) entry under 're-'.

resistance genes (*plant pathol.*) See GENE-FOR-GENE CONCEPT.

resistance integron See INTEGRON.

resolvase See the entries RECOMBINASE and TRANSPOSABLE ELEMENT (figure legend).

resolvase-based IVET (RIVET) See the entry IVET.

resolving gel See SDS–PAGE.

resorufin A compound whose fluorescence is monitored in an assay for pyrophosphate: see PIPER PYROPHOSPHATE ASSAY KIT.

respiration (*energy metab.*) A specific energy-converting process in which a substrate is metabolized with the involvement of an exogenous electron acceptor (such as oxygen); ATP is characteristically generated by oxidative phosphorylation.

In *anaerobic* respiration (a process carried out e.g. by some bacteria in the absence of oxygen) the electron acceptor may be e.g. a substance such as fumarate, nitrate, selenate, sulfur or sulfate.

(cf. FERMENTATION.)

respiratory syncytial virus A virus of the genus *Pneumovirus*, family PARAMYXOVIRIDAE (q.v).

response regulator See entry TWO-COMPONENT REGULATORY SYSTEM.

restriction (*DNA technol.*) Cleavage of dsDNA by one or more types of RESTRICTION ENDONUCLEASE.

In DNA technology, particular type(s) of restriction endonuclease are chosen in order to cleave at specific sites in the target molecule(s), as required.

(See also MODIFICATION.)

restriction digest The result of a procedure, used in a variety of protocols, in which target molecules are exposed to cleavage by a RESTRICTION ENDONUCLEASE or by more than one type of restriction endonuclease.

Restriction digests may be involved e.g. in TYPING bacteria (methods such as AFLP, RFLP ANALYSIS and also SUBTRACTED RESTRICTION FINGERPRINTING); RESTRICTION LANDMARK GENOMIC SCANNING; SAMPL; SELECTOR-BASED MULTIPLEX PCR.

(See also SAGE.)

restriction endonuclease (restriction enzyme; REase, RE) An endonuclease which binds to dsDNA, commonly at a specific sequence of nucleotides (termed the *recognition sequence* or *recognition site*), and which typically cuts each strand once *if* certain bases in the recognition sequence are *un*methylated (for some REs) or methylated (for other REs); according to the enzyme, the terminal regions at a cleaved site are either STICKY ENDS or BLUNT-ENDED DNA.

The methylation of dsDNA referred to above is carried out *in vivo* by methyltransferases ('methylases') – enzymes which add methyl groups to specific bases (the N6 of adenine, the C5 of cytosine) in prokaryotic DNA (see, for example, DAM GENE). Note that restriction (cutting) and methylation may be carried out by the same enzyme, i.e. in some cases REs have a dual role. (See also MODIFICATION.)

The heterogeneity of REs precludes a single definition that covers all the REs. Thus, some REs appear to have no precise recognition site while others have several recognition sites; some REs bind to palindromic sequences, while others bind to asymmetric sites; some need two copies of the recognition site (to be functional) while most need only one copy; some cut within the site while others cut outside it; some REs cut on both sides of the recognition site – i.e. excising part of the target molecule; some *nicking enzymes* (see later) resemble REs but cut only one strand.

A given sequence of nucleotides may be recognized by two or more different REs, derived from different species – see ISOSCHIZOMER.

Under non-optimal conditions, the specificity of an RE, for a given recognition sequence, may become less stringent, i.e. the RE may be able to bind to certain other sequences – as well as to its primary sequence. This phenomenon is called STAR ACTIVITY.

For some types of experiment it may be necessary to direct the activity of an RE to a *particular* recognition site, even when other copies of the given site are freely available elsewhere in the molecule: see e.g. ACHILLES' HEEL TECHNIQUE and PROGRAMABLE ENDONUCLEASE.

Restriction endonucleases are obtained from a wide range of prokaryotic microorganisms. Thermostable REs have been obtained from certain thermophilic microorganisms; one such

RESTRICTION ENDONUCLEASE: some examples

Enzyme	Type	Recognition sequence (5′→3′); cutting site (↓)	Notes
AatII	IIP	GACGT↓C	
AccI	IIP	GT↓MKAC	Isoschizomers: XmiI, FblI
AccIII	IIP	T↓CCGGA	Isoschizomers: BseAI, BspEI, Kpn2I, MroI
AflII	IIP	C↓TTAAG	Isoschizomers: BfrI, BspTI
AgeI	IIP	A↓CCGGT	
AluI	IIP	AG↓CT	Blunt-ended cut Recognition sites common in genomic 'Alu sequences' Isoschizomer: MltI
AlwNI	IIP	CAGNNN↓CTG	Isoschizomer: CaiI
AocI	IIP	CC↓TNAGG	
ApaI	IIP	GGGCC↓C	Neoschizomers: Bsp120I, PspOMI
ApoI	IIP	R↓AATTY	
AscI	IIP	GG↓CGCGCC	Useful e.g. in RLGS (restriction landmark genomic scanning)
AsuNHI	IIP	G↓CTAGC	
AvaI	IIP	C↓YCGRG	Activity ∼doubled at 45°C Isoschizomers: Ama87I, BsoBI, Eco88I, NspIII
AvaII	IIP	G↓GWCC	Isoschizomers: Bme18I, Eco47I, SinI
BamHI	IIP	G↓GATCC	Star activity in e.g. buffers of low strength, or >5% glycerol (v/v) Isoschizomer: BstI
BanII	IIP	GRGCY↓C	Isoschizomers: Eco24I, EcoT38I
BbeI	IIP	GGCGC↓C	
BclI	IIP	T↓GATCA	Isoschizomers: FbaI, Ksp22I
BfiI	IIS	5′-ACTGGG(5N)↓ 3′-TGACCC(4N)↑N	Cleavage site variability reported in the upper strand
BglI	IIP	GCCNNNN↓NGGC	Isoschizomer: Tsp8EI BglI was reported to generate an increased number of ribotypes in the species *Vibrio cholerae* (see entry RIBOTYPING)
BglII	IIP	A↓GATCT	
BpmI	IIG,E,S	5′-CTGGAG(16N)↓ 3′-GACCTC(14N)↑NN	
Bpu1102I	IIP	GC↓TNAGC	Isoschizomers: BlpI, Bsp1720I, CelII, EspI
BseAI	IIP	T↓CCGGA	
BsgI	IIG,E,S	5′-GTGCAG(16N)↓ 3′-CACGTC(14N)↑	
BsmI	IIS	5′-GAATGCN↓ 3′-CTTAC↑GN	Isoschizomers: BsaMI, PctI
BsmFI	IIG,S	GGGAC(10/14)	Used e.g. in serial analysis of gene expression (see entry SAGE)
Bsp106I	IIP	AT↓CGAT	
BspCI	IIP	CGAT↓CG	

(*continued*)

RESTRICTION ENDONUCLEASE (*continued*)

Enzyme	Type	Recognition sequence (5′→3′); cutting site (↓)	Notes
BspMI	IIS	5′-ACCTGC(4N)↓NNNN 3′-TGGACG(8N)↑	
BssHII	IIP	G↓CGCGC	Used e.g. for cutting genomic DNA in CpG islands Isoschizomers: BsePI, PauI
BstEII	IIP	G↓GTNACC	Isoschizomers: BstPI, Eco91I
BstNI	IIP	CC↓WGG	Isoschizomers: BptI, BseBI, BstOI, MvaI Neoschizomers: AjnI, EcoRII, Psp6I, PspGI Cleavage occurs even if the internal cytosine is methylated
BstXI	IIP	CCANNNNN↓NTGG	Isoschizomer: BstHZ55I
BstYI	IIP	R↓GATCY	Isoschizomers: MflI, XhoII
Bsu36I	IIP	CC↓TNAGG	Cutting site used e.g. in a baculovirus-related vector [BioTechniques (2006) 41(4):453–458]
CfoI	IIP	GCG↓C	Isoschizomer: HhaI Neoschizomer: Hin6I
Cfr9I	IIE,P	5′-C↓CCGGG 3′-GGGCC↑C	
Cfr10I	IIF,P	5′-R↓CCGGY 3′-YGGCC↑R	
ClaI	IIP	AT↓CGAT	Isoschizomers: BanIII, BspDI, BspXI, Bsu15I
CspI	IIP	CG↓GWCCG	
CvnI	IIP	CC↓TNAGG	Isoschizomers: AxyI, Bse21I, Bsu36I, Eco81I
DdeI	IIP	C↓TNAG	Isoschizomer: BstDEI
DpnI	IIM,P	G(m^6A)↓TC	Activity of the enzyme *requires* methylation at the N6 position of adenine Isoschizomer: MalI
DraI	IIP	TTT↓AAA	Isoschizomers: AhaIII, SruI
DraIII	IIP	CACNNN↓GTG	
EagI	IIP	C↓GGCCG	Cutting sites e.g. in CpG islands
Eam1104I	IIS	5′-CTCTTCN↓NNN 3′-GAGAAGNNNN↑	Used e.g. in the Seamless® cloning kit
Ecl18kI	IIP	↓CCNGG	Mechanism for determining specificity: EMBO J (2006) 25:2219–2229
Eco47III	IIP	AGC↓GCT	Blunt-ended cut. Star activity in high levels of enzyme Isoschizomer: AfeI
Eco57I	IIE,G	5′-CTGAAG(16N)↓ 3′-GACTTC(14N)↑NN	
EcoKI	I	random	[DNA binding: Nucleic Acids Res (2009) 37(6):2053–2063]
EcoO109I	IIP	RG↓GNCCY	Isoschizomer: DraII Neoschizomer: PssI
EcoP15I	III	CAGCAG(25/27)	Used in super SAGE (see SAGE)
EcoRI	IIP	G↓AATTC	Star activity in e.g. buffers of low strength, >5% glycerol (v/v), pH >8 Isoschizomers: none reported See also MunI

(*continued*)

RESTRICTION ENDONUCLEASE (*continued*)

Enzyme	Type	Recognition sequence (5′→3′); cutting site (↓)	Notes
EcoRII	IIP,E	↓CCWGG	Isoschizomers: AjnI, Psp6I Neoschizomers: BptI, BseBI, MvaI Sensitive to methylation. *Two* copies of the recognition site are needed for activity of EcoRII
EcoRII-C	IIP-type		Cleaves *single* copies of the recognition sequence in a methylation-sensitive way [BioTechniques (2005) 38(6):855–856]
EcoRV	IIP	GAT↓ATC	Blunt-ended cut [mechanism: Biochemistry (2009) 48:9061–9075] Isoschizomer: Eco32I
EcoR124I	IC	random	[Mechanism of translocation: EMBO J (2006) 25:2230–2239]
FokI	IIS	GGATG(9N)↓NNNN CCTAC(13N)↑	Used in ZINC-FINGER NUCLEASES
FseI	IIP	GGCCGG↓CC	Rare-cutter
FspI	IIP	TGC↓GCA	Isoschizomers: AviII, NsBI
HaeII	IIP	RGCGC↓Y	Isoschizomers: Bsp143I, BstH2I
HaeIII	IIP	GG↓CC	Isoschizomers: BsnI, BsuRI, PhoI
HhaI	IIP	GCG↓C	Isoschizomer: CfoI Neoschizomers: Hin6I, HspAI
HincII	IIP	GTY↓RAC	Enzyme used in strand displacement amplification as a *nicking* enzyme (see SDA) Isoschizomer: HindII
HindIII	IIP	A↓AGCTT	
HinfI	IIP	G↓ANTC	Star activity e.g. if Mg^{2+} is replaced by Mn^{2+} or in high levels of glycerol
HinP1I	IIP	G↓CGC	Isoschizomers: Hin6I, HspAI Neoschizomers: BstHHI, CfoI, HhaI [Mechanism of action: Nucleic Acids Res (2006) 34(3):939–948]
HpaI	IIP	GTT↓AAC	Isoschizomer: KspAI
HpaII	IIE	C↓CGG	Isoschizomers: HapII, MspI
HpyHI	IIP	CTNAG	Cleavage site unknown
KpnI	IIP	GGTAC↓C	Star activity if >5% glycerol (v/v) Neoschizomers: Acc65I, Asp718I
Kpn2I	IIP	T↓CCGGA	
MboI	IIP	↓GATC	Isoschizomers: DpnII, NdeII, Sau3AI Neoschizomer: BstKTI
MboII	IIS	5′-GAAGA(8N)↓ 3′-CTTCT(7N)↑N	
MfeI	IIP	C↓AATTG	
MluI	IIP	A↓CGCGT	Isoschizomer: Bbi24I
MmeI	IIG	5′-TCCRAC(20N)↓ 3′-AGGYTG(18N)↑	Used e.g. in one method for generating shRNAs
MnlI	IIS	5′-CCTC(7N)↓	Slow cleavage reported for single-stranded DNA

(*continued*)

RESTRICTION ENDONUCLEASE (*continued*)

Enzyme	Type	Recognition sequence (5′→3′); cutting site (↓)	Notes
		3′-GGAG(6N)↑	
MroI	IIP	T↓CCGGA	
MscI	IIP	TGG↓CCA	Isoschizomers: BalI, MluNI
MseI	IIP	T↓TAA	Isoschizomer: Tru9I
MspI	IIP	C↓CGG	Isoschizomers: BsiSI, HapII, HpaII
MunI	IIP	C↓AATTG	Produces sticky ends that can hybridize with those formed by EcoRI Isoschizomer: MfeI
NaeI	IIE,P	GCC↓GGC	Blunt-ended cut. Site preference: the (single) NaeI site in phage λ DNA is reported to be cut more slowly than other NaeI sites [DNA looping dynamics: Nucleic Acids Res (2006) 34:167–174] Isoschizomer: PdiI Neoschizomers: MroNI, NgoMIV
NarI	IIE	GG↓CGCC	Site preference: the (single) NarI site in phage λ DNA is reported to be cut more slowly than other NarI sites. [DNA looping dynamics: Nucleic Acids Res (2006) 34:167–174] Isoschizomer: Mly113I Neoschizomers: BbeI, DinI, EheI, KasI, SfoI
NciI	IIG,P	CC↓SGG	Isoschizomer: BcnI
NcoI	IIP	C↓CATGG	Isoschizomer: Bsp19I
NdeI	IIP	CA↓TATG	Isoschizomer: FauNDI
NdeII	IIP	↓GATC	Isoschizomers: BfuCI, BssMI, DpnII, Sau3AI Neoschizomer: BstKTI
NgoAIV	IIP	G↓CCGGC	Isoschizomer: NgoMIV Neoschizomers: NaeI, PdiI
NheI	IIP	G↓CTAGC	Isoschizomer: AsuNHI Neoschizomer: BmtI
NlaIII	IIP	CATG↓	Used e.g. in serial analysis of gene expression (see entry SAGE)
NotI	IIP	GC↓GGCCGC	‘Rare-cutting’ enzyme; useful e.g. for cutting genomic DNA into large fragments Used in RLGS (restriction landmark genomic scanning) Isoschizomer: CciNI
NruI	IIP	TCG↓CGA	Isoschizomers: Bsp68I, BtuMI
NsiI	IIP	ATGCA↓T	Isoschizomers: EcoT22I, Mph1103I, Zsp2I
NspI	IIP	RCATG↓Y	Isoschizomers: BstNSI, XceI
NspV	IIP	TT↓CGAA	Isoschizomers: AsuII, BstBI, Csp45I
PacI	IIP	TTAAT↓TAA	Used e.g. in the AdEasy™ adenoviral vector system
PaeI	IIP	GCATG↓C	
PalI	IIP	GG↓CC	Isoschizomers: HaeIII, BsuRI
PinAI	IIP	A↓CCGGT	Isoschizomers: AgeI, BshTI, CspAI

(*continued*)

RESTRICTION ENDONUCLEASE (*continued*)

Enzyme	Type	Recognition sequence (5′→3′); cutting site (↓)	Notes
PmeI	IIP	GTTT↓AAAC	Isoschizomer: MssI Used e.g. in the AdEasy™ adenoviral vector system
PspAI	IIP	C↓CCGGG	Isoschizomer: XmaI Neoschizomer: SmaI
PstI	IIP	CTGCA↓G	Isoschizomer: BspMAI
PvuI	IIP	CGAT↓CG	Buffer conditions important for optimal activity. Cleavage of supercoiled plasmids may require high concentrations of enzyme Isoschizomers: BpvUI, MrvI, Ple19I
PvuII	IIP	CAG↓CTG	Star activity e.g. with >5% (v/v) glycerol and low-salt buffer Used as a programmable endonuclease when linked to a TFO (i.e. triplex-forming oligonucleotide) [see Nucleic Acids Res (2005) (22):7039–7047]
RcaI	IIP	T↓CATGA	Isoschizomers: BspHI, PagI
RsaI	IIP	GT↓AC	Isoschizomer: AfaI Neoschizomer: Csp6I
RsrII	IIP	CG↓G(A/T)CG	Isoschizomers: CpoI, CspI
SacI	IIP	GAGCT↓C	
SacII	IIE	CCGC↓GG	Reported to cleave sites in DNA of phages φX174 and λ more slowly than in other recognition sites Isoschizomers: KspI, SgrBI, SstII
SalI	IIP	G↓TCGAC	Star activity if glycerol is in high concentration, or with high levels of enzyme; needs salt concentration >100 mM
SanDI	IIP	GG↓GWCCC	
Sau3AI	IIE	↓GATC	Isoschizomers: DpnII, MboI, NdeII Neoschizomer: BstKTI Activity reported to be decreased when the target DNA is under mechanical tension (approx. 10-fold decrease at 0.7 pN) [Proc Natl Acad Sci USA (2006) 103(31):11555–11560]
ScaI	IIP	AGT↓ACT	Isoschizomers: AssI, BmcAI, ZrmI
SfiI	IIF,P	GGCCNNNN↓NGGCC	Rare cutter
SgrAI	IIP	5′-CR↓CCGGYG 3′-GYGGCC↑RC	
SmaI	IIP	CCC↓GGG	Neoschizomers: Cfr9I, XmaI, XmaCI
SnaBI	IIP	TAC↓GTA	Isoschizomers: BstSNI, Eco105I
SpeI	IIP	A↓CTAGT	Isoschizomers: AhlI, BcuI
SphI	IIP	GCATG↓C	Isoschizomers: BbuI, PaeI
SrfI	IIP	GCCC↓GGGC	Blunt-ended cut Ligation of SrfI-cut ends is reported to be promoted e.g. by 15% polyethylene glycol (PEG)
SspI	IIP	AAT↓ATT	Inhibited by glycerol >5%
SstI	IIP	GAGCT↓C	Isoschizomers: SacI, Psp124BI

(*continued*)

RESTRICTION ENDONUCLEASE (*continued*)

Enzyme	Type	Recognition sequence (5′→3′); cutting site (↓)	Notes
			Neoschizomer: EcoICRI
SstII	IIP	CGGC↓GG	Isoschizomers: KspI, SacII, SgrBI
StuI	IIP	AGG↓CCT	Isoschizomers: AatI, Eco147I, SseBI
StyI	IIP	C↓CWWGG	Isoschizomers: Eco130I, ErhI
SuiI	IIP	G↓CWGC	
SunI	IIP	C↓GTACG	Isoschizomers: BsiWI, PspLI
SwaI	IIP	ATTT↓AAAT	Rare cutter
TaqI	IIP	T↓CGA	Isoschizomers: Tsp32I, TthHB8I
ThaI	IIP	CG↓CG	Isoschizomers: AccII, Bsh1236I, BstUI, MvnI
Tsp509I	IIP	↓AATT	Used e.g. in SUICIDE POLYMERASE ENDONUCLEASE RESTRICTION (q.v.)
TspMI			Isochizomer of XmaI (q.v.)
UbaF13I	IIS	GAG(6N)CTGG	Cleavage site unknown
UbaF14I	IIS	CCA(5N)TCG	Cleavage site unknown
VspI	IIP	AT↓TAAT	Isoschizomers: AseI, PshBI
XbaI	IIP	T↓CTAGA	Star activity with high levels of enzyme and with DMSO. Sensitive to methylation
XhoI	IIP	C↓TCGAG	Isoschizomers: StrI, TliI, PaeR7I
XhoII	IIP	R↓GATCY	
XmaI	IIP	C↓CCGGG	Isoschizomer (thermostable): TspMI [Appl Microbiol Biotechnol (2006) 72: 917–923]
XmaIII	IIP	C↓GGCCG	Isoschizomers: BstZI, EclXI, Eco52I
XmnI	IIP	GAANN↓NNTTC	

(thermophilic) enzyme is used for signal generation in REAL-TIME PCR (see the entry FLAG).

RE-encoding genes in previously uncharacterized genomes have been identified during the cloning of fragments derived from SHOTGUN SEQUENCING. Thus, some of these fragments can be cloned in *Escherichia coli* – but some cannot. Of the latter, some contain RE-encoding genes which, if expressed in *E. coli*, would cause lethal cleavage of the genome; such genes were identified, for example, in genomes of strains of *Bacillus* and *Methylococcus* spp [Nucleic Acids Res (2009) 37(1):e1].

Nomenclature

The currently recommended nomenclature for REs (and the methyltransferases) [Nucleic Acids Res (2003) 31(7):1805–1812] avoids the (former) use of italics. For example, EcoRI (instead of *Eco*RI) is used to refer to an RE from *Escherichia coli* strain R, the I indicating one particular RE from strain R. Sometimes the prefix 'R' is used – e.g. R.EcoRI, in which 'R' indicates a restriction – rather than a methylating – enzyme; enzymes which incorporate both restriction and methylating functions may have the prefix RM. Some other examples of named REs are given in the table.

Online information on REs

For online information on REs (and methyltransferases) see details in the entry REBASE.

Classification of REs

REs have been classified into four main groups (types I–IV):

Type I REs. This type of RE is a multisubunit enzyme with subunits for restriction, methylation and site-specificity (the latter specifying the recognition sequence); *ATP-dependent* cleavage occurs at a non-specific site outside the recognition sequence. An example of a type I RE is EcoKI; some others:

Eco394I, Eco826I, Eco851I, Eco912I. EcoKI was reported to dimerize when binding to DNA [Nucleic Acids Res (2009) 37(6):2053–2063].

The activity of type I REs has been confirmed as catalytic (i.e. enzymic) rather than stoichiometric [Nucleic Acids Res (2009) doi: 10.1093/nar/gkp195].

(See also HSD GENES.)

Type II REs. These REs typically recognize a specific sequence of nucleotides, and both of the strands are cleaved, in an *ATP-independent* manner, at fixed locations within or near the recognition site – leaving a 5′-phosphate terminus and a 3′-hydroxyl terminus on both sides of the cut.

Precise cutting of DNA makes the type II RE very useful in recombinant DNA technology (and more than 3500 type II REs have been characterized). A number of type II REs have been engineered to have new binding and cleavage characteristics [Nucleic Acids Res (2009) doi: 10.1093/nar/gkp535].

Subtypes of the type II REs include:

IIA. These REs recognize asymmetric sequences.

IIB. These REs cleave on both sides of the recognition site. Examples include BaeI and BcgI. The recognition and cutting sites of BaeI can be written:

$$(10/15)AC(N_4)GTAYC(12/7)$$

in which N_4 refers to four (unspecified) nucleotides and Y is either C or T. 10/15 means that the strand shown is cut 10 nucleotides upstream and the complementary strand is cut 15 nucleotides upstream. 12/7 means that the strand shown is cut 12 nucleotides downstream and the complementary strand is cut 7 nucleotides downstream.

IIC. In these REs the cutting and methylation functions are combined in one polypeptide; they include members of other subtypes.

IIE. These REs cut the recognition site but need to interact with another copy of that site for 'activation'. [DNA looping by two-site restriction enzymes: Nucleic Acids Res (2006) 34(10):2864–2877.] Studies on various 'two-site' REs have shown that these enzymes are sensitive to mechanical tension in target DNA, enzymic activity being lost when the DNA is subjected to a few piconewtons of stress [Proc Natl Acad Sci USA (2006) 103(31):11555–11560].

IIF. These REs interact with, and cut, two copies of the recognition site.

IIG. Methylation and cutting functions in one protein; may be affected by *S*-adenosyl-L-methioneine. (See e.g. MMEI.)

IIM. These REs cut at fixed locations with a specific pattern of methylation – that is, they carry out *methylation-dependent* cleavage. An example is DpnI (see DPNI).

IIP. These REs include a large number that recognize symmetric (i.e. palindromic) sequences and cleave at fixed sites in, or immediately adjacent to, their recognition sites.

One example of a type IIP RE is EcoRI, whose recognition and cutting (↓) sites are:

5′-G↓AATTC-3′

in one strand, and

3′-CTTAA↑G-5′

in the complementary strand. Note that, in this case, there are staggered cuts which leave *sticky ends*: 5′-AATT overhangs on each strand.

IIS. The type IIS REs cut *outside* the recognition site. These enzymes have particular uses in DNA technology – see e.g. DNA CUTTING AND ASSEMBLY. An example of a type IIS RE is SAPI (q.v. for details of recognition/cutting site).

Type IIS enzymes are used e.g. in procedures such as APSR and the STICKY RICE method. (See also SEAMLESS CLONING KIT.)

Note. A given type II RE may belong to more than one subtype.

Type III REs. These REs interact with both copies of a non-palindromic sequence that are specifically orientated; their ability to cleave depends on *ATP-dependent* translocation and occurs at a specified distance from one of the sites. Examples of type III REs include EcoP1I and EcoP15I.

Studies on EcoP15I have pointed to the roles of both DNA looping and translocation in a model for restriction by type III restriction endonucleases [EMBO J (2007) 26(16):3815–3825]. Studies which looked at the way in which the type III enzymes deal with their paired recognition sites concluded that looping is 'both highly unlikely and unnecessary' [Proc Natl Acad Sci USA (2009) 106(6):1748–1753].

Type IV REs. These REs recognize a *methylated* site but – unlike the type IIM REs – their cutting sites apparently lack specificity.

Putative REs. The existence of these REs (which have a prefix 'P') is inferred e.g. by comparing an open reading frame (ORF) in a sequenced genome with the sequence of a known RE.

Nicking enzymes. These enzymes cut ('nick') only one of the strands of a DNA duplex; at least some of these enzymes resemble restriction endonucleases – e.g. BstNBI.

The nicking enzymes have an 'N' prefix, e.g. N.BstNBI.

Some (normal) REs can be made to act as nicking enzymes by using DNA with chemically modified bases: see e.g. SDA.

At a low pH, the type IIS restriction enzyme BfiI becomes a nicking enzyme, selectively cutting a single strand.

restriction endonuclease analysis of plasmid DNA See REAP.

restriction enzyme *Syn*. RESTRICTION ENDONUCLEASE.

restriction enzyme analysis *Syn*. DNA FINGERPRINTING.

restriction enzyme-independent cohesive ends See STICKY RICE.

restriction fragment Any fragment cut from a larger molecule of DNA by the action of a RESTRICTION ENDONUCLEASE (or by the action of more than one restriction enzyme).

restriction fragment length polymorphism analysis See RFLP ANALYSIS.

restriction landmark genomic scanning (RLGS) A technique

used e.g. to detect changes in the pattern of DNA methylation (in a given genome) or to compare the methylation patterns in different tissues within the same individual.

In an early approach, suitably prepared genomic DNA was first digested with RESTRICTION ENDONUCLEASE NotI, and the 5′ overhangs (at cleavage sites) were filled in with labeled (^{32}P) nucleotides. The end-labeled (i.e. ^{32}P-labeled) DNA was digested with another restriction enzyme, EcoRV, after which the first phase of gel electrophoresis was carried out. DNA-containing gel was subsequently cut out and digested with the restriction enzyme MboI; after this digestion the next phase of gel electrophoresis was carried out. This was followed by analysis of the products.

In studies on methylation patterns in genomic DNA, NotI is a particularly useful enzyme because its recognition site (5′-GC↓GGCCGC-3′) occurs primarily in CPG ISLANDS: >85% of NotI cutting sites within the (human) genome are found in CpG islands. This is significant in that CpG islands are often associated with the promoter regions of genes, and aberrant patterns of methylation in such regions have been noted in cancer-related silencing of tumor-suppressor genes; hence, it is of particular interest to examine the methylation patterns in fragments that include CpG islands. In this context, it is NotI (rather than EcoRV) which determines the specific (CpG-associated) locations of the restriction fragments; hence, NotI is called the *landmark* enzyme (or the restriction landmark) in this procedure.

Note that NotI is a methylation-sensitive enzyme, i.e. it will not cleave its recognition site if an internal cytosine residue is methylated. Consequently, an *aberrantly* methylated cutting site will not be cleaved and end-labeled – so that a particular band of RLGS fragments will be absent at the location in the gel which is normally occupied by a band; that is, absence of a given band may indicate methylation at the corresponding landmark site. Conversely, an increase in the number of NotI-cut fragments may indicate loss of methylation at particular site(s), or perhaps gene amplification.

While most RLGS studies have been carried out with NotI, some studies have used an additional enzyme: AscI (cutting site: GG↓CGCGCC).

A useful follow-up is to prepare a library by cloning RLGS fragments, thus permitting further studies on particular loci of interest.

[Uses of RLGS: Cancer Biol Ther (2008) 7(8):1182–1190; Carcinogenesis (2008) 29(8):1623–1631.]

A form of *virtual* RLGS (vRLGS) involves the generation of theoretical (computer-based) fragments [BMC Genomics (2007) 8:446].

restriction map In relation to a given DNA molecule or fragment: any diagram showing the positions of the recognition sequences of RESTRICTION ENDONUCLEASES; other features, e.g. origin of replication, particular genes, direction of transcription etc. may also be shown on the map in order to give orientation.

(See also FLASH CHEMILUMINESCENT GENE MAPPING KIT.)

restriction-minus cells (*DNA technol.*) Genetically engineered bacteria in which some, or all, of the restriction systems have been deleted.

An example of a restriction-minus strain is XL1-Blue MRF′ (Stratagene, La Jolla CA): Δ(*mcrA*)183, Δ(*mcrCB-hsdSMR-mrr*)173, *endA1*, *supE44*, *thi-1*, *recA1*, *gyrA96*, *relA1*, *lac*[F′ *proAB*, *lacI*^{q}*ZΔM15*, Tn*10*(Tet^{r})].

Examples of the use of XL1-Blue MRF′ include: studies on translational regulation in archaeans [PLoS ONE (2009) 4(2): e4484]; preparation of new analogs of gilvocarcin anti-tumor agents [Chembiochem (2009) 10(2):278–286]; a study on the horizontal transfer of genes among strains of *Escherichia coli* [PLoS Pathog (2009) 5(1):e1000257].

restriction–modification system See MODIFICATION.

reticulate body (RB) See CHLAMYDIA.

retrocopy (of a gene) See RETROGENE.

retrogene (retrocopy) A derivative of a normal gene formed by reverse transcription of an mRNA into DNA and subsequent insertion of the DNA into the genome (with an opportunity for expression); this type of process apparently occurs only rarely in evolution, although it may be a significant influence on the development of (mammalian) genomes [PLoS ONE (2009) 4(3):e5040].

If the inserted sequence has no opportunity for expression, the resulting promoter-less, intron-less and function-deficient entity has been referred to as a *retropseudogene*.

(See also PSEUDOGENE.)

retrohoming See INTRON HOMING.

retron In some bacteria, e.g. strains of *Escherichia*, *Klebsiella*, *Proteus*, *Rhizobium*, *Salmonella*: a (chromosomal) sequence encoding a REVERSE TRANSCRIPTASE similar to that encoded by retroviruses. These bacteria are able to synthesize a hybrid RNA/DNA molecule, called *multicopy single-stranded DNA* (msDNA), in which the 5′ end of ssDNA is covalently bound (via a phosphodiester bond) to the 2′-position of an internal (non-terminal) guanosine residue in the RNA. A single cell may contain as many as 500 copies of msDNA.

Retrons have been named according to the initials of the host species and the number of bases in the DNA moiety of the corresponding msDNA; thus, for example, in *Myxococcus xanthus* the msDNA contains 162 bases, and this molecule is designated Mx162.

As well as retrons, bacteria encode reverse transcriptases in group II introns – and in so-called diversity-generating retro-elements (see the entry DRG) [a diversity of uncharacterized reverse transcriptases in bacteria: Nucleic Acids Res (2008) 36(22):7219–7229].

retroposon *Syn.* RETROTRANSPOSON.

retropseudogene See RETROGENE.

retroregulation A type of post-transcriptional regulation which involves the influence of a *cis*-acting sequence downstream of the given gene. Ongoing transcription of the gene into the *cis*-acting sequence results in degradation of the transcript, thus blocking gene expression; if transcription stops before reaching the *cis*-acting sequence the transcript is available for

translation.

Retroregulation is involved, e.g. during the development of conditions which promote lytic infection in a lysogenic strain of bacteria containing bacteriophage lambda (λ). Under these conditions expression of the *int* gene (encoding integrase) is repressed: transcription of *int* extends beyond its normal limit into a sequence (*sib*); this causes the 3′ end of the mRNA to adopt a secondary structure that is recognized, and cleaved, by RNase III.

retrotranscription Synthesis of a strand of DNA on an RNA template.

(See also REVERSE TRANSCRIPTASE.)

retrotransfer (*bacteriol.*) Transfer of DNA from the recipient cell, in CONJUGATION, to the donor cell.

retrotransposon (retroposon) A TRANSPOSABLE ELEMENT that has a mode of transposition involving an RNA intermediate (and the activity of reverse transcriptase); the intermediate is reverse transcribed, and the DNA product is inserted into the target site.

Retrotransposons are found in a wide range of organisms. One example is the TY ELEMENT of the yeast *Saccharomyces cerevisiae*, and in *Drosophila* specialized non-LTR telomeric retrotransposons have roles in maintaining the chromosomal length (cf. TELOMERASE). In mice, the INTRACISTERNAL A-PARTICLE is common in the genome. [Trypanosomatid retrotransposons: Eukarayotic Cell (2009) 8(10):1532–1542.]

A study on cap-selected RNA transcripts derived from mice and humans concluded that retrotransposon transcription has a major influence on transcriptional output of the mammalian genome [Nature Genetics (2009) 41:563–571].

(See also ACTIVATION-INDUCED CYTIDINE DEAMINASE.)

Retroviridae A family of ssRNA-containing viruses; members of this family are referred to, collectively, as RETROVIRUSES (q.v.).

retroviruses Enveloped, ssRNA viruses (family Retroviridae) which replicate via a dsDNA intermediate.

Retroviruses are found in a wide range of species, including e.g. mammals, birds and reptiles. (See also COPIA ELEMENT.) Infection of a given host by a retrovirus may be symptomless or may be associated with any of a number of diseases which range from pneumonia and tumors to leukemia and AIDS.

The treatment of AIDS involves ANTIRETROVIRAL AGENTS which target specific aspects the retroviral infective cycle.

Transmission of retroviruses may occur horizontally and/or vertically.

The following is a *generalized account* of the retroviruses, although with particular emphasis on HIV-1.

Classification of retroviruses

The retroviruses have been placed in subfamilies on the basis of e.g. the type(s) of pathogenic effect they cause. Viruses of the subfamily LENTIVIRINAE (including HIV-1) replicate in – and kill – host cells.

The subfamily ONCOVIRINAE includes oncogenic viruses.

Viruses of the subfamily SPUMAVIRINAE are characteristically non-pathogenic.

The virion

The heat-sensitive and detergent-sensitive virion, 80–120 nm in diameter, consists of a nucleoprotein core (which includes the viral REVERSE TRANSCRIPTASE) and an envelope of host-derived lipid from which project virus-encoded glycoproteins that determine the types of cell which can be infected by the virus.

All the genetic information is encoded in a single-stranded, positive-sense RNA of about 3–10 kb, according to virus – ~9.2 kb in HIV-1; the virion contains two identical (or nearly identical) copies of this RNA, linked non-covalently at the 5′ ends.

The RNA has regions of TERMINAL REDUNDANCY that are flanked, externally, by a 5′ CAP and a 3′ poly(A) tail.

U5 and U3 are sequences unique to the 5′ and 3′ regions of the RNA, respectively; each is adjacent to its corresponding region of terminal redundancy (see diagram).

Coding regions of the RNA molecule include the *gag*, *pol* and *env* sequences.

The *gag* region encodes proteins of the viral core.

The *pol* region encodes e.g. reverse transcriptase, the viral integrase (involved in integrating the (dsDNA) provirus into the host cell's chromosome), and the viral protease (used for maturation of viral proteins).

The *env* region encodes e.g. major envelope (glyco)proteins gp120 and gp41.

The genome also contains e.g. a so-called ψ (psi) sequence which promotes packaging of viral RNA into virions (during lytic infection by the virus).

Infection of cells

Infection of cells involves interaction between viral envelope proteins and specific type(s) of cell-surface receptor. During infection by HIV-1, the envelope glycoprotein gp120 (which is at the outermost end of the trans-envelope protein gp41) binds to cell-surface antigen CD4; target cells for this virus therefore include e.g. the ($CD4^+$) 'helper' subset of T lymphocytes (i.e. helper T cells). In addition to the CD4 receptor, infection by HIV-1 depends on the presence of certain other cell-surface receptor molecules such as receptors for CXC-type and CC-type chemokines – see the entry HIV-1.

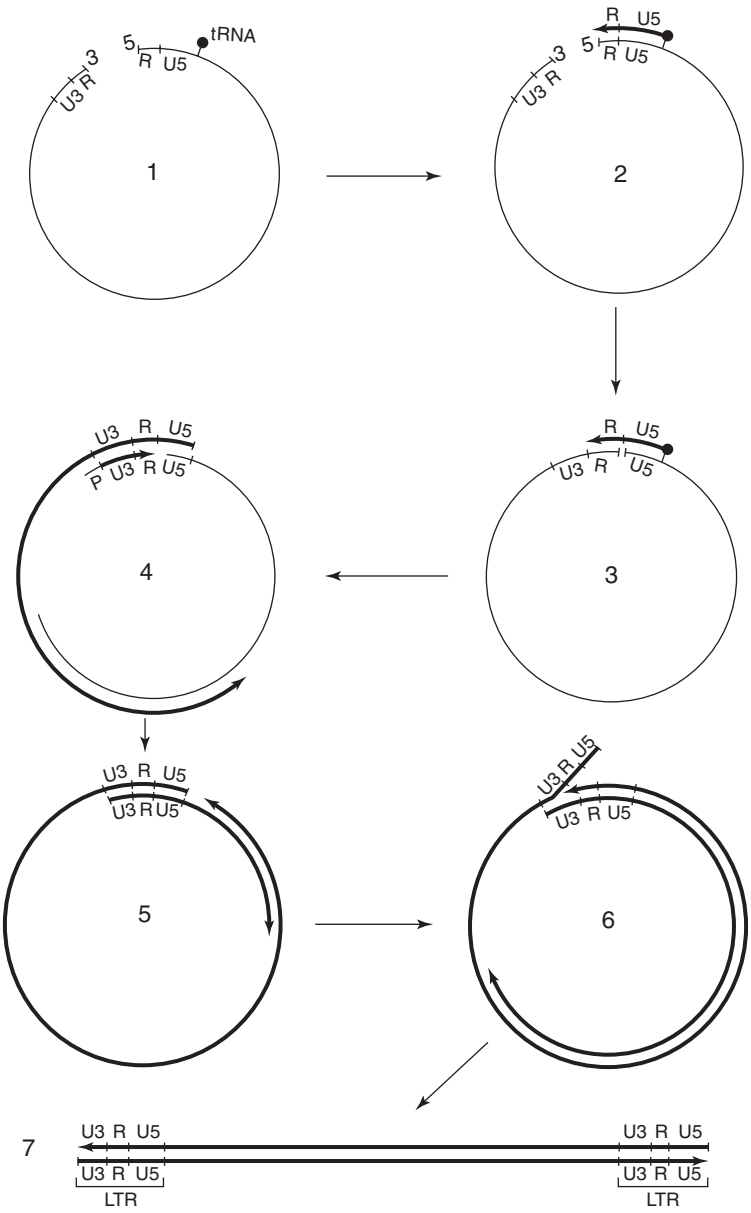
A PCR-based study reported that, during infection of blood leukocytes by HIV-1 *ex vivo*, blood dendritic cells were preferentially infected [J Virol (2007) 81(5):2297–2306].

(See also DC-SIGN.)

Reverse transcription and integration of provirus

Within the cell, viral RNA is converted to DNA by the viral reverse transcriptase. Priming involves a host tRNA molecule which binds at the 3′ boundary of the U5 sequence; in many cases $tRNA^{pro}$ is used, but $tRNA^{lys}$ is used by HIV-1. [Stability of binding of $tRNA^{lys}$ to the HIV genome: Nucleic Acids Res (2009) doi: 10.1093/nar/gkp187.]

Following extension of the (minus) strand of DNA to the 5′ end of the RNA template, the RNA corresponding to the 5′ terminal redundancy is degraded by RNase H activity. The resulting (single-stranded) sequence of DNA then hybridizes



RETROVIRUSES: a model for the synthesis of dsDNA from the ssRNA retroviral genome.

1. The ssRNA retroviral genome.
2. Minus-strand DNA synthesis is initiated, by reverse transcriptase, from a tRNA primer, and synthesis continues to the 5′ end of the (viral) RNA template. (*continued*)

with the 3′ terminal redundancy region at the other end of the RNA template; extension of the DNA strand then continues on this end of the template (see diagram).

Initiation of synthesis of the DNA *plus* strand requires the use of a specific polypurine tract upstream of U3 ('P' in the diagram) as a primer; this particular tract of the HIV-1 RNA genome is resistant to RNase H degradation during reverse transcription. Mispriming of DNA synthesis from *other* polypurine tracts in the HIV-1 genome has been reported to be blocked by the chaperone-like activity of the retroviral nucleocapsid protein [Nucleic Acids Res (2009) 37(6):1755–1766].

As synthesis of the second (i.e. plus) strand of DNA begins at a staggered site, each end of the final dsDNA product has a so-called *long terminal repeat* (LTR) in which U3 and U5 are separated by one copy of the terminal redundancy (as shown in the diagram). LTRs are necessary for subsequent enzymic insertion of retroviral dsDNA into the host cell's genome.

One or more copies of the dsDNA *provirus* may integrate into the host cell's DNA. Integration of the provirus DNA is mediated by retroviral integrase, a tetramer of which forms a stable complex with the two ends of the viral DNA; the two ends of the prophage are inserted, sequentially, into the host cell's chromosome [EMBO J (2006) 25(6):1295–1304]. (This phase of the HIV-1 infection process is the target of certain drugs: see INTEGRASE INHIBITOR.)

Expression of the viral genome (the provirus)

Transcription of proviral DNA, processing of transcripts and the subsequent formation of virions are complex processes involving e.g. synthesis of polycistronic pre-mRNAs and the essential contribution of certain virus-encoded proteins. For example, the small early intronless transcripts can translocate readily from the nucleus to the cytoplasm, but larger intron-containing transcripts, formed later, require the REV PROTEIN (in HIV-1) for translocation to the cytoplasm via the CRM1 nuclear export pathway.

Other virus-encoded proteins (in HIV-1) include e.g. the transcription regulatory factor Tat; the Vif ('viral infectivity factor') protein [functions of Vif: Nucleic Acids Res (2009) doi: 10.1093/nar/gkp226]; and Vpu protein (which promotes efficient budding of the virion during productive infection of the cell).

The maturation of retroviral RNA and post-transcriptional events leading to production of virions were reviewed [Retrovirology (2006) 3:18].

Retroviruses in DNA technology

Recombinant (engineered) retroviral vectors have been used e.g. in clinical trials in GENE THERAPY – and also for various studies involving genetic modification of cultured cells.

So far, most studies have been carried out with replication-deficient retroviruses, but – particularly in the case of therapy for solid tumors – attention is being given to the use of RCR (replication-competent retroviral) vectors: see section below.

Factors favorable to the use of retroviruses:

- retroviruses can effect gene transfer to (sensitive) cells with high frequency;
- once integrated in the cell's genome, the retroviral provirus may be expressed on a long-term basis;
- proviral DNA is transmitted to daughter cells;
- high-titer preparations of infectious, replication-deficient viral particles can be produced;
- retroviruses can be pseudotyped to control the host range.

Factors unfavorable to the use of retroviruses:

- most retroviruses infect only actively dividing cells; however, lentiviruses can infect both dividing and non-dividing cells (and are used as vectors);
- the retroviral provirus integrates randomly into the DNA of the host cell – with potential lethality; thus, in a gene therapy setting, a knock-out insertion in a tumor-suppressor gene may lead e.g. to tumorigenesis (see also INSERTIONAL ONCOGENESIS);
- in a laboratory setting, retroviral particles are rather labile.

Replication-competent retroviral (RCR) vectors

Until quite recently the routine approach used non-replicating

RETROVIRUSES (*continued*)

3. The 5′ end-region of the RNA template (labeled 'R' for terminal redundancy) is degraded (removed) by RNase H activity (the reverse transcriptase has RNase H activity). This leaves the 3′ end of the nascent DNA strand unpaired, so that it can base-pair with the 'R' region at the 3′ end of the viral RNA.
4. As the DNA strand elongates, the RNA template strand is degraded by RNase H activity. A short polypurine tract of RNA, at 'P', is resistant to RNase H; it acts as a primer for the synthesis of the plus-strand of DNA (on the minus-strand DNA template).
5. Minus-strand DNA synthesis reaches the end of the viral RNA template, providing a template for ongoing synthesis of the DNA plus strand.
6. Minus-strand DNA synthesis continues beyond the original genome length, and copies the U3–R–U5 sequence.
7. Plus-strand DNA synthesis (at step 6) has continued to the end of the U3–R–U5 sequence. Hence, the complete dsDNA molecule contains a U3–R–U5 sequence at each end; the U3–R–U5 sequence is termed the *long terminal repeat* (LTR). This dsDNA molecule is inserted into the genome of the host cell by a process involving the viral integrase.

Figure reproduced from *Dictionary of Microbiology and Molecular Biology*, 3rd Edition Revised, page 658, Paul Singleton & Diana Sainsbury (2006) John Wiley & Sons Ltd, Chichester UK [ISBN 978-0-470-03545-0] with permission from the publisher.

retroviral vectors – the so-called RDR (replication-deficient retroviral) vectors. Although some success has been achieved using this approach in specific cases, results have been somewhat disappointing in terms of clinical benefit e.g. in therapy directed against solid tumors; this is, perhaps, not surprising because at least some of the cells *inside* a solid tumor are unlikely to be accessible to infecting viruses given the existing techniques available for gene delivery.

Attention is now being focused on the use of replication-competent retroviral (RCR) vectors which – in solid tumors – may infect adjacent tumor cells when a given cell has been lysed by the replicating retroviral vector.

In the context of RCR vectors, a number of studies using murine leukemia virus (MLV) have shown promising results.

The use of RCR vectors is, however, associated with some potential problems. Thus, for example, if the vector carries a transgene, in addition to its own genome, then the unnaturally extended viral genome can adversely affect the efficiency of viral replication.

A further potential problem with transgene-carrying RCR vectors is that they may undergo recombination – giving rise to mutants which lack the transgene and which replicate more efficiently than the original (therapeutic) vector; were such an event to occur, then mutant viruses could rapidly reduce the efficacy of the original vector. Hence, design features of the *replicating* vector must be carefully investigated and optimized in order to reduce the risk of this kind of problem. Work has been carried out to investigate the influence of the vector characteristics on the mechanism of recombination and on the emergence of mutant subpopulations in the replicating retroviral vectors [BMC Mol Biol (2009) 10:8].

Rett syndrome An X-linked neurodevelopmental disorder (in humans) commonly resulting from mutation in *MECP2*, a gene (Xq28) encoding methyl-CpG-binding protein 2, but it may also result from mutation in gene *CDKL5* (Xp22), which encodes a cyclin-dependent kinase.

Specific mutations in the *MECP2* gene were reported to be associated with different levels of severity [Neurology (2008) 70(16):1313–1321].

(See also GENETIC DISEASE (table).)

Rev protein A regulatory protein encoded by human immunodeficiency virus type 1 (HIV-1). The activity of Rev protein is essential for ongoing productive infection by HIV-1; thus, Rev is required e.g. for transporting various intron-containing mRNAs (encoding envelope and certain other proteins) from the nucleus to the cytoplasm. Rev protein binds to a complex secondary structure (called the Rev response element, RRE) in relevant transcripts and mediates their transport from the nucleus.

The Rev protein itself incorporates a nuclear localization signal (NLS) and a nuclear export signal (NES).

Rev activity was found to be inhibited by a dsRNA-binding protein (nuclear factor 90) [Retrovirology (2006) 3:83].

The Rev protein has been studied in a live-cell assay for the simultaneous monitoring of expression and interaction of proteins [BioTechniques (2006) 41(6):688–692].

(See also transdominant negative proteins in the entry GENE THERAPY.)

***REV3* gene** See DNA POLYMERASE ζ.

***REV7* gene** See DNA POLYMERASE ζ.

reverse genetics Essentially: working/studying in the direction genotype → phenotype, as opposed to the usual (or 'classic') approach: phenotype → genotype. Experimental procedures used in reverse genetics include (for example) KNOCK-OUT MUTATIONS and RNAI.

reverse gyrase A protein, found in various species of bacteria and archaeans, which can introduce *positive* supercoiling into a ccc dsDNA molecule. It consists of a helicase-like domain fused with a type I TOPOISOMERASE. Reverse gyrase is found in most/all organisms which grow >70°C, and is generally associated with organisms that live at high temperatures; the protein was (but is apparently no longer) regarded as specific for *hyper*thermophilic organisms.

In organisms that contain both gyrase and reverse gyrase, plasmids are reported to be *negatively* supercoiled.

Various functions have been suggested for reverse gyrase, including stabilization of the genome at high temperatures – although one study reported that it is not absolutely essential for growth at high temperatures (in the organisms examined).

Studies on a bacterium, *Nautilia profundicola*, from submarine hydrothermal environments found that expression of the gene for reverse gyrase increased 100-fold when the temperature rose by 20°C – suggesting an adaptive role [PLoS Genet (2009) 5(2):e1000362].

Studies on an archaean, *Sulfolobus solfataricus*, reported an interaction between reverse gyrase and the organism's translesion DNA polymerase, SsoPolY/Dpo4 – a member of the Y FAMILY of DNA polymerases [Nucleic Acids Res (2009) doi: 10.1093/nar/gkp386].

reverse hybridization The procedure (used e.g. in some diagnostic tests) in which target sequences of nucleic acid bind to a set of *immobilized* probes in an array, i.e. the reverse of the procedure in which the target is immobilized. (cf. REVERSE SOUTHERN HYBRIDIZATION.)

(See also LINE PROBE ASSAY.)

reverse mutation *Syn.* BACK MUTATION.

reverse primer See the entry for FORWARD PRIMER.

reverse Southern hybridization A method used to detect a specific sequence of DNA, in solution, by hybridization to a labeled probe. The probe is initially modified with a psoralen derivative such that, when bound to its target and exposed to ultraviolet radiation, the probe is covalently cross-linked to the target.

reverse splicing Splicing of RNA into a DNA target: the term used e.g. to describe insertion of intronic RNA into recipient DNA during retrohoming by group II introns – see the entry INTRON HOMING.

reverse transcriptase (RNA-dependent DNA polymerase) An enzyme which can synthesize a strand of DNA on an RNA template. Reverse transcriptases are encoded e.g. by retro-

viruses as an essential requirement for their intracellular development and chromosomal integration, and are also encoded by some retrovirus-like elements, such as the TY ELEMENT in the yeast *Saccharomyces cerevisiae*.

Genes encoding these enzymes also occur in some bacteria, including *Escherichia coli* (see RETRON).

DNA polymerase I of *E. coli* was reported to have (poorly processive) reverse transcriptase activity.

A reverse transcriptase synthesizes the DNA strand 5′-to-3′ from a short primer. Divalent cations are needed for enzymic activity.

The r*Tth* enzyme (Perkin Elmer/Applied Biosystems) can function as a reverse transcriptase in the presence of Mn^{2+}; on chelation of Mn^{2+} (with EGTA), the enzyme can behave as a DNA polymerase in the presence of Mg^{2+}. Alternatively, this enzyme can be used with a specialized single-buffer system.

At least some reverse transcriptases are multifunctional – e.g. the reverse transcriptase of the HIV-1 virus can behave, in addition, as a DNA-dependent DNA polymerase and as an RNase H enzyme; in the latter case, it is able to degrade the RNA strand in a DNA/RNA hybrid. [Interconversion of the reverse transcriptase of HIV-1 between RNase H and DNA polymerase modes: Nature (2008) 453:184–189.]

The lack of 3′-to-5′ exonuclease (proof-reading) activity in a reverse transcriptase results in error-prone DNA synthesis.

A reverse transcriptase may be inhibited by agents such as AZT (see NUCLEOSIDE REVERSE TRANSCRIPTASE INHIBITORS) as well as NON-NUCLEOSIDE REVERSE TRANSCRIPTASE INHIBITORS. (See also SURAMIN.)

reverse transcriptase PCR (rtPCR, rt-PCR, RT-PCR; *formerly* RNA PCR) A PCR-based procedure for making copies of a sequence from an RNA template; rtPCR has various uses – including studies on gene expression (e.g. detecting mRNAs that are in low concentration in cell lysates), and studies on RNA viruses.

In an early 'two-tube' protocol, a strand of DNA is initially synthesized on the RNA target by REVERSE TRANSCRIPTASE activity at ∼50°C. The RNA strand is then degraded (e.g. by RNase H) and a complementary strand of DNA is synthesized on the first strand, forming a dsDNA template which is then amplified by routine PCR methodology.

The above protocol, carried out at ∼50°C, can be problematic if the target RNA contains secondary structures that are stable at these temperatures and that are able to inhibit DNA synthesis by (physically) blocking the polymerase. In a more recent protocol, rtPCR is conducted at higher temperatures using a thermostable enzyme (e.g. RTTH DNA POLYMERASE) which is able to act as both reverse transcriptase (at 60–70°C) and DNA polymerase under appropriate conditions.

A different method for amplifying an RNA target sequence (NASBA) was found to be less inhibited by factors that were inhibitory in an rtPCR assay of noroviruses in environmental waters [Appl Environ Microbiol (2006) 72(8):5349–5358].

reverse translation *Syn*. BACK TRANSLATION.

reverse two-hybrid system A TWO-HYBRID SYSTEM in which transcriptional activation of the reporter system is inhibitory to growth – see e.g. the *URA3* negative selection mechanism in YEAST TWO-HYBRID SYSTEM.

RF (1) Recombination frequency.

(2) The replicative form of the genome which is found in a certain stage of infection in some types of bacteriophage: see e.g. PHAGE M13.

(cf. RFI.)

RFI The *supercoiled* replicative form of the genome in certain types of bacteriophage: see e.g. PHAGE M13.

RFLP analysis Restriction fragment length polymorphism analysis: a method used for comparing *related* sequences of nucleotides (e.g. wild-type and mutant forms of a particular sequence of DNA) by comparing the subfragments produced when these molecules are cleaved by the same RESTRICTION ENDONUCLEASE; different numbers and/or sizes of fragments are produced from a given molecule if that molecule has lost or gained a restriction site (e.g. through a point mutation) or if it has undergone any insertion or deletion of nucleotides. The sets of subfragments from each sample are compared by gel electrophoresis. Differences between samples, in terms of the number and/or position of restriction sites or the existence of insertions or deletions, are manifested by the development of bands at different positions in the gel.

Various types of sample can be examined by this method – for example: viral genomes, plasmids, PCR-generated copies of particular sequences of nucleotides in bacterial and other genomes.

One factor in the sensitivity of RFLP analysis is the choice of restriction endonuclease; this, in turn, is influenced by the composition of the target sequence. Any variation among the samples is more likely to be detected if the chosen enzyme is one which has a greater number of restriction sites within the target DNA; the number of cutting sites of a given restriction endonuclease can vary greatly according to the composition of the target sequence – see e.g. the figure (parts A, B and C) in the entry GC%.

RFLP analysis has various uses which include the typing of bacteria (comparison of corresponding sequences of nucleotides from various strains) and the detection of mutations or polymorphisms (or methylation of specific bases) by detecting their effects on the recognition site of a given restriction enzyme.

The use of this method for TYPING bacteria presupposes an appropriate level of stability in the target sequence. This may not be the case in some instances. For example, in certain strains of *Mycobacterium tuberculosis* the validity of results from RFLP analysis was questioned because it appeared that the target sequence was evolving too rapidly for the reliable interpretation of results on a short-term basis (a few years).

One potential problem with RFLP analysis is that samples of DNA derived directly from cells may have a pattern of modification (i.e. methylation) which is inhibitory to certain types of restriction endonuclease – thus precluding the use of

these enzymes in the technique. To avoid this problem, the target sequence can be copied (from genomic DNA) by PCR and the resulting amplicons used as sample DNA for RFLP analysis; as the PCR-derived DNA is synthesized *in vitro* it is non-methylated (see PCR-RFLP ANALYSIS).

RFLP analysis has been used e.g. for identifying sequence variations in the (human) *SLC26A5* gene [PLoS ONE (2009) 4(6):e5762], and for studying microbial communities in clean and hydrocarbon-contaminated soil [Open Microbiol J (2009) 3:75–86].

RGD motif See INTEGRIN (β_2).

***rglA* gene** *Syn*. MCRA GENE.

rglB *Syn*. MCRBC.

Rh1B helicase See the entry DEGRADOSOME.

Rhizobium radiobacter See the entry AGROBACTERIUM.

rhLF Recombinant human lactoferrin: see GLYCOSYLATION.

***rho* factor** (ρ factor) In bacteria: a protein which is required for termination of transcription at a *rho*-dependent terminator.

***rho*-dependent terminator** In bacteria: a sequence at the 3′ end of a transcript that signals termination of transcription; termination depends on interaction with a protein *rho* factor.

(cf. RHO-INDEPENDENT TERMINATOR.)

***rho*-independent terminator** In bacteria: a sequence of nucleotides, at the 3′ end of a transcript, that signals termination of transcription; it includes a GC-rich PALINDROMIC SEQUENCE and a stretch of consecutive uridine residues. Unlike the *rho*-dependent terminator, this sequence can mediate termination in the absence of a *rho* factor.

***rho*-zero (ρ^0) cells** Cells which lack mitochondrial DNA; they are used e.g. in the preparation of CYBRIDS.

ribavirin 1-β-D-ribofuranosyl-1,2,4-triazole-3-carboxamide: a broad-spectrum antiviral agent which is used therapeutically against e.g. the hepatitis C, Lassa fever and SARS viruses.

Ribavirin was reported to inhibit GTP synthesis and viral RNA-dependent RNA polymerase.

The mechanism of the *in vivo* antiviral activity of this agent remains obscure. One study (involving hepatitis C virus) was carried out with patients undergoing joint ribavirin and interferon treatment; the results suggested that the effects of such treatment resulted, at least partly, from the induction of lethal mutations in the virus [J Virol (2009) 83(11):5760–5764].

RibEx See RIBOSWITCH.

RiboGreen® A fluorescent dye used e.g. for quantitation of RNA in solution (Molecular Probes Inc., Eugene OR/Invitrogen, Carlsbad CA). For quantitation of RNA, the sensitivity of RiboGreen is reported to be at least 1000-fold better than that obtainable by conventional ULTRAVIOLET ABSORBANCE measurement at 260 nm.

Enhancement of fluorescence is also shown when the dye binds to DNA.

[Uses (e.g.): Mol Vision (2009) 15:534–544; Plant Physiol (2009) 150(1):402–415.]

(See also DNA STAINING and PICOGREEN ASSAY.)

ribonuclease Any of various enzymes which cleave RNA – see under RNASE for specific examples.

Ribonucleases that have been used for sequencing RNA are mentioned in the entry MAXAM–GILBERT METHOD.

An *endoribonuclease* acts at internal sites while an *exoribonuclease* cleaves terminal nucleotides.

Ribonucleases may need to be inactivated in solutions etc. in order to prevent loss of sample RNA. Inactivation can be achieved e.g. by DIETHYLPYROCARBONATE and by products such as RNA*secure*™ (Ambion, Austin TX); this commercial product can be used for treating almost any type of solution, including those containing Tris, and there is no requirement for post-treatment autoclaving.

ribonuclease protection assay A method used e.g. for detecting a specific type of mRNA in a sample that contains total RNA. Briefly, a labeled antisense oligonucleotide is allowed to hybridize with the required mRNA in solution. Then, the addition of RNase degrades *single*-stranded RNA, leaving the hybridized (double-stranded) target mRNA intact. The RNase is inactivated, and the double-stranded target molecule is isolated and examined e.g. in a (denaturing) polyacrylamide gel.

Nucleases used in this type of assay include e.g. mung bean nuclease and endonuclease S_1.

ribonuclease T_1 See the entry RNASE T1.

ribonucleic acid See the entry RNA.

ribonucleotide reductase An enzyme which catalyzes the conversion of ribonucleotides to deoxyribonucleotides.

RiboPrinter Microbial Characterization System™ See the entry RIBOTYPING.

riboregulation A term that refers to RNA-mediated regulation of gene expression, including e.g. control by MICRORNAS and regulation by a RIBOSWITCH mechanism.

ribosomal frameshifting See FRAMESHIFTING.

ribosome density mapping (RDM) A method which is used to determine the number of ribosomes associated with particular regions of an mRNA molecule; the procedure involves site-specific cleavage of polysomal mRNAs and separation of the fractions in a sucrose density gradient. The products are then analyzed.

ribosome rescue A term sometimes used to refer to the process of TRANS TRANSLATION.

riboswitch A regulatory sequence within an mRNA molecule; on binding a particular metabolite, the riboswitch undergoes a structural change that may e.g. inhibit translation. That part of a riboswitch which binds the metabolite is referred to as an *aptamer*; the sequence which undergoes structural change is called the *effector element* or *expression platform*. (Synthetic riboswiches have been used for regulating gene expression in mammalian and other cells.)

Riboswitches are found in prokaryotes, and they have also been reported in certain eukaryotes (e.g. fungi, plants).

The bacterial adenine riboswitch aptamer (referred to as the 'A box') regulates gene expression positively when it binds adenine [core requirements of the adenine riboswitch aptamer for ligand binding: RNA (2007) 13(3):339–350].

[Ligand-dependent folding of the three-way junction in the purine riboswitches: RNA (2008) 14(4):675–684.]

[Mutational analysis of the purine riboswitch aptamer: Biochemistry (2007) 46(46):13297–13309.]

[Ligand recognition determinants of guanine riboswitches: Nucleic Acids Res (2007) 35(16):5568–5580.]

[Riboswitch-dependent gene regulation and its evolution in the plant kingdom: Genes Dev (2007) 21(22):2874–2879.]

[Screening, with a flow-cytometry-based approach, for synthetic riboswitches that are activated by theophylline: Nucleic Acids Res (2009) 37(1):184–192.]

[Design principles for riboswitch function: PLoS Comput Biol (2009) 5(4):e1000363.]

Online source

RibEx: http://www.ibt.unam.mx/biocomputo/ribex.html

ribothymidine See NUCLEOSIDE.

ribotype See RIBOTYPING.

ribotyping (*microbiol.*) A method for TYPING bacteria which exploits inter-strain variability in the RRN OPERON (encoding 16S, 23S and 5S rRNA).

The *rrn* operon is useful as a target in ribotyping for three main reasons:

- it is found in all bacteria;
- most species of bacteria contain more than one copy of the *rrn* operon (*Escherichia coli* has seven copies);
- the intergenic spacers are commonly of variable length.

Essentially, the genomic DNA of a given strain is cleaved by a restriction endonuclease and the fragments are separated into bands by gel electrophoresis. A labeled DNA (or RNA) probe, complementary to a sequence within the *rrn* operon, is then allowed to bind to those bands of fragments containing the target sequence. Commonly, between three and six bands are labeled by the probe. The number and position of bands produced from any given strain will depend on the following factors:

- the number of *rrn* operons in the genome;
- the lengths of the intergenic spacers;
- the precise distribution of cutting sites of the restriction endonuclease used in the procedure; loss or gain of recognition sites – e.g. through point mutations – in the sample DNA will affect the fingerprint by altering the pattern of labeled bands in the gel.

A strain defined by ribotyping is called a *ribotype*.

[Molecular genetic basis of ribotyping: Clin Microbiol Rev (2008) 21(2):262–273.]

In one study on *Vibrio cholerae* the authors referred to high levels of recombination between the *rrn* operons in cells of this species (10^{-5} per cell per generation); they pointed out that any given rcombinational event could undergo reversal over extended periods of time and suggested that ribotyping may not be a suitable method for determining *evolutionary* relationships between strains. Nevertheless, in general, ribotyping is often used successfully for differentiating strains of various species of bacteria in a short-term setting.

The study on *V. cholerae*, referred to above, also found that recombination among *rrn* operons often resulted in a gain or loss of cutting sites for the restriction endonuclease BglI; this has suggested a reason why the range of ribotypes detected by this enzyme is often greater than that obtained with other types of restriction enzyme.

Ribotyping has been used to characterize a range of species – including e.g. *Shigella boydii* (using PvuII as the restriction enzyme) [BMC Res Notes (2008) 1:74], and *Streptococcus pyogenes* [BMC Genomics (2009) 10:198].

Attempts to accelerate the ribotyping procedure have lead to the development of PCR RIBOTYPING, and also to an automated approach (the RiboPrinter Microbial Characterization System™, marketed by DuPont, Wilmington DE, USA). The RiboPrinter has been used in studies on various species – e.g. strains of *Listeria monocytogenes* (using EcoRI as restriction enzyme) [Appl Environ Microbiol (2008) 74(15):4954–4961] and strains of *Corynebacterium striatum* (also using EcoRI) [Emerg Infect Dis (2009) 15(1):75–78].

ribozyme Any RNA molecule which can function in a way that is usually associated with a (protein) enzyme; the ribozymes have roles that include nuclease and RNA–RNA ligase (see also APTAZYME).

Ribozymes occur in both prokaryotic and eukaryotic cells, and they are also found in certain viruses (see also VIROID); in cells, ribozymes have functions such as the processing of RNA precursors (see e.g. RNASE P) and the self-splicing of introns.

(Less well known are the DEOXYRIBOZYMES.)

In DNA technology, ribozymes have been used in various ways, and in various types of combination, both *in vitro* and *in vivo* (see e.g. GENE THERAPY).

When complexed with polyethyleneimine (PEI), ribozymes form small particles which have been used for transfection studies. (See also MELITTIN.)

A ribozyme may also be linked to an APTAMER to form a structure in which the activity of the ribozyme is regulated by the binding of a specific molecule to the aptamer.

A functional HAMMERHEAD RIBOZYME has been expressed in cells of *Escherichia coli* from a DNA NANOCIRCLE.

Some ribozymes (e.g. one from the ciliate protozoan *Tetrahymena*) are thermostable, and there has been interest in the relationship between sequence/structure and thermostability.

In studies on stabilizing mutations in the *Tetrahymena* ribozyme, it was reported that the effect on thermostability of a given mutation was highly dependent on context – thus, for example, both of the mutations A269G and A304G (close together on the ribozyme) had to be present at the same time in order to promote stability. In general, the conclusion from these studies was that mutation to stability in RNA (which may depend e.g. on the presence of rare double mutations) is less likely to occur than it is in proteins, in which stabilizing (single-site) mutations can accumulate [RNA (2006) 12(3): 387–395].

(See also ALLOSTERIC NUCLEIC ACID ENZYMES.)

rifampicin See RIFAMYCINS.

rifampin See RIFAMYCINS.

rifamycins Macrocyclic antibiotics whose target is the β sub-

unit of the (bacterial) DNA-dependent RNA polymerase, and which therefore inhibit transcription in sensitive strains. Rifamycins act against certain Gram-positive bacteria (including mycobacteria) and some Gram-negative bacteria (e.g. strains of *Brucella*, *Chlamydia* and *Legionella*).

In *Mycobacterium tuberculosis* resistance to the rifamycins may result from mutation in the *rpoB* gene which encodes the β subunit of RNA polymerase.

The rifamycins include e.g. rifampicin (= rifampin).

(See also ANTIBIOTIC and LINE PROBE ASSAY.)

Rift Valley fever virus A member of the genus *Phlebovirus* (of the family BUNYAVIRIDAE) which causes an acute febrile disease (in man and animals) in parts of Africa; transmission of the disease commonly involves certain insects (usually mosquitoes).

(See also AMBISENSE RNA.)

right end (of phage Mu DNA) See the entry PHAGE MU.

right-handed helix (of DNA) A helix which, from an axial (i.e. end-on) direction, winds away from the viewer in a clockwise fashion.

rimantadine See the entry AMANTADINE.

rinderpest virus A virus of the genus *Morbillivirus* within the family PARAMYXOVIRIDAE (q.v.).

rINN Recommended international non-proprietary name.

RISC RNA-induced silencing complex: the effector complex in RNA INTERFERENCE (q.v.).

RISE-resistant tuberculosis Tuberculosis which is resistant to treatment with, at least, rifampin, isoniazid, streptomycin and ethambutol.

ritonavir See the entry PROTEASE INHIBITORS.

RIVET Resolvase-based IVET, or recombination-based IVET: see the entry IVET.

RLGS RESTRICTION LANDMARK GENOMIC SCANNING.

RLU Relative light unit – see e.g. ACCUPROBE.

R–M system See the entry MODIFICATION.

RMCE RECOMBINASE-MEDIATED CASSETTE EXCHANGE.

RNA Ribonucleic acid, a NUCLEIC ACID which, in cells, has two major roles: (i) to act as an intermediate in the synthesis of proteins (see MRNA), and (ii) to carry out a wide range of gene-regulatory functions (see NON-CODING RNAS); in some viruses (e.g. RETROVIRUSES) the genome consists of RNA.

Both isothermal and temperature-cycling methods are used for RNA AMPLIFICATION (q.v.).

Enzymes associated with RNA include e.g. RNA POLYMERASE and REVERSE TRANSCRIPTASE (see also the entries for POLYNUCLEOTIDE PHOSPHORYLASE and RNASE H).

(See also RIBOZYME, RNA EDITING, RNA INTERFERENCE.)

(cf. DNA, GNA, PNA, and TNA.)

RNA amplification (*in vitro*) Any method used for amplifying, i.e. copying, a (typically) defined sequence of ribonucleotides in an RNA molecule. (cf. DNA AMPLIFICATION and NUCLEIC ACID AMPLIFICATION.)

Methods used for RNA amplification include NASBA (isothermal) and REVERSE TRANSCRIPTASE PCR (which involves temperature cycling).

(See also APTAZYME.)

RNA chaperone See e.g. FINOP SYSTEM and RNA CHAPERONE ACTIVITY.

RNA chaperone activity (RCA) An effect, mediated by certain types of protein, in which the structure/conformation of an RNA molecule is targeted in such a way that the function/activity of the RNA is promoted; RCA may help to stabilize the normal, functional structure of an RNA molecule or may assist in the resolution of a misfolded RNA molecule.

The ability to mediate RCA has been noted in a number of proteins. The RCA of ribosomal protein L1 is reported to be conserved in all three domains of life [Nucleic Acids Res (2007) 35(11):3752–3763].

Whether or not a given protein is involved in RCA can be assessed in various ways. Thus, for example, the ability of a protein to promote the re-structuring of RNA into an active, catalytic form can be assessed by determining the protein's ability to promote the particular form of (RNA-mediated) catalysis in the presence of the given RNA molecule and its substrate.

A different approach to the study of RCA involved a FRET-based assay of the effect of specific proteins on annealing and strand displacement of two (complementary) 21-mer RNAs [BioTechniques (2007) 43(3):304–310].

RNA degradosome *Syn*. DEGRADOSOME.

RNA-dependent DNA polymerase A DNA polymerase that is able to synthesize a strand of DNA on an RNA template: see REVERSE TRANSCRIPTASE.

RNA-dependent RNA polymerase An RNA POLYMERASE that uses an RNA template; these enzymes are encoded by some viruses (e.g. members of the family PARAMYXOVIRIDAE). (In some cases the cell's DNA-dependent RNA polymerase may assume the role (in association with a viral protein) [J Virol (2005) 79(13):7951–7958].)

RNA-dependent RNA polymerases occur in eukaryotes – in which they are associated e.g. with RNA INTERFERENCE.

[Role in *Arabidopsis* PLoS ONE (2009) 4(3):e4971.]

(cf. REPLICASE and TRANSCRIPTASE.)

RNA-directed DNA methylation (RdDM) A process in which methylation of DNA is mediated by a molecule of RNA. It may be involved in GENE SILENCING (e.g. in *Transcriptional gene silencing*: see the entry SIRNA), but has been reported to produce transcriptional *activation* in plants [Proc Natl Acad Sci USA (2009) 106(5):1660–1665].

RNA–DNA chimeric oligonucleotide *Syn*. CHIMERAPLAST.

RNA editing A normal *in vivo* process, in animals, plants and some microorganisms, in which new transcripts are generated from existing ones.

Mechanistically distinct forms of RNA editing are carried out in different types of organism.

In one example, *un*edited transcripts of apolipoprotein B, a component of chylomicrons (intestinal lipoproteins), encode a protein of 4536 amino acid residues; editing, by APOBEC-1 (a CYTIDINE DEAMINASE), deaminates a specific cytidine residue in the mRNA, creating a stop codon that reduces the

product length to 2152 residues. This involves the mRNA-binding factor ACF/ASP (i.e. APOBEC-1 complementation factor, or APOBEC-1-stimulating protein); regulation of the editing of this mRNA may involve phosphorylation of ACF.

The activity of the mammalian enzyme AID (ACTIVATION-INDUCED CYTIDINE DEAMINASE) may be another example of RNA editing.

Transcript editing also involves deamination of adenosine to inosine by dsRNA-specific, structure-directed enzymes of the ADAR group: adenosine deaminase acting on RNA; the inosine is read as guanosine. [Recognition of A-to-I editing sites: Nucleic Acids Res (2009) doi: 10.1093/nar/gkp731.]

In trypanosomatids (a group of eukaryotic microorganisms) a unique type of RNA editing occurs in many or most of the mitochondrial mRNAs. This involves insertion and deletion of uridylates by a complex of proteins (see EDITOSOME) directed by small 'guide' RNA molecules (gRNAs). Most of the gRNAs are encoded by DNA minicircles (which occur in the KINETOPLAST).

RNA extraction (from blood) See e.g. PAXGENE BLOOD RNA KIT.

RNA-induced silencing complex See RNA INTERFERENCE.

RNA interference (RNAi) A form of post-transcriptional gene silencing (PTGS) found in plant, animal and insect cells; the phenomenon appears to have evolved as a natural antiviral response which is triggered by the presence of intracellular double-stranded RNA (dsRNA). In this process, long molecules of dsRNA (more than several hundred nucleotides) are cut into regular-sized fragments by a dsRNA-specific RNase-III-like endonuclease referred to as DICER; these fragments of dsRNA, referred to as *small interfering RNA* (= siRNA: see SIRNA), are reported to be about 21–28 nucleotides in length, depending e.g. on organism. One of the strands of ds siRNA, in association with a multiprotein assembly (which includes a protein of the ARGONAUTE family), forms the *RNA-induced silencing complex* (RISC); RISC may then cleave a specific target mRNA. Hence, mRNAs of various specificities may be cleaved as a result of the activity of the various molecules of siRNA derived from the original, long molecule of dsRNA. (Note that the cleavage of a target mRNA is also a feature of many *microRNAs*: see the entry MICRORNAS.) Long dsRNAs can also trigger the *interferon* response in mammalian cells; this involves a dsRNA-dependent protein kinase (referred to as PKR) which can cause a generalized inhibition of viral and cellular protein synthesis.

siRNAs are also involved in the phenomenon referred to as *transcriptional gene silencing* (TGS): see the entry SIRNA.

Synthetic siRNA molecules are used for experimental gene silencing (see the entry SIRNA), and RNAi is one approach to GENE THERAPY (q.v.).

A SHORT HAIRPIN RNA can be synthesized within cells and then cleaved (*in vivo*) to form a functional siRNA. This can be arranged by using an shRNA-encoding vector – such as pBLOCK-iT^{TM}3-DEST (see the table in the entry GATEWAY SITE-SPECIFIC RECOMBINATION SYSTEM) – that is introduced into cells by an appropriate transfection procedure.

[RNAi (mechanisms, applications): BioTechniques (2008) 44(5):613–616.]

(See also LIGHT-ACTIVATED RNAI.)

RNAi in mosquitoes

For RNAi in mosquitoes see the entry ARBOVIRUSES.

Absence of RNAi in the malaria parasite (Plasmodium)

It has been reported that an RNAi system has not been found in the malaria parasite, *Plasmodium* – which apparently does not have the necessary enzymes [Nucleic Acids Res (2009) doi: 10.1093/nar/gkp239].

Selection of potent siRNAs from an siRNA library

siRNA libraries include certain siRNAs that are more potent than others, for a given target. A simplified method has been proposed for selecting potent siRNAs from an siRNA library. In this method, the target gene, in human embryonic kidney cells (HEK cells), was fused 3′ to the gene encoding cytosine deaminase. A culture of the recombinant HEK cells was then exposed to a population of *Escherichia coli* which had been transformed with an siRNA library. siRNA-encoding DNA was introduced into the HEK cells when they internalized the *E. coli* cells. (The efficiency of bacterial internalization in the HEK cells was reported as two/three bacteria per HEK cell.) The cells of *E. coli* lysed within the HEK cells, releasing the siRNA-encoding DNA for subsequent transcription. The cell culture was washed and re-fed with a medium that contained (i) gentamicin – to kill the extracellular bacteria – and (ii) 5-fluorocytosine.

If a given HEK cell contained an siRNA which has only weak activity against the target gene, then translation of the (two-gene) transcript could be unhindered – and this would result in production of cytosine deaminase and conversion of 5-fluorocytosine to the *lethal* substrate 5-fluorouracil; hence, HEK cells containing poorly inhibitory siRNAs were killed by the 5-fluorouracil. However, in HEK cells that contained potent siRNAs – which degraded the transcript – no cytosine deaminase could be produced – so that these cells survived. siRNA-encoding DNA was then extracted from the surviving cells for further study [Nucleic Acids Res (2009) 37(1):e8].

RNA interference pathway The gene-regulatory pathway that involves small RNA molecules such as the siRNAs (see RNA INTERFERENCE) and MICRORNAS – both of which have e.g. a developmental stage involving the DICER enzyme.

RNA isolation Separation of total RNA, or a specified fraction of RNA, from cells; thus, for example, poly(A)-tailed mRNA molecules may be isolated from a cell lysate by adsorption to a poly(T) matrix.

Isolation of RNA from blood may be carried out e.g. with the PAXGENE BLOOD RNA KIT.

The isolation of RNA from source material can be delayed, if necessary, for an extended period of time (e.g. weeks) by placing the material (e.g. tissue samples) in RNALATER (q.v.) and then storing at room temperature, 4°C or −20°C.

RNA ligase Any enzyme with RNA–RNA ligase activity.

RNA modification pathways Database MODOMICS [Nucleic

Acids Res (2009) 37(Database issue):D118–D121].

RNA PCR An early name for REVERSE TRANSCRIPTASE PCR.

RNA polymerase (RPase) Any enzyme which can polymerize ribonucleoside 5′-triphosphates on a DNA or RNA template; however, this term usually refers to a DNA-dependent RNA polymerase, i.e. an enzyme that uses a DNA template.

(See also entry RNA-DEPENDENT RNA POLYMERASE; cf. APTAZYME.)

In *Escherichia coli*, the bulk of RNA synthesis is carried out by a single multi-component RPase; in this enzyme, the three types of component (α, β, β′), together with a SIGMA FACTOR (σ), form the functional *holoenzyme*: $\alpha_2\beta\beta'\sigma$. This enzyme is sensitive to e.g. rifampicin.

The *E. coli* primase (*dnaG* gene product) is an RPase used for synthesis of the RNA primers that are extended by DNA synthesis to form the Okazaki fragments produced during the replication of DNA. Primase is resistant to rifampicin.

RPases of members of the ARCHAEA appear to resemble the eukaryotic (nuclear) RPases (see below) rather than bacterial enzymes.

In general, eukaryotes have three multi-component nuclear RPases:

RPase I (= RPase A) synthesizes most rRNA in the nucleolus.

RPase II (RPase B) synthesizes pre-mRNA – and at least some of the transcripts that mature into MICRORNAS.

RPase III (RPase C) synthesizes e.g. tRNAs and 5S rRNA. (It was proposed that a link may exist between RPase III activity and cancer [Nature Rev Cancer (2008) 8:911–914].)

At least one of these enzymes also synthesizes snRNAs for pre-mRNA SPLICING.

The (eukaryotic) RPases are not inhibited by rifamycins.

In eukaryotes, the mitochondrial and chloroplast RPases are distinct from the nuclear enzymes.

RNA-recognition motif See SR PROTEINS.

RNA sequencing See the note in MAXAM–GILBERT METHOD.

RNA splicing (pre-mRNA) See the entry SPLICING.

RNA staining See e.g. RIBOGREEN.

(See also DNA STAINING and ULTRAVIOLET ABSORBANCE.)

RNAi An abbreviation used for RNA INTERFERENCE.

RNA*later*™ An aqueous reagent (Ambion, Austin TX) used for storing tissue samples and other specimens from which RNA is to be subsequently extracted after e.g. days or weeks.

In this product, the samples may be stored at 25°C ('room temperature') for up to 1 week, at 4°C for up to 1 month, and at −20°C for an indefinite period.

When the RNA is required, the sample is removed from the reagent and is treated in the same way as a freshly harvested sample.

The use of this reagent has various advantages – e.g. (i) it protects cellular RNA from endogenous RNases, (ii) it allows extraction of RNA – at a convenient time – from samples collected at different times, and (iii) it permits the overnight transport of samples from one laboratory to another at 'room temperature'.

RNA*later*™ was used e.g. for the analysis of a coral larval transcriptome [BMC Genomics (2009) 10:219]; for studying the gene expression profile of bladder tissue [BMC Genomics (2009) 10:199]; and studies on iron-related genes in human brain tissue [BMC Neurosci (2009) 10:36].

RNAqueous™ technology A commercial approach (Ambion, Austin TX) used for isolating total RNA from various types of tissue or cell without the need for phenol–chloroform extraction. Essentially, tissues are disrupted by the chaotropic agent guanidinium thiocyanate – which also inactivates the endogenous nucleases. Following dilution of the lysate with an ethanol solution, the whole is transferred to a glass-fiber filter which binds RNA; three consecutive washes are used to remove proteins and DNA etc. and the RNA is then eluted.

Various products, incorporating RNAqueous™ technology, are designed for particular applications – e.g. several kits are suitable for rtPCR, and an automated kit is suitable for high-throughput working.

RNAqueous™ was used e.g. for studies on antisense transcription in *Arabidopsis* [PLoS Genet (2009) 5(4):e1000457]; studies on up-regulation of mRNA in neuropathy [Mol Pain (2009) 5:14]; studies on non-coding RNAs in *Caenorhabditis elegans* [Nucleic Acids Res (2009) 37(3):999–1009].

(See also NUCLEIC ACID ISOLATION.)

RNase RIBONUCLEASE: see the entries below for some specific examples, and see also BARNASE and BINASE.

(See the entry RIBONUCLEASE for general information.)

RNase III An endoribonuclease which cleaves double-stranded RNA (e.g. in hairpin and stem–loop structures). The enzyme is involved e.g. in the processing (maturation) of rRNA from precursor molecules (see e.g. RRN OPERON) and regulation of gene expression.

The (active) RNase III of *Escherichia coli* is a homodimer. dsRNA is cut into fragments which have 5′-phosphate and 3′-hydroxyl termini and short 3′ overhangs.

Factors that determine the cleavage sites of RNase III have been reported to include structure (e.g. helix length, often two helical turns) and sequence. Based on some earlier observations it was suggested that this enzyme might recognize cleavage sites by the *absence* of particular base pairs within discrete regions of double-helical RNA – referred to as the *proximal box* and the *distal box*. However, specific base pair sequence elements may act as (positive) recognition determinants [the characterization of sequence determinants and antideterminants of processing reactivity for a minimal substrate of *Escherichia coli* RNase III: Nucleic Acids Res (2006) 34:3708–3721].

(See also DICER and DROSHA.)

RNase binase See the entry BINASE.

RNase Cocktail™ A product (Ambion, Austin TX) consisting of a mixture of highly purified ribonucleases: RNase A and RNase T_1; it can be used e.g. for eliminating RNA in plasmid minipreps.

RNase E In *Escherichia coli*: an (endo)ribonuclease involved e.g. in the formation of 5S rRNA from pre-rRNA.

(See also BL21 STAR, CELL-FREE PROTEIN SYNTHESIS, DEGRADOSOME and RNASE P.)

In *Escherichia coli*, ribosomal protein L4 was reported to be involved in regulating the activity of RNase E [Proc Natl Acad Sci USA (2009) 106(3):864–869].

RNase H An RNase which (specifically) degrades a strand of RNA that is hybridized to a strand of DNA, e.g. the strand of RNA hybridized to FIRST STRAND cDNA. The RNase H from *Escherichia coli* is an endoribonuclease that forms products with 3′-hydroxyl and 5′-phosphate termini.

RNase H is used in the preparation of cDNAs, and is also used e.g. in NASBA.

The nuclease activity of retroviral reverse transcriptases (absent in a recombinant RT from Moloney murine leukemia virus) can also degrade the RNA in RNA/DNA hybrids.

RNase P An endoribonuclease – consisting of protein and RNA – involved in generating 5′ termini in tRNAs; its catalytic activity is mediated by the RNA moiety. [Catalysis: Nucleic Acids Res (2009) doi: 10.1093/nar/gkp775.]

[Human RNase P: Nucleic Acids Res (2007) 35(22):3519–3524.]

[*Escherichia coli* RNase P processes tRNA without RNase E: Nucleic Acids Res (2007) 35(22):7614–7625].

In contrast to bacterial RNase P (which contains one RNA component and one protein component), the RNase P of yeast (*Saccharomyces cerevisiae*) contains nine protein subunits as well as the RNA moiety. Despite this, one study has suggested that the minimal kinetic mechanism of cleavage of pre-tRNA is similar for both enzymes [RNA (2009) 15(2):224–234].

RNase protection assay See entry RIBONULEASE PROTECTION ASSAY.

RNase R (*bacteriology*) A strongly processive 3′-to-5′ exoribonuclease which appears to be multifunctional. In the Gram-negative species *Pseudomonas putida* the absence of RNase R was found to correlate with higher levels of many mRNAs – suggesting that it has a role in mRNA turnover [J Bacteriol (2008) 190(18):6258–6263].

RNase R has a role in TRANS TRANSLATION (q.v.).

RNase T In *Escherichia coli*: an exoribonuclease that specifically removes the terminal AMP residue from the 3′ end of an uncharged (non-aminoacylated) tRNA.

RNase T_1 (RNase T1) An endoribonuclease, encoded by the fungus *Aspergillus oryzae*, which specifically cleaves single-stranded RNA 3′ of guanosine residues (Gp↓N), thus forming guanosine 3′-phosphate and oligonucleotides with 3′ terminal phosphate.

(See G-LESS CASSETTE and MAXAM–GILBERT METHOD.)

RNA*secure*™ See RIBONUCLEASE.

rne131 See the entry BL21 STAR.

Robust-LongSAGE See SAGE.

rolling circle A mode of nucleic acid synthesis which occurs in some *circular* molecules of DNA and RNA.

In (ds) DNA molecules, the process begins with a NICK in one strand; the free 3′ end is then extended by a DNA polymerase, using the un-nicked strand as template. The 5′ end of the nicked strand is progressively displaced as polymerization continues.

Rolling circle synthesis occurs e.g. in many small circular bacterial plasmids, the nicking being mediated by a plasmid-encoded *Rep protein*; plasmid-length pieces of the displaced strand may then be circularized and converted to (dsDNA) plasmids.

In bacteriophage φX174 (which has a single-stranded, ccc DNA genome), nicking of the (double-stranded) parental replicative form (RF) is followed by rolling circle synthesis (involving the bacterial DNA polymerase); genome lengths of the displaced strand are excised, circularized and converted to progeny RFs.

In phage λ, rolling circle synthesis leads to the formation of a continuous multi-genome (double-stranded) CONCATEMER that is later cut at the (repeating) *cos* sites into genome-sized sections. (The concatemer is double-stranded as the displaced strand is used as a template for the synthesis of the complementary strand.)

Rolling circle synthesis is also involved in the replication of viroids: small circular ssRNA molecules found in plants (see VIROID).

In vitro, rolling circle synthesis has been used in various applications, e.g. for the transcription of a DNA NANOCIRCLE.

The efficiency and specificity of rolling circle amplification were reported to be improved by a mutant form of the SSB (single-strand binding protein) from the bacterium *Thermus thermophilus* [Nucleic Acids Res (2006) 34(9):e69].

Real-time, single-molecule observation of the rolling circle mode of DNA replication has been reported [Nucleic Acids Res (2009) 37(4):e27].

Rom protein See COLE1.

Rop protein See COLE1.

Rotor-Gene™ 6000 See HRM.

RPA (1) Recombinase polymerase amplification: an isothermal method for amplifying specific DNA sequences that involves recombinase-mediated binding of primers to targets coupled with strand-displacement DNA synthesis [PLoS Biol (2006) 4(7):e204].

Essentially, in the presence of ATP, the primers form complexes with molecules of the recombinase UvsX (from phage T4), and primer–recombinase complexes search the double-stranded sample DNA for sequences that are complementary to the primers. When primer-binding sites are located, the recombinase mediates strand exchange: the binding of a given primer displaces a loop of ssDNA which is stabilized by the binding of single-strand binding proteins (SSBPs).

Disassembly of the recombinase permits the bound primer to be extended by a (strand-displacing) DNA polymerase.

As primer extension continues, SSBPs stabilize the grow-strand of displaced DNA. Similar, repeated activity by both of the (opposing) primers leads to exponential amplification of the target.

Ongoing activity in RPA depends on the presence of agents

that mediate a balance in the development and disassembly of the primer–recombinase complexes.

Reactions in which the template is either absent or in low concentration were reported to lead to primer-dependent artefacts. To avoid this, the detection of products is achieved by a probe-based system; the probe includes an abasic-site mimic flanked by a fluorophore-labeled nucleotide and a quencher-labeled nucleotide. The *intact* probe exhibits low-level fluorescence, but hybridization of the probe to its complementary sequence permits the abasic-site mimic to be recognized, and cleaved, by *Escherichia coli* endonuclease IV (Nfo), resulting in a measureable rise in fluorescence, as the fluorophore is no longer quenched. The increase in fluorescence is exploited in monitoring products. As cleavage by Nfo depends on stable pairing between the probe and its corresponding binding site, this arrangement promotes specificity in product detection.

RPA is reported to be capable of multiplex operation, the method having been employed e.g. to amplify simultaneously targets from several different strains of methicillin-resistant *Staphylococcus aureus* (MRSA).

(2) Replication protein A (see e.g. OKAZAKI FRAGMENT).

RPase RNA POLYMERASE.

***rpmI* gene** See OPERON.

***rpoB* gene** See RIFAMYCINS.

***rpoD* gene** See SIGMA FACTOR.

***rpoH* gene** See ESCHERICHIA COLI (table).

***rpoS* gene** See SIGMA FACTOR.

***rpsL* gene** (*syn. strA* gene) In *Escherichia coli*: a gene encoding the ribosomal protein S12; mutation in this gene may lead to resistance to the antibiotic streptomycin (for which S12 is the target protein).

RRBS (1) REDUCED REPRESENTATION BISULFITE SEQUENCING.

(2) Relative reciprocal BLAST score (used in the context of bioinformatics).

RRE Rev protein response element: see REV PROTEIN.

***rrn* operon** (in bacteria) An OPERON encoding the three species of rRNA (16S, 23S, 5S) which is transcribed (as a single pre-rRNA transcript) in the order 16S→23S→5S. The genes are not contiguous: a so-called *intergenic spacer region* (ISR) is found between each pair of coding sequences. The length of an ISR can vary, even in different copies of the *rrn* operon in the same chromosome, and this feature has been exploited in certain forms of TYPING (PCR-RIBOTYPING and RIBOTYPING).

Bacteria usually contain multiple copies of the *rrn* operon, but single copies are reported to occur e.g. in *Mycobacterium tuberculosis* and *Tropheryma whipplei*.

Escherichia coli contains seven copies of the *rrn* operon. The pre-rRNA molecule forms secondary structures (through base-pairing between inverted repeats) in which the 16S and 23S components form loops in separate stem–loop structures; the 16S and 23S precursor molecules are liberated by the action of an endoribonuclease (RNase III).

(See also LEUCINE-RESPONSIVE REGULATOR PROTEIN.)

RSV Respiratory syncytial virus: see PARAMYXOVIRIDAE.

RT-PCR (rt-PCR, rtPCR) REVERSE TRANSCRIPTASE PCR.

rt-PCR (RT-PCR, rtPCR) REVERSE TRANSCRIPTASE PCR.

Rta (*brlf-1* gene product) See the entry ZEBRA.

rtPCR (rt-PCR, RT-PCR) REVERSE TRANSCRIPTASE PCR.

r*Tth* DNA polymerase A 94-kDa recombinant DNA polymerase (Perkin Elmer/Applied Biosystems) that is derived from the thermophilic bacterium *Thermus thermophilus*.

In a special buffer system this enzyme can behave as both a DNA-dependent DNA polymerase and a reverse transcriptase.

The enzyme can function in the range 60–70°C for reverse transcription, and it is used e.g. in REVERSE TRANSCRIPTASE PCR.

(cf. RTTH DNA POLYMERASE XL.)

Examples of use: studies on the use of siRNA for inhibiting HCV genotype-4 replication [Virol J (2009) 6:13], and (as a reverse transcriptase) for studies on hepatitis C virus (HCV) [World J Gastroenterol (2009) 15(2):240–244].

r*Tth* DNA polymerase XL A recombinant, hybrid DNA polymerase (Perkin Elmer/Applied Biosystems) which is derived from the polymerases of two thermophilic bacteria – *Thermus thermophilus* and *Thermococcus litoralis*.

This enzyme was designed to have optimized synthesis and PROOF READING functions, and has been used e.g. for LONG PCR.

[Uses (e.g.): Am J Pathol (2008) 172(2):417–429; Int J Biol Sci (2008) 4(6):397–405; PLoS ONE (2009) 4(4):e5304; and PLoS ONE (2009) 4(1):e4160.]

RyhB A bacterial sRNA: see SRNAS.

S

S (1) A specific indicator of ambiguity in the recognition site of a RESTRICTION ENDONUCLEASE, or in any other sequence of nucleic acid; thus, for example, in ↓GTSAC (enzyme Tsp45I) the 'S' indicates G or C.

(2) L-Serine (alternative to Ser).

***S*-adenosyl-L-methionine** A methyl group (CH_3–) donor that is used in various reactions.

S end (of phage Mu DNA) See PHAGE MU.

***S* gene** (of phage λ) A gene whose product is involved in lysis of the host cell during the lytic phase. A mutant form of the *S* gene is used in some lambda-based vectors in order to delay lysis and (hence) to increase the yield by allowing more time for phage maturation.

(See also PHAGE LAMBDA.)

S phase (of the cell cycle) That phase of the (eukaryotic) cell cycle in which the genome is replicated.

(See also RE-REPLICATION.)

S_1 nuclease *Syn*. ENDONUCLEASE S1 (q.v.).

10Sa RNA *Syn*. tmRNA – see the entry TRANS TRANSLATION.

Saccharomyces A genus of YEASTS; the species *S. cerevisiae* is widely used as an experimental organism (and is also used in the manufacture of alcoholic beverages and fermented products).

S. cerevisiae occurs as single cells which may be haploid or diploid, depending on the stage of the life cycle. Both haploid and diploid cells reproduce asexually by a process known as *budding* (cf. SCHIZOSACCHAROMYCES).

Sexual reproduction involves fusion of two (haploid) cells of appropriate MATING TYPE. The diploid zygote gives rise to diploid somatic (vegetative) cells. Diploid cells form a major phase in the life cycle. Diploid cells undergo meiosis to form four *ascospores* within a structure called an *ascus*; two of the ascospores are of the α mating type, while two are of the **a** mating type. Ascospores develop into haploid somatic cells.

The existence of both haploid and diploid somatic cells in the life cycle has been exploited in complementation studies in which e.g. the diploid cells can be used to study the effects of different alleles derived from parental haploids.

Cells of the AB1380 strain of *S. cerevisiae*, and derivatives, have been used e.g. as hosts for YEAST ARTIFICIAL CHROMOSOMES. These cells have mutations in genes *TRP1* and *URA3* and require both tryptophan and uracil in growth media; the corresponding (wild-type) genes in YACs are therefore useful for the selection of YAC-containing cells (in tryptophan-deficient and uracil-deficient media). The cells of this strain also have a mutation in a gene whose product is involved in adenine metabolism, leading to *red pigmentation*; this defect is bypassed by the suppressor gene *SUP4* which, in a YAC, is involved in a mechanism for indicating the presence/absence of an insert (see YEAST ARTIFICIAL CHROMOSOME).

S. cerevisiae is used in the YEAST TWO-HYBRID SYSTEM for studying protein–protein interactions.

[Yeast-based technologies in genomics and proteomics (a review): BioTechniques (2006) 40(5):625–644.]

Some relevant entries: ARS, CSRE, CYTOTRAP TWO-HYBRID SYSTEM, FLP, GAL1, GENOME DELETION PROJECT, HIRUDIN, I-SCEI, INTEIN, KILLER FACTOR, TY ELEMENT and UAS.

***Saccharomyces cerevisiae* Genome Deletion Project** See entry GENOME DELETION PROJECT.

SafeBis DNA DNA which has been treated with BISULFITE but which has not been subsequently desulfonated; SafeBis has been used e.g. in studies on DNA methylation [Nucleic Acids Res (2007) 35(1):e4].

safety cabinet (*syn*. sterile cabinet) A cabinet which provides a partially or completely enclosed space in which certain procedures can be carried out; some cabinets are designed specifically to provide effective containment of dangerous pathogens – with the object of protecting laboratory staff and the environment.

Class I cabinets are accessed for working via an opening in the front, below a glass viewing panel. A fan draws air into the cabinet, via a front opening, and air leaves the cabinet via a high-efficiency particulate air filter (a HEPA filter); the rate of inflow of air is such that aerosols do not leave the cabinet via the front opening.

Class I cabinets are used e.g. for work on pathogens such as *Coxiella burnetii*, *Francisella tularensis* and *Mycobacterium tuberculosis*. In addition to the use of a class I cabinet, work with this type of pathogen requires the use of gloves and also the existence of *plenum* ventilation in the laboratory – i.e. a constant inflow of air to the laboratory, with filtration of the exhausted air through a HEPA filter.

Class II cabinets are used primarily to protect the materials in the cabinet from environmental contamination. Filtered air constantly flows down onto the work surface and leaves the cabinet after further filtration. Access is from the front, below a glass viewing panel. These cabinets (which are common in microbiological laboratories in universities) are used e.g. to avoid contamination when inoculating media.

Class III cabinets are high-security cabinets that are totally enclosed and gas-tight. Air is filtered before entry and before discharge. Manipulation of materials etc. within the cabinet is conducted by means of (arm-length) rubber gloves that are fitted securely into the front of the cabinet. The interior of the cabinet is viewed via a glass panel. Access to the cabinet is via a separate two-door sterilization/disinfection chamber.

Class III cabinets are used when handling hazardous pathogens such as the Ebola virus, Lassa fever virus and Marburg virus.

SAGE Serial analysis of gene expression: a method for detecting, identifying and quantitating genes expressed by a given type of cell, under given conditions; it involves the analysis of a collection of short (~10-nucleotide) sequences (referred to as SAGE tags or diagnostic tags) – each tag being derived

from the transcript (mRNA) of a given gene. SAGE may be used without prior knowledge of the nucleotide sequences of genes; a particular gene may be identified via its diagnostic tag.

SAGE is used e.g. for discovering new genes, for studies on the up- and down-regulation of genes under different conditions, and for establishing links between various aspects of a cell's physiology; in the latter context, studies on transcript levels were able to reveal e.g. links between phosphate metabolism and other aspects of physiology in the smut fungus *Ustilago maydis*.

The basic procedure for SAGE is essentially as follows.

Initially, total mRNA is isolated from the cells or tissue and is reverse transcribed (with BIOTINylated oligo(dT) primers) to cDNAs. Double-stranded cDNAs are then prepared.

Double-stranded cDNAs are digested with a RESTRICTION ENDONUCLEASE – such as NlaIII – which cleaves the cDNA molecules more than once, i.e. a 'frequent-cutting' enzyme is used. (This enzyme is often called the 'anchoring enzyme'.)

The *biotinylated ends* of the cleaved cDNA molecules are captured e.g. by STREPTAVIDIN-coated DYNABEADS.

The Dynabead-bound fragments are then separated into two subpopulations.

In one of the subpopulations, the free end of each fragment (i.e. the end formed by cleavage with the anchoring enzyme) is ligated to a dsDNA oligonucleotide linker (linker A); this linker contains the recognition site of a type IIS RESTRICTION ENDONUCLEASE.

In the second subpopulation, the free ends are ligated to a different linker (linker B); however, this linker also contains the recognition site of the same type IIS restriction enzyme.

Both subpopulations are then subjected to the activity of the type IIS restriction enzyme. This enzyme cuts some distance from its recognition site; this generates a number of free fragments (released from the Dynabeads), each of which consists of (i) the linker, and (ii) a short sequence of nucleotides (9–14 bp) from the (contiguous) cDNA fragment. The sequence of nucleotides derived from a given cDNA fragment is called the *signature tag* of the corresponding gene.

Any overhang produced by type IIS restriction is filled in e.g. by the Klenow fragment.

The two subpopulations are mixed.

The tags are ligated together by blunt-end ligation. Hence, in the reaction mixture, structures of the following type are formed:

$$\text{linker A–(signature tag)}_2\text{–linker B}$$

These structures are amplified by PCR – using primer-binding sites in each of the two types of linker.

The PCR products are subjected to the restriction enzyme used initially to cleave the cDNAs (the anchoring enzyme). This results in the excision of both (terminal) linkers, leaving a pair of ligated signature tags (the *ditag*).

Ditags from the reaction mixture are ligated serially to form long molecules of DNA, each of these molecules containing many tags. These molecules can be cloned and sequenced.

The relative abundance of any given mRNA in the *original* sample is reflected in the number of corresponding tags in the final SAGE product (i.e. the long molecule of ligated ditags). A signature tag identifies a particular gene – so that the level of expression of a given gene can be assessed from the final SAGE product.

Computer-based analysis of sequence data (from the final product) can be used to identify particular genes.

Various modifications of the basic scheme described above have been devised:

MicroSAGE uses minute quantities of sample material. The principle is similar to that described, although e.g. mRNAs in the sample bind directly to oligo(dT)-coated magnetic beads (i.e. prior to reverse transcription).

LongSAGE uses a different restriction enzyme (e.g. MMEI) to obtain a longer signature tag (this enzyme cuts deeper into the contiguous cDNA fragment).

Robust-LongSAGE offers a further series of modifications.

Variations in the method also include 5′ LongSAGE and 3′ LongSAGE.

SuperSAGE uses a type III restriction enzyme (EcoP15I) to increase the size of tags to 26 bp; superSAGE facilitates the identification of a gene from databases such as GenBank®.

salicylate (inducer of gene expression) Salicylate has been used to induce gene expression in a commercial method (see the CASCADE EXPRESSION SYSTEM), and was used in a control circuit for gene regulation in *Salmonella* [Nature Methods (2007) 4:937–942].

Salmonella A genus of Gram-negative, rod-shaped, generally motile bacteria of the family Enterobacteriaceae (see entry ENTEROBACTERIA); the species *S. typhi* is the causal agent of typhoid.

The GC% of chromosomal DNA is 50–52.

The salmonellae (i.e. members of this genus) can be grown on (e.g.) nutrient agar and MacConkey's agar; media used for enrichment include e.g. selenite broth and tetrathionate broth. Most strains do not ferment lactose.

Linguistic note

The name *Salmonella* derives from the name of an American bacteriologist, D. E. Salmon; the correct pronunciation of the genus name is therefore 'Salmon-ella' – the first 'l' is silent.

***Salmonella*/microsome assay** *Syn.* AMES TEST.

SAM *S*-adenosyl-L-methionine (a methyl donor).

same-sense mutation Any mutation which fails to change the identity of the amino acid that is specified by a given codon.

A same-sense mutation may not be equivalent to a SILENT MUTATION if it introduces a CODON BIAS.

SAMPL Selective amplification of microsatellite polymorphic loci, *or* selectively amplified microsatellite polymorphic loci: a method for detecting variations (in length) of microsatellite sequences within genomic DNA.

The genomic DNA is initially digested with two types of RESTRICTION ENDONUCLEASE, using enzymes that do not cut

within the particular microsatellite sequence of interest. The resulting fragments are ligated to adaptors. Pre-amplification is carried out by PCR with primers having a single selective 3′ nucleotide.

Another round of PCR amplification is carried out with one of the former primers and a (labeled) SAMPL primer that is complementary to the given microsatellite sequence; the two primers in the second round of PCR therefore amplify those microsatellite sequences present in amplicons formed during pre-amplification. After gel electrophoresis, the products can be sequenced – the location of a given microsatellite DNA in the genome being indicated by its flanking sequence(s).

sampling (DNA) See, for example, BUCCAL CELL SAMPLING and FORENSIC APPLICATIONS.

Sandhoff disease A GENETIC DISEASE similar to TAY–SACHS DISEASE (q.v.) but resulting from a deficiency in the activity of the β-subunit of hexosaminidase A.

Sanger's method (DNA sequencing) See DIDEOXY METHOD.

SapI A (type IIS) RESTRICTION ENDONUCLEASE; recognition/cutting site:

5′-GCTCTTCN↓NNN
3′-CGAGAAGNNNN↑

SalI is used e.g. in the STICKY RICE procedure.

Sapphire-II™ See CHEMILUMINESCENCE ENHANCER.

saquinavir See PROTEASE INHIBITORS.

sarcoma A type of CANCER which develops from connective tissue (cf. CARCINOMA).

SARS virus (detection) A rapid, sensitive method for detecting the SARS virus (an ssRNA coronavirus) in clinical samples exploited (genome-specific) PADLOCK PROBES; on circularization, these probes act as templates for isothermal replication of target sequences by the ROLLING CIRCLE mechanism.

SATB1 Special AT-rich sequence-binding protein 1: a protein which binds to MATRIX-ATTACHMENT REGIONS and mediates CHROMATIN remodeling and regulation of gene expression; when bound to a matrix-interacting MAR, SATB1 can recruit various factors, e.g. histone acetylases (HATs) and/or histone deacetylases (HDACs), that may affect gene expression e.g. by modifying histones or remodeling nucleosomes.

SBE See SINGLE-BASE EXTENSION.

SBP2 SECIS binding protein 2: see the entry SECIS.

scaffolding proteins (in phage assembly) Proteins which form a temporary (structural) framework during assembly of (e.g.) a phage head; they do not form part of the completed phage. The scaffolding proteins may be subsequently degraded – as e.g. in phages λ and T4; they are re-cycled e.g. in phages P22 and φ29.

scanning force microscopy *Syn*. ATOMIC FORCE MICROSCOPY.

SCC*mec* In strains of MRSA: a mobile genetic element (name apparently derived from staphylococcal cassette chromosome *mec*) that includes the MECA GENE complex and the *ccr* gene complex.

Different versions of SCC*mec* have been found in environmental isolates of MRSA, and SCC*mec* elements have been classified into six types (I–VI) on the basis of differences e.g. in the recombinase gene complex and in a gene complex that includes regulatory factors of the *mecA* gene.

SCC*mec* has been exploited for TYPING isolates of MRSA.

[SCC*mec*-based typing of MRSA: J Antimicrob Chemother (2009) 63(5):873–876; Emerg Infect Dis (2009) 15(5):727–734.]

SCE Single-cell ELECTROPORATION (q.v.).

SCF (stem cell factor) See KIT.

scFv (or scFV) A SINGLE-CHAIN VARIABLE FRAGMENT (q.v.).

Schizosaccharomyces A genus of fungi in which the morphological forms are single cells that reproduce by *fission* (rather than *budding*: cf. SACCHAROMYCES) or hyphae that fragment into spores. Some strains of these fungi are homothallic (i.e. self-fertile); others are heterothallic.

Studies on the effects of mutations in replication-initiation genes, and checkpoint genes, in the species *S. pombe* found that the resulting replication stress, and inappropriate mitosis, resulted in the production of reactive oxygen species (ROS) and cell death [J Cell Sci (2006) 119(1):124–131].

The species of *Schizosaccharomyces* are distinguished e.g. by the number of ascospores formed in the asci and by the ability to ferment specific sugars.

schlieren system A system used for representing a refractive index gradient.

Schneider's medium A liquid medium used for the culture of insect cells; it includes various inorganic salts (e.g. magnesium sulfate, potassium chloride, sodium chloride), D-glucose, trehalose, malic acid, succinic acid, and a number of amino acids.

SCID Severe combined immunodeficiency: see entry GENETIC DISEASE (table).

SCIDX1 See GENETIC DISEASE (table).

(See also INSERTIONAL ONCOGENESIS.)

scintillation proximity assay See SPA.

SCNT Somatic cell nuclear transfer: removal of the chromosomes, and polar body, from an unfertilized oocyte (a process known as *enucleation*) and their replacement with chromosomes from a somatic (donor) cell.

An oocyte which has undergone SCNT may be used e.g. to produce an embryonic clone which may serve as a source of embryonic stem cells (see STEM CELL).

As an alternative, an SCNT-treated oocyte may be used to initiate the process of whole-animal cloning, as follows.

Oocytes are collected and matured *in vitro* to metaphase II. The metaphase plate, and the POLAR BODY, are then removed (*enucleation*) and discarded, forming a *cytoplast*. Enucleation can be carried out by micromanipulation. (More recently, the procedure referred to as *hand-made cloning* has been used; this does not require the use of micromanipulation [example of use: J Assist Reprod Genet (2008) 25(1):13–16].)

Note: an oocyte used for SCNT must be a *mature* oocyte, and its ability to participate in SCNT is limited to a narrow timeframe.

The next stage is the formation of a 'reconstructed embryo', a structure in which the cytoplast encloses the chromosomes of a somatic cell from the donor animal (i.e. the animal to be cloned). (The somatic cell should be in the same phase of the cell cycle – the G_1 or G_0 phase – as the oocyte from which the cytoplast was derived.)

The reconstructed embryo is then activated – by a chemical and/or electrical procedure – in order to initiate development (in culture) to a BLASTOCYST (q.v.).

The blastocyst is transferred to a surrogate female animal for further development into live offspring (i.e. a *clone* of the donor animal).

Since the original use of this method (to produce 'Dolly the sheep' [Nature (1997) 385:810–813]), it has been used for the cloning of various other species of mammal – including cat, dog, horse, mouse and rabbit.

(See also RECLONING.)

Scorpion probe A PROBE analogous to a MOLECULAR BEACON PROBE used for monitoring the products of PCR [e.g. Nucleic Acids Res (2000) 28(19):3752–3761]. The probe resembles a molecular beacon, with fluorophore and quencher in the stem section – but it has a single-stranded 3′ extension. During the (heat) denaturation stage, both the target DNA *and* the stem section of the probe become single-stranded; on cooling, the probe's 3′ extension binds to target DNA and is subsequently extended, copying the target sequence. During the subsequent denaturation and cooling, the *loop* region of the probe hybridizes with a complementary sequence in the probe's extended 3′ region; in this configuration, the fluorophore and quencher are separated – so that the probe exhibits fluorescence under suitable excitation.

***Sc*Pol4** See the entry X FAMILY.

***Sc*Rad51** The RAD51 protein of *Saccharomyces cerevisiae*.

screening Any *selective* procedure in which individual cells or molecules with given characteristics are distinguished and/or isolated from the other individuals in the same population.

For examples of screening see BLACK–WHITE SCREENING, blue–white screening (described in the entry PBLUESCRIPT), COLONY HYBRIDIZATION, ELISPOT, METAGENOME, RECACTIVE, SIGEX and SSCP ANALYSIS.

screening test (*immunol.*, *microbiol.*) (1) Any test that is used to examine specimens in order to detect a particular type of antibody, antigen or microorganism. Screening tests are often simple and inexpensive but are frequently not highly specific. As a positive result in a screening test may be followed-up by other – more specific – test(s), a proportion of false-positive results may be tolerated in a screening test.

Examples of nucleic-acid-based screening tests include e.g. AMP CT and LINE PROBE ASSAY.

(2) Any test designed to assign a given, unknown organism to one of a possible range of categories or taxa.

SCS system Separate-component-stabilization system: see the entry CCD MECHANISM.

SCS110 A strain of *Escherichia coli* (Stratagene, La Jolla CA) which lacks the *dam* (adenine methylase) and *dcm* (cytosine methylase) functions (see DAM GENE and DCM GENE); it can be used e.g. for cloning DNA which is to be subsequently cut by a methylation-sensitive restriction enzyme (such as XbaI).

SD See the entry SEGMENTAL DUPLICATION.

SDA Strand displacement amplification: an isothermal method for copying a given sequence of nucleotides in a sample of DNA. (See also NUCLEIC ACID AMPLIFICATION.) Originally, SDA was carried out at an operating temperature of ~40°C; later, the operating temperature was raised to 50°C or above (thermophilic SDA, tSDA) with the object of improving the specificity of the method.

SDA – shown diagrammatically in the figure – is explained in a step-wise fashion in the legend. Part I of the figure shows the target-generation phase; Part II shows the amplification phase.

SDA is characterized by two distinctive features:

First, the use of a particular RESTRICTION ENDONUCLEASE that repeatedly *nicks* a (chemically modified) recognition site in the double-stranded intermediate products; this continually regenerates sense and antisense strands which are displaced from the dsDNA products by a strand-displacing DNA polymerase.

Second, use of a strand-displacing DNA polymerase (such as the Klenow fragment: see DNA POLYMERASE I) that is able to displace an existing strand from dsDNA by extending the 3′ terminus at a nick upstream of that strand – and which can also displace a strand by extending the 3′ end of the *bumper primers* B_1 and B_2 (shown in the diagram). Bumper primers are used only in the first phase of SDA (the target generation phase); they 'kick-start' events by displacing strands produced by extension of the primers S_1 and S_2 on the initial *target* DNA (e.g. genomic DNA) – B_1 and B_2 have no binding sites on amplicon-length products.

Restriction endonuclease

The restriction endonuclease used in SDA is HincII, and the specific recognition site employed is:

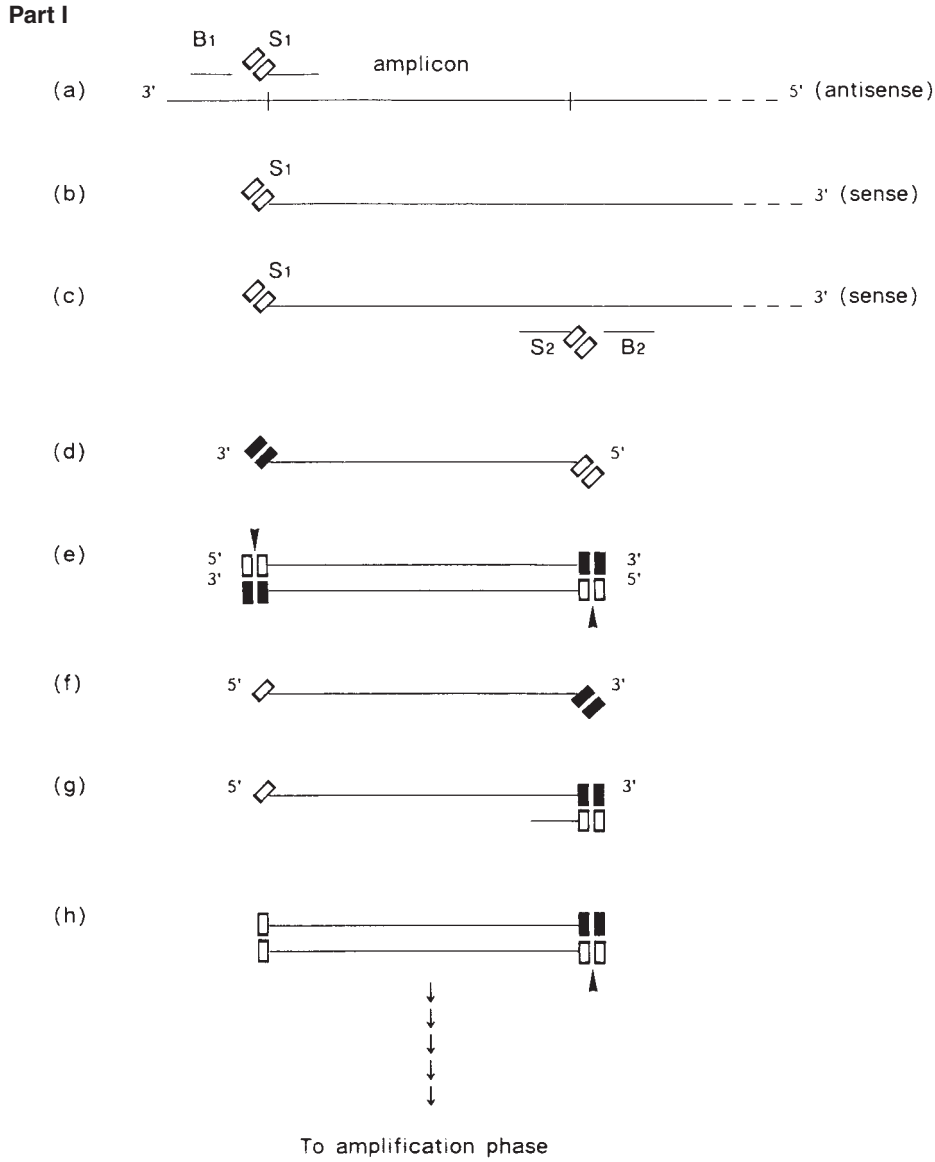
5′-GTT/GAC-3′
3′-CAA/CTG-5′

Normally, HincII makes a double-stranded (blunt-ended) cut, as shown by the oblique (/), but this activity does not occur in SDA because, *in each recognition site*, one of the two strands (▮▮ in the figure) is chemically modified and cannot be cut; the other strand (□□ in the figure) is *nicked* by HincII.

Both of the primers, S_1 and S_2, have a 5′ terminal tag: 5′-GTTGAC-3′; this *unmodified* sequence – when present in a (duplex) recognition site – can be nicked by HincII between the 'T' and 'G' nucleotides.

In an ongoing reaction, a *nickable* recognition site originates either from a primer (which contains unmodified nucleotides), or it is formed by the regeneration of a strand through 3′ extension from a nicked site (see below).

An *un-nickable* strand in a recognition site arises because, in SDA, all strand synthesis occurs in a reaction mixture that



SDA (strand displacement amplification) – part I: the target-generation phase (diagrammatic).

Target DNA is initially denatured to the single-stranded state prior to SDA.

(a) The antisense (3′-to-5′) strand of the target duplex is shown with the amplicon (i.e. the sequence of nucleotides to be amplified) delimited by two short vertical bars. For clarity of presentation, and economy of space, only one strand of the target DNA is considered here;

contains a chemically modified form of deoxyadenosine triphosphate: α-thiophosphoryl dATP (dATPαS). As a consequence, when a complementary strand is synthesized from a primer (see figure, Part I (c) and (d)), the newly synthesized strand (ending in 3′-CAACTG . . . . . .) includes two modified nucleotides, in the AA positions, which are immediately adjacent to the cutting site; because of the presence of the two modified nucleotides, *this* strand cannot be cleaved by HincII. When a primer binds to the 3′-CAACTG . . . . . sequence, it forms a so-called 'hemi-modified' site (or *hemiphosphorothioate* site) which is susceptible to nicking, only, by HincII in the *primer* strand between 'T' and 'G' .

Later, during the amplification phase, a hemi-modified site is formed when the (recessed) 3′ end of a template strand is extended (see figure, Part II (c), (d) and (f), (g)).

In the amplification phase, a nick between the 'T' and 'G' is followed by extension from the 3′ end of the nick, and this displaces the strand downstream of the nick. *Importantly*, the strand formed by such extension (5′-GTT→) incorporates a modified nucleotide (dATPαS) at the 'A' site in the sequence GAC – yet, *despite this*, the HincII site that is regenerated in the newly synthesized strand (i.e. 5′-GTTGAC. . ..3′) is still nickable because the *one* modified nucleotide (in GAC) does not interfere with the nickability of the sequence GTTGAC.

SDA has been used e.g. for the identification of *Chlamydia trachomatis* (BD ProbeTec™ ET assay; Becton Dickinson, Franklin Lakes NJ, USA) [BMC Infect Dis (2009) 9:16].

SDS SODIUM DODECYL SULFATE.

SDS–PAGE A special form of GEL ELECTROPHORESIS (sodium dodecyl sulfate–polyacrylamide gel electrophoresis) which is used e.g. for determining the molecular weight of a protein.

For determining molecular weights, proteins are denatured with SODIUM DODECYL SULFATE – an anionic detergent that binds (cooperatively) to proteins and gives rise to rod-like SDS–polypeptide complexes in which surface charges on the protein are masked by SDS. Most proteins bind SDS in proportion to their molecular weight, so that the electrophoretic characteristics of a given protein will be related to its molecular weight. Electrophoresis is carried out with the proteins under study – with one or more proteins of known molecular weight which act as standards for comparison.

For separating proteins in a mixture, maximum resolution can be achieved by using the Laemmli system in which the proteins in a sample are first compressed into a thin band by electrophoresis within a *stacking gel* before entering the so-called *resolving* (= *running* or *separating*) *gel* in which they are subjected to the sieving action of a gel with a smaller pore size.

Seamless® cloning kit A kit (Stratagene, La Jolla CA) that can be used e.g. for expression cloning, for constructing chimeric genes, and for site-specific mutagenesis; it involves facilitated insertion of the fragment of interest into a circular vector.

Initially, the fragment of interest, and the functional region

SDA (*continued*)

corresponding events occur on the complementary strand. A primer (S_1) has bound at the 3′ end of the amplicon. The 5′ end of this primer carries a tag that includes one strand of the HincII recognition sequence (5′-GTTGAC-3′) (shown by the symbol □□). A bumper primer (B_1 – see entry) has bound upstream of primer S_1. Extension of S_1 will produce a sense strand on the antisense template; this sense strand will be displaced by the extension of B_1.

(b) The sense strand which has been displaced, by bumper primer B_1, from the original template strand.

(c) Primer S_2 has bound to the 3′ end of the amplicon on the sense strand. Like primer S_1, its 5′ end is tagged with the recognition sequence of HincII. The bumper primer B_2 has bound upstream of primer S_2. Extension of S_2 will produce an antisense copy of the amplicon which is tagged at both ends with a (single-stranded) HincII recognition sequence; this strand will be displaced by bumper primer B_2.

(d) The antisense strand which has been displaced from stage (c) by bumper primer B_2. Notice that the HincII sequence at the 3′ end of this strand, having been synthesized with dATPαS (see entry), is 'modified' (as indicated by the symbol ▮▮); this (modified) HincII sequence (when subsequently part of a double-stranded amplicon) is not susceptible to the restriction endonuclease HincII. At the other (5′) end of the strand, the HincII sequence is not modified because it originated in primer S_2 (which was synthesized with normal nucleotides). Primer S_1 (not shown) can bind at the 3′ end of this strand – the HincII sequence in S_1 hybridizing with the modified sequence in the strand. Extension of S_1 yields the double-stranded amplicon shown at (e).

(e) Each end of this amplicon has a 'hemi-modified' (hemiphosphorothioate) recognition sequence for HincII. A hemi-modified site can be nicked in the *non*-modified strand (as shown by the arrowheads). Nicking can therefore occur in the 'upper' strand or the 'lower' strand of this (double-stranded) amplicon. For clarity, nicking in only the *upper* strand is considered here. Nicking is followed by extension of the 3′ end of the nick – resulting in displacement of the nicked strand, which is shown at (f).

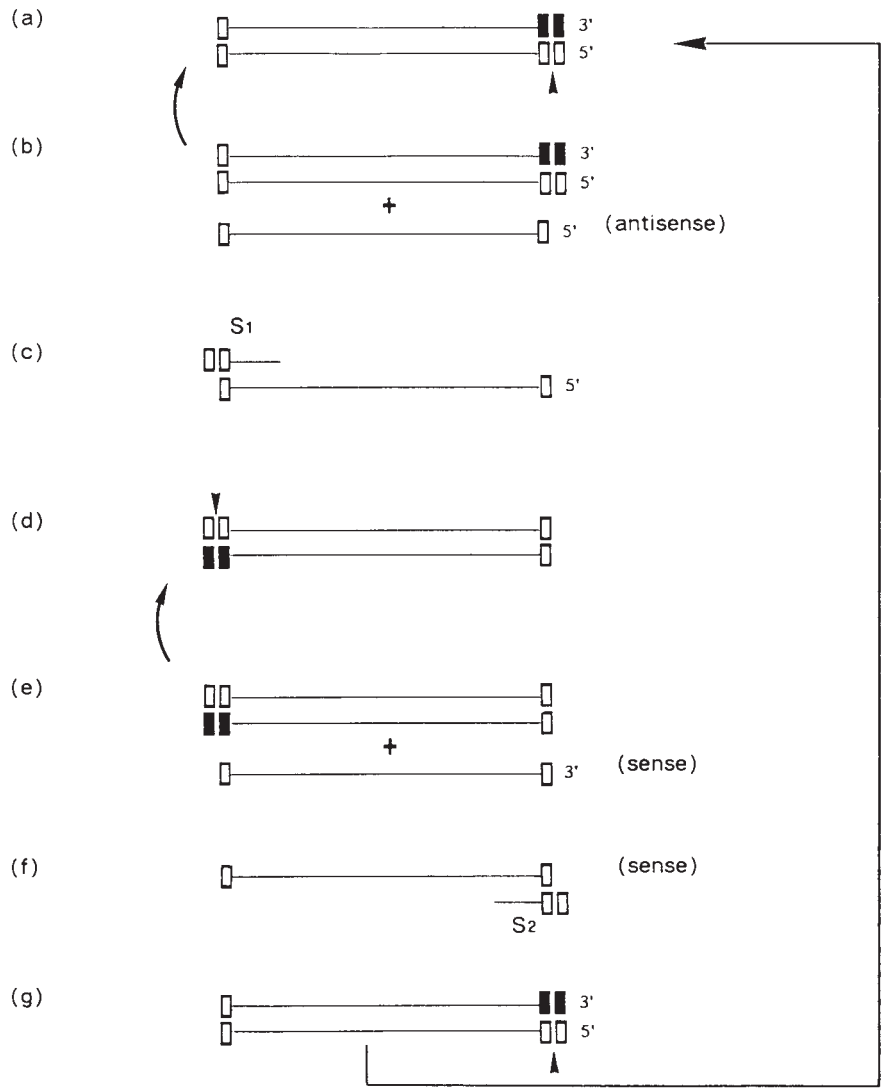
(f) The strand displaced from (d) following nicking and extension of the 3′ end of the nick.

(g) Primer S_2 has bound to the displaced strand and will be extended to form the double-stranded product shown at (h).

(h) This product feeds into the amplification phase (see part II of the figure).

In the above scheme, at stage (e), nicking of the *upper* strand, only, was considered. Nicking of the *lower* strand, and its displacement, produces a single-stranded product that binds primer S_1; extension of primer S_1 leads to the formation of a double-stranded product that feeds into the amplification phase.

Part II



SDA (strand displacement amplification) – part II: the amplification phase (diagrammatic).

(a) A double-stranded product derived from the target-generation phase. In this product the antisense (3′-to-5′) strand contains a nickable HincII restriction site (indicated by the arrowhead); this nickable site was derived from the tag in primer S_2 (see part I (g)). When this restriction site is nicked, DNA synthesis proceeds from the 3′ end at the nick site; during synthesis, the polymerase displaces the downstream section of the strand, producing an antisense copy of the target sequence which is flanked, on each side, by half a restriction

of the vector, are both amplified (separately) by PCR with the use of specialized primers. Each primer has a 5′ terminal 'tag' containing the recognition sequence of a (type IIS) restriction endonuclease, Eam1104I (see table in the entry RESTRICTION ENDONUCLEASE); this enzyme cuts at a fixed distance from its recognition sequence, a feature which allows the fragment to be cloned without introducing additional nucleotides.

The PCR reaction uses a (high-fidelity) DNA polymerase that minimizes the introduction of mutations.

The last five cycles of the PCR reaction are carried out with 5-methyl-dCTP, thereby modifying, and thus protecting from subsequent cleavage, any Eam1104I sites which happen to be within the fragment or vector sequence.

After PCR, the (linear, double-stranded) products from both the fragment and the vector are mixed and digested with the restriction enzyme Eam1104I. Eam1104I binds to each of the 5′ termini (at each end of the dsDNA amplicons) and cleaves that part of the amplicon derived from the primer's 5′ tag – i.e. that part containing the enzyme's recognition and cutting sites; thus, these sequences are not in the final products. The final products consist of (linear) copies of the insert and the vector, all of which have compatible sticky ends and are able to give rise to circularized molecules.

Circularized molecules, containing the vector and fragment components, are ligated with T4 DNA ligase and can be used to transform cells of *Escherichia coli*.

Chimeric genes may be constructed by using two different fragments, each fragment contributing to the final product.

SECIS The selenocysteine insertion sequence: a *cis*-acting sequence, within an mRNA molecule, which directs the insertion of a selenocysteine residue at a UGA codon. (An mRNA that contains a SECIS sequence is called a *selenoprotein mRNA* or a *selenocysteine mRNA*.)

(See also SELENOPROTEOME.)

In bacteria, the SECIS sequence appears to occur normally immediately downstream of the relevant UGA codon.

In (at least) a number of eukaryotes (including humans) the SECIS sequence is located in the 3′ untranslated region (3′-UTR) of a selenoprotein mRNA. In these eukaryotes, the so-called 'translational recoding' of UGA, i.e. recoding to change from a stop codon to a codon which specifies selenocysteine, depends not only on a SECIS element in the mRNA but also certain *trans*-acting factors which include a specific tRNA, a specialized elongation factor, and the *SECIS binding protein 2* (SBP2).

Interestingly, in (the protozoan) *Euplotes crassa*, the UGA codon in selenoprotein mRNAs was reported to encode either cysteine *or* selenocysteine according to the position of the UGA codon within the mRNA: see the entry GENETIC CODE.

[Novel structural determinants in (human) SECIS elements modulate translational recoding of UGA as selenocysteine:

SDA (*continued*)

site. Moreover, DNA synthesis from the 3′ end of the nick site regenerates the original duplex. Both of these products – the antisense copy and the regenerated duplex – are shown at (b). Notice that the restriction site in the regenerated duplex is still nickable; this is because the regenerated site contains only one modified nucleotide (in the 5′-GAC-3′ part of site) and this does not affect nickability (see entry).

(b) Products formed from the double-stranded structure at (a) by nicking, strand displacement and regeneration of the duplex. This process is repeated cyclically, with continual regeneration of the original duplex and the production of many copies of the antisense strand.

(c) One of the antisense strands from (b). Each of these strands can bind primer S_1. Extension of S_1 – and of the (recessed) 3′ end of the antisense strand – produces the product shown at (d). During extension of the (recessed) 3′ end of the antisense strand, a modified, non-nickable HincII restriction site is formed (▮▮); the site is non-nickable because the three terminal nucleotides in this strand (3′-CAA-5′) include two modified nucleotides (dATPαS).

(d) This product has a nickable restriction site (arrowhead) derived from primer S_1. Like the product at (a), it can undergo a cyclical process of nicking, strand displacement and regeneration of the duplex, displacing many copies of a single-stranded product. Note that this single-stranded product is a *sense* strand of the target sequence (whereas the strand formed from (a) is an antisense strand). The sense strand and regenerated duplex are shown at (e).

(e) The products from (d). The sense strand will be able to bind primer S_2.

(f) The sense strand from (e) has bound primer S_2. DNA synthesis (i.e. extension from S_2, and also extension from the recessed end of the sense strand) gives rise to the product shown at (g). (DNA synthesis at the 3′ end of the sense strand produces a non-nickable restriction site for the reason given under (c), above.)

(g) This product is equivalent to the one shown at (a).

Summarizing, double-stranded products from the target-generation phase undergo cyclical nicking, strand displacement and regeneration of the original duplex, forming numerous copies of antisense and sense strands of the target duplex. Antisense strands bind primer S_1 and form the (double-stranded) products that yield sense strands; sense strands bind primer S_2 and form the (double-stranded) products that yield antisense strands. Nickable HincII restriction sites are provided by the 5′-tags on S_1 and S_2 primers (which are present in large numbers in the reaction mixture)..

Parts I and II are modified from Figures 5.3 and 5.4, pages 134–137, in *DNA Methods in Clinical Microbiology* [ISBN 07923-6307-8], Paul Singleton (2000), with the kind permission of Springer Science and Business Media.

Nucleic Acids Res (2009) doi: 10.1093/nar/gkp635.]

[SECISaln, a web-based tool for structure-based alignments of eukaryotic SECIS elements: Bioinformatics (2009) 25(5): 674–675.]

SECIS binding protein 2 (SBP2) See SECIS.

SECISaln A web-based tool: see SECIS.

second-generation sequencing system Any of the procedures referred to in MASSIVELY PARALLEL DNA SEQUENCING.

second strand (of cDNA) See the entry FIRST STRAND.

secondary homothallism (homoheteromixis) (*fungal genetics*) In some (heterothallic) fungi: a seemingly homothallic state, involving a self-fertile thallus, which can occur when a spore incorporates nuclei of two compatible mating types – leading to a self-fertile thallus.

In such cases the compatibility of nuclei may be based on a one-locus two-allele system (the process then being termed *homodimixis*) or a one-locus or two-locus *multi*-allele system (*homodiaphoromixis*).

secretion (of proteins) (*DNA technol*.) In the production of recombinant proteins (by engineered cells) it is advantageous, when possible, to arrange for the target protein to be secreted into the medium. This avoids the use of cell lysis to recover the target protein; lysis releases a large number of unwanted proteins (together with the target protein) – making recovery of the target protein more complicated and expensive.

When recombinant proteins are synthesized in *Escherichia coli*, recovery of the product may be facilitated e.g. by fusing the gene of the target protein with the gene of THIOREDOXIN – see OVEREXPRESSION (*Secreted/intracellular proteins*).

In some cases it may be possible to promote the secretion of a target protein by incorporating an N-terminal SIGNAL SEQUENCE. Thus, nucleotides encoding this sequence can be included in the expression vector – and fused to the inserted gene of the target protein; hence, when expressed, the fusion product will consist of the target protein carrying the signal sequence. This approach may or may not work, depending on protein, cells and conditions.

The *solubility* of the target protein can often be enhanced e.g. by using a vector which encodes a solubility-enhancing entity to form a fusion product with the protein of interest – see e.g. CHAMPION PET SUMO VECTOR (and also SOLUBILITY ENHANCEMENT TAG).

segmental duplication (SD) In (eukaryotic) genomic DNA: a sequence of up to 300–400 nucleotides in length, copies of which (with a high degree of sequence similarity) occur at different locations in the genome.

About 5% of the human genome is reported to consist of SDs.

Segmental duplications tend to be found around the centromeres, and adjacent to telomeres; they are also reported to be found at the breakpoints of chromosomal rearrangements.

(See also DISPERSED REPEAT.)

segmented genome Any *viral* genome that consists of two or more pieces of nucleic acid.

Segmented genomes occur e.g. in bunyaviruses (negative-sense/ambisense ssRNA), and also in the influenza viruses A and B (negative-sense ssRNA), the reoviruses (dsRNA) – and in bacteriophage φ6 (dsRNA).

segregation (of plasmids) (*syn*. partition) See PLASMID.

segregation lag A PHENOTYPIC LAG due to the time needed for segregation of a newly acquired allele. For example, a mutant allele in a bacterial cell that contains multiple chromosomes may not be expressed in the presence of the wild-type alleles; an opportunity for expression may occur, for example, if the mutant chromosome has been segregated by cell division to a cell that contains no wild-type alleles.

selectively amplified microsatellite polymorphic loci (selective amplification of microsatellite polymorphic loci) See the entry SAMPL.

selector See SELECTOR-BASED MULTIPLEX PCR and SELECTOR PROBE.

selector-based multiplex PCR A form of PCR in which a range of different target sequences can be amplified – in one assay – with a single pair of primers (cf. conventional MULTIPLEX PCR); this avoids e.g. the need to find a temperature suitable for hybridization of all the primer–target pairs, and avoids the potential PRIMER–DIMER problem [Nucleic Acids Res (2005) 33(8):e71].

For each of the targets to be amplified the reaction mixture contains a target-specific *selector*. A selector consists of two oligonucleotides: one (the *selector probe*) of ∼70 nt, and the other (the *vector oligo*) of 34 nt; the vector oligo hybridizes to the *central* region of the selector probe, so that the selector probe has two terminal single-stranded *target-complementary* regions, each ∼18 nt. The central, i.e. duplexed, region of the selector incorporates a primer-binding motif; this motif is the same in all the selectors (regardless of their target specificity) and is designed for a common ('universal') primer pair.

Sample DNA is first digested with appropriate restriction enzymes. The digested DNA is then mixed with the pool of selectors (of different target specificities), and denatured; this allows the two target-specific parts of each selector probe to hybridize to target sequences, if present, on a given restriction fragment.

In some cases, the two terminal, target-specific parts of the selector probe will hybridize to the two *ends* of a restriction fragment. When this happens, the *fragment's* two termini are juxtaposed to the *vector oligo* in such a way that the vector oligo can undergo ligation, at both ends, to the fragment – forming a single, circular molecule. In other cases, a target-specific end of a selector probe will hybridize to an *internal* sequence in a restriction fragment, thus forming a branched structure; when this happens, the free end of the fragment can be cleaved by a STRUCTURE-SPECIFIC ENDONUCLEASE – thus leaving the vector oligo (as before) ready for ligation to the ends of the fragment to form a circular molecule.

Following ligation, any non-circularized DNA is degraded by an exonuclease. The collection of circularized templates is then ready for amplification by PCR using a universal primer pair (whose binding sites are incorporated in the sequence of

the vector oligo).

selector probe A PROBE used e.g. in SELECTOR-BASED MULTIPLEX PCR (q.v. for a description)

Selector probes were used for targeted molecular analyses [Nucleic Acids Res (2009) 37(1):e7].

selenocysteine insertion sequence (SECIS) See SECIS.

selenocysteine mRNAs *Syn*. SELENOPROTEIN MRNAS.

selenoprotein mRNAs (selenocysteine mRNAs) mRNAs that include a SECIS element; during translation, a SECIS element encodes the insertion of selenocysteine at a UGA codon (but see the item on *Euplotes crassa* in the entry GENETIC CODE).

selenoproteome The total number of selenocysteine-containing proteins encoded by a given genome; in humans the number is reported to be ~26.

(See also SECIS.)

SELEX See APTAMER.

self-activation (*syn*. autoactivation) (two-hybrid systems) Activation of a reporter system without prior interaction between the bait and prey proteins.

self-reporting primer (in PCR) A fluorophore-labeled primer hybridized, via a 5′ tag, to a quencher-labeled PNA (Q-PNA) that quenches the primer's fluorescence; the PNA is displaced during the formation of double-stranded amplicons – so that the primer's fluorophore can fluoresce with appropriate excitation.

self-sustained sequence replication (3SR) An early system for the *in vitro* amplification of nucleic acids; it was based on the retroviral mode of replication [see: Proc Natl Acad Sci USA (1990) 87:1874–1878].

This system is reflected in current methods such as NASBA and TMA.

semi-conservative DNA synthesis See DNA REPLICATION.

Sendai virus A virus of the family PARAMYXOVIRIDAE (q.v.).

SENP The designation given to a number of SUMO proteases – SENP1, SENP2, SENP3 etc. – homologous to the Ulp family of proteases: see SUMOYLATION.

sense strand (of DNA) See the entry CODING STRAND.

separate-component-stabilization system (SCS system) See the entry CCD MECHANISM.

Sequenase® A modified form of the DNA polymerase from bacteriophage T7 which lacks 3′→5′ exonuclease activity but which has strand-displacement activity.

The processivity of the enzyme is promoted by complexing it with THIOREDOXIN from *Escherichia coli*.

sequence capture Any approach used e.g. for the isolation of sequence-specific fragments of single-stranded nucleic acid in solution. For example, a BIOTINylated probe may hybridize to the required fragments – the probe then being captured by streptavidin-coated DYNABEADS which are held immobilized by a magnetic field while the other (unwanted) fragments are removed by washing.

sequence conversion A phrase sometimes used to refer to the treatment of genomic DNA with BISULFITE (an initial stage in procedures used to examine the methylation status of cytosine residues in DNA); such treatment is used to convert non-methylated cytosine residues to uracil residues.

sequence-specific endonuclease Any ENDONUCLEASE which cleaves only at a target site containing a specific sequence of nucleotides. The enzymes in this category include e.g. most of the RESTRICTION ENDONUCLEASES.

(cf. STRUCTURE-SPECIFIC ENDONUCLEASE.)

sequence-tagged site (STS) A unique, known sequence of nucleotides (several hundred base-pairs) in a given fragment of genomic DNA.

STSs can be useful in genomic mapping; thus, for example, they may help to align cloned, tagged fragments into an overlapping series: e.g. if a specific STS is found in two different fragments they may share an overlapping sequence.

A given STS may be identified by means of an EXPRESSED SEQUENCE TAG.

sequencing (of DNA) See DNA SEQUENCING.

sequencing (of RNA) See the note regarding the use of specific endoribonucleases for sequencing RNA in the entry MAXAM–GILBERT METHOD.

sequencing by oligonucleotide ligation and detection system (SOLiD™) See the entry SOLID TECHNOLOGY.

sequencing-by-synthesis Any DNA SEQUENCING procedure in which the sequence in the template strand is determined by sequentially adding nucleotide(s) to a PRIMER hybridized to the template strand; as the added nucleotides base-pair with their *complementary* nucleotides (C to G, A to T etc.) in the template strand, the unknown sequence in the template strand can be deduced from the sequence of nucleotides that extend the primer.

Sequencing-by-synthesis is used in the DIDEOXY METHOD, in PYROSEQUENCING, in SOLEXA SEQUENCING and in SOLID TECHNOLOGY. However, the mode in which nucleotides are added to the primer differ in these methods. Thus, e.g. in the dideoxy method, nucleotides are added sequentially to each primer until a (chain-terminating) dideoxynucleotide is added – at which point synthesis is terminated. By contrast, in the Solexa method, every nucleotide added is a chain-terminating nucleotide, but a reversible one; hence, the synthesis of DNA proceeds by one base at a time. In both of these methods the nucleotides are added by a DNA polymerase – but in SOLiD technology the nucleotides are added by a ligase; moreover, in the SOLiD approach, the nucleotides are added in *pairs*.

SEQUEROME See the entry BLAST.

SERE *Salmonella enteritidis* repetitive element: a repetitive sequence found in the genomic DNA of *S. enteritidis*; it has been used for typing (see e.g. REP-PCR).

serial analysis of gene expression See the entry SAGE.

serine proteinase Any proteinase which has a serine residue at the active site.

Examples of serine proteinases include ENTEROKINASE and THROMBIN.

serine recombinase See the entry RECOMBINASE.

***SERPINB5* gene** See EPIGENETIC ALLELIC RATIO ASSAY.

SET SOLUBILITY ENHANCEMENT TAG.

sex chromatin *Syn*. BARR BODY.

sex determination assay See entry DNA SEX DETERMINATION ASSAY and SRY-BASED ASSAY.

sex-linked chromosome *Syn*. HETEROSOME.

sex-linked disorder *Syn*. X-LINKED DISORDER.

sex pilus *Syn*. PILUS.

sex sorting (of spermatozoa) (*vet*.) A method which is used for controlling the sex of the offspring of domestic animals; this method involves sorting the spermatozoa, in a given sample, into two populations according to whether they contain an X CHROMOSOME or a Y CHROMOSOME.

In this method, spermatozoa are stained with a fluorophore, e.g. HOECHST 33342, and then examined in a FLUORESCENCE-ACTIVATED CELL SORTER. Those spermatozoa which contain an X chromosome bind more fluorophore compared with the spermatozoa containing a Y chromosome – and thus give a higher level of fluorescence on excitation by a laser.

Spermatozoa pass singly through the instrument. The signal (i.e. level of fluorescence) from each is monitored by a computer. Each spermatozoan leaves the detection tubule, singly, within a micro-droplet.

Under computer guidance, a charge is given to each micro-droplet – either positive or negative – depending on whether it contains an X chromosome (higher-level fluorescence) or a Y chromosome (lower-level fluorescence), respectively. The charged micro-droplets then pass between a pair of electrostatically charged plates – so that they are directed into either one or the other of two collecting vessels.

Sex sorting is carried out commercially in the USA by the company XY Inc (website: www.xyinc.com/).

sexual PCR *Syn*. DNA SHUFFLING.

Sf9 cells Ovarian cells from the lepidopteran insect *Spodoptera frugiperda* which are widely used for baculovirus-mediated overexpression of recombinant proteins. These cells can be grown in monolayers or in suspension, and growth occurs in serum-free media.

Sf9 cells have been used for the highly efficient production of recombinant adeno-associated virus (AAV) vectors [Proc Natl Acad Sci USA (2009) 106(13):5059–5064].

(See also MIMIC SF9 INSECT CELLS.)

***sfiA* gene** See the entry SOS SYSTEM.

SFM-adapted Refers to any cell line which has been adapted to growth in *serum-free* media.

Sgal See the entry BLACK–WHITE SCREENING.

SgrS A bacterial sRNA: see the entry SRNAS.

Shiga-like toxins Toxins (SLT-I, SLT-II, also called Stx1, Stx2 or 'verocytotoxins' VT1, VT2) produced by strains of enterohemorrhagic *Escherichia coli* (see EHEC).

The shiga-like toxins are so-called because they are very similar to the dysentery-causing toxin (Shiga toxin) produced by *Shigella dysenteriae*.

Stx2 has been assayed by IMMUNO-PCR (q.v.).

Shine–Dalgarno sequence See the entry 5′-UTR.

SHOM Sequencing by hybridization to oligonucleotide microchips: a method used e.g. for sequencing PCR-amplified DNA by hybridization to an array of oligonucleotides immobilized on gel-based microchips.

The method was used for diagnostic investigations, e.g. the detection of mutations in chromosomal DNA of patients with β-thalassemia.

short hairpin RNA (shRNA) A short, engineered molecule of RNA (∼60 nucleotides in length) which forms a stem–loop structure and which can be cleaved *in vivo* (within cells) to an SIRNA able to induce gene silencing (RNA INTERFERENCE).

shRNAs can be synthesized by *in vivo* transcription from an appropriate DNA target sequence in a plasmid or viral vector. The DNA target sequence for the vector may be synthesized, chemically, as two ssDNA (complementary) oligonucleotides that are hybridized before ligation into the vector [example of construction of a knockdown plasmid: Mol Biol Cell (2008) 19(2):722–734].

In vivo transcription of shRNAs from an appropriate DNA target sequence often employs a promoter that is suitable for RNA polymerase III. The use of RNA polymerase III for this purpose was reported to be advantageous in that transcription ends predictably at the second thymidine residue in a run of more than four thymidine residues in the template.

A further advantage of the RNA polymerase III promoter is that it supports transcription in many types of cell; shRNAs can be expressed constitutively in such cells. However, the RNA polymerase II promoter can also be used for expressing shRNAs, and a system based on RNA polymerase I was used for a species-specific expression vector for shRNA [Nucleic Acids Res (2007) 35(2):e10].

Once synthesized *in vivo*, an shRNA is cleaved, and the product (siRNA) enters the RNA interference pathway.

A method for generating shRNA *libraries* involves random enzymatic digestion of relevant cDNA molecules by DNase I. The resulting fragments are ligated to hairpin adaptor molecules (one adaptor at each end) which contain the recognition sequence of the RESTRICTION ENDONUCLEASE MmeI. This endonuclease cleaves at a site 20 nt in the 3′ direction from its binding sequence (see the table in the entry RESTRICTION ENDONUCLEASE), i.e. in each case, it cuts the *fragment* at a fixed distance from the (closed, hairpin) end of the adaptor. The resulting pieces – all the same length, and consisting of both adaptor and fragment DNA – therefore contain a hairpin adaptor at one end and a 2-nt 3′ overhang at the cut end of the fragment.

Another adaptor is then ligated to the sticky end of each of the fragments. PCR amplification of the resulting structures yields dsDNA products, each containing the original hairpin sequence and target DNA derived from the original cDNA. Products are then inserted into vectors at a site downstream of an RNA polymerase III promoter.

A ligand-controlled aptamer–shRNA fusion transcript has been used to regulate gene expression in mammalian cells [RNA (2006) 12:710–716].

shRNAs have been used to knock down the expression of several target genes within the same cell in an arrangement using a tetracycline-regulated RNA polymerase II promoter

[BioTechniques (2006) 41(1):64–68].

[An shRNA vector for high-efficiency RNAi in embryonic stem cells: BioTechniques (2007) 42(6):738–743.]

(See also vector pBLOCK-iT^{TM}3-DEST in the entry GATEWAY SITE-SPECIFIC RECOMBINATION SYSTEMS (table).)

short non-coding RNAs (small non-coding RNAs) See SRNAS.

short patch repair *Syn.* UVRABC-MEDIATED REPAIR.

short sequence repeats (SSR) A non-specific term that may apply e.g. to MICROSATELLITE DNA or to minisatellite DNA.

(See also SNR.)

short tandem repeat See STR.

shotgun sequencing A method used for sequencing genomes (or large molecules of DNA). A sample of the DNA is cut extensively with a RESTRICTION ENDONUCLEASE and a large number of the fragments are cloned and sequenced. Because of overlap among the fragments, sequence data obtained from individual clones can be amalgamated to reveal long, continuous sequences within the genome. However, usually there are gaps in the genome's sequence because some parts of the genome may not have been cloned. A missing sequence may be found by using PCR to copy that sequence from genomic DNA – see e.g. the entry PACE.

shRNA See the entry SHORT HAIRPIN RNA.

shuffling (of DNA) See DNA SHUFFLING.

shufflon In e.g. the IncIα plasmid R64: a set of DNA sequences which can invert, singly or in groups, to form complex rearrangements that may influence the specificity of cell–cell contact by the host cell in conjugation. The shufflon encodes a recombinase that mediates the rearrangements.

Similar recombinase-encoding mobile DNA elements occur in other plasmids. Moreover, a sequence encoding an enzyme similar to the R64 shufflon recombinase was reported in a prophage within the genome of the Gram-negative bacterium *Pseudomonas fluorescens* [BMC Microbiol (2009) 9:8].

shuttle vector (bifunctional vector) Any CLONING VECTOR, or other type of vector, which is capable of replication in more than one type of organism – for example, in (the bacterium) *Escherichia coli* and (the yeast) *Saccharomyces cerevisiae*; the vector must include an appropriate origin of replication for each type of organism.

(See also BACMID.)

sialidase *Syn.* NEURAMINIDASE.

sickle-cell anemia An inherited disorder involving the β-chain of hemoglobin; the mis-shaped erythrocytes (red blood cells) are eliminated, leading to (sometimes fatal) anemia.

Sickle-cell trait refers to the generally less serious condition in those heterozygous for the gene; it may be asymptomatic, and it confers a degree of resistance to falciparum malaria.

[Correction of the sickle-cell mutation in embryonic stem cells: Proc Natl Acad Sci USA (2006) 103(4):1036–1040.]

(See also THALASSEMIA.)

sickle-cell trait See SICKLE-CELL ANEMIA.

SIDD Stress-induced duplex destabilization: destabilization at a particular region in dsDNA (a SIDD site) – promoting strand separation – in response to a given level of superhelicity.

In the genome of *Escherichia coli*, SIDD sites appear to be closely associated with promoters. Transcription from the *E. coli ilv*P_G promoter was reported to be preceded by destabilization at an upstream SIDD site following the binding of the INTEGRATION HOST FACTOR.

In the *E. coli* K-12 genome, strong SIDD sites were found to be overrepresented in upstream regions of genes encoding transcription regulators; moreover, upstream regions of genes that respond directly to physiologic or environmental stimuli were reported to be destabilized to a greater extent than are those of genes that are not involved in such responses [PLoS Comput Biol (2008) 4(1):e17].

side-by-side model (of DNA) A model of (double-stranded) DNA in which the two strands are alongside one another in parallel fashion rather than being wound around each other in a helical fashion. In this model, double-stranded DNA occurs as alternating short stretches of right-handed and left-handed DNA, each of the short stretches being five nucleotides long.

siderophilins Ferric iron-chelating glycoproteins that are found in vertebrates; they include LACTOFERRIN and transferrin.

SIGEX Substrate-induced gene expression screening: a method in which induction of gene expression – by making available substrates of the encoded proteins – is used as one approach for isolating catabolic genes from metagenome libraries.

sigma (σ) factor In bacteria: a type of protein that binds to the RNA POLYMERASE and confers on that enzyme the ability to initiate transcription from particular class(es) of PROMOTER. Bacteria typically (or always) encode more than one type of sigma factor; thus, by synthesizing particular type(s) of sigma factor at appropriate times, a bacterial cell can switch activity to specific class(es) of promoter and, in this way, control the expression of particular type(s) of gene.

In *Escherichia coli* the main sigma factor is σ^{70} – encoded by gene *rpoD*; it can promote transcription from most types of promoter, although in some cases transcription does not occur in the absence of certain regulatory factors – as, e.g., in CATABOLITE REPRESSION.

The *E. coli* sigma factor σ^{S} (encoded by *rpoS*) has roles e.g. in stationary-phase events and in regulating certain responses to stress.

signal amplification An approach used e.g. for detecting and/or quantitating a given DNA or RNA target sequence within a sample. In one method, each target sequence in the sample is linked to an enzyme which, with a suitable added substrate, generates an amplified signal whose intensity can indicate the number and/or location of target sequences in the sample.

Signal amplification avoids one major problem associated with TARGET AMPLIFICATION methods (such as PCR), i.e. the contamination of a reaction mixture with AMPLICONS from earlier assays.

A disadvantage of signal amplification is that it provides no copies of the target sequence for analysis.

One example of a signal amplification method is the BDNA ASSAY; another is TYRAMIDE SIGNAL AMPLIFICATION.

A different form of signal amplification is used in a scheme

for raising the detection sensitivity of MOLECULAR BEACON PROBES (q.v.).

signal peptide *Syn*. SIGNAL SEQUENCE.

signal sequence (signal peptide) In certain exported or secreted proteins: a specific sequence of (N-terminal) amino acid residues that directs the protein into, or through, the cell's cytoplasmic membrane; the signal sequence may be cleaved from the protein during its translocation through the membrane.

Signal sequences are found in prokaryotes and eukaryotes, although protein translocation occurs by distinct processes in the two types of cell.

In *Escherichia coli* (and in other Gram-negative bacteria) signal-sequence-dependent translocation is involved in the type II mode of export/secretion of proteins. In some cases a protein may also include other sequence(s) that help to guide it to the correct destination; for example, a protein targeted to the membrane (as opposed to a *secreted* protein) may include a *membrane anchor sequence* which binds the mature protein on or within the membrane. In the Braun lipoprotein (which links peptidoglycan to the *E. coli* outer membrane) the signal sequence is cleaved and replaced by a fatty acid residue; this post-translational modification appears to target the protein to its correct destination in the cell envelope.

In the so-called Tat system, which is involved in exporting *folded* proteins across the cytoplasmic membrane, the signal sequence characteristically includes two consecutive arginine residues (hence Tat: twin arginine translocation).

Some other modes of protein export/secretion in the Gram-negative bacteria are not characterized by an N-terminal signal sequence. For example, secretion via the type I system in *E. coli* (involving a so-called ABC exporter) is characterized by a C-terminal secretion sequence.

(See also SECRETION (of proteins).)

signal transducers and activators of transcription (STATs) In mammalian cells: a family of cytoplasmic proteins that are activated by phosphorylation when certain types of signaling molecule (e.g. interferons, interleukins) bind to cell-surface receptors. The binding of a signaling molecule initially activates a tyrosine kinase of the JAK (Janus kinase) family which then activates a STAT.

There are at least seven different types of STAT. When activated, STATs form homo- or heterodimers, according to the type of STAT, which then promote transcription of relevant gene(s).

signal transduction pathway Any pathway which involves the sequential transmission of a signal along a series of two or more components – often involving activation of particular components e.g. by phosphorylation.

For an example of a method used for investigating signal transduction pathways see e.g. PATHDETECT SYSTEMS.

(See also TWO-COMPONENT REGULATORY SYSTEM.)

signature-tagged mutagenesis (STM) An approach used e.g. for detecting those genes (in a pathogen) which are essential for the pathogen's growth in a host organism.

The principle of STM is explained diagrammatically in the figure (page 334).

The method has been used successfully e.g. for identifying a PATHOGENICITY ISLAND in *Salmonella typhimurium*. It was also used to identify genes, in *Mycobacterium tuberculosis*, involved in parasitism of human macrophages [Infect Immun (2007) 75(1):504–507]; for identifying genes involved in the survival/pathogenesis of *Burkholderia pseudomallei* [Infect Immun (2007) 75(3):1186–1195]; and for identifying genes that are expressed anaerobically by *Pseudomonas aeruginosa* [J Bacteriol (2008) 190(8):2739–2758].

The possibility of a false-positive result in STM may arise e.g. if (during the *random* mutagenesis) a transposon inserts into a *non*-virulence gene in a cell which (by chance) already contains an (inactivating) mutation in a virulence gene; with this scenario, the given tag would be absent (or low level) in the 'recovered' pool, thus giving a false-positive indication for the particular gene inactivated by the transposon.

In a non-medical context, the STM principle has been used to study the survival of certain bacteria (e.g. *Desulfovibrio*) in anaerobic sediment [Appl Environ Micobiol (2005) 71:7064–7074].

Another method – IVET – is used e.g. to detect those genes in a pathogen which are induced (i.e. which become active) during infection of the host animal. Like STM, this method is also used in a non-medical context.

silencing (of genes) See GENE SILENCING.

silencing (of miRNAs) See e.g. ANTAGOMIR.

silent mutation Any mutation lacking a phenotypic effect. (cf. SAME-SENSE MUTATION.)

Note that these mutations may not be 'silent' in DNA-based identification tests (see e.g. SSCP ANALYSIS).

silent origin See e.g. ARS.

silica (in gene delivery) See GENE-DELIVERY SYSTEM.

Silwet L-77 (in transfection) A type of detergent which is used in the FLORAL DIP METHOD.

simian foamy virus See SPUMAVIRINAE.

simian virus 40 See the entry SV40.

simple sequence repeats A non-specific term which may apply e.g. to MICROSATELLITE DNA.

The term was also used to include single-nucleotide repeats (SNRs) [J Clin Microbiol (2006) 44(3):777–782].

simple transposon (class II transposon) See TRANSPOSON.

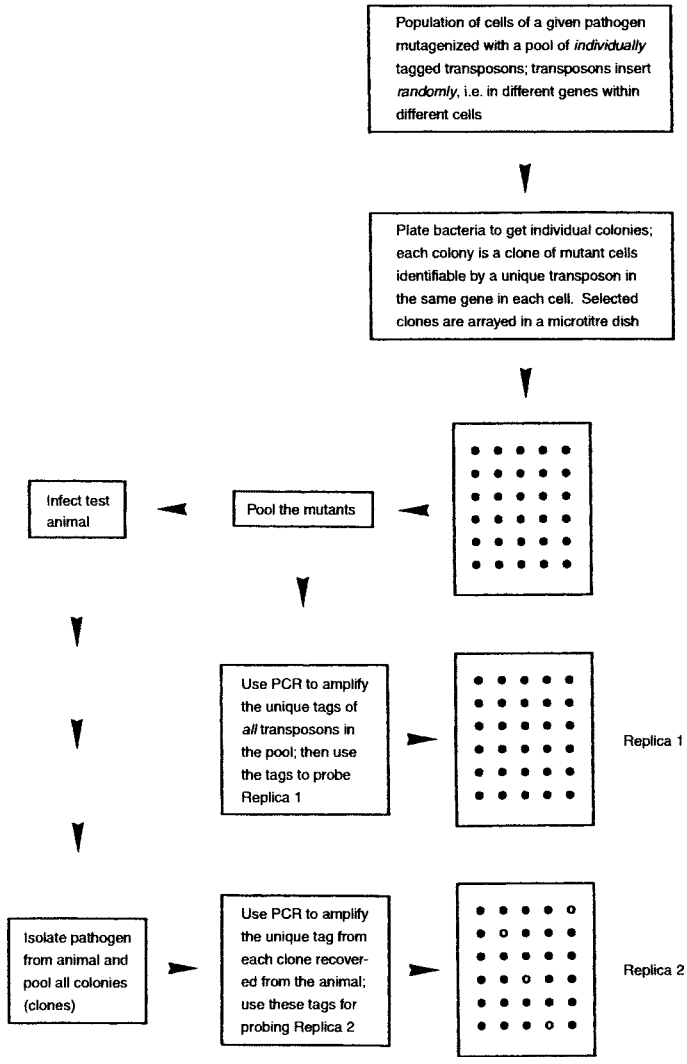
Sindbis virus A virus of the genus ALPHAVIRUS that can cause mosquito-borne disease characterized (e.g.) by arthralgia and rash. A recombinant Sindbis virus was used in an experiment to study the role of RNA interference in mosquitoes: see the entry ARBOVIRUSES.

(See also SUBGENOMIC MRNA.)

SINE Short interspersed element: any one of a number of repetitive elements, each typically <500 nt long, dispersed in the genomes of at least some mammalian species (including the human genome). The SINEs include ALU SEQUENCES.

(cf. LINE.)

single-base extension (SBE) A method used e.g. to test for the presence of a given base at a specific location in a sample of



SIGNATURE-TAGGED MUTAGENESIS: a method used e.g. for detecting virulence-associated genes in a bacterial pathogen. The figure shows the general principle diagrammatically.

A unique sequence ('tag'), about 40 base-pairs in length, is inserted into each of a population of transposons, i.e. the tag in a given transposon is different from that in other transposons. Each tag is flanked, on both sides, by a primer-binding site; these primer-binding sites are the same in all transposons. The transposons are used to mutagenize a population of cells of the pathogen, and they insert – randomly – into different genes in different cells.

The mutagenized cells are plated to form individual colonies. In a given colony of mutant cells, each cell contains the same uniquely tagged transposon in the same gene. A number of colonies are chosen, and an inoculum from each colony is arrayed, separately, in a microtiter dish.

(*continued*)

DNA.

Essentially, the presence of the base – at a specific location in a strand – is determined by hybridizing a primer whose 3′ terminal nucleotide *can be extended* with a nucleotide which base-pairs with the specific base in question. Extension of the primer with such a nucleotide argues for the presence of the *complementary* nucleotide at the specific site in the template (sample) strand. As the primer is extended with a single dye-linked ddNTP, i.e. *chain-terminating* dideoxyribonucleoside triphosphate, extension can be detected via the dye-labeled product (or, as in QEXT, by a FRET-based reaction).

This procedure can be used e.g. to detect a point mutation at a specific location.

(See also entry METHYLATION-SPECIFIC SINGLE-BASE EXTENSION.)

single-base repeat *Syn*. Single-nucleotide repeat: see SNR.

single-beam gradient force optical trap An alternative name for OPTICAL TWEEZERS.

single-cell electroporation See ELECTROPORATION.

single-cell *in vivo* fluorescence See DFRS PLASMID.

single-cell PCR PCR in which the sample DNA is derived from a single cell – see e.g. PGD.

single-chain antibody See the following entry: SINGLE-CHAIN VARIABLE FRAGMENT.

single-chain variable fragment (scFv, scFV etc.) A recombinant polypeptide transcribed and translated from an engineered sequence encoding the (antigen-binding) V_H and V_L regions *of a given antibody* – these two coding regions being linked by a sequence encoding a short synthetic peptide; an scFv has the structure:

$$V_H\text{–linker–}V_L$$

scFv-displaying phages can be prepared by fusing the scFv-encoding sequence with the gene of a phage coat protein such as pIII; the fusion protein is then incorporated into nascent phage particles that are released from the bacterial host cell.

An scFv may be displayed on a recombinant phage e.g. for use in one form of IMMUNO-PCR (q.v.).

A phage that displays multivalent anti-CD30 scFv has been used for gene transfer to Hodgkin cell lines The multivalent scFv mediated initial phage-to-cell binding. After uptake of the phage, the transfected cells processed the (ss) DNA and transiently expressed a eukaryotic gene cassette insert [BMC Mol Biol (2008) 9:37].

An scFv was used for targeting a recombinant herpesvirus to $CD38^+$ human leukemic cell lines and to cells of $EGFR^+$ tumors in mice [Gene Therapy (2008) 15:1579–1592].

An *in vivo*-expressed scFv can be engineered to incorporate specific effector function(s), and it can be designed to remain *intracellular* (cf. INTRABODY). An scFv, targeting the (cell-surface) CD147 molecule of human embryonic kidney cells (cell line 293A), was effective *within* these cells; the purpose of this study was to down-regulate CD147, i.e. to reduce the cell-surface expression of CD147 by sequestering this molecule *intracellularly* (that is, by scFv–CD147 binding) [BMC Biotechnol (2008) 8:5].

[A compact phage display human scFv library for selection

SIGNATURE-TAGGED MUTAGENESIS: (*continued*) (The dish shown in the figure has 30 wells, but larger dishes are normally used.) Two replica 'blots' of the array are made on membranes (for subsequent DNA hybridization studies); in these blots the cells are lysed and their chromosomal DNA is exposed and fixed to the membrane.

Cells are taken from each of the wells and pooled. The pool is used in two ways. First, it provides an inoculum for infection of the test animal. Second, cells from this ('input') pool are lysed, and PCR (polymerase chain reaction) is used – with labeled primers – to amplify the unique tags of all transposons in the pool. The amplified, labeled tags are then used as probes on one of the replica blots (Replica 1); in this blot, each unique tag should hybridize with the DNA from cells containing the corresponding transposon. This 'pre-screening' process on Replica 1 checks for efficient amplification of the unique tags.

The pathogen is then recovered from the test animal by plating an appropriate specimen. The resulting colonies are pooled (forming the 'recovered' pool), and PCR is used (with labeled primers) to amplify the tag from each clone in this pool. The amplified, labeled tags are then used to probe the second replica blot (Replica 2).

Considering the original, mutagenized cells (in the 'input' pool), the cells of interest are those which, through a (transposon-mediated) mutation, are *unable* to grow within the test animal, i.e. cells whose virulence has been lowered. Such cells will be absent, or few in number, in the 'recovered' pool – compared with cells which have grown normally in the test animal. Hence, the signature tags of these 'virulence-attenuated' cells will be present in the input pool but absent (or rare) in the recovered pool; consequently, such cells can be identified by hybridization in Replica 1 but an *absence* of (or weak) hybridization in Replica 2 (see figure).

In a given, 'virulence-attenuated' clone, the relevant gene can be identified from the inserted transposon. This gene can be isolated and sequenced for further study.

Figure reproduced from *Bacteria in Biology, Biotechnology and Medicine*, 6th edition, Figure 8.22, pages 256–257, Paul Singleton (2004) John Wiley & Sons Ltd, UK [ISBN 0-470-09027-8] with permission from the publisher.

of antibodies to a wide variety of antigens: BMC Biotechnol (2009) 9:6.]

(See also DIABODY and MINIBODY.)

single-molecule amplification In several of the methods used for DNA SEQUENCING: amplification (by PCR) of individual fragments of DNA. In the GS FLX method, the amplification takes place when the (bead-bound) target fragments of DNA are contained within droplets of an *emulsion* (PCR reaction mixture in oil). In SOLEXA SEQUENCING, amplification takes place when the fragments are attached to a surface within a flow cell. In SOLID TECHNOLOGY, amplification of the (bead-bound) fragments occurs within an emulsion (as occurs in the GS FLX method).

single-molecule DNA manipulation/visualization Techniques of various types in which individual molecules of DNA are manipulated by certain physical force(s) and viewed by ultra-sensitive fluorescence-based methods. These techniques are useful e.g. for investigating DNA–protein interactions and for studying the physical properties of the DNA molecule itself.

The kinds of approach used in these techniques include e.g. DNA COMBING (molecular combing), SURFACE TETHERING, SURFACE-INDEPENDENT TETHERING (which involves the use of OPTICAL TWEEZERS), and techniques that involve the use of MAGNETIC TWEEZERS.

single-molecule nanopore DNA sequencing A method which was proposed for the (nanopore-based) sequencing of nucleic acids: see the entry DNA SEQUENCING.

single-nucleotide polymorphism See SNP.

single-nucleotide repeat See SNR.

single-primer extension See SPEX.

single-strand-annealing protein See ET RECOMBINATION and LAMBDA (λ) RED RECOMBINATION.

single-strand binding protein (SSB protein; helix destabilizing protein) Any of the various proteins which bind specifically and co-operatively to single-stranded DNA in processes such as DNA replication and repair; they may stabilize and protect ssDNA but apparently do not unwind a duplex.

(See also ROLLING CIRCLE.)

single-strand conformation polymorphism analysis See SSCP ANALYSIS.

single-strand rescue See HELPER PHAGE.

single-strand-specific nuclease Any of a category of nucleases that degrade single-stranded DNA or single-stranded regions in double-stranded DNA; this kind of enzyme is used e.g. to remove terminal *overhangs* in fragments of DNA which have been cut by some types of restriction endonuclease (removal of overhangs being achieved without significantly affecting the double-stranded regions). (See also NESTED DELETIONS.)

Another useful feature of these enzymes is their ability to nick one strand at the site of a single-base-pair mismatch in (heteroduplex) DNA – which is exploited e.g. in ECOTILLING and in the CEL I NUCLEASE MISMATCH ASSAY.

(See also CLEAVASE.)

The single-strand-specific nucleases include those that have similar zinc-dependent active-site motifs – e.g. endonuclease CEL I (derived from celery), endonuclease P_1 (from a fungus, *Penicillium citrinum*), ENDONUCLEASE S1 (from the fungus *Aspergillus oryzae*), and the MUNG BEAN NUCLEASE (from *Phaseolus aureus*).

siRNA Small interfering RNA: an effector molecule involved in RNA INTERFERENCE (q.v.).

Making/delivering siRNAs

For experimental purposes, siRNAs can be made in a number of ways, for example:

(a) Chemical synthesis (commercially available) according to specified target sequences.

(b) *In vitro* transcription of target sequences using e.g. T3 or T7 polymerase, i.e. enzymes that can work with simple transcription signals. (A target sequence can be initially amplified by PCR with primers that incorporate a T3 or a T7 promoter sequence in a 5′ tag.)

(c) Cleavage of dsRNA with bacterial RNase III.

(d) RNAi may be induced in target cells by *in vivo* expression of *short hairpin RNAs* (see SHORT HAIRPIN RNA) which are then cleaved, *in situ*, to siRNAs. *In vivo* transcription can be achieved e.g. with the cell's RNA polymerase III. The use of RNA polymerase III may be facilitated by means of an inducible minimal RNA polymerase III promoter which can be activated in the presence of doxycycline – i.e. induction in a dose-dependent manner is achieved by the administration of doxycycline [Nucleic Acids Res (2006) 34(5):e37].

Using siRNAs

siRNAs are used for various *in vitro* and *in vivo* applications such as GENE SILENCING/KNOCKDOWN, including studies on potential GENE THERAPY systems in animal models.

Studies have been carried out on both normal and modified siRNAs, on the efficacy of siRNAs, and also on various ways in which siRNAs can be introduced into cells.

Single-stranded siRNAs can induce RNAi but are generally thought to be less effective. However, ss siRNAs modified by boranophosphate were reported to exhibit high-level potency [Nucleic Acids Res (2006) 34(9):2773–2781].

[siRNAs for distinguishing between genes differing by one nucleotide: PLoS Genetics (2006) 2(9):e140.]

In one method for intracellular delivery, siRNA was linked, via a streptavidin bridge, to an APTAMER that was known to bind to prostate tumor cells; the conjugate (siRNA+aptamer) was taken up by the cells within 30 minutes following simple addition of conjugate to the cells. siRNA-mediated inhibition of gene expression in this way was reported to be as efficient as that produced by using lipid-based internalization reagents [Nucleic Acids Res (2006) 34(10):e73].

(See also PAMAM vectors in the entry DENDRIMER.)

Problems, including off-targeting, and the uses of chemical modification and asymmetric siRNAs

Within their target cells, siRNAs have the potential to cause certain problems. For example, they may silence *non*-targeted genes (so-called *off-targeting*), particularly under high-dose conditions. There is also the possibility that they may trigger

inappropriate immune responses in the whole-animal setting. Moreover, the siRNA itself may be subjected to degradation by the target cell's RNases.

Such problems have been addressed by the use of chemical modification. Chemical modification of siRNAs may achieve e.g. improved stability against the cell's nucleases, a reduced incidence of off-target effects – and also improvement in the pharmacodynamics [chemical modification of siRNAs for *in vivo* use: Oligonucleotides (2008) 18:305–320].

Studies have also been carried out with siRNA constructs in which the two strands are of different length. In one study, a construct consisted of a guide strand of 19 nt base-paired to a passenger strand of 15 nt; these *asymmetric interfering RNAs* (aiRNAs) were, in most cases, said to be at least as efficient as the conventional siRNAs. Experiments in which cells were transfected with aiRNAs – and controls with normal siRNAs – indicated (from gene expression data) that fewer off-target effects were produced by the aiRNAs (compared to siRNAs); thus, it appeared that the short (15 nt) passenger strand in the aiRNAs was less likely than the full-length passenger strand to give rise to off-target effects.

Other modified forms of the basic siRNA molecule are also reported to exhibit silencing activity. These include a three-component structure which consists of the guide strand base-paired to two (separate) pieces of passenger strand; this *small internally segmented interfering RNA* (sisiRNA) can accomodate significant chemical modification, which could be useful for flexibility of design.

Although such modified forms of siRNA are potentially of great use, conflicting results were found in some studies, and further work is needed to establish the rules which govern the design of these engineered constructs.

[siRNA asymmetry: Gene Therapy (2009) 16:827–829.]

Selection of potent siRNAs

A method for selecting potent siRNAs from an siRNA library is described in the entry RNA INTERFERENCE.

[Assessment of siRNA potency by a luciferase assay: BioTechniques (2007) 42(5):599–606.]

Transcriptional gene silencing (TGS)

siRNAs can also mediate a process known as transcriptional gene silencing (TGS) in mammalian and other cells (e.g. cells of the small plant *Arabidopsis*, and the fission yeast *Schizosaccharomyces pombe*); transcriptional gene silencing refers to the phenomenon in which transcription (rather than translation) is inhibited (cf. PTGS). In human cells, proteins of the ARGONAUTE family may be implicated in this process.

In some cases, TGS is reported to involve siRNA-mediated methylation of DNA as the silencing mechanism – so-called RNA-directed DNA methylation (RdDM).

In other cases, the mechanism of TGS apparently involves methylation of histones (promoting the formation of heterochromatin).

Transcriptional gene activation

As well as gene silencing (either by RNAi or TGS), siRNAs are able to mediate transcriptional gene activation. Thus, one report refers to the siRNA-directed (ARGONAUTE-dependent) transcriptional activation of the *p21* gene (in human cells) which resulted from post-transcriptional silencing of an antisense transcript directed against expression of the *p21* gene [PLoS Genet (2008) 4(11):e1000258].

sis (*SIS*) An ONCOGENE identified in simian sarcoma virus. The c-*sis* product is the B chain of platelet-derived growth factor (PDGF). The viral gene encodes an almost-identical B chain, which is the transforming agent in virus-infected cells.

sisiRNA Small internally segmented interfering RNA: see entry SIRNA.

site-directed cross-linking (in DNA) See e.g. TRIPLEX DNA.

site-directed mutagenesis (or site-specific mutagenesis) (*DNA technol.*) Any form of engineered MUTAGENESIS involving change(s) in target molecules at *pre-determined* site(s). Older methods of *in vitro* site-directed mutagenesis (e.g. the D-loop mutagenesis procedure) have been replaced by simpler ones: see e.g. QUIKCHANGE SITE-DIRECTED MUTAGENESIS KIT and MEGAPRIMER MUTAGENESIS; and see EXSITE SITE-DIRECTED MUTAGENESIS KIT, GENETAILOR SITE-DIRECTED MUTAGENESIS SYSTEM.

Some studies have made use of the PHAGE M13 system (see diagram).

A deletion can be made by using INVERSE PCR to copy all except a specific sequence of the target; the amplicons therefore provide a shortened form of the original target sequence. The product can then be subjected to further amplification.

Insertions can be introduced e.g. by adding 5′ extensions to the primers used in PCR.

Site-directed mutagenesis in vivo

Site-directed mutagenesis *in vivo* (i.e. within living cells) has been carried out by various methods, including e.g. the use of RECOMBINEERING, and the use of ZINC-FINGER NUCLEASES. Knock-outs have been carried out in bacteria and eukaryotic cells by the TARGETRON approach.

[Site-directed mutagenesis in animal models (progress and prospects – a review): Gene Therapy (2009) 16:581–588.]

Site-specific recombinase systems – for example, the CRE–LOXP SYSTEM and the FLP system – are commonly used for making specific manipulations within living cells.

(See also GENE TARGETING and MIRAGE.)

site-specific biotinylation See BIOTIN.

site-specific cleavage (of proteins) Cleavage of proteins at sites between specific amino acid residues. Some of the enzymes that can carry out site-specific cleavage are exploited in DNA technology – e.g. for separating fusion proteins or removing a tag from a recombinant protein.

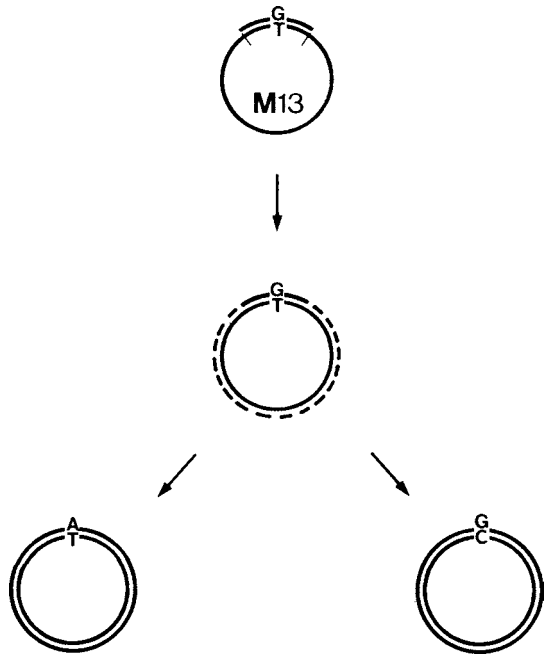
(See e.g. EKMAX, SUMO protease in SUMOYLATION, TEV protease in TEV; see also entries ENTEROKINASE, SORTASE, and THROMBIN.)

site-specific cross-linking (in DNA) See e.g. TRIPLEX DNA.

site-specific deletion A specific KNOCK-OUT MUTATION.

site-specific mutagenesis *Syn.* SITE-DIRECTED MUTAGENESIS.

site-specific recombination (SSR) A form of RECOMBINATION between *specific* dsDNA sequences in the same molecule, or



SITE-DIRECTED MUTAGENESIS (oligonucleotide-directed mutagenesis) using a phage M13 cloning vector: creating a point mutation at a known site in a given gene (diagrammatic).

Phage M13 is used to make single-stranded copies of the given gene. Initially, the (double-stranded) gene is inserted into a circular, double-stranded form of M13 (see entry PHAGE M13); replication of M13 (in a bacterial host) produces a single-stranded copy of the gene *incorporated in each of the ccc ssDNA genomes of M13*. One such genome, carrying the target gene, is shown in the diagram (*top*).

Part of the target gene (a sequence of 15–20 nucleotides) – its size exaggerated for clarity – is shown between two bars (*top*); within this small sequence we may wish to replace deoxythymidine (T) with deoxycytidine (C). To do this, we initially synthesize copies of a *complementary* sequence: an oligonucleotide, 15–20 nucleotides long, containing one mismatch at the required site – deoxyguanosine (G) opposite deoxythymidine (T). Each oligonucleotide is then allowed to bind to a copy of the target gene (*top*) where it can prime *in vitro* DNA synthesis (*dashed line*, *center*). The resulting dsDNA molecules, each with the single mismatch, are inserted into bacteria e.g. by transformation. Within the cells, DNA repair mechanisms correct the mismatch – sometimes replacing T with C, sometimes replacing G with A (deoxyadenosine). Even without repair, replication of the mismatched duplex will produce some molecules with the required mutant sequence (*bottom*, *right*) as well as wild-type molecules (*bottom*, *left*).

Figure reproduced from *Bacteria in Biology*, *Biotechnology and Medicine*, 6th edition, Figure 8.20, page 253, Paul Singleton (2004) John Wiley & Sons Ltd, UK [ISBN 0-470-09027-8] with permission from the publisher.

in different molecules, in which there is neither synthesis nor degradation of DNA; this (ATP-independent) process is mediated by a site-specific RECOMBINASE (which recognizes *only the specific target sequences*). The two sequences of dsDNA that are recognized by a given site-specific recombinase are not necessarily identical.

Examples of site-specific recombination: (i) integration of the PHAGE LAMBDA genome into the bacterial chromosome, and excision of the genome – the phage attachment site, *att*P, not being identical to the bacterial attachment site (*att*B); (ii) PHASE VARIATION in a *Salmonella* filament protein; (iii) the CRE–LOXP SYSTEM; (iv) the FLP system; and (v) insertion of a cassette into an INTEGRON.

(See also entry PHAGE BXB1.)

site-specific transposition (*DNA technol.*) In some cases, when a transposon's target site is predictable, transposon-mediated recombination (engineering) can be carried out by exploiting the target-binding sites of the relevant transposase – see e.g. the preparation of an expression bacmid in the entry BAC-TO-BAC.

(See also the entry TN7.)

six-histidine tag (6×His tag; His tag) A peptide – consisting of six consecutive histidine residues – that is encoded by some expression vectors; the sequence encoding the 6×His tag is fused to the gene encoding the protein of interest, the protein and tag therefore being expressed as a FUSION PRODUCT. The 6×His tag enables detection of the fusion product by anti-His antibodies, and aids purification of the protein by a NICKEL-CHARGED AFFINITY RESIN.

[Use (e.g.): J Biol Chem (2009) 284(13):8866–8876; PLoS Pathog (2009) 5(9):e1000579; Mol Syst Biol (2009) 5:295; Proteins (2009) 76(3):772–777.]

size fractionation (*DNA technol.*) Any procedure in which a population of DNA molecules/fragments – present in a range of sizes – is treated so as to obtain physical separation of the molecules/fragments according to size.

For example, a population of fragments, in a range of sizes, may be subjected to electrophoresis in an agarose gel; a particular band of fragments, of appropriate size, can be excised, purified (by removal of the agarose) and used for the required purpose.

Size fractionation may be used e.g. for selecting fragments of appropriate size(s) from a restriction-digested genome in order to use such fragments in cloning vectors which accept inserts only within a certain range of sizes.

skin cancer See MELANOMA, ULTRAVIOLET RADIATION and XERODERMA PIGMENTOSUM.

Skyline (DNA notation) See entry DNA SKYLINE.

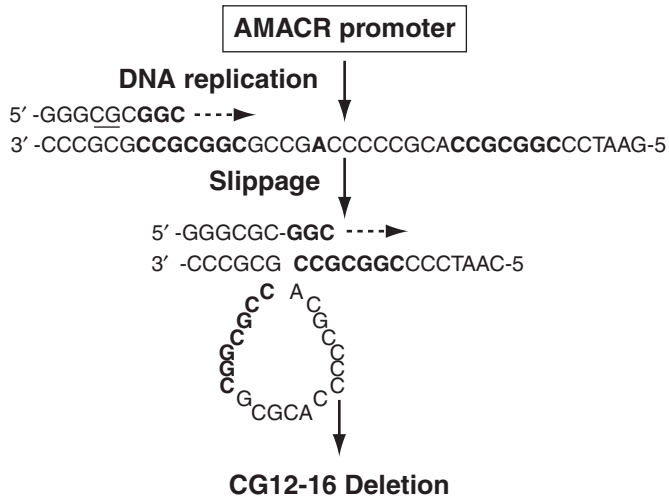
Sleeping Beauty See the entry RECONSTRUCTED TRANSPOSABLE ELEMENT.

sliding A movement made by certain DNA-binding proteins – translocation lengthwise along a DNA duplex until the appropriate binding site is reached. For example, sliding of the *lac* repressor protein (see LAC OPERON) along the duplex occurs until it reaches the operator sequence.

slipped-strand mispairing (SSM) Mispairing between strands during DNA replication in susceptible regions of the genome; regions susceptible to SSM are characterized e.g. by multiple tandemly repeated short sequences of nucleotides – such as MICROSATELLITE DNA.

SSM can apparently result in an increase or a decrease in the (normal) number of repeated subunits in a given sequence of microsatellite DNA. In humans, 'microsatellite expansion' (i.e. an increase in the number of subunits at a given locus) can cause disease when associated with particular genes (e.g. FRAGILE X DISEASE, HUNTINGTON'S DISEASE).

Changes in such sequences are reported to have a range of effects in bacteria. For example, they may promote PHASE



SLIPPED-STRAND MISPAIRING: the mechanism proposed to account for a 20-bp deletion in a CpG island in the promoter region of *AMACR*. Figure reproduced from 'Deletion hotspots in *AMACR* promoter CpG island are *cis*-regulatory elements controlling the gene expression in the colon', Figure 3B, by X. Zhang and colleagues, PLoS Genetics (2009) 5(1):1000334.

VARIATION in *Neisseria*. In *Helicobacter pylori*, *mutY*, which encodes a DNA excision-repair function, contains a sequence of eight consecutive adenine residues that seem to be subject to slipped-strand mispairing; SSM is able to cause frameshift mutations which eliminate gene function, and this was interpreted as a mechanism that contributes to phase-variable base excision repair in this organism [J Bacteriol (2006) 188(17): 6224–6234]. In *Escherichia coli* (strain O157:H7), slipped-strand mispairing may be responsible for a proportion of the instability in the VNTR (variable number of tandem repeats) loci [J Bacteriol (2006) 188(12):4253–4263].

Studies on the poly-T (promoter) tract of *sabA* in *H. pylori* suggested that this sequence undergoes SSM [Microbiology (2008) 154(8):2231–2240].

The diagram illustrates the proposed mechanism involved in a 20-bp deletion in a CpG island in a promoter region.

Slipped-strand mispairing is believed to be involved in the maintenance of tandem repeat arrays in mitochondrial DNA in species of halibut (*Hippoglossus* and *Reinhardtius*) [BMC Genomics (2008) 9:10].

slippery sequence In an mRNA molecule: a sequence of nucleotides – downstream of the initiation codon – which causes slippage of the ribosome along the mRNA. Slippage, which may occur in an upstream (5′) or a downstream (3′) direction, is an integral part of the mechanism of FRAMESHIFTING.

slot blot See the entry DOT BLOT.

SLT-I, SLT-II Shiga-like toxins: see EHEC.

***slyD* gene** See the entry CELL-FREE PROTEIN SYNTHESIS.

small interfering RNA (siRNA) See RNA INTERFERENCE; see also SIRNA.

small non-coding RNAs See SRNAS.

small nuclear ribonucleoprotein particles See SNRNAS.

small nuclear RNAs See SNRNAS.

small protein B (SmpB) See the entry TRANS TRANSLATION.

small RNAs See SRNAS.

small silencing RNAs See RNA INTERFERENCE, SIRNA and also MICRORNAS.

small t antigen (of SV40 virus) See T ANTIGEN.

small ubiquitin-like modifier See SUMOYLATION.

smart probe A (singly-labeled) PROBE based on the principle of the MOLECULAR BEACON PROBE. The 5′ end of this (stem–loop) probe carries a fluorophore that, in the *unbound* probe, is quenched by guanosine residues in the (3′) complementary stem region. The fluorophore becomes unquenched when the probe binds to its target sequence.

[Example of use: Nucleic Acids Res (2006) 34(13):e90.]

smear-positive specimen See the entry AMTDT.

Smith–Lemli–Opitz syndrome An autosomal recessive syndrome, involving (e.g.) mental retardation, reported to be due to defective cholesterol metabolism resulting from a mutation in gene *DHCR7* (encoding 3β-hydroxysterol-Δ7 reductase). (See also GENETIC DISEASE (table).)

SmpB Small protein B – see the entry TRANS TRANSLATION.

SMT3 The SUMO protein in *Saccharomyces cerevisiae*: see the entry SUMOYLATION.

Smt4 (Ulp2) See SUMOYLATION.

sncRNAs Small, non-coding RNAs: see the entry SRNAS.

snip-SNP Any SNP that affects RESTRICTION ENDONUCLEASE recognition site(s).

[Program for identifying snip-SNPs in reference sequences: BMC Genetics (2006) 7:27.]

SNP Single-nucleotide polymorphism: in the genome of an individual, a variant nucleotide which differs from that *at the corresponding site* in most other members of the population – but which occurs in at least 1% of other individuals within that population; for example, if 1% of the population has the sequence . . .TTTCAT. . . but the majority have . . .TTCCAT. . . then the minority group have an SNP at the third nucleotide in this sequence.

SNPs are common in the human genome. They are found approximately once every 100 to 300 bases, making a total of several million SNPs in the whole human genome. Most of the recorded SNPs in humans involve a C-to-T change.

SNPs which occur with frequencies below ∼5% have been classified as 'rare'. Those that occur with frequencies >5% are said to be 'common'.

SNPs are found in coding and non-coding sequences. Many SNPs have no detectable effect, but some may influence the response e.g. to infectious agents (bacteria, viruses etc.) and to chemicals, including drugs; this characteristic is helping to shape pharmaceutical/biomedical research: the comparison of SNP patterns in groups of drug-resistant and drug-sensitive patients may reveal important links between specific SNP(s) and the efficacy of a given drug – see 'association studies' in the entry SNP GENOTYPING.

SNP patterns may also be helpful in elucidating the polygenic nature of certain diseases in which a given gene is only partly responsible for the pathology.

The NCBI website on single-nucleotide polymorphisms is accessible at:
http://www.ncbi.nlm.nih.gov/projects/SNP/

Because of their potential uses in biomedicine, SNPs are being actively characterized and recorded in public databases by various government and commercial organizations. However, one report suggested that some of the SNPs recorded in public databases may be associated with *editing* sites rather than genuine SNPs [Nucleic Acids Res (2005) 33(14):4612–4617].

SNPs, which are found in both prokaryotic and eukaryotic microorganisms, have been used e.g. for phylotyping and for classification, and have been exploited in genome mapping. [Global phylogeny of *Mycobacterium tuberculosis* based on SNP analysis: J Bacteriol (2006) 188 (2):759–772.]

SNP genotyping Any procedure in which samples of genomic DNA are examined with the object of detecting and recording specific SNPs (see SNP; cf. SNP MAPPING).

Large collections of validated SNPs are needed in order to facilitate the so-called 'association studies' in which particular traits – such as susceptibility to a given disease, resistance to a drug etc. – may be linked to an established pattern of SNPs.

For example, one study reported that an SNP (G→A) at +299 in the resistin gene may influence the susceptibility to type 2 diabetes in a population studied [J Clin Biochem Nutr (2009) 44(1):104–110]. Data from a different study suggested that some of the SNPs evaluated were associated with the risk of ovarian cancer, although the level of observed effects was too little to detect associations with individual SNPs [PLoS ONE (2009) 4(6):e5983].

(Association studies are also carried out in the context of COPY NUMBER VARIANTS.)

Problems in separating true from spurious associations may be assisted by the GENOTYPE IMPUTATION approach.

SNPs are also studied in a veterinary setting: for example, genome-wide survey of SNP variation in sheep [PLoS ONE (2009) 4(3):e4668]; development/characterization of a high-density SNP genotyping assay for cattle [PLoS ONE (2009) 4(4):e5350]; design of a high-density SNP genotyping assay in the pig [PLoS ONE (2009) 4(8):e6524].

SNP genotyping also has uses in a forensic context: see the entry FORENSIC APPLICATIONS.

Methods for detecting SNPs

Various methods have been developed for the simultaneous genome-wide typing of thousands of SNPs. One such method is MOLECULAR INVERSION PROBE GENOTYPING (q.v.).

For detecting small numbers of SNPs (e.g. in studies on specific candidate genes) use can be made of a modified form of ALLELE-SPECIFIC PCR which uses T_m-shift primers. In this method [BioTechniques (2005) 39(6):885–893], the alleles of a given gene are assayed by using two allele-specific forward primers, each primer containing a 3′ terminal base which is complementary to a given SNP in one of two allelic variants of the gene; the same reverse primer is used with both of the forward primers. The forward primers are differentiated by incorporating a 14-nt GC-rich 5′ tag in one primer, and a 6-nt GC-rich 5′ tag in the other; the products of each primer can thus be distinguished by their THERMAL MELTING PROFILES, i.e. the products (dsDNA amplicons) formed from one primer are characterized by a melting temperature (T_m) which differs from that of the products formed from the other primer. As a consequence, melting curve characteristics of a given product will indicate which of the two (allele-specific) primers gave rise to that product. (An earlier procedure, in which only one primer was tagged, was found to be problematic in that, in some cases, the primers were amplified unevenly.)

A novel approach to the detection of SNPs was suggested by the finding that linear dsDNA fragments and oligonucleotides form highly stable RecA-mediated terminal/subterminal *triple*-stranded structures if the oligos are complementary to the 5′-phosphate end of a strand in the dsDNA fragments; the stability of this structure is markedly reduced by a single mismatched base – regardless of its location within the structure. In this scheme, therefore, the (sample) dsDNA fragments can be effectively probed by sequence-specific oligonucleotides.

A number of rare SNPs have been detected by the method known as ECOTILLING.

Detection and quantitation of SNPs can also be achieved by multiplex QEXT (q.v.).

SNP mapping A phrase used to refer to a procedure in which a given gene is mapped by using SNPs as genetic *markers* (cf. SNP GENOTYPING).

In so-called interval mapping, the object is to locate a gene of interest in the interval between two specific SNPs.

SNR Single-nucleotide repeat: a sequence of nucleotides *of one type*, at a given genetic locus, that may vary in length (i.e. in number of constituent nucleotides) in different members of a population; variation in length may result e.g. from SLIPPED-STRAND MISPAIRING. Mutation rates at SNR loci are reported to be high, apparently reflecting the ease with which events such as slipped-strand mispairing may occur.

SNRs may be found at various loci within a given genome. SNRs in the genome of *Bacillus anthracis* (causal agent of anthrax) have been exploited for the TYPING of this pathogen (an organism for which various other typing procedures have proved to be inadequately discriminatory) [J Clin Microbiol (2006) 44(3):777–782].

SNRs are also called 'mononucleotide repeats' and 'single-base repeats'.

snRNAs (small nuclear RNAs) RNA molecules, <300 nt, rich in uridylic acid residues, and capped (see CAP), present in the eukaryotic nucleus. There are six types, designated U1–U6.

snRNAs complex with proteins to form *small nuclear ribonucleoprotein particles* (snRNPs; snurps) which are involved in the SPLICING of pre-mRNAs; they function e.g. as *external guide sequences* (EGSs) in the (human) splicing reaction, i.e. helping to select a particular splice site.

snRNPs See SNRNAS.

snurps See SNRNAS.

SOB system The equivalent of the SOS SYSTEM in the Gram-positive bacterium *Bacillus*.

SOC medium (S.O.C. medium) A medium used for incubating bacteria following transformation. The medium contains (per liter): tryptone 20 g, yeast extract 5 g, glucose 3.6 g, together with sodium chloride (10 mM), potassium chloride (2.5 mM), magnesium chloride (10 mM), magnesium sulfate (10 mM).

Sodalis A genus of Gram-negative bacteria found as intra- and inter-cellular symbionts in tissues of the tsetse fly (*Glossina*); these bacteria can be grown *in vitro*.

(See also BUCHNERA, WIGGLESWORTHIA, WOLBACHIA.)

sodium butyrate (*DNA technol.*) An inhibitor of histone deacetylases (HDACs) (see HDAC).

sodium dodecyl sulfate (SDS) ($CH_3(CH_2)_{11}OSO_3Na$) An agent which binds co-operatively to proteins via the hydrophobic part of the molecule.

SDS is used for e.g. denaturing proteins, determining the molecular weight of proteins, and for separating components of protein–nucleic acid complexes.

(See also SDS–PAGE.)

sodium polyanetholesulfonate (SPS) An ANTICOAGULANT; it can inhibit PCR.

Elimination of SPS has been achieved in several ways, e.g.

with benzyl alcohol. An alternative method (used for bacteria grown in a blood-culture medium) is to sonicate the bacteria in the presence of charcoal and guanidinium hydrochloride [J Clin Microbiol (2007) 45(6):1927–1935].

SOE PCR See SPLICE OVERLAP EXTENSION PCR.

Soj protein See PAR LOCI.

solenoid See CHROMATIN.

Solexa sequencing An approach to MASSIVELY PARALLEL DNA SEQUENCING (Illumina Inc.) in which surface-attached target fragments (within a flow cell) are amplified – thus forming separate colonies (i.e. clones) of replicated fragments. Single-stranded templates are then sequenced by using fluorescently labeled, reversibly chain-terminating nucleotides: following the addition of each nucleotide to the sequencing primer, the label is read and removed – leaving a free 3′ end for addition of the next labeled nucleotide. Thus, the primer is extended one base at a time. The flow cell has eight separate lanes and a total capacity of over 40 million reads, with an average read length <40 bases. Reads as short as this can cause problems e.g. when incorrect bases are recorded in individual reads; two sets of read data, obtained with the Solexa system, were analyzed with the object of detecting biases in error positions and rates [Nucleic Acids Res (2008) 36(16):e105].

Solexa was adapted for diagnosis of 'trisomy 21' from fetal DNA in maternal plasma [Proc Natl Acad Sci USA (2008) 105(51):20458–20463]. [Other uses (e.g.): Nucleic Acids Res (2008) 36(19):e122; BMC Genomics (2009) 10:422 & 439.]

SOLiD™ technology 'Sequencing by oligonucleotide ligation and detection': a system (Applied Biosystems) used in a form of MASSIVELY PARALLEL DNA SEQUENCING. This system is reported to generate >20 Gb of sequence per run.

In this method, the library of template DNA molecules can be prepared in either of two ways. The 'fragment library' can be prepared by ligating adaptors to each end of the fragments of genomic DNA. Alternatively, a *mate-paired* library can be prepared. This involves initially ligating adaptors to the ends of the fragments. These molecules are circularized – and then cut to leave an internal adaptor sequence flanked by regions of the original fragment; adaptors are then ligated to the ends of *these* fragments. A mate-paired library is intended to offer greater precision in the characterization of structural variation across the genome.

The target molecules are bound to beads and then subjected to emulsion PCR (see entry GS FLX for details of emPCR). In a later stage, the denaturation of DNA is followed by a bead-enrichment process to exclude any beads lacking appropriate template DNA. The 3′ ends of the (bead-bound) templates are then modified so as to permit their covalent binding to a glass slide (the 5′ end of each template being bound to the bead via the adaptor sequence). When bound to the slide the templates are ready for sequencing.

The sequencing primer binds to the adaptor sequence at the 5′ (bead) end of each template. Templates are then exposed to a set of differentially labeled di-base probes that offer various two-base combinations to complement the first two unknown bases in the template. For example, one set of probes may be 3′-GG–label[a], 3′-CA–label[b], 3′-GC–label[c] and 3′-CT–label[d] – probing, respectively, for 5′-CC, 5′-GT, 5′-CG and 5′-GA in the template. (The four labels – a to d – are four fluorophores with distinct emission spectra, i.e. one may fluoresce yellow, another red etc.) If a given probe binds to the two unknown bases in the template it is ligated to the 5′ end of the primer; the label is then cleaved, leaving both nucleotides ligated to the primer. These two nucleotides are identified by the color of the fluorescence of the cleaved label. Primer extension thus occurs two bases at a time in the 3′-to-5′ direction. Such ligation and cleavage is repeated cyclically with different sets of probes; the number of cycles determines the read length in a given run.

After the first run, the product (i.e. the extended primer) is removed from the template. Another primer is hybridized to the template with a starting point staggered by 1 nucleotide, and a second round of sequencing is carried out. Five rounds of primer re-set are completed for each sequence. In the case of mate-paired libraries, the entire process is repeated with a primer complementary to the internal adaptor.

solid-phase PCR (on-chip PCR) A form of PCR involving the extension of primers on an immobilized complementary copy of the target sequence.

Two types of primer are used. The *immobilized* primers are covalently attached, via their 5′ end, to a solid support. These primers are complementary to the 3′ end of the given target sequence. When these primers bind the target sequence they are extended (3′ →) using the target sequence as template; *an immobilized primer therefore develops into a complementary copy of the target sequence*. The target fragment dissociates from the (extended) immobilized primer during temperature cycling.

The second type of primer is not immobilized, i.e. it is free in solution; it is labeled, and its sequence is *identical* to the 5′ end of the target sequence. This primer binds to the 3′ end of the (extended) immobilized primer and uses it as a template; extension of the *labeled* primer therefore forms a copy of the target, which subsequently dissociates from its (immobilized) template during temperature cycling. Hence, ongoing binding and extension of the labeled primers amplifies the target.

[Use of on-chip PCR for identifying bacteria from clinical samples: J Clin Microbiol (2004) 42(3):1048–1057.]

The on-chip approach is also used for *in situ* solid-phase DNA amplification which involves the generation of *colonies* of target DNA (each of which is a cloned product); in this context, benzene-1,3,5-triacetic acid (BTA) was reported to be useful for attaching (5′-aminated) oligonucleotides to an aminosilanized glass surface [Nucleic Acids Res (2006) 34 (3):e22].

solid-phase sequencing See e.g. PRIMER WALKING.

solubility enhancement tag (SET) An (engineered) peptide on a protein expressed from a vector – the purpose of which is to enhance the solubility of the tagged protein. The sequences encoding the SET and the protein of interest (both present in

the vector) are jointly expressed as a FUSION PRODUCT. This system is marketed by Stratagene (La Jolla CA, USA) under the name VariFlex™.

There are various types of SET – e.g. SET1, SET2 – each with a solubility-enhancing effect which may differ from that of other SETs, for a given protein. Although the solubility of many proteins can be enhanced by SETs, this is not possible for all proteins.

[Uses of VariFlex™ SETs (e.g.): J Bacteriol (2007) 189 (17):6213–6221; Differentiation (2008) 76(10):1081–1092.]

(See also CHAMPION PET SUMO VECTOR and Q-TAG (sense 1).)

solution hybridization Hybridization of complementary sequences wholly within the liquid phase (none of the sequences being immobilized).

solvatochromic fluorescent dye A fluorescent dye that is able to change the intensity and/or emission characteristics of its fluorescence in response to local changes in the conditions – i.e. conditions at the site where it is bound to a given molecule.

Such dyes have been used e.g. to monitor protein–protein interaction – the (protein-bound) dye typically responding to (i.e. 'sensing') a decrease in polarity caused by the exclusion of water from the protein–protein interfacial region.

Oligonucleotide–peptide interactions were sensed by using a fluorescent dye (2-(2-furyl)-3-hydroxychromone) which is covalently bound to the N-terminal of the peptide. Interaction between peptide and oligonucleotide causes marked changes in the two emission bands of the dye; these changes correlate with the structure of the oligonucleotide–peptide complex – suggesting that changes in the dye's fluorescence pattern are site-specific. Measurement of the emission bands of this dye allowed a number of binding sites to be distinguished in the oligonucleotide [Nucleic Acids Res (2009) 37(3):e25].

somatic cell Any cell forming part of the structure, or body, of an organism.

(cf. GAMETE.)

somatic cell hybridization See CELL FUSION.

somatic cell nuclear transfer See SCNT.

somatic stem cell See STEM CELL.

SopA (in the F plasmid) See F PLASMID.

sortase An enzyme, encoded by many types of Gram-positive bacteria; four classes (A–D) are distinguished – the sortases in a given class being characterized by particular function(s).

A sortase was initially identified in *Staphylococcus aureus*. This enzyme catalyzes site-specific cleavage of certain types of secreted protein. This type of protein is cleaved between a threonine residue and a glycine residue, threonine then being bound covalently to the cell-wall polymer peptidoglycan; this activity thus results in the *tethering* of the protein residue to the cell envelope. Proteins that are substrates for this enzyme are characterized by a C-terminal *cell wall sorting signal*; in *S. aureus* this is reported to include the consensus sequence LPXTG (note glycine as the terminal amino acid residue in this sequence). (*Bacillus anthracis* is reported to use NPKTG as the cell wall sorting signal.)

PROTEIN A is an example of a tethered protein in *S. aureus*.

In *Bacillus anthracis* (causal agent of anthrax) a sortase is required for processing the IsdC protein to form a cell-bound fragment needed for the scavenging of heme iron [J Bacteriol (2006) 188(23):8145–8152].

Sortases are also involved e.g. in aspects of sporulation and in the assembly of fimbriae (see FIMBRIA).

Sos A guanosine nucleotide exchange factor (encoded by gene *sos*).

(See also CYTOTRAP TWO-HYBRID SYSTEM.)

SOS box See SOS SYSTEM.

SOS chromotest An assay for DNA-damaging agents (genotoxins) which uses an engineered test strain of *Escherichia coli* containing the *sulA* gene (see SOS SYSTEM) fused with the *lacZ* gene (encoding β-galactosidase); *lacZ* is thus used as a reporter gene, being expressed when the test strain of *E. coli* is exposed to any compound which triggers expression of *sulA*. Expression of *lacZ* may be registered with a substrate such as ONPG.

The strain of *E. coli* used in the SOS chromotest is deficient in DNA excision repair, and its cell envelope is made more permeable to certain chemicals. These cells also synthesize the enzyme alkaline phosphatase constitutively; this is useful for excluding the possibility that the compound under test may inhibit protein synthesis (this possibility being excluded by colorimetric assay of alkaline phosphatase).

Use of gene *sulA* for detecting genotoxins was reported to be a less sensitive procedure than a ColD *cda* promoter-based system [Appl Environ Microbiol (2005) 71(5):2338–2346].

The SOS chromotest has been used e.g. to examine extracts of fungus-fermented rice [J Clin Biochem Nutr (2008) 43(2): 118–125].

(See also AMES TEST.)

SOS system (in bacteria) A system which mediates the cell's response to damaged DNA, or inhibition of DNA replication, caused e.g. by ULTRAVIOLET RADIATION, QUINOLONE ANTIBIOTICS (such as ciprofloxacin), or alkylating agents.

In *Escherichia coli* the SOS system involves approximately 30 genes. The SOS response includes e.g. inhibition of cell division and an increase in the cell's capacity for DNA repair (associated with increased mutagenesis). (The SOS response also promotes induction of phages in lysogenic bacteria – see LYSOGENY.)

In *E. coli* the SOS genes are normally inhibited by LexA: a transcriptional repressor which binds to a specific consensus sequence (the *SOS box*) in the promoter region of relevant genes/operons. (Some SOS genes are expressed at low levels under normal conditions.) In the presence of damaged DNA, the RECA protein promotes autocatalytic cleavage of LexA, leading to expression of the SOS response.

The product of SOS gene *sulA* (= *sfiA*) inhibits polymerization of the FtsZ protein. FtsZ proteins normally polymerize to form a ring-shaped structure (the *Z ring*), midway in the cell, as a preliminary stage in the development of the pre-

division septum. SulA therefore inhibits septum formation and, hence, inhibits cell division. (Growth may continue as septum-less *filaments*.) Resumption of normal septation takes place rapidly when repression of *sulA* is re-imposed by LexA following a return to normal conditions. Rapid resumption of septation is promoted by inactivation of SulA by the product of the *lon* gene, the Lon protease, a heat-shock protein whose production can be induced by some of the factors that trigger the SOS response; SulA is more stable in *lon* mutants, and even low-level induction of the SOS response in these cells may lead to lethal filamentation.

Various DNA repair systems are enhanced/expressed in the SOS response. One of these systems is an error-prone process (= *mutagenic repair*) resulting in increased levels of mutation in surviving cells; this process involves certain types of DNA polymerase (induced specifically during the SOS response) which carry out so-called TRANSLESION SYNTHESIS which is characterized by loss of base-pairing specificity at lesions in the DNA.

In *Staphylococcus aureus* the SOS response was reported to be triggered by the (fluoro)quinolone antibiotic ciprofloxacin, with concomitant de-repression of sixteen genes [J Bacteriol (2007) 189(2):531–539].

Induction of the SOS system by ciprofloxacin was shown to regulate transcription from the *qnrB2* promoter which controls a gene involved in resistance to fluoroquinolones [EMBO Rep (2009) 10(8):929–933].

Southern blotting Following electrophoresis of fragments of DNA in a gel strip: transfer of the fragments from the gel to a sheet of nitrocellulose (or to a different type of membrane) by capillary action; following the transfer the bands of fragments have the same relative positions as they had in the gel strip.

First, dsDNA fragments in the gel are denatured to ssDNA by alkali. The gel is then placed onto an (absorptive) wick in contact with a neutral buffer. A nitrocellulose sheet is placed over the gel strip; this is overlaid by a stack of (absorptive) paper towels pressed onto the nitrocellulose by a weight.

Buffer rises, via the wick, and passes into, and through, the gel and then through the (permeable) nitrocellulose filter into the paper towels; this upward flow of the buffer transports the fragments from the gel onto the surface of the nitrocellulose filter. The filter is then removed and is 'baked' (∼70°C) under a vacuum to bind the DNA to the surface of the filter.

The bound fragments can be probed, probe–fragment binding (i.e. *Southern hybridization*) indicating the presence of a given target sequence in a particular band of fragments in the original gel strip.

APT paper can be used instead of the nitrocellulose; this incorporates 2-aminophenylthioether. Before use, this reagent is modified to a reactive diazo derivative which binds ssDNA *covalently* to the paper (avoiding the requirment for baking). Probes can be removed from APT paper and a different set of probes can be used on the same sheet.

Southern hybridization Probing a membrane which has been prepared by SOUTHERN BLOTTING (q.v.).

(See also REVERSE SOUTHERN HYBRIDIZATION.)

Southwestern blotting A procedure used e.g. for detecting the presence of DNA-binding proteins in a cell lysate, and identifying their binding sites in genomic DNA.

The sample is first subjected to gel electrophoresis, and the *proteins* in the gel are transferred to a matrix. Proteins on the matrix are probed with specific, labeled DNA sequences, and protein–DNA binding is detected via the label. Any protein, and/or the DNA attached to it, can be isolated and analyzed.

(cf. CHROMATIN IMMUNOPRECIPITATION.)

Sox2 (transcription factor) See STEM CELL.

sp A designation used for an *unspecified* species, e.g. *Quercus* sp. (See also SPP.)

Sp1 Specificity protein 1 (also known as stimulatory protein 1) A (eukaryotic) transcription factor that binds to GC-rich sequences in the promoter regions of various genes.

SpA *Syn.* PROTEIN A.

SPA Scintillation proximity assay: a method for detecting the binding – or close apposition – of two molecules, one being radioactively labeled and the other having the properties of a *scintillant* (a scintillant being a substance that produces light when in close proximity to a source of radioactivity). Detection of a light signal is regarded as an indication of binding between the two given molecules. Thus, the principle of SPA resembles that of FRET, although the mechanism is different.

SPA has a range of uses, and commercial kits are available for particular applications. The procedure has been used e.g. to assay the binding of human tumor suppressor protein p53 to DNA; in this assay, the binding of p53 to radioactively (^{3}H) labeled DNA was detected by (i) an anti-p53 antibody, and then (ii) PROTEIN A linked to scintillant-containing beads [BioTechniques (2006) 41(3):303–308].

SPA has also been used for studying lysosome/phagosome targeting [Mol Biol Cell (2006) 17(4):1697–1710]; for the measurement of RNA methylation (using radiolabeled RNA) [Nucleic Acids Res (2009) 37(4):e32]; and for studying intracellular interactions with APOBEC3G [Retrovirology (2009) 6:56].

(See also DNA–PROTEIN INTERACTIONS.)

spacer arm In certain types of construct: a link which joins two molecules but which allows them to move freely relative to one another – for one example see the entry BIOTIN.

A spacer arm is sometimes called a 'linker'. (cf. LINKER.)

special AT-rich sequence-binding protein 1 See entry SATB1.

specialized transducing particle (STP) See TRANSDUCTION.

specialized transduction See TRANSDUCTION.

spermatozoa (in sex sorting) (*vet.*) See SEX SORTING.

SPEX Single-primer extension: an approach that is reported to provide more accurate sequence data from *ancient* DNA than can be obtained by the usual PCR methodology. In SPEX, the primers are targeted to defined locations on selected template strands.

It was claimed that SPEX gives strong evidence that C→U-

type changes are the primary cause of *post mortem* miscoding lesions in ancient DNA [Nucleic Acids Res (2007) 35: 5717–5728]. [Hominin DNA: Nucleic Acids Res (2009) doi: 10.1093/nar/gkp.897.]

(See also ANCIENT DNA.)

spheroplast An osmotically sensitive entity which is prepared from a cell by removal of its cell-wall materials.

Spheroplasts of Gram-positive bacteria can be prepared e.g. by incubating the cells in a solution containing EDTA, LYSOZYME and Tris buffer.

Spheroplasts of the yeast *Saccharomyces cerevisiae* can be prepared e.g. by treating the cells with ZYMOLYASE.

Spiegelmer (L-RNA) See L-RNA.

spinning-disk microscopy See IN VIVO FLUORESCENCE.

spinoculation Centrifugally assisted viral infection of cells.

Optimized spinoculation (centrifugation at 1200 *g*/2 hours, at 15^{2}C) was used for improved coinfection of pseudotyped retroviral vectors [J Biomed Biotechnol (2009) doi: 10.1155/2009/901079].

splice See the entry SPLICING.

SPLICE 'Swift PCR for ligating *in vitro* constructed exons' – a rapid *in vitro* method that is used for obtaining spliced coding sequences from genomic DNA; thus, a number of exons can be joined – eliminating the intervening introns – to form the complete coding sequence of a given gene.

The basic scheme is shown in the diagram.

This method has been used successfully for genes with 5 or 6 exons but may also be useful for larger and more complex genes (including those with a greater number of exons), given appropriate protocols; for example, for larger genes it may be necessary to use more than two rounds of PCR.

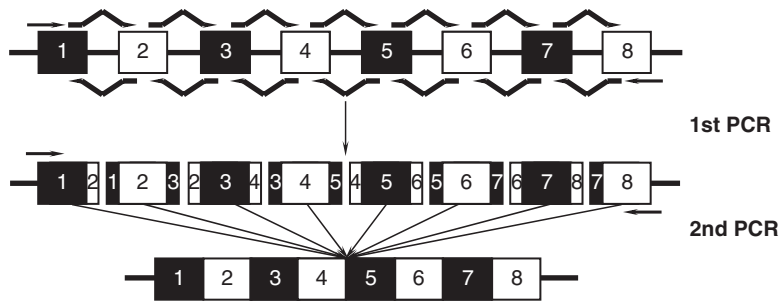
SPLICE was used, for example, to produce a single piece of DNA, with the full-length coding sequence of a five-exon opsin gene, from the genomic DNA of *Eulemur fulvus* (the brown lemur) [BioTechniques (2007) 43(6):785–789]. [See also Curr Biol (2007) 17:R161–R163 for other examples of the method.]

This procedure may have additional uses – for example, the production of splice variants and other modified constructs as well as the insertion of reporter functions etc.

splice overlap extension PCR (SOE PCR) A form of PCR that can be used to fuse two sequences of DNA – present on the same molecule or on separate molecules – without the use of restriction endonucleases. In order to be fused by SOE PCR, the 3′ end of one of the sequences must have a short region of homology with the 5′ end of the other. (If the two sequences do not have a common region of homology, such a region can be incorporated by carrying out a preliminary PCR (for each of the sequences) in which a suitable region is inserted (in a 5′ tag) in the primers.)

Initially, each of the two sequences is amplified, separately, by PCR; in each case the amplicons include the homologous region referred to.

Amplicons from both reactions are mixed and subjected to a further round of PCR. In this case, the primers are designed to bind to the *non*-homologous end of each of the two sequences. During the reaction, hybridization (at the region of homology) occurs between single strands of the two types of amplicon; subsequently, strand synthesis occurs by extension of the 3′ end of each amplicon using the other amplicon as a template. This results in the formation of a hybrid (chimeric) product which includes both of the original sequences. PCR,



SPLICE (swift PCR for ligating *in vitro* constructed exons): a method used for obtaining spliced coding sequences from genomic DNA.

In the 1st round of PCR, the internal primers (all 30 nt) are designed to bind to those specific sequences which are to be joined; thus, one half of a primer binds to the 3′ terminal region of a given exon while the other half of that primer binds to the 5′ end of the following (downstream) exon.

The required products of the 1st PCR (center row in figure) are purified. A mixture of equimolar concentrations of these products is then used in the 2nd round of PCR with only the external (i.e. outermost) primers.

Figure reproduced by courtesy of Wayne L Davies and colleagues, UCL Institute of Ophthalmology, London, UK.

using primers which bind to the *ends* of this double-stranded hybrid product, then amplifies the required hybrid sequence.

SOE PCR was used e.g. to produce hybrid mouse–human monoclonal antibodies by taking advantage of the homology between analogous regions of murine and human immunoglobulin genes [BioTechniques (2005) 38(2):181–182]. It has also been used e.g. for studies on conjugation in *Clostridium perfringens* [J Bacteriol (2008) 190(14):5075–5086]; for the genomic analysis of *Listeria monocytogenes* [Appl Environ Microbiol (2008) 74(13):3921–3934]; for work on β-catenin signaling PLoS ONE (2009) 4(2):e4310]; for studies on the interaction between *Pseudomonas aeruginosa* kinases [Genes Dev (2009) 23(2):249–259]; and for a study on the induction of a plant metalloproteinase [BMC Plant Biol (2009) 9:83].

splice site See ACCEPTOR SPLICE SITE and DONOR SPLICE SITE.

(See also SPLICING.)

splice-switching oligonucleotide (SSO) Any oligonucleotide designed specifically to modify a given SPLICING event – for example, to correct disease-causing aberrant splicing of the pre-mRNA of a given gene. The manipulation of splicing has been carried out frequently within cultured cells, and has also been used in (live) animal models. Positive results have been reported in clinical trials in patients with Duchenne muscular dystrophy.

[Therapeutic potential of splice-switching oligonucleotides: Oligonucleotides (2009) 19(1):1–14.]

spliced leader See TRANS SPLICING (of transcripts).

spliceosome A large, multi-component complex involved in SPLICING pre-mRNA to mature mRNA in (eukaryotic) cells.

In (human) cells the spliceosome includes e.g. snRNPs (see SNRNAS), SR PROTEINS, RNA helicase and other factors; it contains DEAD-BOX PROTEINS. The assembly of the spliceosome occurs by sequential addition of the components, some being added in pre-formed complexes.

The trypanosomatid spliceosome is involved in the process of TRANS SPLICING (q.v.).

[Evolution of the spliceosome: PLoS Comput Biol (2009) 5(3):e1000315.]

splicing (of pre-mRNA) The process in which non-coding sequence(s) in a SPLIT GENE are excised from the primary RNA transcript of the gene during formation of the final, mature mRNA. Splicing involves cleavage of both ends of a given intron, at precise phosphodiester bonds, and the formation of a phosphodiester bond between the exon on the 5′ side of the intron and that on the 3′ side of the intron.

Splicing, of course, occurs naturally in living cells. Under *in vitro* conditions, the introns in a split gene can be removed from isolated genomic DNA e.g. by the method called SPLICE (see entry).

(See also ALTERNATIVE SPLICING, CRYPTIC SPLICING and EXON TRAPPING.)

The mechanism of splicing is different in different types of split gene.

Bacterial introns are primarily, or solely, *autocatalytic* (i.e. self-splicing) – meaning that the RNA has intrinsic catalytic properties which are used to excise intronic RNA.

Introns of phage T4 are autocatalytic *in vitro*. The intron in the T4 thymidylate synthase gene was reported to require a host factor for efficient intracellular splicing.

Splicing in human pre-mRNAs

The splicing of *human* pre-mRNAs is a complex, multi-step process involving development of a SPLICEOSOME. One early requirement is the need to select an appropriate (5′) DONOR SPLICE SITE and a (3′) ACCEPTOR SPLICE SITE in respect of each intron; this process of selection appears to be facilitated by the *cis*-acting regulatory elements within the pre-mRNA – the so-called *exonic splicing enhancers* (ESEs) and *exonic splicing silencers*. ESEs often bind to the 'RNA-recognition motif' of SR PROTEINS; such binding may help to define the appropriate splice sites e.g. by promoting the binding of other components of the spliceosome to the arginine–serine region of the SR protein and/or by antagonizing effect(s) of adjacent silencer sequences. Such positive selection of splice sites is needed e.g. to counteract the effect of *pseudo-exons* that are flanked by potential splice sites.

An early event is duplex formation between U1 SNRNA and the 5′ splice site. Other components, including snRNAs, join the complex; for example, U2 binds at the *branch point* near the 3′ end of the intron. Subsequently, other snRNAs join the spliceosome and some are destabilized or displaced. The 5′ end of the intron is bonded to a nucleotide at the branch point and the intron is excised in the form of a tailed loop (*lariat*) – a phosphodiester bond being formed between the 3′ end of one exon and the 5′ end of the next. [Development and fate of the lariat in the splicing process: Nucleic Acids Res (2009) 37(3):891–902.]

U1-*independent* splicing may contribute to the regulation of ALTERNATIVE SPLICING events in certain pre-mRNAs [Nucleic Acids Res (2009) 37(6):1907–1914].

In addition to splicing, another post-transcriptional event is *capping* (see CAP) and a further event is POLYADENYLATION.

A tool has been proposed for identifying splicing motifs in human DNA – and for predicting the effects of mutations on splicing signals [Nucleic Acids Res (2009) doi: 10.1093/nar/gkp215].

Splicing in pathogenesis and therapy

Faults in the splicing process, either inherited or due to a new mutation, may cause disease – e.g. through lack or deficiency of the normal product. Thus, aberrant splicing was reported to be responsible for the Axenfeld–Rieger syndrome (in some patients); variation in the severity of the splicing fault might reflect the range of manifestations in this syndrome.

One approach to problems in splicing is the use of so-called *splice-switching oligonucleotides* (SSOs); SSOs can correct splicing faults by re-directing the splicing machinery to select normal splice sites. This approach was reported to have given positive results in trials on patients with Duchenne muscular dystrophy. [Therapeutic potential of splice-switching oligonucleotides (review): Oligonucleotides (2009) 19(1):1–14.]

In a different approach, a (normal) splicing reaction, which

is needed for synthesis of an essential retroviral glycoprotein, was inhibited by the use of certain indole derivatives. These compounds, which acted by selectively inhibiting a particular SR protein involved in the synthesis of the glycoprotein, were able to protect mice from retrovirus-mediated pathogenesis [PLoS ONE (2009) 4(2):e4533].

splicing (of proteins) See INTEIN.

splinker ('sequencing primer linker') A synthetic oligodeoxynucleotide containing INVERTED REPEAT sequences that form a (double-stranded) hairpin structure which contains a recognition site for a RESTRICTION ENDONUCLEASE. Splinkers are used e.g. for sequencing DNA restriction fragments.

A given target restriction fragment is treated with alkaline phosphatase to de-phosphorylate the 5′ ends. Then, using an appropriate ratio of splinkers to fragments, a splinker (whose 5′ end *is* phosphorylated) is ligated to each fragment. That strand of the fragment which is *not* ligated to the splinker is removed (by heating and rapid cooling), leaving the primed splinker-bound strand. The primer, i.e. 3′ end of the splinker, is extended (in a reaction mixture containing dideoxy chain-terminating nucleotides), and each of the newly synthesized strands is released from its template strand by cleavage at the splinker's internal restriction site and heat denaturation.

This procedure avoids e.g. the need to synthesize a specific primer.

split-Cre A method reported to detect the *simultaneous* activity of two different promoters within the same cell. The method is based on COMPLEMENTATION of two fragments of the Cre recombinase (see CRE-LOXP SYSTEM) which, together, form a functional enzyme that activates a recombination-dependent reporter system.

Each of two parts of the coding region of the Cre gene was fused (separately) to a sequence encoding a 'protein interaction domain' (derived from yeast transcription factor GCN4). These two sequences – transcribed from the two promoters being monitored – were translated as fusion products which bound together, via their protein interaction domains, to form a functional Cre recombinase. Cre then activated a reporter gene by excising a sequence (bracketed by *loxP* sites) which had been blocking reporter activity [PLoS ONE (2009) 4(1): e4286].

split gene (interrupted gene) Any gene whose coding sequence is interrupted by one, several or many *non-coding* sequences of nucleotides (*intervening sequences*: see INTRON) that do not encode any part of the gene product. Intronic sequence(s) are represented in the *primary* RNA transcript (*pre-mRNA*) – but are normally spliced out (see SPLICING) during formation of the mature mRNA (*messenger RNA*, *mRNA*). The mRNA of a split gene thus consists of covalently linked sequences that correspond to the *coding* sequences (*exons*) of the gene.

(See also ALTERNATIVE SPLICING.)

In higher eukaryotes, many or most genes are split genes. Split genes also occur in eukaryotic microorganisms, such as *Dictyostelium* and *Saccharomyces*, and all (protein-encoding) genes in the slime mold *Physarum polycephalum* apparently contain introns.

Some archaeal, bacterial and viral genes are split genes.

Examples of intron-less human genes include those encoding interferons α and β.

The (human) insulin gene contains one intron. By contrast, certain genes contain 50 to 100 introns.

split intein See INTEIN.

Spodoptera frugiperda The armyworm, cell cultures of which are used for the synthesis of recombinant proteins encoded by baculovirus expression vectors – see the entry BACULOVIRUS EXPRESSION SYSTEMS.

spoligotype See SPOLIGOTYPING.

spoligotyping A method used for simultaneous detection and TYPING of strains of the *Mycobacterium tuberculosis* complex [original description of spoligotyping: J Clin Microbiol (1997) 35:907–914].

Essentially, spacer sequences within the chromosomal DR (direct repeat) region of the strain under test are amplified by PCR; the amplicons are hybridized to a set of oligonucleotides that represent spacer sequences of the reference strain (i.e. *M. tuberculosis* H37Rv). (The name 'spoligotyping' derives from a contraction of the phrase 'spacer oligotyping'.) The pattern of matching, and mis-matching, of the test strain's spacer sequences (amplicons) with the set of reference oligonucleotides is called the *spoligotype* of that strain.

Spoligotyping was used e.g. for grouping survey strains of *Mycobacterium tuberculosis* [BMC Microbiol (2009) 9:39] and for molecular characterization of *M. tuberculosis* in New Delhi [PLoS ONE (2009) 4(2):e4540].

***spoT* gene** In *Escherichia coli*: a gene that encodes an enzyme which can synthesize and degrade the alarmone ppGpp.

Spot42 A bacterial sRNA: see the entry SRNAS. Spot42 appears to regulate the *galK* gene (involved in galactose metabolism) in *Escherichia coli*.

spp Species (plural); a single (unnamed) species is designated 'sp' – e.g. *Quercus* sp (a species of *Quercus*).

***Sp*Pol4** A DNA polymerase in the X FAMILY.

SPS SODIUM POLYANETHOLESULFONATE (q.v.), a blood ANTICOAGULANT which inhibits PCR.

Spumavirinae A subfamily of viruses (family Retroviridae: see RETROVIRUSES) which occur in various species of mammal; these viruses are apparently non-pathogenic. In tissue culture the spumaviruses produce characteristic cytopathic effects – including the development of syncytia and the formation of a highly vacuolated ('foamy') cytoplasm within infected cells, hence the name 'foamy viruses'.

The spumaviruses include e.g. bovine syncytial virus, feline syncytial virus, and the human and simian foamy viruses; the transmission of simian foamy virus (SFV) from macaques to humans was reported some years ago.

spumaviruses Viruses of the subfamily SPUMAVIRINAE.

SR proteins In the SPLICING of pre-mRNA: certain proteins involved in the choice of splice sites.

The N-terminal region contains an *RNA-recognition motif*; the C-terminal region consists mainly of alternating residues

of serine and arginine – the designation SR deriving from the (one-letter) symbols for these two amino acids.

Some engineered systems that are designed to modify gene expression interfere with the function of SR proteins. Thus, for example, a number of antisense oligomers were designed to induce exon skipping in gene *DMD* (which is defective in Duchenne muscular dystrophy); subsequent analysis of these oligomers showed that one of the features of the *successful* (active) oligomers was the ability to inhibit binding of certain SR proteins [Mol Therapy (2009) doi: 10.1038/mt.2008.287].

Interaction between the SR protein SF2/ASF and ESEs (i.e. exonic splicing enhancers) during splicing was reported to be selectively inhibited by certain types of indole derivative; the inhibition of SF2/ASF inhibits (splicing-dependent) synthesis of a retroviral glycoprotein (thus inhibiting viral replication), and this was able to protect mice from a challenge with MLV (murine leukemia virus) [PLoS ONE (2009) 4(2):e4533].

src (*SRC*) An ONCOGENE first detected in Rous sarcoma virus (RSV). RSV is an avian virus which can cause experimental tumors in some mammals. The v-*src* product exhibits TYROSINE KINASE activity.

The product of c-*src* is expressed only weakly in a normal cell.

Differences in the C-terminal regions of viral and cellular gene products may account for the transforming ability of the former.

SRF SUBTRACTED RESTRICTION FINGERPRINTING.

sRNAs (small non-coding RNAs, *or* short non-coding RNAs, *or* small RNAs) A wide range of small RNAs, both prokaryotic and eukaryotic, which do not encode proteins but which, in many cases, are known to have a gene-regulatory role.

(cf. LONG NON-CODING RNAS.)

(See also NON-CODING RNAS.)

Both the eukaryotic and prokaryotic sRNAs appear to exert their regulatory functions primarily at the post-transcriptional level, typically by controlling translation or stability of target mRNAs; although they are typically inhibitory, some sRNAs are reported to exert a positive effect on gene function.

Eukaryotic sRNAs

This group includes microRNAs and the effector molecules of RNA INTERFERENCE, the siRNAs – see entries MICRORNAS and SIRNA, respectively. (*Note the difference between siRNA and sRNA*.)

Prokaryotic sRNAs

Examples of prokaryotic sRNAs have been known for many years (see the entries COLE1 PLASMID, FINOP SYSTEM and R1 PLASMID), but only quite recently has it become clear that the prokaryotes contain a large number of gene-regulatory RNAs that are involved in cellular responses to stimuli as diverse as oxidative stress, phosphosugar stress, QUORUM SENSING and iron limitation.

The prokaryotic sRNAs are reported to be ~50 to several hundred nucleotides in length (compare miRNAs), and most appear to bind to the mRNA 5′-UTR (compare miRNAs).

Many sRNAs bind strongly to the so-called Hfq protein, an RNA chaperone which is thought to protect sRNAs that have not bound to their target mRNAs. Moreover, the Hfq protein may promote sRNA–mRNA binding. One report referred to the formation of a ribonucleoprotein complex that consists of the sRNA (SgrS) bound to an enzyme (RNase E) through the Hfq protein; this complex mediates the silencing of a target mRNA that encodes a membrane component of the glucose-specific phosphoenolpyruvate phosphotransferase system (the PTS system) [Mol Microbiol (2006) 61(4):1013–1022].

mRNA-inhibitory sRNAs seem to be the most abundant in bacteria. Among these are MicF (involved in osmotic stress), OxyS (oxidative stress), RyhB (metabolism of iron) and SgrS (phosphosugar stress). However, *up*-regulation by an sRNA is seen with DSRA – which promotes the translation of sigma factor σ^S. (See also entries for GCVB, QRR1 and SPOT42.)

[IntaRNA: efficient prediction of bacterial sRNA targets – incorporating both target site accessibility and seed regions: Bioinformatics (2008) 24(24):2849–2856.]

[High-throughput, kingdom-wide prediction and annotation of bacterial sRNAs: PLoS ONE (2008) 3(9):e3197.]

[sRNA map: genomic maps for sRNAs, their regulators and their targets in microbial genomes: Nucleic Acids Res (2009) 37(suppl 1):D150–D154.]

Engineered bacterial sRNAs

Engineered RNAs have been utilized for GENE SILENCING in eukaryotic cells (e.g. SHORT HAIRPIN RNA), and similar molecules (artificial *trans*-encoded sRNAs; = atsRNAs) have been used in bacteria [Nucleic Acids Res (2009) doi: 10.1093/nar/gkp447].

sRNAs tagged with APTAMERS have been expressed *in vivo* [Nucleic Acids Res (2009) doi: 10.1093/nar/gkp719].

***SRY*-based assay** (for gender determination) An assay that may be used to determine the gender of a source of (human) DNA. This assay involves PCR amplification of a specific marker in the *SRY* gene – which encodes a transcription factor involved in testes determination. One study reported that no amplicons were detected from female template DNA at a concentration of 10 ng/μL – while amplicons were reliably recorded from male DNA at levels as low as 25 pg [J Forensic Sci (2009) 54(3):551–555].

(See also DNA SEX DETERMINATION ASSAY and FORENSIC APPLICATIONS.)

SSB protein SINGLE-STRAND BINDING PROTEIN (q.v.).

SSCP analysis (single-strand conformation polymorphism analysis) A technique that is used for comparing single-stranded samples of *related* nucleic acid by comparing their electrophoretic speeds within a polyacrylamide gel; even a one-base difference between strands may be detectable if it affects the intra-strand base-pairing (i.e. the conformation) in such a way that it causes a change in electrophoretic mobility.

Such is the specificity of this method that, in some cases, a result can indicate a particular type of mutation at a precise locus (i.e. a specific nucleotide).

Note that SSCP analysis uses a *non*-denaturing gel; this is used in order to preserve (or promote) the intra-strand base-

pairing on which differentiation depends.

Samples are often strands ~100–300 bases in length.

SSCP analysis has been used e.g. for TYPING bacteria, for screening bacterial resistance to antibiotics [J Clin Microbiol (2006) 44(3):970–975], and for screening the *MFN2* gene in patients with a clinical diagnosis of Charcot–Marie–Tooth disease [BMC Med Genet (2006) 7:53].

More recently, the method was used for the rapid differentiation of yeast species [Appl Environ Microbiol (2008) 74(9): 2604–2611]; for detecting SNPs [BMC Genet (2009) 10:10]; for studying mutations in the *TP53* gene [J Exp Clin Cancer Res (2009) 28(1):17]; screening cDNA for mutations [BMC Mol Biol (2009) 10:5]; SNP analysis [Haematologica (2009) 94:1301–1306]; for studies on coral [PLoS ONE (2009) 4(7): e6364] and on G6PD mutations [Malaria J (2009) 8:152].

PCR-SSCP

PCR-SSCP is a convenient approach that has a range of uses. In this approach, PCR is used simply to provide the single-stranded targets.

Samples for use in SSCP may be conveniently prepared by ASYMMETRIC PCR. Alternatively, samples can be prepared by using a standard form of PCR, with *one* of the primers 5′-phosphorylated, and digesting the amplicons with the lambda (λ) phage exonuclease (see EXONUCLEASE). A further option is to carry out PCR with *one* of the primers biotinylated; the strands formed by extension of this primer can subsequently be isolated from the (denatured) amplicons by a streptavidin-based system.

For detecting resistance-conferring mutations, PCR is used to amplify the particular region of a genome that is associated with resistance to a given antibiotic; then, the single-stranded amplicons are compared with those produced from wild-type (antibiotic-sensitive) strains.

PCR-SSCP has been used, for example, to detect strains of *Mycobacterium tuberculosis* in which mutation in the *rpoB* gene (which encodes the β-subunit of RNA polymerase) has conferred resistance to the anti-tuberculosis drug rifampicin (rifampin) whose target is RNA polymerase; in this organism the mutation-prone region of the *rpoB* gene was amplified by PCR and the amplicons examined as described.

Potential problems with SSCP analysis

When studying antibiotic resistance with SSCP analysis, one problem is that a SILENT MUTATION in an antibiotic-sensitive strain may produce a conformation which simulates that of an antibiotic-resistant strain, yielding a false-positive result [see, for example, J Clin Microbiol (1997) 35:492–494]. However, uncertainty can be resolved by sequencing the test sample.

SSCP analysis is used to monitor the distribution of strains of the (ssRNA, positive sense) hepatitis C virus (HCV); this procedure involves examining DNA copies of the viral RNA. However, this approach has been questioned as it was found that DNA copies of different HCV strains do not necessarily exhibit distinct patterns of electrophoretic migration [J Virol (2009) 83(17):9018–9021].

(See also DGGE.)

ssDNA Single-stranded DNA.

SSM See SLIPPED-STRAND MISPAIRING.

SSO See SPLICE-SWITCHING OLIGONUCLEOTIDE.

SSR (1) See SITE-SPECIFIC RECOMBINATION.
(2) See SHORT SEQUENCE REPEATS.
(3) See SIMPLE SEQUENCE REPEATS.

SsrA RNA *Syn*. tmRNA – see the entry TRANS TRANSLATION.

ssRNA Single-stranded RNA.

SssI (M.SssI) A methyltransferase from *Spiroplasma* (species MO1) which is reported to be specific for 5′-CpG-3′ sites (in purified DNA); SssI methylates the cytosine residue at the 5-position.

stacking gel See SDS–PAGE.

staining (of DNA) See DNA STAINING.

standard virus See DEFECTIVE INTERFERING PARTICLE.

staphylococcal cassette chromosome *mec* See SCCMEC.

Staphylococcus A genus of Gram-positive bacteria that include pathogens and commensals; the GC% of the genomic DNA in the type species, *S. aureus*, is ~33.

The cells are cocci (spheres) ~1 μm in diameter; they often occur in clusters, and generally contain orange/yellow (carotenoid) pigments. The staphylococci are facultatively anaerobic (i.e. they are able to grow with or without oxygen) and chemoorganoheterotrophic; their carbon sources include a range of 'sugars' (i.e. carbohydrates). These organisms are often halotolerant – meaning that their tolerance of salt (sodium chloride) is often greater than that typical of other bacteria. Staphylococci are catalase-positive.

The staphylococci are classified according to whether they produce, or fail to produce, the enzyme coagulase (which coagulates, i.e. clots, the plasma of certain species); coagulase converts the plasma protein fibrinogen to fibrin (which forms the clot). *S. aureus* typically produces coagulase (encoded by the COA GENE).

(See also PANTON–VALENTINE LEUCOCIDIN.)

S. aureus is a well-known pathogen. Problems of antibiotic resistance are found e.g. in the methicillin-resistant *S. aureus* strains (see the entry MRSA).

DNA-based TYPING of *S. aureus* has been carried out by various methods: see e.g. entries ACCUPROBE, MLVA, MLST, PFGE and SCCMEC.

star activity A change in the *specificity* of certain RESTRICTION ENDONUCLEASES (for example, BamHI and EcoRI) in non-optimal conditions.

A measure of star activity is given by the FIDELITY INDEX.

start point *Syn*. START SITE.

start site (of a gene) The nucleotide (referred to as +1) which corresponds to the site at which transcription of a given gene begins; it determines the 5′ terminal nucleotide of the coding sequence in an mRNA molecule.

The start site is usually a purine, often adenosine.

Nucleotides *downstream* from the start site (i.e. in the 5′ → 3′ direction) are numbered +2, +3, +4 etc.; those *upstream* from the start site (in the 3′ → 5′ direction) are numbered –1, –2, –3 etc.

(See also PRIMER EXTENSION ASSAY.)

STATs SIGNAL TRANSDUCERS AND ACTIVATORS OF TRANSCRIPTION.

stavudine See NUCLEOSIDE REVERSE TRANSCRIPTASE INHIBITORS.

steel factor (stem cell factor) See KIT.

stem-and-loop In a strand of nucleic acid: a structure which has been formed by base-pairing between (complementary) sequences of nucleotides, the hybridized regions forming the (double-stranded) stem and the intervening (single-stranded) region forming the loop. This kind of structure is used e.g. in MOLECULAR BEACON PROBES.

Stem–loop probes in a microarray were reported to increase the *specificity* of hybridization with targets [BioTechniques (2008) 44(1):119–126].

stem cell Any of certain types of cell which have the capacity for indefinite *self-renewal* and the ability to differentiate into other kind(s) of cell; on division, a stem cell gives rise either to two daughter stem cells (as occurs *in vitro* with *embryonic* stem cells in culture: see below) or to one stem cell and a cell that undergoes differentiation (as occurs e.g. in *hematopoietic* stem cells, which give rise to blood cells). Moreover, *in vitro*, differentiation may follow an appropriate stimulus.

In the adult animal, particular types of stem cell provide a reserve, being able to replace the cells (blood, skin etc.) that are lost during the life of the individual.

Different categories of stem cell have been distinguished on the basis of their origin and on the extent of their capacity for differentiation into other types of cell (but see later).

Embryonic stem (ES) cells (pluripotent stem cells)

The *pluripotent* embryonic stem cells, found in the inner cell mass of a BLASTOCYST, have the highest innate capacity of all the stem cells for differentiation; they have the potential to form any type of cell in the body. When grown *in vitro* for long periods of time, ES cells maintain pluripotency: progeny cells retain the potential for multilineage differentiation.

Maintenance of pluripotency in ES cells appears to involve continual interactions between multiple nuclear factors – this achieving a balance in which some interactions are inhibitory or antagonistic and others are positive or cooperative; as well as promoting pluripotency, this represses the genes involved in differentiation. Important factors involved in maintaining pluripotency include Oct4, Nanog and Sox2; although these particular transcription factors are found in both human and mouse ES cells, some of the other molecular markers of ES cells are found only in mouse ES cells or only in human ES cells. As well as their roles in maintaining pluripotency, some of these factors have important roles in the early development of the embryo.

Regulation of self-renewal and differentiation in embryonic stem cells, as well as in the germline and somatic tissue stem cells, involves contributions from the microRNAs (miRNAs) [Nature Rev Mol Cell Biol (2009) 10:116–125].

Nanog is a major component in the network of interactions that maintain pluripotency. Moreover, in cultured *mouse* ES cells, engineered expression of Nanog enhances self-renewal and it also permits ongoing growth of these cells without the (usual) need for leukemia inhibitory factor (see also FEEDER CELLS). Nanog exists as a homodimer in its interactions with other proteins in the pluripotency network [Proc Natl Acad Sci USA (2008) 105(17):6326–6331].

Oct4 also has a major role in pluripotency but is not needed for self-renewal in mouse *somatic* stem cells [Cell Stem Cell (2007) 1(4):403–415]). It is required for development prior to blastocyst formation – as early as the two-cell stage [PLoS ONE (2008) 3(12):e4109].

Sox2 has roles that are reported to include maintenance of appropriate levels of Oct4.

Transcriptional programs in ES cells, adult tissue stem cells and (human) cancer cells were compared, and an ES cell-like program was reported in a number of epithelial cancers [Cell Stem Cell (2008) 2(4):333–344].

Gene expression profiles in (human) ES cells and oocytes were found to show commonality in the genes linked to pluripotency and chromatin remodeling [BMC Genomics (2009) 10:10].

[Human embryonic stem cells (10 years on) (an overview): Lab Invest (2009) doi: 10.1038/labinvest.2008.162.]

Somatic stem cells

In this category, some authors place the *fetal stem cells* (from the fetus) and *adult stem cells* (found in various tissues, such as bone marrow). These cells have been regarded as having a capacity for differentiation lower than that of the pluripotent ES cells – with the capacity of fetal stem cells being greater than that of adult stem cells; they apparently differentiate into only a limited range of types of cell and have been described as *multipotent*. The '*tissue-specific*' *stem cells* normally give rise to only one type of cell.

Plasticity of stem cells

The more or less rigid distinction between the different types of stem cell may be somewhat artefactual. For example, some (adult) 'tissue-specific' stem cells were reported to have been re-programed to form cells of a different lineage. Moreover, the PIGGYBAC transposition system has been used to transfect human and mouse fibroblasts with genes encoding the transcription factors c-Myc, Klf4, Oct4 and Sox2 – which resulted in stable conversion of the transgenic (re-programed) fibroblasts to *induced pluripotent stem cells* (iPS cells); these cells expressed markers characteristic of pluripotent cells and also demonstrated the capacity to undergo differentiation [Nature (2009) doi: 10.1038/nature07863].

Stem cells in DNA technology and therapeutics

Stem cells can be obtained e.g. by the procedures used in the initial stage of whole-animal cloning: see the entry SCNT.

ES cells have been studied by various techniques, including RNA interference (RNAi) [shRNA vector for high-efficiency RNA interference in ES cells: BioTechniques (2007) 42(6): 738–743].

Mouse ES cells are widely used for studying gene function – they can be modified *in vitro* (e.g. by GENE TRAPPING) and

then inserted into an (isolated) blastocyst for re-implantation and development in a surrogate female animal.

Human stem cells have potential for use in *cell replacement therapy* in diseases such as hepatitis and Parkinson's disease – in which a particular subset of the body's cells have become non-functional. In such diseases the loss of functionality may be reversed by introducing into the body a healthy population of appropriate cells. Research in this area has been facilitated by the decision (in the USA, January 23, 2009) to permit the Geron company to carry out human clinical trials of ES cell-based therapy.

stem cell factor (steel factor) See KIT.

stem–loop See STEM-AND-LOOP.

sterilant Any agent used for STERILIZATION.

sterile cabinet *Syn*. SAFETY CABINET.

sterilization Any procedure which is used to ensure the death/permanent inactivation of all forms of life, including bacterial endospores, that are present on or within objects or materials, or within environments, subjected to that process.

One common method of sterilization involves treatment in an AUTOCLAVE.

Chemicals which have been used as *sterilants* (i.e. agents used for sterilization) include ethylene oxide, glutaraldehyde and β-propiolactone; in each case, physical conditions must be carefully controlled in order to achieve maximum activity of the sterilant.

Note. The phrase 'partial sterilization' is meaningless because sterility is an 'all-or-nothing' phenomenon.

Sterne strain (*Bacillus anthracis*) See BACILLUS ANTHRACIS.

steroid Any of the large group of biologically active tetracyclic compounds (which contain a tricyclic phenanthrene skeleton) that include e.g. bile acids, cholesterol and the sex hormones estrogens and androgens.

sticky ends (or cohesive ends) Complementary, single-stranded terminal regions present on dsDNA molecules. For example, the following shows a pair of sticky ends consisting of two 3′ five-base extensions:

```
5′-----NNNNNATCTG-3′          NNNNN----3′
3′-----NNNNN          3′-TAGACNNNNN----5′
```

The single-stranded regions can hybridize so that, if required, the two terminal regions can be ligated (see LIGATION).

Sticky ends can be generated e.g. by the staggered cleavage produced by *certain* types of RESTRICTION ENDONUCLEASE (e.g. EcoRI). Note that the single-stranded region may be a 3′ extension or a 5′ extension, depending on the mode of cutting of the particular restriction enzyme used.

Single-stranded extensions are also called *overhangs*.

Some restriction enzymes form BLUNT-ENDED DNA.

(See also the entry POLYLINKER.)

Compatible sticky ends can be produced (i) when DNA is cut by the same enzyme, or by an isoschizomer, or (ii) when DNA is cut by an unrelated enzyme that creates hybridizable sequences; for an example of the latter type of enzyme see MunI in the table in entry RESTRICTION ENDONUCLEASE.

sticky RICE A method used for adding STICKY ENDS (cohesive ends) to (blunt-ended) PCR amplicons – without cutting them with a restriction enzyme – so that the amplicons can then be inserted, directionally, into specific vector molecules. 'RICE' derives from 'restriction enzyme-independent cohesive ends' [BMC Biotechnol (2007) 7:52].

The RICE approach is outlined below.

The vector molecule contains two recognition sites for the type IIS restriction enzyme SapI – at separate locations on the molecule. SapI has the following recognition/cutting sites:

```
5′-GCTCTTCN↓NNN
3′-CGAGAAGNNNN↑
```

After cutting by SapI (at *both* locations), the vector has two terminal regions with 5′ 3-nucleotide overhangs, as follows:

```
. . .GAG                    5′-GCTTCT. . .
. . .CTCCGG-5′                    AGA. . .
```

(Note that the identity of the nucleotides in the 5′ overhangs can be chosen as required – they are represented by NNN in the enzyme's recognition sequence.)

The blunt-ended PCR amplicons are generated in a reaction which uses two (5′-phosphorylated) primers as follows:

```
5′-GGCCTT. . . . . .
                    . . . . . .TTCG-5′
```

producing amplicons:

```
5′-GGCCTT . . . . . .AAGC-3′
3′-CCGGAA . . . . . .TTCG-5′
```

The SapI-digested vectors and PCR amplicons are mixed. To this mixture are added:

- T4 DNA polymerase (which has 3′→5′ exonuclease activity as well as polymerase activity);
- The nucleotides dATP and dTTP;
- T4 DNA ligase.

In this reaction mixture, the vector's termini are modified to:

```
. . .GA                     5′-GCTTCT. . .
. . .CTCCGG-5′                    AAGA. . .
```

as a result of the (dual) action of the T4 DNA polymerase – which excises one G residue from the 3′ end of the left-hand terminal, and adds one A residue to the 3′ end of the right-hand terminal.

The (blunt-ended) PCR amplicons are modified to:

5′-GGCCTT . . . . . AA
AA . . . . . TTCG-5′

– again, as a result of the activity of the T4 DNA polymerase, which, in the presence of dATP and dTTP, excises the C and G residues from the 3′ ends of the amplicons, leaving 5′ overhangs as shown.

The 5′ overhangs on vector molecules and on amplicons are now complementary, so that the amplicons can be inserted, in a known orientation, into the vector molecules – using the 5′ phosphorylated ends of the amplicons for ligation by the T4 DNA ligase.

(See also IN-FUSION.)

stimulator sequence See FRAMESHIFTING.

stimulatory protein 1 (eukaryotic transcription factor) See SP1.

STM SIGNATURE-TAGGED MUTAGENESIS.

Stoffel fragment A recombinant form (the C-terminal region) of *Taq* polymerase which has greater thermal stability than the parent enzyme; it lacks exonuclease activity.

The Stoffel fragment has been used e.g. for AP-PCR.

Stokes' shift For any given fluorescent molecule or particle: the difference between the maximum absorbance wavelength and the maximum emission wavelength.

A large Stokes' shift is characteristic e.g. of the QUANTUM DOT.

stop codon See NONSENSE CODON.

STP Specialized transducing particle: see TRANSDUCTION.

STR (short tandem repeat) In genomic DNA: one of a number of sequences of DNA, each sequence consisting of tandem repeats of a unit containing 2–6 bases; for example, an STR may consist of eight consecutive four-base units:

. . . . TCTATCTATCTATCTATCTATCTATCTA. . . .

STRs are common in the genome. In a given STR (i.e. one found at a specific location on a particular chromosome), the *number* of repeat units varies in different individuals; this has made STRs useful in genetic mapping and also in systems for establishing human identity (see e.g. CODIS).

[STR database: Nucleic Acids Res (2001) 29:320–322.]

Male-specific STRs (on the Y chromosome) may be useful e.g. in sex-based forensic investigations.

STRs are found in plant genomes as well as in the genomes of other animals (see e.g FORENSIC APPLICATIONS).

***strA* gene** *Syn*. RPSL GENE.

strain (*microbiol*.) Within a given species: a collection of those individuals that have certain characteristic(s) in common and which can be distinguished from all the other members of the species on the basis of such characteristic(s). The existence of strains reflects the heterogeneity in a species.

(See also TYPING.)

strand A chain of ribonucleotides or deoxyribonucleotides (of no specified length).

A double-stranded molecule of nucleic acid consists of two strands that are hybridized by base-pairing between bases in one strand and (complementary) bases in the other.

strand displacement Displacement of an existing strand, from its complementary strand, by an (upstream) DNA polymerase that is actively synthesizing a new strand on that same complementary strand. Not all DNA polymerases have the ability to carry out strand displacement.

Strand displacement can be initiated from a NICK, the polymerase extending the 3′ terminus and simultaneously displacing the existing strand from the 5′ end of the nick. This activity is involved e.g. in strand displacement amplification (see the entry SDA). (Strand displacement is also involved in the LAMP procedure for nucleic acid amplification.)

Alternatively, strand displacement can be achieved by the use of a *bumper primer*. A bumper primer binds to a single-stranded sequence, upstream of a double-stranded region – so that extension of the bumper primer's 3′ terminus displaces the existing complementary strand.

strand displacement amplification An isothermal method for DNA amplification; the method is described in the entry SDA.

strand-separation pin (in a helicase) See HELICASE.

StrataClean™ resin Silica-based particles used e.g. for the removal of *Taq* polymerase activity following PCR, and for removing restriction endonucleases from restriction digests. It is also used for concentrating proteins.

[Uses (for concentrating proteins): Proc Natl Acad Sci USA (2008) 105(24):8470–8475; and Plant Physiol (2009) 150(1): 416–423.]

(cf. NICKEL-CHARGED AFFINITY RESIN.)

streptavidin A protein (synthesized by the bacterium *Streptomyces avidinii*) that binds strongly to BIOTIN. Its uses include the detection of biotinylated probes; for this purpose, it may be labeled with an enzyme (e.g. ALKALINE PHOSPHATASE) or with a fluorophore so that, when bound to biotin, it can signal the presence of biotin (and, hence, the probe) by expression of the label.

Streptavidin has also been used e.g. for separating biotin-labeled, single-stranded amplicons from unlabeled amplicons following PCR.

(See also PROXIMITY LIGATION ASSAY.)

The protein was involved in developing the QUANTUM DOT PROBE.

streptococcal nuclease See STREPTODORNASE.

streptodornase (streptococcal nuclease) Any of various serologically distinct forms of DNASE produced by *Streptococcus* spp, e.g. strains of Lancefield group A; the activity of these enzymes is promoted e.g. by Ca^{2+} or Mg^{2+}, and their products (unlike those of staphylococcal thermonuclease) have terminal 5′-phosphate groups (cf. DNASE I).

Streptodornase is reported to be a *phage*-encoded virulence factor in at least some strains of group A streptococci [Infect Immun (2003) 71(12):7079–7086].

streptolydigin A complex, heterocyclic antibiotic that inhibits bacterial RNA polymerase.

[Structural basis of activity: Mol Cell (2005) 19:655–666.]

streptomycin An AMINOGLYCOSIDE ANTIBIOTIC used e.g. as a selective agent for bacteria containing a vector which carries a streptomycin-resistance gene.

streptonigrin An (iron-dependent) antibiotic/antitumor agent reported to cause strand breakage in DNA in the presence of oxygen and a reducing agent, and to be inhibited e.g. by superoxide dismutase and other scavengers of free radicals.

streptozotocin 2-deoxy-2-(3-methyl-3-nitrosoureido)-D-glucopyranoside: an agent which is bactericidal for (certain) Gram-positive and Gram-negative bacteria that are actively growing (cf. PENICILLIN); bacterial susceptibility to this agent depends on the presence of a functional transport (uptake) system for *N*-acetylglucosamine (which is also involved in the uptake of streptozotocin). The agent is internalized as streptozotocin 6-phosphate, which is subsequently converted to diazomethane (a highly toxic and mutagenic compound).

Non-growing cells are unaffected by streptozotocin. Growing cells are killed – but are not *lysed* so that (unlike the cells killed by penicillin) their cytoplasmic contents are not shed immediately into the environment. For this reason, streptozotocin can be more useful than penicillin for the isolation of AUXOTROPHIC MUTANTS (q.v.).

Streptozotocin is diabetogenic in laboratory animals.

stress-induced duplex destabilization See SIDD.

stringency In the hybridization (annealing, base pairing) of two single strands of nucleic acid: the level of constraint imposed by factors such as temperature, concentration of electrolyte, and the presence of compounds (such as formamide) which influence base–base interaction.

High-stringency conditions (e.g. high temperatures) require that the two strands have a high degree of complementarity (i.e. few, or no, mismatched bases) in order for annealing to take place.

Low-stringency conditions (e.g. lower temperatures) may permit annealing to occur between two (partly mis-matched) strands even when such strands would not hybridize at higher temperatures.

Appropriate levels of stringency can be used e.g. to control the specificity with which probes and primers hybridize to their target sequences.

stringent control (of plasmid replication) See PLASMID.

stringent response See *relA* in ESCHERICHIA COLI (table).

stroboscopic lighting (for *in vivo* fluorescence) See IN VIVO FLUORESCENCE.

structure-dependent endonuclease *Syn*. STRUCTURE-SPECIFIC ENDONUCLEASE.

structure-specific endonuclease (or structure-dependent endonuclease) Any ENDONUCLEASE which cleaves only at sites which conform to a particular type of structure. For example, certain enzymes cleave a phosphodiester bond – in a *single strand* of nucleic acid – *specifically* at the junction where the two strands of a duplex molecule have separated (bifurcated) into two single strands; one paper [Science (1993) 260:778–783] noted that these enzymes seem to reach the cleavage site by moving from the free end of a 5′ extension to the point of bifurcation of the duplex, where cleavage takes place. One example of this kind of enzyme is CLEAVASE (q.v.).

Certain bacterial DNA *polymerases* (including *Taq* DNA polymerase) – which are also known as 5′ exonucleases – can also carry out the type of structure-specific cleavage referred to above. This is exploited e.g. in SELECTOR-BASED MULTIPLEX PCR (q.v.) for cleaving the 5′ free end (so-called 'flap') of the restriction fragment following the hybridization of a selector probe to an *internal* sequence in the fragment. More generally, this structure-specific endonuclease activity can be used to cleave, at a specific site, any single strand of nucleic acid; this is achieved by hybridizing an oligonucleotide to the strand in such a way that it converts the required cleavage site into the type of structure cleavable by the enzyme.

(cf. SEQUENCE-SPECIFIC ENDONUCLEASE.)

(See also FLAP ENDONUCLEASE 1.)

STS SEQUENCE-TAGGED SITE.

Stuart's transport medium See TRANSPORT MEDIUM.

STUB1 *Syn*. CHIP ('C-terminus of Hsp70-interacting protein'): see the entry CHIP.

stuffer See REPLACEMENT VECTOR.

Stx1, Stx2 Shiga-like toxins: see the entry EHEC.

subcloning CLONING a DNA fragment that has been cleaved, by restriction enzymes, from a larger fragment; in general, the object of subcloning is to facilitate studies on the larger fragment by examining individual, smaller pieces.

subgenomic mRNA (of an RNA virus) An mRNA, produced in an infected eukaryotic cell, which is shorter than the viral genome itself. Subgenomic mRNAs are needed because, in a eukaryotic cell, translation generally cannot be initiated from an *internal* (i.e. non-terminal) site in a transcript (as would be required in a full-genome-length mRNA containing multiple genes); a translation-initiation site located internally in a viral genome may be situated terminally in a subgenomic mRNA. (cf. POLYPROTEIN.)

Subgenomic mRNAs may be monocistronic, bicistronic or polycistronic.

In paramyxoviruses the negative-sense monopartite ssRNA genome is transcribed by an RNA-dependent RNA polymerase (present in the virion) into subgenomic mRNAs encoding some (early) viral proteins. Subsequently, the viral genome is copied into genome-length *positive*-sense RNAs; these RNAs can be copied into negative-sense strands which are used for (i) transcription of the subgenomic mRNAs that encode viral structural proteins and (ii) incorporation, as the genome, into new virions.

In SINDBIS VIRUS, the positive-sense ssRNA monopartite genome functions as mRNA for translation of the early, non-structural, viral proteins. Later, negative-sense ssRNAs are synthesized, and subgenomic mRNAs which encode the late viral proteins are transcribed from them [see also PLoS ONE (2009) 4(3):e4772].

substrate-induced gene expression screening See SIGEX.

subtracted restriction fingerprinting (SRF) A technique for

TYPING bacteria involving (i) the digestion of genomic DNA with two types of restriction endonuclease, (ii) end-labeling, with biotin or digoxigenin, by filling in the 5′ overhangs, (iii) differential capture (and elimination) of a proportion of the fragments (those labeled with biotin), and (iv) electrophoresis and detection of the remaining fragments (those labeled with digoxigenin) in agarose gels. Strains are typed on the basis of the patterns of fragments formed in the electrophoresis gels.

subtraction (of genes) See GENE SUBTRACTION.

subtractive hybridization A procedure for isolating particular mRNAs which are synthesized in a given type of cell under given conditions. Essentially, this involves initially isolating mRNAs from the target cells and then removing (subtracting) those mRNAs which are also found in other types of cell (and/or in the same type of cell under different conditions) – thus leaving the required mRNAs.

In one approach, mRNAs are first isolated from the target cells. Separately, mRNAs are isolated from other type(s) of cell (the 'subtractor cells'), and *these* mRNAs are converted to first-strand cDNAs by using (biotinylated) poly(T) primers; the mRNAs are eliminated, and the single-stranded cDNAs – bound to superparamagnetic beads (e.g. DYNABEADS) – are allowed to hybridize to mRNAs of the target cell population. The cDNA/mRNA hybrids are then removed magnetically – leaving the target-cell-specific mRNAs.

While this account indicates the principle of subtractive hybridization, various modifications/refinements are employed, including e.g. the use of PCR amplification to enrich differentially expressed transcripts.

suicide polymerase endonuclease restriction (SuPER) During PCR: a method for enhancing the amplification of minor (e.g. low-copy-number) templates by the elimination of dominant templates present in the same reaction mixture; this approach may thus be an alternative to PCR CLAMPING.

Essentially, within the reaction mixture (containing the unwanted, dominant template as well as the required template), the following activities occur:

- annealing of PCR primers to the *unwanted* template under stringent conditions;
- primer elongation by DNA polymerase;
- targeted restriction of the double-stranded amplicons, thus specifically eliminating copies of the unwanted template.

These activities depend on the establishment of appropriate operational conditions (e.g. temperature) in the reaction mixture. If temperatures are too low the specificity may be lost; if they are too high the efficiency of the restriction enzyme may be reduced, allowing survival of unwanted templates.

[SuPER: Appl Environ Microbiol (2005) 71:4721–4727.]

suicide vector Any vector which encodes the mechanism for its own elimination from a population of cells after carrying out its specific function(s) in those cells.

A suicide vector may be used, for example, to introduce a mutant sequence into a bacterial gene. The vector is inserted into the cell by transformation, or electroporation, and after integration of the given sequence into the chromosome (by homologous recombination) the vector becomes superfluous. If the vector is a plasmid that exhibits temperature-sensitive replication (i.e. unable to replicate at, say, ∼43–44°C), it can be easily eliminated from the cell simply by raising the temperature to the non-permissive level; ongoing growth of the cells will therefore result in a progressive loss of the plasmid. (The bacterial strain used must, of course, be able to maintain growth at the higher temperature.)

In one approach, a suicide plasmid carrying a *recA* gene (whose expression was required *temporarily* in the host cell) also carried a gene encoding a genetically modified form of the enzyme phenylalanyl-tRNA synthetase. This variant form of the enzyme has a relaxed substrate specificity in that it can mediate the incorporation into proteins of the phenylalanine analog *p*-chlorophenylalanine – which is lethal for the cell; hence, when the plasmid has fulfilled its function, any cells in which the plasmid has *not* been lost spontaneously are killed when the cells are grown on a medium that contains *p*-chlorophenylalanine [BioTechniques (2005) 38(3):405–408].

***sulA* gene** (*sfiA* gene) See SOS SYSTEM.

sulfamethoxazole See ANTIBIOTIC (synergism).

SUMO Small ubiquitin-like modifier: see SUMOYLATION.

SUMO-1, SUMO-2 etc. See SUMOYLATION.

SUMO-ome The sumoylated proteome: the range of proteins, in a given organism, which are subjected to sumoylation.

SUMO proteases See SUMOYLATION.

sumoylation A form of reversible, post-translational modification, analogous to ubiquitinylation, in which a UBIQUITIN-like protein (SUMO: small ubiquitin-like modifier) is covalently bound to a specific lysine residue in a target protein; as in the ubiquitinylation process, sumoylation involves three types of enzyme: E1 (activation), E2 (conjugation) and E3 (ligation) (see the entry PROTEASOME).

Target proteins for sumoylation include e.g. certain transcription factors, the effects of sumoylation depending on the target protein. For example, in a number of cases it represses a transcriptional activator, while for some proteins it affects subcellular localization.

Sumoylation may (indirectly) affect the stability of IκB (a regulatory factor for NFκB) and it is involved e.g. in histone modification and other cellular functions.

Sumoylation of a (mouse) transcription factor was reported to affect its *in vivo* gene-targeting activity [PLoS ONE (2009) 4(5):e5542].

[Genetic and proteomic evidence for roles of *Drosophila* SUMO in cell cycle control, Ras signaling, and early pattern formation: PLoS ONE (2009) 4(6):e5905.]

One report indicated that the Ebola (Zaire) virus promotes sumoylation-induced inhibition of the interferon response in a mouse model, thus helping to explain how this virus is able to cause rapidly overwhelming disease [PLoS Pathog (2009) 5(6):e1000493].

Sumoylation sites*, *types of SUMO protein*, *desumoylation Many, but not all, sumoylation sites in proteins are associated

with a consensus motif (K-X-E, or K-X-E/D) – in which the lysine (K) residue is preceded by a hydrophobic amino acid residue.

Mammals have four types of SUMO, designated SUMO-1, SUMO-2, SUMO-3 and SUMO-4; there is a confusing range of synonymous designations. These four types of protein may have distinct roles as they appear to modify different sets of target proteins; the expression of SUMO-4 may be limited to the kidney. The yeast *Saccharomyces cerevisiae* has only one type of SUMO, which is designated SMT3. Various similar proteins exist in plants.

Desumoylation (i.e. reversal of sumoylation), involving the removal of the SUMO moiety, is mediated by specific types of enzyme: the SUMO proteases (also called *isopeptidases* or cysteine proteases).

The desumoylating enzyme Ulp1 was first isolated from *S. cerevisiae*. Mutation in this enzyme was reported to give rise to lethal cell-cycle defects in *S. cerevisiae*.

Another yeast protein, Ulp2 (= Smt4), which is found in the nucleus, and certain proteins from vertebrates (SENP1, 2, 3 etc.), have desumoylation activity.

All the known SUMO proteases contain a highly conserved sequence in the C-terminal domain. However, the N-terminal domain shows little homology between enzymes; a study on the Ulp2 protein of *S. cerevisiae* found that only small parts of the N-terminal domain play essential roles – suggesting a reason for the low-level sequence conservation found among the proteins in this family [Mol Biol Cell (2009) 20(8):2196–2206].

[SUMO (regulating the regulator): Cell Division (2006) 1: 13; SUMOsp (a web server for predicting sumoylation sites in proteins): Nucleic Acids Res (2006) 34(Web Server issue): W254–W257.]

SUMO proteins in DNA technology

SUMO proteins, and proteases, have been exploited in DNA technology. For example, a SUMO protein has been used as a fusion partner in order to increase solubility of a recombinant target protein: see e.g. CHAMPION PET SUMO VECTOR.

SUP4 (*Saccharomyces cerevisiae*) See SACCHAROMYCES and YEAST ARTIFICIAL CHROMOSOME.

SuPER SUICIDE POLYMERASE ENDONUCLEASE RESTRICTION.

super-integron See INTEGRON.

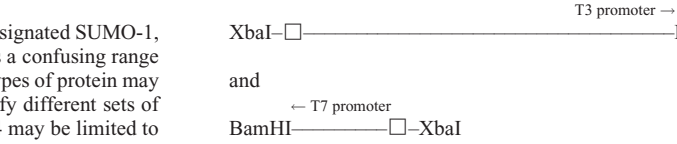
SuperCos I vector A (7.9-kb) COSMID vector (Stratagene, La Jolla CA) used e.g. for the preparation of cosmid libraries and for COSMID WALKING. It includes the ColE1 (plasmid) origin of replication and an SV40 origin.

The SuperCos I vector can accomodate inserts of 30–42 kb.

Within the MCS, the BamHI cloning site is flanked by T3 and T7 promoters, arranged in opposite orientation. Flanking the promoters are NotI recognition sites, permitting excision of the promoter-flanked insert as a whole.

Ampicillin can be used for selection in bacteria, and G418 sulfate for selection in mammalian cells.

SuperCos contains two *cos* sites which are close together in the vector and separated by a cutting site for XbaI. When cut by BamHI and XbaI, the vector yields two sections:



in which □ represents a *cos* site. (Notice that each section is flanked by dissimilar sticky ends, preventing circularization.) This type of design avoids the need for dephosphorylating the vector DNA (which would otherwise be needed in order to inhibit the formation of cosmid multimers). The fragment or gene inserts between the two BamHI sites.

The SuperCos I vector has been used e.g. for engineering a gene cluster [Nucleic Acids Res (2008) 36(17):e113], and for preparing a cosmid library [Environ Microbiol (2009) 11(5): 1038–1055].

superfemale See the entry CHROMOSOME (section: *Number of chromosomes per cell*).

(See also BARR BODY.)

superinfection exclusion See SUPERINFECTION IMMUNITY.

superinfection immunity The phenomenon in which a cell that is already infected with a particular virus resists infection by another virus of the same (or similar) type. For example, a bacterium which is already lysogenized with PHAGE LAMBDA (see also LYSOGENY) exhibits superinfection immunity to an incoming lambda genome because the incoming genome is subject to repression by the repressor of the incumbent prophage.

If *Salmonella typhimurium* is lysogenized by phage P22, a phage-encoded function causes modification of the receptor sites for P22 on the surface of the host bacterium, preventing further infection with phage P22; this phenomenon is called *superinfection exclusion*.

'Superinfection immunity' also refers to the phenomenon in which a *plasmid* of a given incompatibility (Inc) group cannot co-exist stably in a cell that contains a resident plasmid of the same Inc group (see the entry PLASMID).

superintegron See INTEGRON.

superoxide dismutase (in fragile X disease) See FMRP.

superparamagnetic Refers to the characteristic of certain types of material in which magnetic properties are exhibited only when they are placed in a magnetic field; such materials have no residual magnetic properties after they have been removed from the magnetic field.

(See also DYNABEADS.)

SuperSAGE See SAGE.

suppressor mutation Any mutation which, at least to a certain extent, reverses the effects of an earlier (*primary*) mutation at a *different site*. (cf. BACK MUTATION.) If the original wild-type phenotype is not fully restored, the result is referred to as a *pseudo-wild* phenotype.

An *intragenic suppressor* is a mutation which occurs in the

gene in which the primary mutation occurred. For example, a primary frameshift mutation may be suppressed by a nucleotide insertion/deletion that restores the correct reading frame (although the nucleotide sequence between the two mutations remains mutant).

An *intergenic suppressor* occurs outside the gene containing the primary mutation. Typically, it affects the function of the tRNA that normally binds to the given wild-type codon; the mutant tRNA may e.g. insert an amino acid at the site of a nonsense mutation (a *nonsense suppressor*), or it may insert an alternative amino acid at the site of a mis-sense mutation (a *mis-sense suppressor*).

An *ochre suppressor* mutation (in a tRNA gene) suppresses the effects of an ochre mutation or, less effectively, an amber mutation, by inserting an amino acid at the mutant codon.

An *amber suppressor* suppresses the effects of an amber mutation (see NONSENSE CODON).

Indirect suppression circumvents the effects of the primary mutation by giving the cell alternative function(s), e.g. activating an alternative metabolic pathway which can replace that disabled by the primary mutation.

suramin A drug used e.g. for the treatment of trypanosomiasis: 8-(3-benzamino-4-methyl-benzamino)naphthalene-1,3,5-trisulfonic acid; the drug is not effective in the later stages of the disease as it does not pass the blood–brain barrier.

Suramin inhibits the REVERSE TRANSCRIPTASE of various retroviruses and is also reported to affect the alternative pathway of complement fixation.

surface-independent tethering (of single-molecule DNA) One of a range of techniques used for SINGLE-MOLECULE DNA MANIPULATION/VISUALIZATION. In this approach, one end of the sample DNA molecule is bound to a (micrometer-sized) dielectric (e.g. polystyrene) sphere which is maintained in a fixed (three-dimensional) position in a flow cell by OPTICAL TWEEZERS; when held in this way, the DNA molecule can be flow-stretched away from the surface in the flow cell.

This approach avoids some problems associated with techniques such as DNA COMBING and SURFACE TETHERING, viz. interaction between the DNA molecule and surface(s), and complications arising from Poiseuille flow in the flow cell.

Surface-independent tethering has been especially valuable e.g. for studying the mechanical properties of DNA.

surface-obligatory conjugation See CONJUGATION.

surface tethering (of single-molecule DNA) A technique, used in SINGLE-MOLECULE DNA MANIPULATION/VISUALIZATION, which provides some improvements on DNA COMBING. In this approach, one end of the DNA molecule is attached (e.g. by a biotin/streptavidin linkage) to a surface inside a flow cell, and stretching is achieved by a flow of liquid through the cell. By attaching a small bead to the free end of the DNA molecule, the drag on the bead will be greater than the drag on the DNA itself – thereby making the stretching force similar along the length of the DNA molecule.

In this technique, the molecule of DNA has only one point of attachment, i.e. the end that is fixed to the surface.

Compared with combing, this approach gives some control over the stretching force, although it is necessary to ensure a smooth flow through the cell for the whole of the experiment in order to maintain tension at a given level.

As in the DNA combing approach, many DNA molecules can be examined simultaneously in the same assay. A number of DNA molecules arranged in parallel fashion (side-by-side) in a flow cell has been called a *DNA curtain*; such a 'curtain' may be viewed within a single field of view.

Uses of surface tethering

This approach has been used e.g. to study the movement of the RAD51 protein along dsDNA, the results suggesting that RAD51 translocates in the form of a multimer. The technique was also used to study translocation, along DNA, of proteins involved in the BASE EXCISION REPAIR system, in particular, to investigate the way in which damaged bases in DNA are detected by these proteins.

(cf. SURFACE-INDEPENDENT TETHERING.)

survivin A member of the *inhibitor of apoptosis* (IAP) family of proteins, homologs of which are found in a wide range of species. It has multiple roles e.g. in embryonic development; for example (as reflected in higher animals), survivin in the zebrafish (*Danio rerio*) was reported to be a key regulator of various processes (such as angiogenesis, hematopoiesis and neurogenesis) [BMC Dev Biol (2009) 9:25].

Survivin is also associated with certain pathophysiological roles [review: Am J Path (2006) 169(1):1–11], and is characteristically upregulated in cancers (in which it is a target for chemotherapeutic approaches).

(See also METHYLATION (of DNA).)

SV40 Simian virus 40: a small non-enveloped icosahedral virus in which the genome is covalently closed circular dsDNA (~5.3 kbp). Within cells the viral DNA occurs in the nucleus. (See also MINICHROMOSOME.) The virion is resistant to lipid solvents and low pH.

SV40 is a member of the POLYOMAVIRUSES.

SV40 was first isolated from the rhesus monkey (*Macaca mulatta*) – in which it is frequently found (in latent form) in kidney cells. This virus causes a characteristic vacuolation in kidney cells from the African green monkey (*Cercopithecus aethiops*). Although commonly non-pathogenic in the natural host, SV40 can e.g. induce tumors in newborn rodents (which lack the immunocompetence of adult animals). Cells transformed by SV40 contain viral DNA integrated in the host's genome, apparently at non-specific sites.

A strong promoter sequence from the genome of SV40 is used e.g. in a variety of systems (e.g. expression vectors) in which effective transcription is required.

Components of the SV40 genome are exploited in various systems: see, for example, COS CELLS, EXCHANGER SYSTEM, EXON TRAPPING, PROMOTER, SUPERCOS I VECTOR.

SWGDAM The Scientific Working Group on DNA Analysis Methods.

swine flu See the entry INFLUENZAVIRUS.

swivelase See TOPOISOMERASE (type I).

SYBR® Green I A fluorescent dye which (bound to dsDNA) has an emission maximum at 520 nm on excitation at 497 nm. (The dye also binds to single-stranded DNA, but in this case the level of fluorescence is significantly lower compared with dsDNA.)

This dye is widely used e.g. for staining DNA in gels and for monitoring the formation of products in real-time PCR. A study comparing dyes used for real-time PCR reported that, unlike SYBR Green I, the dyes SYTO-13 and SYTO-82 (i) do not inhibit PCR, (ii) do not bind preferentially to GC-rich sequences, and (iii) do not influence the melting temperature (T_m) [Nucleic Acids Res (2007) 35(19):e127].

[DNA intercalation and surface binding by SYBR Green I (determination of structure and methodological implications): Nucleic Acids Res (2004) 32(12):e103.]

SYBR Safe™ DNA gel stain A fluorescent stain (Molecular Probes/Invitrogen) which, compared with ethidium bromide, is reported to be much less mutagenic (Ames test) and more sensitive.

The bound stain has an emission maximum at ~530 nm.

symmetric methylation A term used to refer to the methylation of DNA bases when the methylated residues are present in a symmetric arrangement within the sequence recognized by the methylating enzyme. For example, the methyltransferase M.SssI (recognition sequence 5′-CG-3′; 3′-GC-5′) methylates both (symmetrically placed) cytosine residues.

syn Abbreviation for SYNCLINAL.

synapsis (in homologous recombination) See HOMOLOGOUS RECOMBINATION.

synaptosomes Sheared-off pre-synaptic nerve endings formed e.g. by homogenization of brain tissue. Synaptosomes, which contain mitochondria, have been used e.g. in the preparation of CYBRIDS.

synclinal (*syn*) Of a nucleotide: the conformation in which the 6-position in a purine, or the 2-position in a pyrimidine, is the shortest distance from the oxygen atom within the sugar ring. (cf. ANTICLINAL.)

synonymous codon See CODON.

syntenic genes Genes on the same chromosome in an organism that has a multichromosome genome; the term may also have the connotation that the genes exhibit evolutionary stability in their locations on that chromosome.

synthesis (of DNA) See entries DNA AMPLIFICATION and DNA REPLICATION.

synthetase *Syn*. LIGASE.

systematic evolution of ligands by exponential enrichment (a technique usually referred to as SELEX) See APTAMER.

SYTO® dyes An extensive family of fluorescent cyanine dyes (Molecular Probes, Eugene OR) which stain DNA and RNA; a given SYTO® dye may fluoresce e.g. blue, green, orange or red.

SYTO® dyes have a low intrinsic fluorescence (quantum yield commonly <0.01) when not bound to nucleic acid. The quantum yield commonly increases to >0.4 when the dyes are bound to nucleic acids.

SYTO® dyes penetrate the cell membranes of most types of (living and dead) cell – including mammalian and (both Gram-positive and Gram-negative) bacterial membranes.

A given SYTO® dye may differ from others in the family e.g. in binding affinity, cell permeability, and enhancement of fluorescence on binding to nucleic acids – as well as in excitation and emission spectra.

See entry SYBR GREEN I for notes on the use of SYTO-13 and SYTO-82 as indicator dyes in real-time PCR.

For other DNA-staining dyes see DNA STAINING (see also LIVE/DEAD BACLIGHT BACTERIAL VIABILITY KIT).

T

T L-Threonine (alternative to Thr).

t antigen (small t antigen of SV40 virus) See entry below.

T antigen (large T antigen) (of SV40 virus) A protein, encoded by virus SV40, that is able to transform ('immortalize') some types of primary cell in culture, and which can induce tumors in experimental animals. Neoplastic transformation by SV40 apparently involves the targeting of tumor suppressors by the T antigen. The *small t antigen* of SV40 (the 't antigen') may contribute to transformation in some types of cell under some conditions.

T-cell-tropic strains (of HIV-1) See HIV-1.

T-DNA See CROWN GALL.

T-even phages A category of morphologically similar, virulent, enterobacterial BACTERIOPHAGES in which the genome is dsDNA; they include phages T2, T4 and T6. When present in the same host cell, the genomes of these phages can undergo recombination.

T-odd phages Phages T1, T3, T5 and T7. T3 and T7 are related dsDNA-containing enterobacterial phages.

T3 RNA polymerase A phage-derived RNA polymerase useful e.g. for *in vitro* transcription.

T4 DNA ligase See LIGASE.

T7 promoter See PROMOTER.

T7 RNA polymerase A phage-derived RNA polymerase useful e.g. for *in vitro* transcription.

T5000^{TM} Biosensor System A system (Ibis Biosciences) which is used e.g. for identifying (human) subjects and distinguishing between strains of microorganisms on the basis of nucleic acid analysis.

Specific target regions of nucleic acid (of differential value in identification) are amplified by PCR, and the amplicons are then characterized by mass spectrometry. Mass spectrometry permits rapid determination of the base composition of the amplicons, i.e. their A, C, G and T content – thus providing a high-resolution 'signature' on which identification is based.

In human forensics, the T5000^{TM} system has been used e.g. for examining samples of mitochondrial DNA (mtDNA) – a role in which it is sufficiently sensitive to reveal a difference of a single nucleotide between samples of mtDNA.

In microbial identification, the system can operate either on a broad-range basis or by using specific assay kits for genotyping particular strains.

The T5000^{TM} is also valuable for epidemiological studies – e.g. in tracing a chain of infection by comparing strains of the pathogen isolated from different patients.

T5000^{TM} has been used e.g. for monitoring genetic changes in influenza viruses [PLoS ONE (2007) 2(5):e489], indentifying strains of *Campylobacter* [J Clin Microbiol (2008) 46(4): 1220–1225] and for the detection and identification of orthopoxviruses [PLoS ONE (2009) 4(7):e6342].

TA *Taq*-amplified – referring to products formed e.g. in PCR in which the *Taq* DNA polymerase was used for amplification.

(See also EXTENDASE ACTIVITY.)

TA Cloning® kits Kits (Invitrogen, Carlsbad CA) designed for cloning PCR products which have been generated by the *Taq* DNA polymerase. *Taq*-generated products are characterized by 3′ single-base (adenosine) extensions (due to EXTENDASE ACTIVITY), and the (linearized) vectors in TA Cloning® kits have complementary 3′-T overhangs to facilitate insert:vector interaction.

The vectors in these kits have an antibiotic-resistance gene, promoter(s), and primer-binding sites for sequencing.

tag (1) A sequence of nucleotides that may be used to identify a given site on a fragment or molecule of nucleic acid (see, for example, EXPRESSED SEQUENCE TAG).

(2) On a PCR *primer*: a 5′ extension that does not bind to the *primary* target sequence but which is copied during synthesis of the *complementary* strand and, hence, becomes an integral part of the amplicon that is replicated during ongoing cycling. (One or both PCR primers may have a tag.)

PCR primers are tagged for various reasons:

A tag may include e.g. a promoter sequence so that double-stranded amplicons can be subsequently transcribed.

In some cases a tag includes the recognition sequence of a RESTRICTION ENDONUCLEASE – allowing e.g. cleavage of a (double-stranded) amplicon with the formation of a particular type of STICKY END; this approach could be used e.g. if the amplicon needed to have sticky ends compatible with those in a given type of vector.

A GC-rich 5′ primer tag is used to facilitate identification of the primer's products – e.g. in a method for detecting SNPs (see T_m-shift primers in the entry SNP GENOTYPING).

PCR primers with 5′ tags carrying sequences homologous to target sequences may be used for preparing fragments for RECOMBINEERING (q.v.).

(3) A specific *peptide* sequence in a recombinant protein that provides a site for interaction with a given entity, e.g. a site for covalent attachment of another molecule (see e.g. biotin ligase recognition peptide in IN VIVO BIOTINYLATION), or a site for *non*-covalent binding of a molecule/ligand involved in the isolation or purification of the given protein (see e.g. the entries EPITOPE TAGGING, FLAG, PESC VECTORS and SIX-HISTIDINE TAG; see also AFFINITY PROTEIN EXPRESSION AND PURIFICATION and PROTEIN LABELING).

(4) A specific peptide sequence in a recombinant protein that promotes degradation of that protein in a PROTEASOME: see e.g. PEST DOMAIN.

(5) A specific peptide sequence in a recombinant protein that promotes secretion from the producing cell (see SECRETION (of proteins)) or promotes solubility of the target protein (see SOLUBILITY ENHANCEMENT TAG and the CHAMPION PET SUMO VECTOR).

In some cases the tag must be removed from an expressed recombinant protein. This may be facilitated by including a short sequence of nucleotides – e.g. a sequence encoding the recognition site of the TEV protease (see TEV) – between the

gene of interest and the sequence encoding the tag; following (*in vitro*) purification of the given protein, the tag is removed by the activity of the relevant protease.

(6) Each of the pair of specific peptide sequences ('ligation tags') which are used for ligating polypeptides or proteins in the split intein system – see *split intein* in the entry INTEIN.

(See also DIELS–ALDER REACTION.)

Tag protein (in *Escherichia coli*) See DNA REPAIR.

***tagAB* operon** See TWO-COMPONENT REGULATORY SYSTEM.

TAGM (targeted gene methylation) A technique used (*in vitro* and *in vivo*) for the methylation of *particular* bases in DNA.

In one example, a DNA METHYLTRANSFERASE – fused to a DNA-binding protein – was used to methylate the 5-position in cytosine residues at specific CG and GC sites in the yeast genome; targeting of methylation was achieved by the binding of the fusion protein to sites in the DNA recognized by the ZINC FINGER motif of the DNA-binding protein [Nucleic Acids Res (2003) 31(22):6493–6501].

(See also METHYLATION.)

TAIL-PCR Thermal asymmetric interlaced PCR: PCR used for copying an unknown sequence of nucleotides that is adjacent to a known sequence. This method is sometimes used instead of INVERSE PCR, or in conjunction with inverse PCR.

Essentially, the method is an adaptation of ANCHORED PCR in which the sequence-specific primers, when annealed to the (known) sequence, have a melting temperature of about 65°C – while the random (= arbitrary degenerate, or AD) primers, which bind to the unknown sequence, have a melting temperature of about 45°C. Initially, a low-stringency cycle of PCR is carried out during which the random primers may bind to one or more sites in the unknown region. Subsequently, two high-stringency cycles of PCR are alternated with one low-stringency cycle – thus increasing the proportion of products primed by the sequence-specific primers.

More than one type of random (AD) primer may be used in a given assay.

Products obtained from TAIL-PCR are sequenced in order to characterize the region adjacent to the known sequence.

[Uses of TAIL-PCR (e.g.): BMC Plant Biol (2009) 9:77; Microb Cell Fact (2009) 8:34; Plant Physiol (2009) 149(4): 1917–1928; Genetics (2009) 181(4):1369–1385.]

tailing The addition of 3′-terminal nucleotides to a DNA molecule by means of the template-independent enzyme *terminal deoxynucleotidyl transferase*. If nucleotides of only one type are added, the result is a *homopolymer tail*; this type of tail is obtained by using only the required species of nucleoside triphosphate in the reaction mixture.

Tailing is most efficient on 3′ overhangs of STICKY ENDS, but BLUNT-ENDED DNA can also be tailed.

3′ extension by other enzymes

A short chain of deoxycytidines is added to the 3′ end of the FIRST STRAND by the enzyme MMLV reverse transcriptase (see the entry CAPFINDER).

During PCR, the enzyme TAQ DNA POLYMERASE adds an adenosine nucleotide to the 3′ end of a strand in a template-independent way (see EXTENDASE ACTIVITY).

(See also the entries TERMINAL TRANSFERASE-DEPENDENT PCR and TUNEL ASSAY.)

Tamiflu® A commercial (Roche) form of OSELTAMIVIR (q.v.).

tamoxifen A non-steroidal agent (used e.g. for anti-estrogenic therapy in breast cancer) which can act as an inducing agent for the *in vivo* temporal control of an engineered form of the Cre recombinase: see CRE–LOXP SYSTEM.

TAMRA Any of the isomers of carboxytetramethylrhodamine, a fluorophore used e.g. for labeling probes, and in automated sequencing of DNA.

TAMRA can be used as either donor or acceptor in a FRET-based system.

TAMRA has been used e.g. in studies on the modification of lentiviral vectors [BMC Biotechnol (2009) 9:13], and on the detection of anatid herpesvirus 1 [Virol J (2009) 6:71].

TANDEM See QUINOXALINE ANTIBIOTICS.

tandem epitope tagging See EPITOPE TAGGING.

tandem repeat (1) A given sequence of nucleotides repeated one or more times, consecutively, in the same molecule.

(See also ITERON.)

(2) Repeated copies of gene clusters.

***Taq* DNA polymerase** A DNA-DEPENDENT DNA POLYMERASE from the thermophilic bacterium *Thermus aquaticus*; it lacks PROOF READING (i.e. 3′→5′ exonuclease) activity. *Taq* polymerase is stable at temperatures up to ∼95°C.

(cf. STOFFEL FRAGMENT and see also STRUCTURE-SPECIFIC ENDONUCLEASE.)

Recombinant forms of *Taq* DNA polymerase are available commercially (see e.g. PLATINUM TAQ DNA POLYMERASE).

During PCR, *Taq* DNA polymerase can exhibit EXTENDASE ACTIVITY (see also TA CLONING KITS).

Mutant *Taq* DNA polymerase was reported to be resistant to PCR inhibitors, allowing DNA amplification from whole blood and from crude soil samples [Nucleic Acids Res (2009) 37(5):e40].

(See also END-IT DNA END-REPAIR KIT and POLISHING.)

TaqMan® probe A type of commercial PROBE (Applied Biosystems) which can be used for monitoring REAL-TIME PCR and for assessing the number of specific target sequences that were initially present in a reaction mixture.

Each probe is a *target-specific* oligonucleotide that carries a covalently bound fluorophore (the reporter dye) and – closely adjacent on the oligonucleotide – a quencher of fluorescence. In the intact state the probe does not exhibit fluorescence (see also FRET).

TaqMan® probes are added to the reaction mixture in large numbers prior to cycling.

At the primer-binding (annealing) stage, TaqMan® probes bind to their complementary sequence at a (non-terminal) site on the amplicons. Subsequently, during primer extension, the DNA polymerase exerts its 5′-to-3′ exonuclease activity, thus degrading the bound probe; as a consequence the fluorophore (i.e. reporter dye) and quencher become separated – allowing the reporter to fluoresce. Fluorescence increases in intensity

as the number of free (unquenched) reporter molecules rises during ongoing cycling.

In this technology, the DNA polymerase must be able to exert 5′-to-3′ exonuclease activity. A number of commercial enzymes have this ability; the Stoffel fragment does not.

TaqMan® probes are also used in a non-target-specific role – see the entry DIGITAL PCR.

target amplification An approach used for the detection and/or quantitation of a specific DNA or RNA target sequence in a sample; it involves repeated copying of the target sequence, producing millions of copies of that sequence – allowing quantitation (see e.g. REAL-TIME PCR) and detection.

Methods involving target amplification include e.g. NASBA, PCR and SDA.

In practice, the production of large numbers of copies of the target sequence can be problematic. Thus, it may be difficult to prevent them from contaminating subsequent assays; such contamination destroys the value of an assay. This problem is avoided in SIGNAL AMPLIFICATION.

targeted cross-linking (in DNA) See e.g. TRIPLEX DNA.

targeted gene methylation See e.g. TAGM.

targeted trapping A modification of GENE TRAPPING in which the gene-trap vector is flanked by sequences homologous to genomic sequences; this enables insertion of the vector to be directed to a specific location in the genome by homologous recombination.

targetron A (bacterial) gene targeting vector, derived from a group II intron, which consists of a ribonucleoprotein particle (RNP) (see the entry INTRON HOMING for background).

A targetron behaves as a (group II) intron in that it initially inserts RNA into the DNA target site. The IEP reverse transcribes the inserted RNA so that, following a strand-corrective mechanism in the recipient genome, a (dsDNA) copy of the intron comes to occupy the target site.

Recognition of the target site, by the targetron, depends on base-pairing between the RNA and the target site – but also includes a contribution by the IEP; by changing the sequence of the RNA it is possible to obtain a targetron that is capable of inserting the required sequence into any of a wide variety of DNA target sites.

To target a given gene (in bacteria) the components of the targetron (RNA and IEP) can be encoded in a plasmid that is used to transfect the cells. In this way it is possible to introduce a required sequence into either Gram-positive or Gram-negative bacteria.

This approach has been used for the targeted inactivation of certain genes in *Francisella tularensis* (the causal agent of tularemia) [Appl Environ Microbiol (2008) 74(9):2619–2626], and for site-directed mutagenesis in *Clostridium perfringens* [PLoS ONE (2009) 4(6):e5849].

The potential for group II intron-based gene targeting in *eukaryotes* was demonstrated by the microinjection of RNP into cells of *Xenopus laevis* [PLoS ONE (2008) 3(9):e3121].

Tat protein (of HIV-1) See RETROVIRUSES.

Tat protein export system In certain bacteria: the system used for exporting *folded* proteins across the cell membrane.

(See also SIGNAL SEQUENCE.)

TATA box See CORE PROMOTER.

TATAAT (in *Escherichia coli*) See PROMOTER.

Tay–Sachs disease An autosomal recessive neurodegenerative condition which is caused by mutation in the gene encoding the α-subunit of hexosaminidase A – see the table in the entry GENETIC DISEASE.

Tay–Sachs disease and related GM2 gangliosidoses such as SANDHOFF DISEASE are currently incurable. A study on mice with Sandhoff disease was conducted with adeno-associated virus vectors (see entry AAVS) encoding genes for the α and β subunits of hexosaminidase; these genes were introduced by intracranial inoculation. The animals survived for a period of time (>1 year versus <20 weeks) that suggested the possible use of GENE THERAPY for the treatment of human Tay–Sachs disease and related GM2 gangliosidoses [Proc Natl Acad Sci USA (2006) 103(27):10373–10378].

***t*-BDMS** See BUTYLDIMETHYLSILYL-.

***Tb*ZPF3** See PROCYCLIN.

TDM Tissue-specific, differentially methylated (referring to a particular region of DNA).

T-DNA See CROWN GALL.

TE TRANSPOSABLE ELEMENT.

teeth (as a source of DNA) See FORENSIC APPLICATIONS.

TelK A PROTELOMERASE encoded by phage φKO2.

telomerase In eukaryotes: a composite enzyme – consisting of reverse transcriptase and an RNA moiety – whose function is to prevent reduction in the length of telomeres at each round of replication; telomerase extends the G-rich 3′ terminus of the telomere, using the RNA moiety as template. A primer can then bind to the extended 3′ strand and begin synthesis of the complementary (C-rich) strand.

Elimination of telomerase in the protozoan *Trypanosoma brucei* was found to cause shortening of the telomere by 3–6 bp per generation.

In *Drosophila*, chromosome length is maintained e.g. by (i) terminal gene conversion (mechanism apparently unknown); (ii) targeted transposition of non-long-terminal-repeat (non-LTR) RETROTRANSPOSONS (the *telomeric retrotransposons*) to chromosome ends.

Telomerase-independent growth

Stable, telomerase-independent growth has been reported in a strain of the yeast *Saccharomyces cerevisiae*; the ability to propagate stably without telomerase was not affected by loss of the gene *RAD52*. In those cells which became telomerase-independent, such escape from telomerase-dependence was correlated with the *initial length* of telomeres. In these cells it was noted that the composition of the telomeres suggested that recombination had occurred within the telomeric DNA. [Genetics (2009) doi: 10.1534/genetics.109.102939.]

telomere (1) In a (eukaryotic) chromosome: a specialized GC-rich section, at each end of the chromosome, that is involved e.g. in maintaining the length of the chromosome from one cell division to the next; the mechanism usually involves the

enzyme TELOMERASE (q.v.).

(See also QUADRUPLEX DNA and VSG.)

(2) In certain prokaryotes with a *linear* genome (e.g. *Borrelia burgdorferi*), and in some bacteriophages (e.g. N15, φKO2), *telomere* refers to the closed (hairpin) terminal regions of the genome.

In e.g. BORRELIA BURGDORFERI (q.v. for further details), replication of the genome involves production of a circular intermediate form consisting of a chromosome dimer. Within the dimer, junction regions between chromosomes undergo a process called *telomere resolution* that generates the hairpin terminal regions present in each daughter chromosome; this process is mediated by the enzyme PROTELOMERASE.

telomere resolution See TELOMERE (sense 2).

telomeric repeat amplification protocol See TRAP.

telomeric retrotransposon See *Drosophila* in TELOMERASE.

temperate phage See BACTERIOPHAGE.

template-independent polymerase activity See e.g. EXTENDASE ACTIVITY, and *terminal deoxynucleotidyl transferase* in the entry TAILING.

template strand (of a gene) See the entry CODING STRAND.

template switching (1) (*in vitro*) See CAPFINDER.

(2) (*in vivo*) See TRANS TRANSLATION.

temporal control (of gene activation/regulation) See the entry CONDITIONAL GENE ACTIVATION/REGULATION and see also the following entry.

temporal control (of recombinase *in vivo*) See e.g. CRE–LOXP SYSTEM.

temporal temperature-gradient gel electrophoresis See entry TTGE (see also DGGE).

tenofovir See NUCLEOSIDE REVERSE TRANSCRIPTASE INHIBITORS.

terminal deoxynucleotidyl transferase A member of the X FAMILY of DNA polymerases which is used e.g. for TAILING and also in the TUNEL ASSAY and TERMINAL TRANSFERASE-DEPENDENT PCR.

terminal gene conversion In *Drosophila*: one way (mechanism apparently unknown) in which chromosome length is understood to be maintained between cell divisions.

(See also TELOMERASE.)

terminal redundancy In a linear molecule of nucleic acid: the presence of the same sequence of nucleotides at both ends – for example, see 'R' in the figure in the entry RETROVIRUSES.

terminal transferase-dependent PCR A technique, involving PCR, which is used to examine a given sequence of dsDNA – e.g. a gene – for the presence of any obstruction, or barrier, which may stop the activity of DNA polymerase during DNA replication. [Original description: Nucleic Acids Res (1998) 26(7):1807–1811.]

A gene-specific primer is initially extended, repeatedly, to form single-stranded products; these products are subjected to 3′ homopolymer TAILING with ribonucleotides, using the enzyme terminal deoxynucleotidyl transferase. The (tailed) ssDNA products are ligated to a dsDNA linker which has a complementary 3′ overhang; the overhang serves to prime the complementary strand.

During the initial primer-extension phase, any barrier that stops the DNA polymerase will form prematurely terminated ssDNA products. These short products are amplified by PCR and are then displayed by gel electrophoresis; the results are compared with products from the corresponding normal wild-type sequence – which gives rise to full-length products. The comparison indicates the existence/location of any barrier to DNA replication.

Obstructions which can arrest the DNA polymerase during replication include various types of PYRIMIDINE DIMER, including cyclobutyl dimers and (6–4) photoproducts.

This method has been used e.g. for detecting photodimers and strand breaks in the *p53* gene that resulted from exposure to ultraviolet radiation.

termination codon See NONSENSE CODON.

Terrific Broth A (liquid) medium that contains peptone, yeast extract, dipotassium hydrogen phosphate and potassium dihydrogen phosphate. It may be used e.g. as a growth medium to promote high yields of plasmid DNA in transformed cells of *Escherichia coli*.

Tetrahymena thermophila A single-celled, free-living ciliate protozoan suggested as a model organism for studies on the molecular and cellular biology of unicellular eukaryotes.

An important (ciliate) feature is the functional and physical separation of the germline and somatic aspects into a micronucleus and a macronucleus. The DNA in the macronucleus (~104 Mb) consists of some 225 chromosomes containing an estimated 27000 protein-encoding genes, of which 15000 are homologous to genes in other organisms. DNA in the macronucleus has been sequenced [see PLoS Biology (2006) 4(9): e286].

tetraplex DNA *Syn*. QUADRUPLEX DNA.

TEV Tobacco etch virus: a member of the potato virus Y group (the potyviruses).

The TEV *protease* is used in recombinant DNA technology (e.g. for removing tags from recombinant proteins); cleavage occurs in the sequence:

Glu-Asp-Leu-Tyr-Phe-Gln-Ser

between the Gln and Ser residues. The nucleotide sequence encoding this recognition site is included in some expression vectors so that the cleavage site can be inserted in a fusion product between a recombinant protein and a removable tag; this allows the tag to be excised from the recombinant protein by cleavage at the TEV protease recognition site.

Use of the TEV protease for protein analysis

A sequence encoding the cleavage site of TEV protease was inserted, *in vivo*, into *various positions* within the *putA* gene of *Salmonella*; the expressed PutA protein thus contained the protease cleavage site at specific, known locations. This was done in order to investigate the predictions regarding domain organization in the PutA protein.

In the recombinant cells, the TEV protease was encoded by

a plasmid (pGEX-2T). It was induced with IPTG.

Three types of mutant were reported from this experiment:

- Mutants in which no functional PutA product was formed. In these mutants, TEV protease cleavage sites were located in positions that may prevent normal folding of the protein.
- Mutants in which PutA was produced but was not cleaved by the TEV protease – apparently due to lack of accessibility of the protease cleavage sites (which were presumed not to be at the protein's surface).
- Mutants in which PutA was cleaved by the TEV protease. In the presence of proline, one of these mutants exhibited a deficiency in proline dehydrogenase activity, possibly as a result of decreased interaction between the PutA protein and the membrane (such interaction apparently promoting proline dehydrogenase activity); the location of the relevant inserted TEV protease cleavage site – residue 1224 – was close to the region (residues 1204 to 1220) which may be important for membrane binding by PutA.

In another mutant (with insertion of the TEV site at residue 48), *increased* activity of proline dehydrogenase was shown in the absence of proline, suggesting that this particular lesion decreased the ability of PutA to bind to DNA (because PutA–DNA binding represses transcription of the *put* operon).

[TEV protease cleavage site insertions for probing protein structure *in vivo*: BioTechniques (2006) 41(6):721–724.]

TFIIB recognition element See CORE PROMOTER.

TFO Triplex-forming oligonucleotide: see TRIPLEX DNA.

***TFPI2* gene** (as a marker for cancer) See CANCER DETECTION.

TGS Transcriptional gene silencing: see SIRNA.

thalassemia (Cooley's anemia) Any of several types of inherited disorder, involving the α-chain or β-chain of hemoglobin, which usually result in anemia and may involve enlargement of the spleen.

(See also SICKLE-CELL ANEMIA.)

THAM *Syn*. TRIS.

theophylline 1,3-Dimethylxanthine: a competitive inhibitor of cyclic adenosine monophosphate (cAMP) phosphodiesterase, the enzyme that cleaves cAMP to adenosine monophosphate.

Theophylline has been used e.g. as an effector molecule in an engineered system for controlling gene expression within mammalian cells (see RIBOSWITCH).

(See also APTAMER and APTAZYME.)

thermal asymmetric interlaced PCR See TAIL-PCR.

thermal cycler *Syn*. THERMOCYCLER.

thermal melting profile (of dsDNA) Absorption of ultraviolet radiation (260 nm) plotted against temperature, for a given sample of dsDNA; absorption increases with the progressive MELTING of dsDNA molecules and the corresponding rise in the proportion of ssDNA. The temperature at which 50% of dsDNA is dissociated is designated T_m. T_m depends e.g. on GC content: GC pairs, being more stable, tend to raise the T_m. FORMAMIDE destabilizes hydrogen bonding, reducing T_m.

Melting can also be monintored by a dsDNA-bound fluorophore (such as SYBR GREEN 1); thus, fluoresecence falls as the dye is relased with rising temperature – T_m then being the temperature at which the *rate of signal loss* is highest.

Therminator™ DNA polymerase A thermostable DNA polymerase (New England Biolabs) engineered from the original enzyme in a species of *Thermococcus*. It lacks 3′-to-5′ and 5′-to-3′ exonuclease functions, but has strand-displacement activity and can polymerize some atypical types of nucleotide. For example, it was reported to polymerize oligomers of α-L-threose nucleic acid (see TNA) of at least 80 bases in length on a DNA template; however, it was reported to be less efficient with glycerol nucleic acid (see GNA) – being able to polymerize only a few (maximum five) glyceronucleotides on a DNA template [PLoS ONE (2009) 4(3):e4949].

thermocycler (thermal cycler) An instrument used for making the cyclical changes in temperature which are needed e.g. for certain nucleic-acid-amplification techniques (such as PCR). In general, thermocyclers are designed so that any of a range of time/temperature protocols can be selected.

A number of reaction mixtures are processed simultaneously in a thermocycler.

Heating and cooling of the reaction mixtures is achieved in either of two main ways. In one approach (the Peltier system) each reaction mixture is contained within a thin-walled tube which fits snugly inside a hole in a metal block. The block is heated and cooled according to the required protocol. In this system the permitted cycling protocol depends on the rate at which the block is able to undergo heating and cooling. Loss of sample through evaporation may be minimized with an oil overlay – another solution involving the use of a *heated-lid cycler* in which the temperature of the cover is maintained up to e.g. 120°C.

In a different approach to temperature cycling, the sample is sealed into a capillary glass tube that is heated and cooled by a forced-air current. In this procedure, the temperature of the air flowing past the glass capillary tube can be changed very rapidly, so that temperature changes in the reaction mixture can be achieved much more rapidly than is possible with the Peltier system.

In a distinct approach to temperature cycling, heating of the reaction mixture (and monitoring of the temperature) exploits the electrolytic resistance of the mixture itself. In this method heating is accomplished by passing an alternating current of electricity through the reaction mixture, and cooling involves a forced-air flow around the capillary tube that contains the mixture; the temperature of the mixture can be monitored by measuring its resistance – exploiting the correlation between resistance and temperature.

Rapid-cycle PCR can now be carried out with single-cycle times of 20–60 seconds in a total run time of 10–30 minutes; this has followed the use of small sample volumes, improved accuracy in temperature control, and the efficient conduction of heat through the thin-walled glass capillary tubes. *Plastic* capillaries were found to be less suitable than glass for rapid-cycle PCR protocols [BioTechniques (2008) 44(4):487–492].

In many commercial thermocyclers, temperature control in

the samples has been reported to be non-optimal – associated problems including distorted temperature profiles and inter-sample variation [BioTechniques (2008) 44(4):495–505].

thermonuclease See DNASE.

thermophilic SDA See SDA.

theta (θ) replication See CAIRNS MECHANISM.

thi-1 See ESCHERICHIA COLI (table).

thiofusion See entry OVEREXPRESSION (*Secreted/intracellular proteins*).

6-thioguanine (in skin cancer) See ULTRAVIOLET RADIATION.

thioredoxin A ubiquitous, heat-stable protein (~110 amino acid residues) with various roles, including e.g. hydrogen donor in the reduction of ribonucleotides (in the formation of deoxyribonucleotides and, hence, DNA).

Thioredoxin is required for assembly of filamentous phages such as PHAGE M13.

The thioredoxin gene is used e.g. as a fusion partner in the production of recombinant proteins in *Escherichia coli*: see entry OVEREXPRESSION (*Secreted/intracellular proteins*).

Thioredoxin from *Escherichia coli* (encoded by gene *trxA*) promotes processivity in the DNA polymerase from bacteriophage T7. (cf. PCNA.)

Thioredoxin has been used (with T7 DNA polymerase) for studying rolling circle replication [Nucleic Acids Res (2009) 37(4):e27].

three-hybrid system See YEAST THREE-HYBRID SYSTEM.

three-prime array (3′ array) A MICROARRAY in which each of the subsets of probes is designed to interrogate the 3′ region of a given mRNA target. This type of array is widely used for studies e.g. on mammalian gene expression.

(See also POLYADENYLATION.)

α-L-threose nucleic acid See TNA.

threshold cycle (C_t) (in real-time PCR) See REAL-TIME PCR.

thrombin (*DNA technol.*) A serine proteinase; in at least some cases this enzyme is reported to cleave proteins or peptides between arginine and glycine residues. Thrombin is used e.g. for cleaving (separating) components of fusion proteins.

(cf. ENTEROKINASE; see also AFFINITY PROTEIN EXPRESSION AND PURIFICATION and SITE-SPECIFIC CLEAVAGE.)

thymidine See NUCLEOSIDE.

thymidine kinase (TK) ATP:thymidine 5′-phosphotransferase (EC 2.7.1.21). Thymidine kinase is found in both prokaryotic and eukaryotic cells and is encoded by some DNA viruses. It catalyzes the (ATP-dependent) phosphorylation of thymidine to thymidine 5′-monophosphate.

A virus-encoded thymidine kinase is responsible for *in vivo* activation of the antiviral agent ACYCLOVIR.

(See also GANCICLOVIR.)

thymine dimer A type of PYRIMIDINE DIMER which is formed by covalent cross-linking between adjacent thymine residues (in the same DNA strand).

A thymine dimer can be produced e.g. by ULTRAVIOLET RADIATION.

A thymine dimer involves the formation of bonds between the 5-positions of both thymines and between the 6-positions of both thymines; this produces a 4-carbon cyclobutane ring structure, hence the name *cyclobutyl dimer*.

A thymine dimer, which can block DNA replication, may be lethal in the absence of a suitable repair mechanism.

(See also UVRABC-MEDIATED REPAIR.)

The enzyme PHOTOLYASE can cleave thymine dimers and restore a normal nucleotide sequence.

Ti plasmid See CROWN GALL.

TIGR The Institute of Genomic Research.

TIGR microbial database (microbial genomes) A source of information on whole-genome sequences of a wide range of species which is available at:
www.tigr.org/tdb/mdb/mdb.html

tiling array A form of MICROARRAY characterized by probes which collectively cover the *complete* nucleotide sequence of particular region(s) of a genome (or, in some cases, the whole genome); a part of the sequence covered by one probe may also be covered by a part of the adjacent probe. – i.e. a given genomic sequence may be covered by two adjacent probes.

An important feature of a tiling array is that, owing to total coverage of particular genomic region(s), the probes are able to detect previously unknown or unexpected sequences; this is in contrast to microarrays in which probes are designed to be complementary to *specific*, known sequences.

A tiling array may be used e.g. to examine DNA fragments obtained by CHROMATIN IMMUNOPRECIPITATION; these fragments (or sequences copied from them) could be hybridized to tiling arrays that cover particular regions of interest – thus, for example, revealing unknown DNA–protein binding sites.

[Optimized design and assessment of whole-genome tiling arrays: Bioinformatics (2007) 23(13):i195–i204.]

The RNA expression profile of *Mycobacterium leprae* was investigated with a tiling array that contained overlapping 60-mer probes covering the entire genome of the organism (3.3 Mbp) [J Bacteriol (2009) 191(10):3321–3327].

TILLING An acronym for: 'targeting induced local lesions in genomes': a method devised for the detection of chemically induced point mutations. The method was subsequently used for detecting naturally occurring polymorphisms and was re-designated ECOTILLING (q.v. for principle).

time-resolved emission spectra (TRES) An approach used to detect fluorescence from a *specific* source in an experimental system containing more than one source of fluorescence.

The method depends on the difference in the rate of decay of fluorescence from the different sources following a brief (<1 nanosecond) pulse of excitation. That fluorescence which has a longer lifetime is measured only after the shorter-term fluorescence has decayed, so that only the longer-term fluorescence is monitored.

Sources whose activity has a lifetime of the order of microseconds have been developed [time-resolved and two-photon emission imaging microscopy (of live cells) using inert platinum complexes: Proc Natl Acad Sci USA (2008) 105(42): 16071–16076].

tirandamycin An antibiotic apparently similar in activity, and

structure, to STREPTOLYDIGIN.

TIRF microscopy Total internal reflection fluorescence microscopy: a form of fluorescence microscopy in which a fluorophore-labeled specimen on the surface of (e.g.) a glass coverslip is illuminated – from *below* the coverslip – by a laser beam that strikes the coverslip at an angle which exceeds the *critical angle* for total internal reflection (TIR) in the glass coverslip (refractive index of glass is often 1.53). (Note that TIR can occur only when the laser beam travels through the glass towards a medium of *lower* refractive index – such as water (1.33), ethanol (1.36) or a cell (∼1.4). Dried mountants, for example Canada balsam and DPX, have a refractive index similar to that of glass.) If the laser beam enters the coverslip at an appropriate angle, most of the energy is reflected (at the glass–specimen interface) back into the coverslip (i.e. TIR). However, a small proportion of the laser's energy does enter the medium of lower refractive index, and it travels parallel to the glass–specimen interface; this energy, referred to as the *evanescent wave*, decays exponentially with distance – but it continues to be present while the laser beam is in place at the correct angle.

The evanescent wave is able to excite fluorescence from a fluorophore *immediately* above the surface of the coverslip – i.e. within a thin layer of approx. 100 nm in thickness. Thus, if a cell is adherent to the upper surface of the coverslip, it is possible to detect fluorophores specifically *within* the plasma membrane, or in the cytoplasm immediately adjacent to that membrane.

In DNA technology, TIRF microscopy has been used e.g. for imaging fluorophore-labeled molecules of combed DNA (see DNA COMBING).

[Use of TIRF microscopy (with diagram): Methods (2008) 46(3):233–238; Nucleic Acids Res (2009) 37(14):e99.]

tissue-specific stem cell See STEM CELL.

TK (1) THYMIDINE KINASE.

(2) TYROSINE KINASE.

T_m See THERMAL MELTING PROFILE.

T_m-shift genotyping (of SNPs) See SNP GENOTYPING.

T_m-shift primers Primers whose (extension) products can be identified (and differentiated from the extension products of other primers) by their melting characteristics – see e.g. SNP GENOTYPING.

TMA Transcription-mediated amplification: a method used for the isothermal amplification of target sequences in RNA. The principle of TMA is similar to that of NASBA; the commercial applications of TMA and NASBA differ, for example, in the way in which amplified products are detected/quantitated.

Both TMA and NASBA reflect an earlier system for the *in vitro* amplification of nucleic acids ('self-sustained sequence replication') – described by Guatelli et al [Proc Natl Acad Sci USA (1990) 87:1874–1878] – based on retroviral replication.

TMA is the basis of certain diagnostic tests which are used e.g. for detecting the pathogens *Mycobacterium tuberculosis* and *Chlamydia trachomatis* in clinical specimens. These tests involve amplification of conserved sequences in rRNA and use of a probe-based system (*hybridization protection assay*: see the entry ACCUPROBE) for detecting the products.

TMA-based assays include AMTDT (for *M. tuberculosis*) and AMP CT (for *C. trachomatis*).

Compared with (for example) the basic PCR methodology, TMA-based assays have several advantages which include:

- theoretically, an improved capacity to detect small numbers of target organisms owing to the use of rRNA targets (which are abundant in cellular pathogens);
- isothermal amplification, avoiding the need to use thermocycling equipment.

tmRNA Transfer-messenger RNA: an effector molecule in the process TRANS TRANSLATION (q.v.).

Tn Abbreviation for TRANSPOSON.

Tn*1* A TN3-like TRANSPOSON that encodes a β-lactamase.

Tn*2* A TN3-like TRANSPOSON that encodes a β-lactamase.

Tn*3* A ∼5-kb class II TRANSPOSON which includes a gene (*bla*) encoding a TEM-type β-lactamase (which inactivates certain β-LACTAM ANTIBIOTICS). The transposon has identical 38-bp terminal inverted repeats.

Insertion of Tn*3* appears to occur preferentially in AT-rich regions.

Transposition of Tn*3*, involving a *replicative* mechanism (see figure in TRANSPOSABLE ELEMENT), requires two transposon-encoded genes: *tnpA* and *tnpR*, which are associated, respectively, with transposase and resolvase activities. (These two genes (in Tn*3*) are transcribed in opposite directions.)

Transposition gives rise to a 5-bp target-site duplication.

Resolution of the cointegrate involves site-specific recombination between two copies of a so-called *internal resolution site* (IRS) present in the duplicated sequences of Tn*3* in the cointegrate. The resolvase is apparently specific for two IRS sequences in the same replicon and in the same orientation, making the reaction irreversible. [Catalytic residues: Nucleic Acids Res (2009) doi: 10.1093/nar/gkp.797.]

Tn*3* exhibits TRANSPOSITION IMMUNITY.

A number of the transposable elements found in bacteria are related to Tn*3*. All of these elements have, for example, similar inverted repeats of 35–40 bp (and they also create a 5-bp target-site duplication on insertion). All transpose replicatively, with the formation of a cointegrate, and many exhibit transposition immunity. The Tn*3*-like elements include Tn*1*, Tn*2*, Tn*4*, Tn*21*, Tn*501*, Tn*551*, Tn*1721* and Tn*2603*.

Collectively, Tn*3* and the Tn*3*-like elements are sometimes referred to as Tn*A*.

Tn*4* A TN3-like TRANSPOSON encoding resistance to ampicillin and e.g. streptomycin.

Tn*5* A 5818-bp class I TRANSPOSON containing genes which encode resistance to aminoglycoside and other antibiotics. In the transposon, the contiguous antibiotic-resistance genes are flanked by insertion sequences IS*50*L and IS*50*R; these differ slightly in sequence. Each insertion sequence is itself flanked by a pair of 19-bp sequences called the 'outer end' (OE) and 'inner end' (IE); the two OE sequences thus form the two ends of the transposon.

IS*50*L encodes two non-functional proteins.

IS*50*R encodes the transposase, Tnp, and also an inhibitor of transposase, Inh. The frequency of transposition of Tn*5*, which occurs by a cut-and-paste mechanism, is influenced by the relative levels of Tnp and Inh. Another regulatory factor is the cell's Dam methylation system which tends to inhibit transcription from the promoter of the Tnp gene; accordingly, like Tn*10*, transposition of Tn*5* is linked to DNA replication – see the entry TRANSPOSABLE ELEMENT for a rationale.

Transposition of Tn*5* requires (i) synthesis of transposase, Tnp, and (ii) formation of the *transpososome*. The transpososome appears to consist of a transposase *dimer* – with each monomer binding one end of the Tn*5* transposon. The small, multifunctional bacterial protein H-NS apparently interacts with the transpososome and promotes transposition of Tn*5* [Nucleic Acids Res (2009) 37(2):309–321].

The transposase (Tnp) can mediate transposition of Tn*5* as a whole, or it can mediate transposition of IS*50*; transposition of Tn*5* involves both OE sequences, while transposition of IS*50* involves one OE sequence and one IE sequence.

In DNA technology a system for *in vivo* and *in vitro* transposition is based on Tn*5*: see TN*5*-TYPE TRANSPOSITION.

Tn*5*-type transposition TRANSPOSITION that can be carried out wholly *in vitro* with a DNA fragment flanked, on both sides, by the (19-bp) outer end (OE) sequences of transposon Tn*5* (see entry for Tn*5*) and the enzyme transposase [J Biol Chem (1998) 273:7367–7374] – or *in vivo* with the Transposome™ approach (see below).

Insertion into the target DNA occurs at random sites.

(A similar procedure is carried out with PHAGE MU (q.v.).)

The system has a number of uses. For example, as insertion occurs at random sites in a population of plasmids, transposition can generate a range of templates (the insert providing known primer-binding sites) that can be used for sequencing the target molecule.

Commercially available transposase-containing complexes (Transposome™; Epicentre Technologies, Madison WI) can be inserted into cells e.g. by ELECTROPORATION – following which insertions occur randomly. A Transposome™ consists of a stable complex of a transposon and a transposase; when Transposomes are introduced into cells, transposons insert at random sites in the genome – so that this system can be used for TRANSPOSON MUTAGENESIS.

The system has been used e.g. for mutagenesis in *Rickettsia prowazekii* [Appl Environ Microbiol (2004) 70:2816–2822] and *Francisella tularensis* [Appl Environ Microbiol (2004) 70:6901–6904]; identifying a gene for resistance to capreomycin (in *Mycobacterium tuberculosis*) [Antimicrob Agents Chemother (2005) 49:571–577]; studying production of outer membrane vesicles by *Escherichia coli* [J Bacteriol (2006) 188(15):5385–5392]; preparing attenuated strains of *Streptococcus pyogenes* [Infect Immun (2008) 76(7):3176–3186]; studying expression of type 1 fimbriae in *Salmonella enterica* (*Typhimurium*) [BMC Microbiol (2008) 8:126]; for identifying virulence markers in *Francisella tularensis* [PLoS ONE (2009) 4(5):e5463]; for identifying a toxin regulatory locus in *Clostridium perfringens* [PLoS ONE (2009) 4(7):e6232]; and (using a transposon-based mutant library) for detecting genes in *Xanthomonas citri* subsp *citri* [BMC Microbiol (2009) 9:12].

(See also GENETIC FOOTPRINTING.)

Tn*7* A TRANSPOSON which can insert into the chromosome of *Escherichia coli* with a high degree of specificity at the target site *att*Tn*7*; insertion depends on several transposon-encoded proteins which jointly contribute transposase activity.

(See also BACMID.)

In another, distinct, mode of transposition, Tn*7* inserts into certain plasmids and into the *E. coli* chromosome at locations near double-stranded breaks and at the site of termination of DNA replication.

(See also INTEGRON.)

Tn*10* A 9.3-kb class I TRANSPOSON found e.g. in plasmid R100 and encoding resistance to tetracycline. The two ends of the transposon consist of 1.4-kb IS*10* elements designated IS*10*R and IS*10*L; these two elements are similar, but not identical. A transposase, essential for transposition of Tn*10*, is encoded by IS*10*R.

Transposition occurs by a cut-and-paste mechanism.

The insertion of Tn*10* can occur at various sites, although there are apparently certain preferred sites. Insertion results in a 9-bp target-site duplication.

As in Tn*5*, transposition of Tn*10* is influenced by the cell's Dam methylation – see entry TRANSPOSABLE ELEMENT for a rationale. Transposition of Tn*10* is influenced e.g. by the phenomenon of MULTICOPY INHIBITION.

Tn*21* A ∼19-kb TN*3*-like TRANSPOSON, found e.g. in *Shigella flexneri* plasmid R100, which encodes resistance to mercury ions, streptomycin and sulfonamides.

Transposition of Tn*21* is regulated by a putative *modulator protein* encoded by the *tnpM* gene; in this transposon, *tnpA*, *tnpR* and the *tnpM* sequence are all transcribed in the same direction. The modulator protein is thought to influence the activities of both the transposase and resolvase – enhancing the former and inhibiting the latter.

(See also TN*2610*.)

Tn*501* An ∼8-kb TN*3*-like TRANSPOSON, found e.g. in *Pseudomonas* plasmid pUS1, which encodes resistance to mercury ions; frequency of transposition was reported to be enhanced by low levels of mercury.

Tn*551* A ∼5-kb TN*3*-like TRANSPOSON that encodes resistance to erythromycin; it is found e.g. in a *Staphylococcus aureus* plasmid.

Tn*916* See the entry CONJUGATIVE TRANSPOSITION.

Tn*1681* A TRANSPOSON containing the gene for a heat-stable enterotoxin (STa) that is produced by enterotoxigenic strains of *Escherichia coli* (see ETEC).

Tn*1721* An ∼11-kb TN*3*-like TRANSPOSON encoding resistance to tetracycline.

(See also TN*2610*.)

Tn*2410* An 18.5-kb TRANSPOSON, present e.g. in a *Salmonella*

typhimurium plasmid, encoding a β-lactamase and resistance to streptomycin and mercury.

Tn*2603* A ~22-kb TN3-like TRANSPOSON which encodes a β-lactamase and resistance to mercury, streptomycin and sulfonamides.

Tn*2610* A 28.8-kb TRANSPOSON, originally isolated from an *Escherichia coli* conjugative plasmid, which appears to have developed from transposons Tn*1721* and Tn*21* following extensive recombination [Antimicrob Agents Chemother (2006) 50(4):1143–1147].

Tn*4655* A class II TRANSPOSON, found e.g. in *Pseudomonas* plasmid NAH7, which includes genes encoding catabolism of naphthalene. The tyrosine recombinase, TnpI, apparently catalyzes both intra- and intermolecular recombination at *att*I sites [J Bacteriol (2006) 188(11):4057–4067].

(See also TOL PLASMID.)

Tn*A* A designation sometimes used to refer – collectively – to transposon Tn*3* and the Tn*3*-like elements.

TNA (α-L-threose nucleic acid) A nucleic acid analog in which the sugar–phosphate units of the backbone chain contain the unusual sugar α-L-threose.

TNA can hybridize to DNA and RNA, and it can also form double-stranded TNA structures.

The THERMINATOR DNA POLYMERASE (q.v.) was reported to be an efficient DNA-dependent TNA polymerase.

(See also GNA and PNA.)

TNP Transdominant negative protein: see GENE THERAPY.

***tnpA* gene** See TN3.

Tn*phoA* See TRANSPOSON MUTAGENESIS.

TnpI See TN4655.

***tnpM* gene** See TN21.

***tnpR* gene** See TN3.

Tn*YLB-1* transposon See MARINER FAMILY.

tobacco etch virus See TEV.

tobramycin See AMINOGLYCOSIDE ANTIBIOTIC.

toeprinting A procedure used for investigating the series of events that occur during *translation* (i.e. synthesis of a polypeptide from an mRNA template bound to a ribosome). The elongation stage of translation is halted, as required, by the addition of CYCLOHEXIMIDE; this blocks the movement of ribosomes, leaving them at various sites along the length of the (polysomal) transcript. Reverse transcriptase is then used to copy mRNA into cDNA, using a (labeled) mRNA-specific primer; the synthesis of cDNA is stopped by the first ribosome which the transcriptase encounters ('primer-extension inhibition'), resulting in a prematurely terminated ('toeprint') fragment. The resulting products can be examined e.g. by gel electrophoresis. The fragment's length gives an indication of the position of the ribosome on the mRNA.

In toeprinting experiments the primer may be labeled with ^{32}P, but a rapid and convenient procedure can be achieved with fluorescent primers.

[Uses of toeprinting (e.g.): Proc Natl Acad Sci USA (2008) 105(31):10738–10743; Proc Natl Acad Sci USA (2008) 105 (14):5343–5348; PLoS ONE (2008) 3(10):e3460; Mol Cell Biol (2008) 28(13):4227–4239.]

TOL plasmid A large (~117 kb) plasmid, found in some strains of the Gram-negative bacterium *Pseudomonas*, which confers the ability to metabolize certain hydrocarbons.

This plasmid encodes enzymes for a two-stage catabolic pathway for the mineralization of toluene and xylene. Both of these hydrocarbons are initially oxidized to their respective carboxylic acids; this involves enzymes encoded by the so-called *upper* operon of the TOL plasmid. The *meta* operon encodes enzymes which effect aromatic ring fission and the subsequent formation of pyruvate and acetaldehyde (both of which enter the TCA cycle).

[Regulation of operons in the TOL plasmid: EMBO Journal (2001) 20:1–11 (4–6).]

(See also TN4655.)

A TOL plasmid has been reported to encode a homolog of the error-prone DNA polymerase pol V which facilitates the survival of host bacteria when DNA damage has accumulated [J Bacteriol (2005) 187(15):5203–5213].

Expression of the TOL plasmid in *Pseudomonas* was found to be markedly upregulated in response to competition stress in a dual-species consortium, resulting in rapid utilization of available carbon and a massive increase in population density [PLoS ONE (2009) 4(6):e6065].

***tonA* gene** In *Escherichia coli*: a gene which encodes the TonA protein (also called FhuA protein), a component in the outer membrane (the outermost layer of the cell envelope) that acts as a receptor for certain phages, e.g. phage T1.

***tonB* gene** In *Escherichia coli*: a gene which encodes the TonB protein, a periplasmic protein that shuttles between the outer membrane and the cytoplasmic membrane [Mol Microbiol (2003) 49:869–882]; it has various roles in transport – such as the uptake of ferric iron complexes and vitamin B_{12}. TonB is also involved e.g. in the uptake of various colicins.

top strand The upper strand when dsDNA is depicted as two, parallel single strands – an upper strand and a lower strand – the nucleotide sequence of the upper strand being shown in the 5′ to 3′ direction (when reading from left to right across the page).

topo TOPOISOMERASE.

topo cloning *Syn.* TOPOISOMERASE I CLONING.

TOPO® pENTR™ vectors See DIRECTIONAL TOPO PENTR VECTORS.

TOPO TA Cloning® kits Kits from Invitrogen (Carlsbad CA) used for TOPOISOMERASE I CLONING (q.v.) of PCR products which have been synthesized with the TAQ DNA POLYMERASE; these amplicons have single-nucleotide (adenosine) 3′ overhangs due to EXTENDASE ACTIVITY, and the (linearized) cloning vectors in these kits include 3′ single-nucleotide *thymidine* overhangs which facilitate direct ligation to the PCR products.

The vector may include *att* sites; these facilitate transfer of the insert to other vectors within the GATEWAY SITE-SPECIFIC RECOMBINATION SYSTEM.

The vector may include primer sites for sequencing and/or

e.g. a *lacZ*α sequence, permitting *blue–white screening* (see PBLUESCRIPT), as well as antibiotic-resistance genes.

topoisomer See TOPOISOMERASE.

topoisomerase Any enzyme which can interconvert topological isomers (i.e. *topoisomers*) of cccDNA. (The topoisomers of a given molecule differ from one another in their topological properties – e.g. a given circular dsDNA molecule may exist in a supercoiled state or in a relaxed state, both of which are topoisomers of that particular molecule.) The interconversion of topoisomers involves the transient breakage of one or both strands of DNA in a given duplex; the reaction may be intramolecular or intermolecular.

Type I topoisomerases (nick-closing enzymes, relaxing enzymes, swivelases or untwisting enzymes) can relax negative supercoiling. In a ccc dsDNA molecule, one strand is cut, the unbroken strand is passed through the gap, and the break is repaired. The type I topoisomerases (type I topos) include the *Escherichia coli* ω (omega) protein (= topo I) and *E. coli* topo III.

Type II topoisomerases (type II topos) relax both negative and positive supercoiling, although the latter is relaxed more efficiently [chirality sensing by topoisomerase enzymes: Proc Natl Acad Sci USA (2003) 100:8654–8659]. In a ccc DNA duplex, the type II enzyme makes a double-stranded break and passes a double-stranded segment of DNA through the gap before re-sealing the break. The type II topos include *E. coli* topo IV and *gyrase*.

Topo IV deals with unwanted entanglements during DNA replication and appears to be the sole enzyme responsible for the decatenation (unlinking) of replicated chromosomes.

Gyrase is a bacterial tetrameric enzyme that appears to be essential for DNA replication. It can e.g. *introduce* negative supercoiling and, in the absence of ATP, the gyrase from *E. coli* can relax negative supercoiling. This enzyme is the target of certain antibiotics, such as quinolones and novobiocin.

(cf. REVERSE GYRASE.)

Topo IV and gyrase may both oppose the positive supercoiling that is introduced by helicase during DNA replication – gyrase by generating negative supercoiling and topo IV by relaxing positive supercoiling.

topoisomerase I cloning A technique in which the DNA insert/fragment is ligated into the vector molecule by the activity of topoisomerase I in preparation for CLONING. (Note that the phrase 'topoisomerase cloning' is somewhat misleading as, in this method, only *ligation* of the fragment into the vector is mediated by topoisomerase I – the actual *cloning* takes place after the vector has been inserted into cells.)

This method is often used when the DNA insert/fragment is an amplicon prepared by PCR with the *Taq* DNA polymerase; amplicons prepared in this way have a 3′ single-nucleotide deoxyadenosine monophosphate overhang at each end (see EXTENDASE ACTIVITY), and the (linearized) vector molecules are designed to exploit this by having a 3′ single-nucleotide deoxy*thymidine* monophosphate overhang at both ends. At each end of the vector, 3′-phosphate is covalently linked to topoisomerase I; this conserves energy which is subsequently used for bond formation with the insert. (The topoisomerase I dissociates after the bond is formed.) For circularization – i.e. integration of an amplicon into a vector molecule to form a closed, circular product – this ligation system works because amplicons from PCR are usually derived from primers which lack a 5′-phosphate; hence, each 3′-phosphate on the vector molecule can form a phosphodiester bond with the 5′-OH of the amplicon (mediated by topoisomerase I).

Commercial vectors using topo cloning include e.g. vectors in the TOPO TA CLONING KITS; as their name implies, these vectors are designed specifically for those inserts prepared by PCR using *Taq* DNA polymerase.

The DIRECTIONAL TOPO PENTR VECTORS differ from those described above in that they do not contain single-nucleotide deoxythymidine monophosphate overhangs – having, instead, one blunt end and one four-nucleotide overhang, so that the fragment is necessarily inserted into the vector directionally; however, the mode of insertion of the fragment involves the activity of topoisomerase I, as described above.

The topo approach is also used in cloning vectors designed to accept *blunt*-ended fragments: see e.g. ZERO BLUNT.

Topo cloning has been used e.g. for studies on β_1 integrin homologs in zebrafish (*Danio rerio*) [BMC Cell Biol (2006) 7:24] and studies on manganese-oxidizing spores of *Bacillus* in hydrothermal sediments [Appl Environ Microbiol (2006) 72(5):3184–3190]; it was also used for production of human chemokine receptors in *Escherichia coli* [PLoS ONE (2009) 4(2):e4509].

topological winding number *Syn.* LINKING NUMBER.

toroidal DNA See DNA TOROID.

total internal reflection fluorescence microscopy See entry TIRF MICROSCOPY.

total RNA A phrase generally used to refer, collectively, to all species of RNA found in a cell extract, including e.g. mRNA and rRNA etc.

touchdown PCR A type of protocol in which a given PCR assay is repeated a number of times with a stepwise decrease in the annealing temperature. The purpose is to determine the lowest temperature at which priming occurs correctly – i.e. avoiding *mispriming*: the binding of primers to sites that are not exactly complementary to the primer. The procedure thus seeks to promote maximum specificity in primer binding and to avoid the production of non-specific products.

***TP53* gene** (as a marker for cancer) See CANCER DETECTION.

TP228 plasmid See R1 PLASMID.

TPA 12-*O*-tetradecanoylphorbol-13-acetate.
(See also ZEBRA.)

tracking cell lineage (with reporter genes) See CELL LINEAGE TRACKING.

***traD* gene** A gene in the F PLASMID (within the transfer operon) which is needed for conjugative transfer.

trailer *Syn.* 3′-UTR.

***trans*-acting element** A discrete sequence of nucleotides which

encodes a product that can act on *another* molecule; thus e.g. an extrachromosomal plasmid may encode a protein that acts on a chromosomal receptor.

***trans* splicing** (of proteins) See split intein in the entry INTEIN.

***trans* splicing** (of transcripts) The mechanism which is found in trypanosomes (and e.g., in some cases, in the nematode worm *Caenorhabditis elegans*) in which a polycistronic transcript is spliced into a number of monocistronic units – each bearing a short leader sequence (a so-called *spliced leader*, SL) which is subsequently added to the 5′ terminus.

As is the case in pre-mRNA splicing in higher eukaryotes, *trans* splicing is mediated by a complex of ribonucleoproteins and other factors known as the *spliceosome*. Characterization of spliceosome proteins has been carried out in *Trypanosoma brucei* [Eukaryotic Cell (2009) 8(7): 990–1000].

***trans* translation** (ribosome rescue) In (at least some) bacteria (including *Escherichia coli*): a process, mediated by a ribonucleoprotein complex, which enables a cell to clear a stalled ribosome on a (defective) mRNA. (Such mRNAs include the 'nonstop' transcripts which lack an in-frame stop codon.)

A major player in the ribonucleoprotein complex is tmRNA (transfer-messenger RNA) which works in conjunction with a helper protein (referred to as small protein B, SmpB), elongation factor EF-Tu, and guanosine 5′-triphosphate (GTP).

The tmRNA–SmpB complex is targeted by enzyme alanyl-tRNA synthetase, which loads the complex with an alanine residue.

The entire complex (aminoacylated tmRNA, SmpB, EF-Tu and GTP) is able to recognize a stalled ribosome. When this happens, tmRNA enters the (now vacant) A site of the stalled ribosome, and the nascent polypeptide (on the ribosome) is transferred to the alanine residue on tmRNA. At this point the ribosome stops using mRNA as the template – but continues the translation process using a coding sequence on tmRNA. Following this template switching, a further ten amino acids are added to the nascent polypeptide (i.e. ten amino acids encoded by tmRNA); the ribosome then reaches a stop codon on the tmRNA template, and the ribosome is released.

The ten-amino-acid tag (reported as ANDENYALAA) at the C-terminal of the aborted polypeptide is used as a signal for proteolysis. The Lon protease (see the entry LON GENE) is one of several proteases that are involved in the degradation of these aborted polypeptides.

Part of the *trans* translation process involves elimination of the defective mRNA. In addition to tmRNA (and SmpB), this requires the activity of RNase R (q.v.).

Trans translation is a quality-control system that promotes high fidelity in the flow of information from DNA, through RNA, to protein. The activity of tmRNA was reported to be essential for the survival of bacteria under adverse conditions, and for the virulence of some, or all, pathogenic bacteria.

[tmRNA-mediated surveillance: Methods Enzymol (2008) 447:329–358.]

[Role of upstream sequences in selecting the reading frame on tmRNA: BMC Biol (2008) 6:29.]

transcapsidation See PHENOTYPIC MIXING.

transconjugant See CONJUGATION.

transcriptase (1) See REPLICASE.

(2) A (DNA-dependent) RNA POLYMERASE.

transcription factor Any of a range of protein factors that are required for transcription (in both eukaryotic and prokaryotic cells). A transcription factor may have a specific recognition site in the promoter region of a gene (see e.g. BRE; see also CORE PROMOTER). Binding between transcription factors and genomic DNA is studied by techniques such as CHROMATIN IMMUNOPRECIPITATION and EMEA.

In bacteria, transcription involves a SIGMA FACTOR, which binds to the RNA polymerase.

An aberrant transcription factor can cause disease (see e.g. AXENFELD–RIEGER SYNDROME).

Oncogene-associated transcription factors may be potential targets in (individualized) cancer therapy [Cancer Gene Ther (2009) 16:103–112].

An alternatively spliced transcription factor has been used to break the viral latency in HIV-1-infected cells [Proc Natl Acad Sci USA (2009) 106(15):6321–6326].

Engineered transcription factors are used in techniques such as e.g. the YEAST TWO-HYBRID SYSTEM and in similar methods.

transcription-mediated amplification See TMA.

transcription repair coupling factor (TRCF) See MFD GENE.

transcription vector *Syn*. EXPRESSION VECTOR.

transcriptional enhancer See ENHANCER (transcriptional).

transcriptional fusion See GENE FUSION.

transcriptional gene activation See SIRNA.

transcriptional gene silencing (TGS) See SIRNA.

transcriptome The totality of transcripts in (or isolated from) a cell or tissue under given conditions.

An 'atlas' of transcriptomes from different types of cell in the rice plant (*Oryza sativa*) has provided insight into various aspects of the inter-cell relationships [Nature Genetics (2009) doi: 10.1038/ng.282].

Unusually comprehensive data on a bacterial transcriptome were obtained by studying the Gram-positive species *Bacillus anthracis* under a variety of growth conditions. Interestingly, it was found that there is a significant transcriptional heterogeneity among the cells in a clonal, synchronized culture [J Bacteriol (2009) 191(10):3203–3211].

The antisense transcriptome

In addition to protein-encoding (sense-orientated) transcripts, the transcriptomes obtained from (at least) some mammalian genomes (including the human genome) include a significant proportion of antisense transcripts (copied from the opposite strand of genomic DNA); the antisense transcripts are usually considered to have a regulatory role. In a given transcriptome it is necessary to distinguish between the sense and antisense transcripts.

In one method, a genome-wide analysis of antisense transcripts was carried out with an Affymetrix EXON ARRAY. The

procedure was essentially as follows. After exclusion of the rRNA from a total RNA preparation, FIRST STRAND synthesis was carried out with random primers using dUTP during the reverse transcription. RNA was eliminated by RNase H. The ssDNA was treated with uracil DNA glycosylase (in order to remove the uracil residues) and an AP endonuclease (which cuts the strand into fragments at apurinic/apyrimidinic sites). The fragments of ssDNA were then end-labeled, using terminal deoxynucleotidyl transferase, and the fragments were hybridized to the exon array and scanned. [BMC Genomics (2008) 9:27.]

Other approaches have also been proposed for analysis of the transcriptome [e.g. Science (2008) 322(5909):1855–1857; Nucleic Acids Res (2009) doi: 10.1093/nar/gkp596].

transdominant negative protein (TNP) See GENE THERAPY.

transductant See TRANSDUCTION.

transduction Virus-mediated transfer of DNA – often chromosomal or plasmid DNA – from one cell to another; the phenomenon, first observed in bacteriophage/bacterium systems, is also seen during virus infection of eukaryotic cells.

Note. In the literature the term *transduction* is now frequently used, in a confusing way, with various (undefined) meanings; the above definition gives a long-established meaning.

The following account deals specifically with the process of transduction (as defined) in phage/bacterium systems.

Generalized transduction involves the infrequent encapsidation of bacterial DNA (instead of phage DNA) during phage assembly in a lytic infection; this could occur, for example, if chromosomal DNA had been fragmented as a consequence of phage infection. Phage capsids containing bacterial DNA can infect other cells and can donate the DNA to the recipient cell (the *transductant*). In a transductant, the fate of the fragment of (bacterial) DNA may be (i) degradation by the cell's own restriction enzymes; (ii) recombination with chromosomal or plasmid DNA in the transductant – in which case the fragment's gene(s) may be stably inherited and may be expressed (*complete transduction*); (iii) persistence as a stable but non-replicating extrachromosomal fragment (see entry ABORTIVE TRANSDUCTION).

Any given bacterial gene may have a similar probability of being transduced. However, in some cases, certain bacterial genes are more likely than others to be transduced. Thus, for example, in the *Salmonella*/phage P22 system, a sequence in the bacterial chromosome resembles that part of the phage genome which is concerned with packaging of DNA into the phage head; this region of the bacterial chromosome has a greater chance of being transduced.

See also MINI-MU.

Specialized transduction involves the rare aberrant excision of a prophage from the bacterial chromosome following lysogenic infection. In such an excision, the prophage takes with it some adjacent chromosomal DNA – and may leave behind, in the chromosome, DNA from the other end of the prophage. The resulting (recombinant) phage genome can be packaged and assembled normally – forming a *specialized transducing particle* (STP). Even if an STP is deficient for replication (having left behind essential genes), it can nevertheless inject its DNA into a recipient bacterium, thus transferring specific donor DNA to a transductant.

In any given system of specialized transduction, the gene(s) transduced are those bacterial genes that flank the prophage when it is integrated in the chromosome; for example, in the *Escherichia coli*/phage λ system, the genes *gal* (encoding galactose utilization) and *bio* (encoding biotin synthesis) may be transduced as these genes flank the integrated λ prophage.

In a unique form of specialized transduction, the genomes of two phages, VGJφ and CTXΦ, join covalently within cells of the Gram-negative pathogen *Vibrio cholerae*. The resulting (single-stranded) hybrid genome is packaged in a VGJφ capsid to form a particle which can infect fresh cells of *V. cholerae* via receptor sites for phage VGJφ. In this way, the genes for cholera toxin (encoded by phage CTXΦ) can be transduced to *Vibrio cholerae* independently of the (normal) receptor sites for CTXΦ [J Bacteriol (2003) 185:7231–7240].

transductional shortening A phenomenon which is sometimes observed when a *large* plasmid is transferred by transduction from one bacterial cell to another: the transduced plasmid is found to be smaller than the original plasmid – an effect that may be due to the encapsidation of a deletion mutant of the plasmid (which may arise spontaneously at low frequencies).

transfection (1) Any of a range of procedures in which nucleic acid is introduced into cell(s) or organism(s) – e.g. within a viral VECTOR, or in a plasmid vector internalized by methods such as ELECTROPORATION or TRANSFORMATION. (See also the entries for LIPOFECTION, MAGNETOFECTION, MICROINJECTION, NUCLEOFECTION and SPINOCULATION.)

(Some other relevant entries: ENDO-PORTER, MELITTIN and PROTAMINE SULFATE.)

(2) As in sense (1), above, but excluding any process that is mediated by viruses (or bacteria).

(3) Experimental infection of a vector organism with parasitic bacteria [Appl Environ Microbiol (2005) 71:3199–3204].

transformation A process in which exogenous (external) DNA is internalized and expressed by a cell, or by a modified cell such as a permeaplast, spheroplast or protoplast – the process occurring under natural conditions in some cases, although apparently only under *in vitro* conditions in others.

The *transforming* or *donor* DNA may be e.g. a fragment of chromosomal DNA or a plasmid.

Transformation is one approach to TRANSFECTION, and is used e.g. for bacteria, yeasts and mammalian cells.

Transformation in bacteria

Uptake of DNA under natural conditions can occur e.g. in the Gram-negative bacteria *Haemophilus* and *Neisseria* and the Gram-positive bacteria *Bacillus* and *Streptococcus*.

Cells that are capable of taking up DNA are said to exhibit *competence*.

In strains of *Neisseria gonorrhoeae* competence appears to be exhibited constitutively. However, some bacteria exhibit

competence transiently, and the expression of competence is influenced by factors such as the nutritional status of the cells (and/or the phase of growth in batch cultures); for example, competence in *Haemophilus influenzae* may be induced by conditions which limit growth and is promoted by high levels of intracellular cyclic AMP (cAMP).

In both *Bacillus subtilis* and *Streptococcus pneumoniae* the population density (i.e. number of cells per unit volume) is a factor in competence (a manifestation of QUORUM SENSING). Thus, when *B. subtilis* grows to a high population density, a secreted pheromone, ComX, accumulates extracellularly and, at an appropriate concentration, activates a TWO-COMPONENT REGULATORY SYSTEM in the cells; this results in activation of *comS* and other genes whose expression is needed for the development of competence. Earlier work with *Streptococcus pneumoniae* suggested that competence is associated with the transmembrane transport of calcium.

Fragments of chromosomal DNA must be double-stranded and must be above a certain minimum size to be effective in transformation. The fragments initially bind reversibly to the cell surface, but subsequently a proportion of the fragments become irreveribly bound prior to uptake. In *B. subtilis* and *S. pneumoniae* binding is non-specific (i.e. these cells can bind DNA from other species as readily as DNA from the same species). By contrast, species of *Haemophilus* and *Neisseria* bind fragments of DNA in a sequence-dependent manner; for example, *H. influenzae* was reported to internalize only those fragments containing 5′-AAGTGCGGTCA-3′ – a so-called *DNA uptake site*; this sequence occurs often in the genome of *H. influenzae*.

B. subtilis and *S. pneumoniae* take up only one strand of a dsDNA fragment; *H. influenzae* takes up both strands.

Laboratory-induced transformation in bacteria

Hitherto, it was generally believed that *Escherichia coli* does not undergo transformation in nature; accordingly, this lead to the development of *in vitro* methods for creating artificial competency in transformation in this species (see later).

One report described spontaneous transformation in *E. coli* on nutrient-containing agar plates; it was concluded that: (i) transformation in this species is stimulated by agar, and (ii) *E. coli* can be transformed by (intact) double-stranded DNA (i.e. the dsDNA is internalized) [J Bacteriol (2009) 191(3): 713–719].

Methods used for making *E. coli* (and other Gram-negative bacteria) competent for transformation include e.g. treatment by chilling and heat-shock in the presence of calcium ions. Earlier studies indicated that competence acquired by *E. coli* in an ice-cold, calcium-containing solution is associated with complexes of polyhydroxybutyrate/calcium polyphosphate in the cytoplasmic membrane; it was suggested that these complexes may form channels in the membrane, thus facilitating uptake of DNA.

Gram-positive bacteria may be subjected to transformation by first converting the cells to protoplasts. Protoplasts can be induced to internalize DNA e.g. by treatment with POLYETHYLENE GLYCOL (PEG); the transformed protoplasts are then used to regenerate viable cells.

Certain cyanobacteria are able to take up DNA following their conversion to permeaplasts.

Enhanced efficiency of bacterial transformation

One barrier to (inter-species) transformation results from the nucleolytic activity of RESTRICTION ENDONUCLEASES in the recipient cell; this tends to degrade 'foreign' imported DNA. In one study, improved efficiency of bacterial transformation was reported with a method in which, before electroporation, the transforming DNA was *methylated* in engineered cells of *Escherichia coli* which expressed the specific methyltransferases of the target bacterium. Thus, when two type II methyltransferases of *Bifidobacterium adolescentis* were expressed in cells of *E. coli*, the DNA that was methylated in these cells was reported to transform *Bifidobacterium* with an efficiency enhanced five orders of magnitude. Enhanced efficiency of transformation was also reported in *Lactococcus lacti*s when pre-methylation was conducted with a putative type I methyltransferase system in *E. coli*. [The method (plasmid artificial modification): Nucleic Acids Res (2009) 37(1):e3.]

Transformation in eukaryotic organisms

Certain (unicellular) green algae (including *Chlamydomonas reinhardtii*) have a thin, cellulose-free, protein-rich cell wall, and these organisms can be subjected to transformation.

Certain fungi are susceptible to *in vitro* transformation. For yeasts (e.g. *S. cerevisiae*), the cells are generally converted to spheroplasts (because cell walls are inhibitory to uptake). In one approach, spheroplasts are first mixed with DNA in the presence of calcium ions, and PEG is subsequently added. In a different approach, whole cells are treated with lithium ions and PEG; although simpler, this method is reported to be less efficient.

Mammalian cells can internalize DNA that has been overlaid on a cell monolayer in a (buffered) solution containing calcium chloride; the solution forms a DNA-containing coprecipitate which facilitates uptake by the cells. The cells are incubated overnight at 37°C, washed, and incubated in fresh medium. The efficiency of this process may be increased by including a 'carrier DNA' (e.g. calf thymus DNA) with the experimental DNA; the optimal concentration of carrier DNA must be pre-established.

transgene (*DNA technol.*) (1) Any gene from a given individual which has been stably inserted into another and which can be transmitted, by the recipient, to successive generations.
(2) (*DNA technol.*) In a more general sense, any gene which has been transferred by DNA technology to a cell or organism (of either the same of a different species) and which can be expressed in that cell or organism.

transgenesis (*DNA technol.*) (1) A form of genetic engineering: the transfer of a gene from one individual to another in order to achieve a specific permanent change in the recipient that is transmissible to successive generations. In general, this term is reserved for gene transfer between eukaryotic, macroscopic organisms.

In mice, genes may be introduced by MICROINJECTION into the nucleus of a (fertilized) egg; this egg is then cultured *in vitro* and undergoes division – so that a recombinant blastocyst can be subsequently implanted in a female mouse for development. When carried out successfully, all the cells in the developing embryo, including the germ line cells, will carry the transgene; this approach may therefore be chosen in order to avoid the development of a *chimera* (an individual, produced by some methods, in which the transgene does not occur in *all* the cells of the developing embryo). One problem is that there is no control over (i) the *site* of gene insertion or (ii) the *number* of copies of the gene that are inserted into chromosomal DNA; this means that, following development, transgene *expression* is unpredictable. In an ideal techinque, a protocol would be used in which a single transgene can be specifically targeted to a chosen chromosomal location.

In another approach, a retroviral vector, incorporating the transgene, may be used to infect a later, multicellular stage of the developing fertilized egg. This can be successful if all the cells receive the transgene, but a chimera will be produced if one or more of the cells remain uninfected by the retrovirus. One disadvantage of this approach is that the insert carried by retroviral vectors is quite small – and may be insufficient for some purposes (e.g. for the transfer of a large gene). (Another problem relates to the unknown number of copies of the gene inserted into a given cell.)

Cultured *embryonic* STEM CELLS (derived originally from a BLASTOCYST) can be engineered (see e.g. GENE TARGETING) before being inserted into a blastocyst and then implanted in a surrogate female mouse for development. This method has the advantage that it permits *site-specific* modifications to be made, and, as it is an *in vitro* procedure, it is possible to use various approaches to select cells which have undergone the required modification before they are inserted into the blastocyst. However, this method gives rise to chimeras if there are unmodified cells (as well as the genetically modified ones) in the blastocyst; even so, breeding and selection of mice which have undergone this treatment may lead to the production of non-chimeric individuals.

(2) (*DNA technol.*) The transfer of a gene from one cell or organism to another (of the same or a different species) such that the gene can be expressed in the recipient cell or organism, whether or not it is transmitted to the next generation.

transition mutation Any point mutation in which a purine is replaced by a different purine, or a pyrimidine is replaced by a different pyrimidine.

(cf. TRANSVERSION MUTATION.)

translational attenuation (1) One of the modes in which gene expression is regulated. An example is the regulation of gene *pyrC* in *Escherichia coli*; *pyrC* encodes dihydroorotase, an enzyme involved in the synthesis of pyrimidines (and, hence, nucleic acids). Transcription of this gene can start at any of several adjacent nucleotides in the DNA template strand – transcription from a particular nucleotide being influenced by the intracellular availability of pyrimidines (as 'sensed' by the CTP/GTP ratio in the cell). Given the presence of adequate concentrations of pyrimidines, the transcripts are mainly of a kind in which intra-strand base-pairing occurs in the region of the Shine–Dalgarno sequence; base-pairing in this region blocks mRNA–ribosome binding – and therefore blocks gene expression. If the levels of pyrimidines are low, most of the transcripts are synthesized from a slightly different starting point; in these transcripts the nucleotide sequence is such that intra-strand base-pairing does not occur – so that transcripts bind to the ribosomes and dihydroorotase is synthesized.

(cf. *translational control* in an OPERON.)

(2) *Syn.* CODON HARMONIZATION.

translational control (in operons) See OPERON.

translational enhancer See ENHANCER (translational).

translational frameshifting See FRAMESHIFTING.

translational fusion See GENE FUSION.

translational recoding (of UGA) See the entry SECIS.

translesion synthesis Error-prone synthesis (repair) of DNA on a template containing a lesion (e.g. a pyrimidine dimer). This occurs e.g. in *Escherichia coli* during the SOS response (see SOS SYSTEM) when DNA is damaged; it involves specialized DNA polymerases (e.g. DNA polymerase V) and characteristically leads to mutagenesis.

[Eukaryotic translesion polymerases (roles and regulation in DNA damage tolerance): Microbiol Mol Biol Rev (2009) 73(1):134–154.]

[A fluorogenic method for characterization of the activity of DNA polymerases in translesion synthesis: Nucleic Acids Res (2009) doi: 10.1093/nar/gkp641.]

(See also Y FAMILY.)

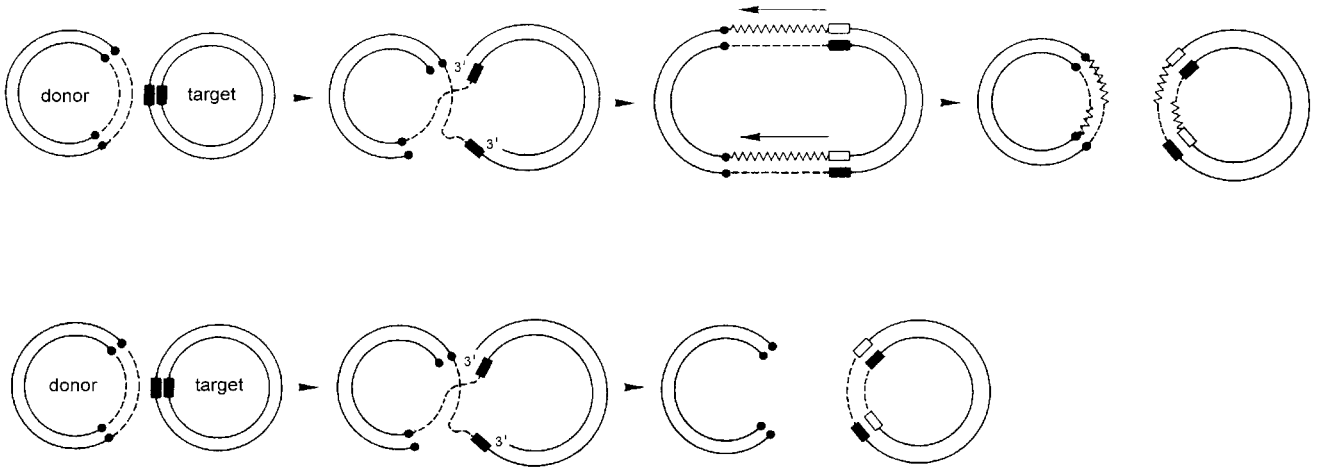
transport medium Any MEDIUM used for the transportation (or temporary storage) of material which is to be examined, later, for the presence of particular type(s) of microorganism.

The main role of a transport medium is to maintain the viability of any organisms in the sample. One common example of this kind of medium is Stuart's transport medium (which is suitable for delicate bacteria such as *Neisseria gonorrhoeae*); Stuart's transport medium contains salts, $<1\%$ agar, sodium thioglycollate and the redox indicator methylene blue.

transposable element (TE; jumping gene) A discrete sequence of DNA – usually present within a larger molecule of DNA, although sometimes in isolation (i.e. excised, or in an *in vitro* preparation) – which has the ability, actually or effectively, to move from one site to another (i) in the same molecule, (ii) in another DNA molecule within the same cell or (iii) in a DNA molecule in another cell; such *transposition* does not depend on extensive sequence homology between the TE and its new location.

A TE that can transpose *in vivo* to a site in *another* cell is referred to as a *conjugative transposon* because it mediates a unique (plasmid-independent) form of conjugation (between a *donor* cell and a *recipient* cell) that is an essential feature of transposition in this type of TE; conjugative transposons are considered in the entry CONJUGATIVE TRANSPOSITION.

This entry deals with TEs *other than* conjugative transposons.



TRANSPOSABLE ELEMENT A simplified, diagrammatic scheme for the replicative transposition (top) and simple ('cut-and-paste') transposition (bottom) of a transposable element (TE).

Replicative transposition. Two circular, double-stranded DNA molecules are shown at the left-hand side. The donor molecule includes a TE (dashed lines), either side of which is an old target site (●); this target site was duplicated when the TE was originally inserted into that molecule (see later). The target molecule has a single target site (▮) where the TE element will be inserted.

An enzyme (*transposase* – not shown) mediates at least the initial stages of transposition. In the target molecule a staggered break has been made at the target site, leaving the 3′ ends recessed. In the donor molecule a nick has been made in each strand of the TE, at opposite ends, and the free ends have been ligated (joined) to the 5′ ends of the target molecule, as shown.

In the next stage, DNA synthesis (zigzag line) has occurred from each 3′ end in the target molecule (arrows). Such synthesis first copies the target site (□) and then continues beyond the target site – using each strand of the TE as template; that is, the TE has been replicated. The end of each newly synthesized strand has been ligated to a free strand-end in the donor molecule. The resulting structure is called a *cointegrate*.

The final stage of replicative transposition involves a *resolvase* – an enzyme, encoded by the TE, which resolves the cointegrate by mediating site-specific recombination at a site in each TE – forming the two molecules as shown.

The donor and target molecules now both contain the sequence of the TE. Note that each of these two sequences contains parts of the original TE (dashed lines) as well as newly synthesized DNA (zigzag lines). Note also that the target site in the target molecule has been duplicated – each of these target sites consisting of one newly synthesized strand (□) and one original strand (▮).

Simple ('*cut-and-paste*') *transposition*. The initial stages are similar to those shown for replicative transposition. However, in this case, DNA synthesis (from the 3′ ends of the target molecule) is required simply to duplicate the target site. The remaining strand-ends of the TE have been cut and ligated to the freshly duplicated target sites, as shown. The phrase *donor suicide* is used if the remainder of the donor molecule is non-viable.

Figure reproduced from *Bacteria in Biology*, *Biotechnology and Medicine*, 6th edition, Figure 8.4, pages 202–203, Paul Singleton (2004) John Wiley & Sons Ltd, UK [ISBN 0-470-09027-8] with permission from the publisher.

TEs occur in both prokaryotic and eukaryotic cells and are found as normal constituents e.g. in chromosomes, plasmids and phage genomes.

Note. Some of the transposable elements which are currently used in DNA technology were originally isolated as inactive (non-functional) sequences of DNA – i.e. they were unable to transpose; these elements were subsequently engineered ('re-animated', reconstructed) such that they can now function as transposable elements: see RECONSTRUCTED TRANSPOSABLE ELEMENT.

If a TE encodes *only* those functions which are necessary for transposition it is referred to as an INSERTION SEQUENCE (q.v.).

If a TE encodes functions (e.g. antibiotic resistance) which are *in addition to* those necessary for transposition it is referred to as a TRANSPOSON (q.v.).

All TEs encode at least one protein: transposase, an enzyme which mediates recombination and which is needed for transposition; some TEs also encode a resolvase (see legend in the figure).

As shown in the figure, there are two *main* types of transposition: (i) a conservative cut-and-paste mechanism, and (ii) a replicative mechanism. These two modes of transposition are not exclusive; thus, for example, some types of insertion sequence (e.g. IS*1111*) excise in a circular, intermediate form prior to insertion at the target site.

The process by which a TE inserts into a new site usually results in the development of a pair of *direct repeats* which flank the newly inserted TE; this is due to duplication of the target site – as shown in the figure. (Note that direct repeats also develop in INSERTION–DUPLICATION RECOMBINATION.)

Note that, in the model of replicative transposition shown in the figure, each of the two *final* molecules includes parts of the original transposon (dashed line) as well as the newly synthesized DNA (zigzag line); accordingly, it is not strictly correct to say that a *copy* of the transposon is inserted at the new site.

In certain TEs in eukaryotes transposition involves an RNA intermediate: see RETROTRANSPOSON.

TEs differ e.g. with respect to the specificity of their *target* site, that is, the site to which the TE can re-locate. Many TEs appear to insert at new sites in a random fashion. It seems likely, however, that even when insertion appears random, a selective process of some kind is involved. By contrast, the transposon Tn*7* has a specific insertion site that is designated *att*Tn*7*, and the insertion sequence IS*5* appears to insert only at locations containing the sequence C(T/A)A(G/A).

Transposons of the MARINER FAMILY are characterized by minimal site-specificity of their target sequences.

In some of the TEs which form a circular intermediate, the target sequence is homologous to the so-called *interstitial junction sequence* (see below), and this has suggested that, in these elements, transposition may be site-specific. Indeed, in the insertion sequence ISHP608 (q.v.) transposition is reported to be site-specific (at the sequence 5′-TTAC-3′).

At least some of the TEs that form a circular intermediate – such as members of the IS*1111* family – have *sub*terminal inverted repeats, unlike other TEs, which are characterized by terminal inverted repeats. In some of these elements, production of the circular intermediate form is reported to involve initial cleavage of one strand at the 3′ end of the TE (using the TE's transposase); the 3′ end is subsequently transferred to a site a few nucleotides upstream of the 5′ end of the TE, on the same strand. This structure leads to the formation of the circular intermediate in which the TE's two inverted repeat sequences are separated by a short stretch of nucleotides: the *interstitial junction sequence*; in some cases, this sequence is reported to provide enhanced facilities for transcription of the TE.

Transposition occurs only rarely *in vivo*. The frequency of transposition depends e.g. on the TE and on the physiological state of the cell. It can also be influenced e.g. by the methylation status of the cell's DNA. For example, transposition of transposon Tn*10* from a chromosomal site in *Escherichia coli* is inhibited by the (normal) Dam methylation of the *E. coli* chromosomal DNA (see DAM GENE); the reason is that Dam methylation inhibits transcription of Tn*10* transposase. This inhibition is transiently lost during chromosomal replication – i.e. immediately after the replication fork has passed the transposase gene but before Dam methylation of the gene has occurred; hence, transposition of Tn*10* tends to be linked to the cell cycle (i.e. to DNA replication). In this transposon, the frequency of transposition is also affected by MULTICOPY INHIBITION.

The effects of transposition in a cell will depend on the site of insertion of the TE. Effects may include e.g. inactivation of a structural gene, inactivation of gene(s) within an operon, or activation of an operon (e.g. when an insertion inactivates a sequence that negatively controls the operon). The presence of insertion sequence IS*3* in the F plasmid (within the *finO* gene) results in derepression of that plasmid for conjugative transfer (see FINOP SYSTEM).

Transposon-mediated mutagenesis is a useful tool in DNA technology: see TRANSPOSON MUTAGENESIS.

The presence of a TE is commonly indicated in text by a double colon. For example, the presence of transposon Tn*3* in the genome of phage lambda (phage λ) is written λ::Tn*3*.

In addition to the main entries INSERTION SEQUENCE and TRANSPOSON, see entries for some individual TEs under 'IS' and 'Tn'.

transposase A type of enzyme required for the transposition of an INSERTION SEQUENCE or a TRANSPOSON (see TRANSPOSABLE ELEMENT); a transposase is also encoded by PHAGE MU (q.v.).

In the *replicative* form of transposition a *resolvase* enzyme is also needed in order to complete the transposition process (by resolving the cointegrate).

transposition See TRANSPOSABLE ELEMENT.

transposition immunity A phenomenon in which the presence

of a given transposon in a replicon inhibits the insertion of a further copy of that transposon into the same replicon. Transposition immunity is exhibited e.g. by transposon Tn*3* and by many of the Tn*3*-like transposable elements.

transposome (1) A term used to refer (apparently) to the sum total of transposable elements/targets in populations of the fruitfly *Drosophila* [PLoS Genetics (2006) 2(10):e165].

(2) A term used to refer to 'nucleoprotein complexes' in the context of transposition of IS*911* [J Bacteriol (2008) 190(18): 6111–6118].

(cf. TRANSPOSOSOME.)

Transposome™ A stable complex, consisting of a transposase and a transposon, which can be inserted into (living) cells and used for TRANSPOSON MUTAGENESIS: see the entry TN*5*-TYPE TRANSPOSITION.

transposon (Tn) A TRANSPOSABLE ELEMENT which encodes functions *in addition to* those required for transposition – i.e. functions such as antibiotic resistance (e.g. in Tn*3*, Tn*5*, Tn*10* and Tn*2410*), resistance to heavy metals (e.g. in Tn*501*) or formation of virulence factors such as toxins (e.g. in Tn*1681*) etc.

(cf. INSERTION SEQUENCE.)

As indicated above, a specific transposon is conventionally designated by the prefix 'Tn' followed by an (*italic*) number.

The *conjugative transposons* mediate plasmid-independent CONJUGATION in bacteria; these transposons differ in various ways from other transposons and are considered separately in the entry CONJUGATIVE TRANSPOSITION.

A *class I* (= type I, or *composite*) transposon consists of the sequence of structural genes flanked, on both sides, by an INSERTION SEQUENCE; the two insertion sequences may or may not be identical. The insertion sequences mediate transposition of the transposon as a whole. One insertion sequence may be able to transpose independently, i.e. it may be able to behave as an isolated transposable element.

A *class II* (= *simple*) transposon consists of the sequence of structural genes flanked by a pair of INVERTED REPEAT sequences.

Transposition may involve a simple, conservative 'cut-and-paste' mechanism or a 'replicative' mechanism, according to the given transposon – see TRANSPOSABLE ELEMENT (figure) for an explanation of these terms. Tn*10* is one example of a transposon that uses the cut-and-paste mechanism. Tn*3* is one example of a transposon that transposes in a replicative way. Transposons in a particular group may characteristically use one or other of these mechanisms – e.g. transposons of the MARINER FAMILY typically use a cut-and-paste mechanism.

The specificity with which a transposon inserts into a target sequence varies with transposon. Some transposons appear to insert more or less randomly. Others (such as members of the *mariner* family) have minimal target specificity. Transposon Tn*7* (see TN*7*) has a highly specific target site.

(See also entries for individual transposons under 'Tn'.)

Transposons in DNA technology

Transposons find a number of uses in DNA technology – for example TRANSPOSON MUTAGENESIS.

Transposons, and derivatives, are also used for transfection and transgenesis. In general, an insert carried by a transposon may be quite small; however, a large sequence (57.5 kb) was inserted into heterologous hosts using this approach [Nucleic Acids Res (2008) 36(17):e113].

(See also TN*5*-TYPE TRANSPOSITION.)

PHAGE MU is also used in transposition-based technology.

transposon library See LIBRARY (sense 6).

transposon mutagenesis MUTAGENESIS mediated by a TRANSPOSON. (See also LIBRARY (sense 6).) A mutation created by the insertion of a transposon can be readily mapped as the transposon provides a known sequence of nucleotides which can be exploited in the design of primers. A general requirement in this method is that the transposon used be able to insert *randomly* into the target replicon; one reason for this is that if (as in some cases) a transposon inserts primarily into certain 'hot spots' then it would be necessary to screen a large number of cells when attempting to detect the required (rare) mutant. The transposons used are therefore ones that insert randomly, or with minimal site-specificity; thus, for example, a derivative of a *mariner* transposon was used in *Francisella tularensis* [Appl Environ Microbiol (2006) 72(3):1878–1885] and also in *Bacillus subtilis* [Appl Environ Microbiol (2006) 72(1):327–333].

Transposon mutagenesis has been used to detect/identify those virulence genes (in a pathogen) which are essential for the pathogen's growth in the host organism: see SIGNATURE-TAGGED MUTAGENESIS.

The method is also used to detect/identify cell-surface or secreted proteins of pathogenic bacteria – i.e. proteins which (because of potential interaction with the host organism) *may* be associated with virulence. Cells of a given pathogen are first mutagenized with transposons containing a promoterless reporter gene encoding the enzyme ALKALINE PHOSPHATASE (AP); these transposons (designated Tn*phoA*) insert randomly into different genes in the population of cells. When plated, these (mutagenized) cells form individual colonies, i.e. each colony arises from a single cell. The colonies are tested with a substrate which is split by AP to produce a colored product. Any colony which gives a positive reaction (i.e. color) therefore indicates a secreted or cell-surface alkaline phosphatase because the substrate can be used only by an extracellular or cell-surface enzyme. As *phoA* in the transposon lacks both promoter and signal sequence, the formation of an active AP by cells of a given colony indicates that Tn*phoA* has inserted, in frame, into the gene of a secreted or cell-surface protein. If this gene is a virulence-associated gene, then the virulence of the cells in the colony may be demonstrably decreased as the gene is likely to have been inactivated by insertion of the transposon; the mutant cells in this colony can be tested for virulence and, if required, the relevant gene can be isolated and sequenced.

Transposons can be either inserted into living cells or used for *in vitro* mutagenization of e.g. a population of plasmids

(see TN5-TYPE TRANSPOSITION).

(See also GENETIC FOOTPRINTING.)

transpososome A complex, which includes a transposase, that is involved in the process of transposition: see e.g. entry TN5.

The H-NS protein was reported to bind to the Tn*10* transpososome, and to stabilize it [Nucleic Acids Res (2009) doi: 10.1093/nar/gkp672].

(cf. TRANSPOSOME.)

transversion mutation Any point mutation in which a purine is replaced by a pyrimidine, or a pyrimidine by a purine.

(cf. TRANSITION MUTATION.)

TRAP Telomeric repeat amplification protocol: a method for gaining information on the activity of TELOMERASE.

A synthetic telomeric repeat sequence is added to a sample lysate and is extended by telomerase. The extended synthetic fragment is then amplified by PCR. On electrophoresis of the products, the extent of addition to the fragment by telomerase is reflected in the length of the PCR products.

The TRAP assay used for plants differs from human TRAP assays e.g. in temperature (room temperature, or 26°C) and in the buffers used.

trastuzumab (Herceptin®) See EPIDERMAL GROWTH FACTOR RECEPTOR FAMILY.

TRCF Transcription repair coupling factor: see MFD GENE.

TRES See TIME-RESOLVED EMISSION SPECTRA.

tribrid system See e.g. YEAST THREE-HYBRID SYSTEM.

Trichoplusia ni The cabbage looper (a lepidopteran) – see e.g. GRANULOSIS VIRUSES and PIGGYBAC.

trichostatin A An inhibitor of histone deacetylases (HDACs). (See HDAC; see also ACETYLATION (of histones).)

Trichostatin was reported to be a key factor in the *in vitro* differentiation of human mesenchymal stem cells – derived from adult bone marrow – towards functional hepatocyte-like cells [BMC Dev Biol (2007) 7:24].

Combination therapy consisting of trichostatin A and oncolytic herpes simplex virus was reported to exhibit enhanced antitumoral (and anti-angiogenic) effects [Mol Ther (2008) 16(6):1041–1047].

Other uses (e.g.): studies on inhibition of Caspase-4 [PLoS ONE (2009) 4(4):e5071], and on the reactivation of murine gammaherpesvirus 68 [PLoS ONE (2009) 4(2):e4556].

trifolitoxin A peptide antibiotic which is active against e.g. the plant-pathogenic bacterium *Agrobacterium*.

trimethoprim See ANTIBIOTIC (synergism).

triostins See QUINOXALINE ANTIBIOTICS.

triple-specific proximity ligation assay (3PLA) See the entry PROXIMITY LIGATION ASSAY.

triple-strand probe A single-stranded PROBE that hybridizes to a complementary region at the 5′ end of one strand in linear dsDNA to form a highly stable triple-stranded structure; such binding is facilitated by the RecA protein. The stability of the triple-stranded structure is greatly reduced by even a single mismatched base, regardless of location, and it was suggested that this system may be useful e.g. for detecting SNPs.

triple-stranded DNA *Syn.* TRIPLEX DNA.

(See also TRIPLE-STRAND PROBE.)

triplex DNA A triple-stranded, helical structure in which a STRAND of nucleic acid is located in the major groove of a DNA duplex – the pyrimidine or purine bases present in the third strand binding to purine bases in the DNA duplex by Hoogsteen base-pairing.

In some cases the third strand in the triplex DNA derives from a sequence close to the acceptor region, in the same DNA molecule; in other cases it may derive from a separate DNA molecule or from a distant region of the same molecule – or it may be a synthetic strand.

The two basic triplex schemes are: (i) the purine motif (Pu) in which the third strand is A-rich or G-rich and binds in an ANTIPARALLEL fashion to the acceptor strand in the duplex; (ii) the pyrimidine motif (Py) in which the third strand is C-rich or T-rich and binds in a parallel fashion.

An intramolecular triplex may develop, for example, in a supercoiled molecule when, on separation of the two strands in a homopurine–homopyrimidine sequence (of more than 10 base-pairs in length), one of the strands forms the third strand of a triplex by binding to a given sequence with Hoogsteen base-pairing. If the third strand is pyrimidine-rich the triplex structure thus formed is called H-DNA; if purine-rich it is called H′-DNA.

Some triplex structures are stable under physiological conditions.

A triplex may also be formed by a synthetic *triplex-forming oligonucleotide* (TFO). TFOs have *in vitro* applications and possibly have uses in GENE THERAPY. They can inhibit transcription, and have been able to down-regulate the expression of a proto-oncogene in HeLa cells.

(See also PROGRAMABLE ENDONUCLEASE.)

The activity of a TFO may be enhanced by using LOCKED NUCLEIC ACIDS [*pyrimidine*: Nucleic Acids Res (2005) 33 (13):4223–4234; *purine*: J Biol Chem (2005) 280(20):20076–20085].

The dye YOYO-1 has been reported to promote formation of triplex DNA.

A PSORALEN-tagged TFO has been used to direct psoralen to the target site for site-directed, light-activated inter-strand cross-linking between pyrimidines in dsDNA.

Targeted gene conversion, conducted in mammalian cells, has been carried out with a triplex-directed approach [Nucleic Acids Res (2009) doi: 10.1093/nar/gkp678].

RNA and PNA in triplex structures

RNA can be involved in triplex structures, as can PNA (see below).

High-affinity targeting of dsDNA by triplex-forming PNA has been achieved with chemically modified PNA oligomers. These oligomers were reported to have a binding affinity capable of >1000-fold modulation – achievable by the regulation of e.g. pH, length of oligomer, net charge and/or substitution of pseudoisocytosine for cytosine [Nucleic Acids Res (2009) doi: 10.1093/nar/gkp437].

triplex-forming oligonucleotide See TRIPLEX DNA.

Tris (TRIS, tris, THAM, Tromethamine) The compound: tris (hydroxymethyl)aminomethane; with hydrochloric acid, Tris forms a (temperature-sensitive) pH buffering system which is effective between approximately pH 7 and pH 9 with a pK_a of 8 at 25°C.

(See also RIBONUCLEASE.)

trisomy (detection) See e.g. the entries DOWN'S SYNDROME and EDWARDS' SYNDROME.

(See also CHROMOSOME ABERRATION.)

trisomy 18 EDWARDS' SYNDROME. (For pre-natal diagnosis of Edwards' syndrome see EPIGENETIC ALLELIC RATIO ASSAY.)

trisomy 21 *Syn*. DOWN'S SYNDROME (q.v.).

trithorax-group genes (*Drosophila*) See POLYCOMB-GROUP GENES.

tritium See e.g. AUTORADIOGRAPHY.

TRIzol® A reagent (Invitrogen, Carlsbad CA) used e.g. in the isolation of total RNA from cells; it consists of a monophasic solution of (i) phenol and (ii) guanidine isothiocyanate.

Tromethamine *Syn*. TRIS.

trophectoderm See the entry BLASTOCYST.

Tropheryma whipplei (former name: *Tropheryma whippelii*) A Gram-positive bacterium associated with Whipple's disease. The organism was once regarded as uncultivable but has been cultured in fibroblasts, and its complete genome sequence is known [Lancet (2003) 361:637–644].

trovafloxacin See QUINOLONE ANTIBIOTICS.

***trp* operon** See OPERON.

TRP-185 protein See AAVS.

truncated gene *Syn*. PSEUDOGENE.

***trxA* gene** The *Escherichia coli* gene encoding THIOREDOXIN.

trypanosomes Parasitic, flagellate protozoans of the genus *Trypanosoma* which cause diseases such as Chagas' disease and sleeping sickness in humans – and which are also responsible for e.g. dourine and nagana in animals.

In addition to their nuclear DNA, these organisms contain KINETOPLAST DNA (in the mitochondria).

The expression of protein-encoding genes in trypanosomes appears generally not to be regulated at the level of initiation of transcription; this is relevant in the design of systems for engineering gene expression in these organisms – see entry EXPRESSION VECTOR.

(See also TRANS SPLICING (of transcripts).)

See the entry VSG for details of ANTIGENIC VARIATION in trypanosomes.

(See also OPERON and RNA EDITING.)

Note. The meaning of *trypanosome* is sometimes extended to include other trypanosomatids – e.g. species of *Leishmania*.

T-S oligo See CAPFINDER.

TSA TYRAMIDE SIGNAL AMPLIFICATION.

tSDA Thermophilic SDA (see the entry SDA).

TSG Tumor suppressor gene.

tsunami victims (identification) See FORENSIC APPLICATIONS (in the section *Identification*).

TTGACA (in *Escherichia coli*) See the entry PROMOTER.

TTGE (temporal temperature-gradient gel electrophoresis) A method – similar in principle to DGGE – in which separation of *related*, *double-stranded* fragments of DNA involves an ongoing rise in temperature (instead of a gradient of chemical denaturing agents) to achieve differential melting of sample fragments.

TTGE has been used e.g. for studies on bacterial population dynamics [Appl Environ Microbiol (2008) 74:5662–5673]; *Lactobacillus rhamnosus* 35 [Appl Environ Microbiol (2008) 74:2679–2689]; and mitochondrial DNA abnormalities in the buccal cells of cigarette smokers [Carcinogenesis (2008) 29: 1170–1177].

TTVI center The Thai Tsunami Victim Identification center: see the entry FORENSIC APPLICATIONS (in section: *Disasters with mass fatalities*).

TUNEL assay Terminal deoxynucleotidyl transferase-mediated dUTP nick end labeling: a method used e.g. for assessing the fragmentation of genomic DNA in APOPTOSIS. (See TAILING for the action of the enzyme terminal deoxynucleotidyl transferase.)

In e.g. apoptosis, fragmentation involves the formation of a large number of breaks in the DNA, thus exposing many free 3′-hydroxyl ends. The extent to which exposed 3′-hydroxyl ends are present can therefore be used as an indication of the level of fragmentation of the DNA.

In the assay, labeled nucleotides are added to the exposed 3′ ends by the enzyme terminal deoxynucleotidyl transferase; the bound label is then measured, giving an indication of the abundance of free 3′ ends. (The label may be a chromophore or a fluorophore.)

[Examples of use: Peptides (2007) 28(7):1383–1389; PLoS Pathogens (2007) 3(11):e161; PLoS ONE (2009) 4(1):e4202 (see Figure 1); Mol Vision (2009) 15:1418–1428; PLoS ONE (2009) 4(8):e6651; J Histochem Cytochem (2009) 57(8): 721–730.]

two-component regulatory system In many bacteria (and in at least some archaeans): a system for regulating genes typically involving (i) a histidine kinase in the cell envelope and (ii) a response regulator protein which, on phosphorylation by the kinase, regulates the expression of particular target gene(s); these two-component systems enable organisms to sense, and respond to, specific environmental factors (such as changes in extracellular osmolarity) – activation of the system occurring via a so-called *signal transduction pathway* (i.e. from the stimulated sensor to the response regulator and thence to the regulatory function).

Two-component systems are involved in the regulation of a wide range of processes. For example, they have been reported to regulate adhesion and oxygen-regulated exotoxigenesis in *Staphylococcus aureus*; uptake of potassium ions (K^+) in *Escherichia coli* (in hypertonic conditions); toxin production in *Clostridium perfringens*; the development of competence in transformation in *Bacillus subtilis*; and membrane permeability in *Pseudomonas aeruginosa*.

(See also ALGR, CROWN GALL and QUORUM SENSING.)

In the high-affinity transport system for K^+ ions in the cell

envelope of *Escherichia coli*, the kinase KdpD, acting as the sensor, apparently responds to the lowered turgor pressure in hypertonic media and phosphorylates the response regulator KdpE – which promotes expression of the *kdpABC* operon; one of the products, KdpB, a K^+-ATPase (potassium pump), mediates (ATP-dependent) uptake of K^+ – which counteracts the hypertonicity.

Some two-component systems are *essential* for viabilitity of the cell. Thus, in some cases, such systems are involved in regulating gene expression in the control of the cell cycle. An example in *Bacillus subtilis* is the YycFG system – in which YycG is the histidine kinase and YycF the response regulator. (A similar system also occurs in other Gram-positive bacteria within the low-GC% category – e.g. *Enterococcus faecalis* and *Staphylococcus aureus*.) In *B. subtilis*, YycF is reported to control genes which regulate the composition of the cell wall – genes which include those in the *tagAB* and *tagDEF* operons associated with biosynthesis of the cell wall polymer teichoic acid.

two-hybrid systems (1) (bacterial) See BACTERIOMATCH TWO-HYBRID SYSTEM and BACTERIAL TWO-HYBRID SYSTEM.
(2) (in *Saccharomyces cerevisiae*) See YEAST TWO-HYBRID SYSTEM and CYTOTRAP TWO-HYBRID SYSTEM.

two-metal-ion catalysis (two-Mg^{2+}-ion catalysis) The proposed involvement of two metal (Mg^{2+}) ions in the substrate recognition and catalytic specificity of DNA and RNA polymerases – and of many nucleases and transposases – [Mol Cell (2006) 22:5–13].

two-micron plasmid (2 μ plasmid, 2 μ circle etc.) A 6318-bp dsDNA multicopy PLASMID which is commonly present in the yeast *Saccharomyces cerevisiae*; typically, more than 50 copies are found in the cell's nucleus. The 2 μ plasmid occurs complexed with proteins to form a chromatin-like structure; it encodes a site-specific recombination system which is used in recombinant DNA technology: see the entry FLP.

two-temperature protocol (in PCR) See 'temperature cycling' in the entry PCR.

Ty element A RETROTRANSPOSON, multiple copies of which occur in the genome of *Saccharomyces cerevisiae*.

At least some of the Ty elements resemble retroviruses; for example, the Ty3 element encodes homologs of the Gag–Pol proteins that, with genomic RNA, can assemble into virus-like particles.

Ty elements transpose to non-homologous sites, forming a 5-bp duplication at the target site; the target sites of Ty1–Ty4 elements tend to be upstream of genes transcribed by RNA polymerase III (often tRNA genes).

type I restriction endonuclease See RESTRICTION ENDONUCLEASE.

type I transposon (class I transposon) See TRANSPOSON.

type II restriction endonuclease See RESTRICTION ENDONUCLEASE.

type II transposon (class II transposon) See TRANSPOSON.

type IIA restriction endonuclease See RESTRICTION ENDONUCLEASE.

type IIB restriction endonuclease See RESTRICTION ENDONUCLEASE.

type IIC restriction endonuclease See RESTRICTION ENDONUCLEASE.

type IIE restriction endonuclease See RESTRICTION ENDONUCLEASE.

type IIF restriction endonuclease See RESTRICTION ENDONUCLEASE.

type IIG restriction endonuclease See RESTRICTION ENDONUCLEASE.

type IIM restriction endonuclease See RESTRICTION ENDONUCLEASE.

type IIP restriction endonuclease See RESTRICTION ENDONUCLEASE.

type IIS restriction endonuclease See RESTRICTION ENDONUCLEASE.

type III restriction endonuclease See RESTRICTION ENDONUCLEASE.

type IV restriction endonuclease See RESTRICTION ENDONUCLEASE.

typing (*microbiol.*) The term 'typing' has two distinct, although related, meanings. In one sense it refers to a form of *classification* in which STRAINS of a given species are categorized according to specified criteria. Thus, for example, if a large number of random clinical isolates of the pathogen *Staphylococcus aureus* are each examined for susceptibility to a range of different phages, these strains will generally be found to differ in their susceptibility to particular phages. Thus, while one strain may be lysed by a certain set of phages, another strain may be lysed by a different set – although some of the phages may lyse both strains. The particular set of phages to which a given strain is susceptible can therefore be used to define (i.e. classify) that strain and to distinguish it from the other strains of that species in the population being studied; this procedure, called *phage typing*, has been used for typing *S. aureus* and for typing certain other bacteria (e.g. *Yersinia enterocolitica*). A species that is typed by this method can be divided (classified) into a number of so-called 'phage types' – although not all strains of a given species may be suitable for phage typing.

Phage typing is only one of many methods used to classify bacteria. There are also various nucleic-acid-based methods – see later.

The second meaning of 'typing' refers to a form of *identification*. Thus, once a species has been typed (i.e. classified) by a given typing method, an *unknown* strain – isolated from a clinical or other sample – may be examined by the criteria of the given typing system to determine its relationship (if any) to other strains of the species. Thus, e.g. (following on from the above example) an unknown clinical isolate of *S. aureus* can be examined for its susceptibility to the particular range of phages used in the given typing system employed in the laboratory; when its pattern of susceptibility is known it may correspond to that of one of the established phage types.

In the context of identification, it should be appreciated that

TYPING (*microbiol.*): some examples of methods[a]

Method	Essential features of method
AFLP	Digestion of genomic DNA with two types of restriction endonuclease followed by ligation of adaptors to fragments. PCR amplification of those fragments which have certain selective bases complementary to base(s) in the primer. Electrophoresis of amplicons to form the fingerprint
AP-PCR	Use of PCR primers of arbitrary sequence to amplify non-specific targets in genomic DNA; this is followed by gel electrophoresis of products and staining to form the fingerprint
Cleavase fragment length polymorphism analysis	Preparation of single-stranded, end-labeled samples of target DNA from each strain, heat-denaturation, and then cooling to a temperature at which intrastrand base-pairing occurs; this is followed by cleavage with a structure-dependent endonuclease (Cleavase™) and electrophoresis of products to form the fingerprint
Cleaved amplified polymorphic sequences	Comparison of strains on the basis of their CAPS markers (see entry CAPS)
DAF	See entry AP-PCR
DGGE	Comparison of DNA samples by two-phase gel electrophoresis, the samples being separated by size in the first phase and by sequence-based melting (i.e. strand separation) in the second phase
DNA fingerprinting	Restriction enzyme digestion of genomic DNA followed by gel electrophoresis and staining of bands to form the fingerprint; alternatively, blotting of gel to membrane and staining of bands on the membrane
ERIC-PCR	A particular form of rep-PCR (see entry)
MLST	Comparison of nucleotide sequences in specific alleles by direct sequencing
MLVA	Comparison on the basis of certain loci that contain tandemly repeated sequences, strains differing e.g. in the number of repeats at a given locus; relevant loci are copied by PCR, and the amplicons from different strains are compared by gel electrophoresis
PCR-RFLP analysis	RFLP analysis (see below) in which the sample DNA is copied from a primary source by PCR
PCR-ribotyping	Amplification of ISRs (intergenic spacer regions) in the 16S–23S region of the *rrn* operon by PCR and examination of products by gel electrophoresis (cf. ribotyping, below)
PFGE	Specialized electrophoresis of *large* restriction fragments of genomic DNA to form the fingerprint; large fragments are formed by so-called 'rare-cutting' restriction enzymes (such as NotI)
RAPD analysis	See entry AP-PCR
rep-PCR	PCR-based amplification in which primers bind to specific repetitive sequences found in genomic DNA (e.g. the ERIC sequence and the REP sequence) and produce amplicons that reflect the distances between consecutive target sequences; this is followed by gel electrophoresis of the amplicons to form a fingerprint
REP-PCR	A particular form of rep-PCR (see above, and entry)
RFLP analysis	Restriction digestion of DNA samples, gel electrophoresis of products, and comparison of the samples by comparing the number and sizes of fragments produced from each sample; the method detects loss/gain of recognition sites (for the given restriction endonuclease used) as well as insertions/deletions
Ribotyping	Digestion of genomic DNA with a restriction endonuclease, separation of fragments by gel electrophoresis, and then use of DNA or RNA probes, complementary to a sequence in the *rrn* operon, which label specific bands containing the target sequence, thereby generating a fingerprint

(*continued*)

TYPING (*microbiol.*) (*continued*)

Method	Essential features of method
SCC*mec* typing	A method for typing strains of MRSA (methicillin-resistant *Staphylococcus aureus*) based on strain-specific differences in the composition of the SCC*mec* element (see entry SCCMEC)
SERE-PCR	A particular form of rep-PCR (see above, and entry)
Spoligotyping	For members of the *Mycobacterium tuberculosis* complex: PCR amplification of spacers in the genomic DR region, hybridization of the amplicons to a set of oligonucleotide probes representing spacers in a reference strain, determination of the pattern of hybridization (the spoligotype)
SSCP analysis	Preparation of single-stranded samples of the given target sequence from each strain, electrophoresis under non-denaturing conditions, and comparison of strains on the basis of differences in electrophoretic mobility due to differences in the pattern of intrastrand base-pairing
Subtracted restriction fingerprinting	Restriction of genomic DNA with two types of restriction endonuclease, end-labeling of fragments with biotin and digoxigenin, selective capture and elimination of biotin-labeled fragments, and electrophoresis of the remaining fragments to form the fingerprint

[a]See individual entries for further information.

no system of typing can prove that a given unknown strain is *identical* to a particular, known strain. To obtain evidence that one strain is identical to another (a clonal relationship) it would be necessary e.g. to compare the complete genomic nucleotide sequences and also to examine extrachromosomal elements (e.g. plasmids) in both of the strains. (See also entry T5000 BIOSENSOR SYSTEM.)

Both aspects of typing – classification and identification – use similar criteria and methods.

Phage typing and e.g. serotyping (which is based on cell-surface antigenic differences between strains) are still viable methods, but most of the current typing procedures are based on nucleic acid criteria.

Nucleic-acid-based techniques include e.g. MLST, MLVA, RFLP ANALYSIS, RIBOTYPING and SSCP ANALYSIS. (See also PFGE and SCCMEC.)

This account refers primarily to the typing of *bacteria*, but other types of microorganism, including viruses, can also be typed by appropriate methods.

The uses of typing

Some uses of typing are indicated below:

- In epidemiology, typing of strains of a pathogen isolated from a number of patients during an outbreak of disease can help to establish the *chain of infection*, i.e. the succession of patients who have been infected from the original organism. The essential premise here is that all patients in the chain of infection will have been infected by progeny of the same ancestral cell and this should be reflected in the high degree of similarity between strains isolated from patients; a typing procedure can be used to demonstrate such a high degree of similarity.
- Typing has been useful for tracing cross-contamination of clinical specimens in the laboratory.
- Typing of isolates of *Mycobacterium tuberculosis* has been used for estimating the relative importance of new infection versus reactivation of latent strains from previous disease. In this case, data from the strains isolated from earlier disease are compared with data from strains that are causing a current infection.
- In forensic investigations, the ability to distinguish between closely related strains and/or to link a particular strain with a distinct source can be useful in providing evidence in cases of sexual abuse, or in cases involving bioterrorism or food poisoning in which the origin of a given strain is highly relevant.

Nucleic-acid-based typing

This refers to the many methods in which strains are typed on the basis of sequences of nucleotides in their DNA or RNA.

In some cases typing involves comparison of the nucleic acid isolated from different strains. Alternatively, a particular sequence of DNA or RNA may be copied (by methods such as PCR) and typing carried out with the amplicons.

Processes used for comparing the nucleic acid sequences of different strains include (i) direct sequencing of a particular region of DNA; (ii) digestion of samples with a restriction enzyme, and comparison of the (resulting) fragments by gel electrophoresis (RFLP ANALYSIS); (iii) probing for specific restriction fragments (RIBOTYPING); and (iv) examination of samples of single-stranded DNA in order to detect strain-specific intra-strand base-pairing (SSCP ANALYSIS). In some methods a particular genomic feature is used for typing: see e.g. REP-PCR, SCCMEC and SPOLIGOTYPING.

A distinct approach to typing concentrates specifically on the *unstable* regions of the genome, i.e. those regions that are likely to offer the maximum opportunity for detecting differences between strains. These regions include the tandemly repeated short sequences that are present in both prokaryotic and eukaryotic genomes. These regions typically have a high rate of mutation, probably because they are more susceptible (than other regions of the genome) to the disruptive changes caused by events such as SLIPPED-STRAND MISPAIRING and recombination. Hence, regions of the genome such as certain VNTRs (see the entry VARIABLE NUMBER OF TANDEM REPEATS) can be used as a basis for comparison of isolates; thus, these regions can be amplified by PCR, with appropriate primers, and different strains can then be compared by comparing the amplified products. This approach is used in methods such as MLVA.

For further details of a number of nucleic-acid-based typing procedures see individual entries for the methods listed in the table.

Conventional versus nucleic-acid-based typing

Nucleic-acid-based typing differs from 'conventional' forms of typing (such as phage typing) in various ways, not least in the type of equipment and facilities required.

- One important difference is that DNA-based typing has the potential to be carried out *directly* from a clinical or environmental specimen – i.e. without the delay involved in isolating the organism in pure culture. This has been achieved with e.g. *Mycobacterium tuberculosis* and *Neisseria gonorrhoeae* and is clearly a significant advantage in the context of certain infectious diseases.
- A further difference is that, unlike conventional methods, DNA-based typing can be affected by *silent* mutations. Such mutations can affect e.g. the binding of probes and primers – and they can also affect intra-strand base-pairing on which certain DNA-based typing methods depend.
- Another point of difference is that, with some conventional methods, certain strains do not give a result, i.e. they are *untypable* by the given procedure; for example, some strains of *Staphylococcus aureus* are not suitable for phage typing. In DNA-based methods, on the other hand, all such phenotypic aberrations are not problematic because these variations are reflected in genomic sequences. Untypability in DNA-based methods can occur e.g. if the strains under test exhibit no difference (i.e. cannot be distinguished) when a given target sequence is used for typing; this situation can be addressed e.g. by re-examining the strains with a method that involves different target sequence(s), or, in the case of RFLP analysis, by using a different restriction enzyme.

tyramide signal amplification (TSA) A procedure used for the ultrasensitive detection of a target nucleic acid (or protein) *in situ* (i.e. at its normal or relevant location within tissues or cells) – a procedure which is more sensitive than the standard form of *in situ* hybridization (FISH).

As in FISH, a probe can be synthesized as a BIOTINylated oligonucleotide. In TSA, when the probe has bound to its *in situ* target sequence it is detected by a STREPTAVIDIN-linked enzyme: horseradish peroxidase (HRP). HRP acts on (added) molecules of e.g. a fluorophore-labeled tyramide, producing highly reactive but short-lived radicals which bind covalently to local residues (particularly tyrosine residues in proteins) in the immediate vicinity of the bound enzyme, thereby depositing molecules of the label at that site. The intensity of the reaction is increased by adding an HRP-conjugated antibody which is specific for the fluorophore (or other label) already deposited at the target site; this extra HRP provides further activation of (added) labeled tyramide molecules, resulting in further deposition of the label and providing an opportunity for generating an increased level of signal.

Tyramine (*p*-(β-aminoethyl)phenol) can be substituted in a number of ways to produce a range of tyramides.

TSA is also called catalyzed reporter deposition (CARD).

[A quantitative evaluation of peroxide inhibitors for TSA-mediated cytochemistry and histochemistry: Histochem Cell Biol (2006) 126(2):283–291.]

Some examples of use: studies on tumor angiogenesis [Mol Cancer (2009) 8:11]; studies on gene regulation [PLoS Biol (2009) 7(2):e1000029]; studies on lysosomal delivery: Mol Biol Cell (2009) 20(4):1223–1240].

tyramine *p*-(β-aminoethyl)phenol (see also TYRAMIDE SIGNAL AMPLIFICATION).

tyrosine kinase Any (ATP-dependent) enzyme (EC 2.7.1.112) which phosphorylates a specific tyrosine residue in a protein.

Tyrosine kinases are important factors in various forms of signal transfer in mammalian cells; they include membrane-associated enzymes (which are activated via specific signal molecules), and cytoplasmic enzymes mediating intracellular signal transfer.

(See entries ABL, COMPETENT CELLS, EPIDERMAL GROWTH FACTOR RECEPTOR FAMILY, ERLOTINIB, FES, FPS, GEFITINIB, GENISTEIN, IMATINIB, ONCOGENE, NEU, SRC, TYRPHOSTIN.)

(See also JANUS KINASES.)

tyrosine recombinase See RECOMBINASE.

tyrphostin A synthetic, monocyclic compound which strongly inhibits EGF-receptor-associated TYROSINE KINASE activity; unlike GENISTEIN, it does not inhibit the binding of ATP.

U

U-box domain (in CHIP) See the entry CHIP.

U1 adaptor A synthetic oligonucleotide that is used for GENE SILENCING. It binds to: (i) a specific coding sequence in pre-mRNA and (ii) the U1 small nuclear RNA (a component in the SPLICING reaction – see SNRNAS); the *tethering* of these two sites is reported to inhibit polyadenylation and lead to the degradation of a target pre-mRNA [Nature Biotechnol (2009) doi: 10.1038/nbt.1525].

U1–U6 Small nuclear RNAs (see SNRNAS) which are involved in the SPLICING of pre-mRNAs.

UAA *Ochre* codon: see NONSENSE CODON.

UAG *Amber* codon: see NONSENSE CODON.

UAS (upstream activation site) In *Saccharomyces cerevisiae*: a transcriptional ENHANCER which functions only if upstream of the promoter.

uAUG context The nucleotide sequence *upstream* of the AUG (initiator) codon; this phrase has been used as synonymous with the KOZAK SEQUENCE [Proc Natl Acad Sci USA (2009) 106(18):7507–7512].

ubiquitin A polypeptide (of 76 amino acid residues) which is widely distributed in eukaryotic cells but apparently absent in prokaryotic cells.

An energy-dependent linking of ubiquitin molecules to earmarked proteins occurs e.g. prior to degradation of a protein in a PROTEASOME (q.v. for further details). (See also CHIP.)

Ubiquitin–protein linking also occurs in various types of cellular process – including regulation of a protein's function and the intracellular localization of proteins.

(See also SUMOYLATION.)

UDP-glucuronosyltransferase deficiency See (e.g.) the entry CRIGLER–NAJJAR SYNDROME TYPE I.

UDS See UNSCHEDULED DNA SYNTHESIS.

UGA *Opal* codon: see NONSENSE CODON and GENETIC CODE.

In some mRNAs, UGA can specify selenocysteine (instead of acting as a stop codon): see the entry SECIS.

UHRF1 In mammals: an accessory factor which, during DNA replication, recognizes hemimethylated DNA; recognition is apparently necessary to enable DNA methyltransferase 1 to methylate the new strand.

UL97 See NUCLEAR LAMINA.

Ulp1 A SUMO protease: see SUMOYLATION.

Ulp2 A SUMO protease: see SUMOYLATION.

ultrasound (used for gene delivery) Ultrasound (40 kHz) was reported to promote the uptake of a plasmid by the (Gram-negative) bacterium *Pseudomonas putida* [Nucleic Acids Res (2007) 35(19):e129].

(See also PEI.)

ultraviolet absorbance In the spectrophotometric quantitation of nucleic acids in solution: absorbance measured at 260 nm (written A_{260}).

At 260 nm, 1 unit of absorbance corresponds to ~40 µg/mL of ssRNA, to ~33 µg/mL ssDNA or to ~50 µg/mL of dsDNA.

The ratio A_{260}/A_{280} may indicate the purity (homogeneity) of a sample of DNA or RNA.

(See also OLIGREEN, PICOGREEN ASSAY and RIBOGREEN.)

ultraviolet radiation (UVR) (effect on DNA) Electromagnetic radiation of wavelength ~300 nm is absorbed by the bases in nucleic acids, and produces a range of photoproducts; these photoproducts may be lethal to the cell (e.g. blocking DNA synthesis if the damage is not repaired) or they may give rise to MUTAGENESIS.

(See also ASEPTIC TECHNIQUE.)

The UVR-induced photoproducts include various types of dimer (see e.g. PYRIMIDINE DIMER; see also PHOTOLYASE).

In *Escherichia coli* UVR-based *mutagenesis* is promoted through the activity of the so-called SOS SYSTEM. Induction of the SOS system (by damage to DNA) leads to the activity of certain *error-prone* DNA polymerases which, when synthesizing DNA through a lesion (i.e. *translesion* synthesis), incorporate incorrect nucleotides (owing to the lack of local base-pairing specificity) and hence give rise to mutations.

The ultraviolet spectrum

The spectrum of ultraviolet radiation has been divided into three bands, designated UV-A, UV-B and UV-C; however, the wavelengths covered by each band are defined in different ways by different authors. A brief survey of the literature shows that UV-B is defined (according to author) as: (e.g.) 280–315 nm; 280–320 nm; 290–320 nm; and 295–320 nm. Similarly, UV-A has been defined as e.g. 315–400 nm, 320–360 nm and 320–400 nm, while UV-C has been defined as e.g. 190–290 nm and <280 nm.

Some authors do not specify what they mean when using a given designation.

Mutagenic effects of ultraviolet radiation

Mutagenic effects that give rise to carcinomas/melanomas in humans have been linked mainly to wavelengths in the region 280–320 nm. Although radiation in this band represents only a small proportion of total solar UVR reaching the planet, its carcinogenic effect is generally considered to be significantly greater than that of UV-A. Even so, radiation in the 340–400 nm range has been reported to induce pro-mutagenic CPDs (cyclobutyl pyrimidine dimers), and oxidative degradation, in the DNA in cultured human fibroblasts.

(See also MELANOMA and XERODERMA PIGMENTOSUM.)

As a special case, patients on long-term anti-cancer therapy with *thiopurine* base analogs are associated with a high risk of skin cancer from UVR of the longer wavelengths. Thus, 6-thioguanine (6-TG), which accumulates in the DNA of these patients, has an absorption maximum at 340 nm (UV-A, by any definition); the interaction between 6-TG and UVR was reported to generate reactive oxygen species that give rise to irreparable transcription-blocking lesions in the DNA, resulting in mutagenic or lethal damage [Nucleic Acids Res (2009) 37(6):1951–1961].

ULYSIS® Reagents used e.g. for the direct labeling of probes (see the entry PROBE LABELING for details).

[Examples of use: BMC Genomics (2008) 9:496; Neoplasia (2009) 11(3):260–268.]

umber codon See NONSENSE CODON.

umuC See DNA POLYMERASE V.

umuD See DNA POLYMERASE V.

underdominance constructs See GMMS.

***ung* gene** (in *Escherichia coli*) See URACIL-N-GLYCOSYLASE.

UNG method (of decontamination in PCR) See AMPLICON INACTIVATION.

univector See VECTOR.

universal nucleotide Any nucleotide containing a base analog which is able to 'pair' with each of the normal bases found in DNA; typically, the interaction of a universal nucleotide with a normal base does not involve hydrogen bonding but rather a contribution to the stability of the base stacking in dsDNA.

Universal DNA base analogs have been used e.g. at certain locations in primers; in general, they are found to be less useful in positions close to the 3′ end of a primer. These analogs have also been used in probes.

Inosine monophosphate is not a universal nucleotide, but it can pair with either C or U (see entry WOBBLE HYPOTHESIS). The inclusion of inosine monophosphate at the 3′ terminus of primers was reported to give an increase in the diversity of sequences copied from environmental samples [Appl Environ Microbiol (2006) 72(11):6902–6906].

universal primer *Syn*. BROAD-RANGE PRIMER.

universal template PCR assay (UT PCR assay) A form of PCR in which amplification can be detected independently of the target sequence which has been amplified. One example of a UT assay is described in the entry DIGITAL PCR.

unlabeled probe Any PROBE which lacks a label but which can nevertheless provide information on its target sequence.

Unlabeled probes can be used e.g. for detecting variation in a given target sequence. Variation is indicated by differences in T_m (see THERMAL MELTING PROFILE) observed if a given probe duplexes with variant forms of the target sequence, the different probe–target duplexes having different values of T_m. (Variation in T_m is also exploited e.g. in the use of T_m-shift primers – see the example in the entry SNP GENOTYPING.)

In general, a probe that is fully complementary to its target sequence will have a higher T_m when compared with the same probe bound to a variant form of the target (with one or more mismatched base-pairs); the actual difference in T_m between a matched and a mismatched probe – for a particular target sequence – is designated ΔT_m°C.

The inclusion of LOCKED NUCLEIC ACIDS (LNAs) in an unlabeled probe was reported to increase the ability of the probe to indicate probe–target mismatches, i.e. LNAs can increase the ΔT_m.

unscheduled DNA synthesis (UDS) A phrase which is used to refer to DNA synthesis that is carried out *specifically* in order to repair damaged DNA, i.e. excluding routine synthesis that occurs during genomic replication. UDS is assayed e.g. in a diagnostic test for XERODERMA PIGMENTOSUM (q.v.).

UDS has also been assayed e.g. to evaluate the response to UV radiation in cells from a patient with Cockayne syndrome [Proc Natl Acad Sci USA (2009) 106(15):6209–6214].

3′-untranslated region See 3′-UTR.

5′-untranslated region See 5′-UTR.

untwisting enzyme See the entry TOPOISOMERASE (type I).

unwinding protein Any protein which can unwind and separate the strands of dsDNA.

(See HELICASE; cf. SINGLE-STRAND BINDING PROTEIN.)

UP element (in *Escherichia coli*) See PROMOTER.

up mutation In a PROMOTER: any mutation which results in an increased level of transcription from that promoter; this type of mutation may e.g. increase the strength of a promoter by causing change(s) in the –10 or –35 site that make it closer to the CONSENSUS SEQUENCE.

A *down mutation* has the opposite effect.

upper operon (in the TOL plasmid) See TOL PLASMID.

upstream See the entry DOWNSTREAM.

upstream active site See UAS.

***URA3* gene** See the entries SACCHAROMYCES and YEAST TWO-HYBRID SYSTEM.

uracil-containing templates (DNA synthesis) See PFUTURBO.

uracil-DNA *N*-glycosylase *Syn*. URACIL-N-GLYCOSYLASE.

uracil-*N*-glycosylase (UNG) In *Escherichia coli*: an enzyme (the *ung* gene product) which cleaves uracil from an aberrant nucleotide in DNA; this is a part of a normal repair function following the spontaneous deamination of cytosine to uracil. Complete repair of a deaminated cytosine is carried out via the BASE EXCISION REPAIR pathway.

UNG is used in AMPLICON INACTIVATION in PCR (to avoid contamination with amplicons from previous assays).

Unlike uracil, 6-sulfonyluracil (formed during the treatment of DNA with BISULFITE) is resistant to UNG; this sulfonated form of DNA has been used (with the UNG method of decontamination) in a PCR-based study of methylation [Nucleic Acids Res (2007) 35(1):e4].

Overexpression of UNG was reported to inhibit replication of human immunodeficiency virus (HIV-1) [Nucleic Acids Res (2009) doi: 10.1093/nar/gkp673].

uracil-specific excision reagent cloning See USER CLONING.

uridine A RIBONUCLEOSIDE.

USER™ cloning Uracil-specific excision reagent cloning: a method (New England Biolabs, Ipswich MA) that is used for cloning a PCR-generated fragment by inserting the fragment, in a ligation-independent manner, into a linearized vector and then using the vector, directly, to transform the host cells.

For the initial generation of the PCR fragments, each of the primers contains one residue of uracil. The resulting blunt-ended PCR amplicons are treated with a mixture of enzymes: uracil-*N*-glycosylase (which removes the uracil residue from each strand of the (double-stranded) amplicons), and another enzyme which facilitates the release of each single-stranded fragment from its location *5′ of an abasic site*. This enzymic activity produces an 8-nucleotide 3′ overhang at both ends of each amplicon; these two 3′ overhangs are designed to have dissimilar sequences.

The linear vector is prepared as two pieces, each of which has one 3′ 8-nucleotide overhang complementary to *one* of the overhangs on the PCR fragment, thus ensuring directional insertion of the fragment into the vector.

Binding between the overhangs on the PCR fragment and those on the fragments of vector is strong enough to allow the hybridized fragments to be inserted, by transformation, into appropriate cells.

One problem with this method arose during generation of the PCR fragments. Thus, the type of (proof-reading) DNA polymerase used for error-free replication of DNA stalls at uracil residues in template strands of DNA. *Taq* DNA polymerase was used – but the absence of 3′-to-5′ exonuclease (proof-reading) activity in this enzyme precludes its use for high-fidelity cloning of DNA. However, the proof-reading enzyme *PfuTurbo*® C_x Hotstart DNA polymerase (Stratagene, La Jolla CA) was subsequently found to be suitable for use in USER™ methodology [Nucleic Acids Res (2006) 34 (18):e122].

The method has been modified by the use of a one-piece cloning vector with (dissimilar) 8-nt 3′ overhangs. The target molecules are prepared with a single deoxyuridine (dU) near the 5′ end of each PCR primer so that the resulting amplicons can be processed to products having 3′ overhangs compatible with those of the vector (thus allowing directional assembly). According to the vector's sequence the procedure can be used for various purposes – such as directional cloning, insertions and deletions. (Target molecules can be generated by primers in which the location of the dU is varied, as required, to suit the 3′ overhangs on other components that may be introduced as inserts etc.) [Nucleic Acids Res (2007) 35(6):1992–2002.]

[Example of use: Glycobiology (2008) 18(10):799–805.]

A derivative method, which employs the USER™ enzyme system and T4 DNA ligase, has been devised for inserting a gene of interest into a linear DNA template in a way suitable for *in vitro* transcription and translation. Essentially, the gene and the two end-pieces are each amplified by PCR, and the three types of fragment are then fitted together, in the correct sequence, by the USER approach; the nicks are sealed by the T4 DNA ligase. A short PCR then amplifies the (three-piece) assembled products, but does not amplify partially assembled products – thereby effectively removing the latter. [Method: Nucleic Acids Res (2009) doi: 10.1093/nar/gkp589.]

UT PCR assay See UNIVERSAL TEMPLATE PCR ASSAY.

3′-UTR (in a eukaryotic mRNA) The 3′-untranslated region: a sequence of nucleotides between the coding region and the poly(A) tail; it includes the signals for polyadenylation (see AAUAAA) and commonly includes the ZIPCODE.

In the so-called selenoprotein mRNAs the 3′-UTR region includes a SECIS element.

3′-UTRs are target sites for many, or most, of the known MICRORNAS; human microRNAs were reported preferentially to target AT-rich 3′-UTRs.

(See also HAMMERHEAD RIBOZYME and RACE.)

5′-UTR (leader sequence; leader) The 5′ untranslated region of an mRNA molecule: that region between the 5′-end of the molecule and the start of the coding sequence.

In prokaryotes, the 5′-UTR includes the *Shine–Dalgarno sequence* which base-pairs with a specific sequence in ribosomal RNA, thereby helping to locate the mRNA correctly for translation. Some 5′-UTRs specify a small peptide (the *leader peptide*) and include an *attenuator sequence*, both of which contribute to regulation of the encoded gene(s) (see the entry OPERON).

Most eukaryotic mRNAs, as well as some viral RNAs, have a CAP at the 5′ terminus.

Studies designed to copy the 5′ end of mRNAs use methods such as RACE and CAP TRAPPER. (See also CAGE.)

UV-A, **UV-B**, **UV-C** Categories of ULTRAVIOLET RADIATION.

UVR ULTRAVIOLET RADIATION.

UvrABC-mediated repair (nucleotide excision repair; short patch repair) An EXCISION REPAIR system which recognizes and repairs DNA damaged by ULTRAVIOLET RADIATION or by certain types of chemical; the kinds of damage repaired by this system include THYMINE DIMERS, photo-products – such as as Pyr(6–4)Pyo – and some types of chemically modified nucleotide.

Some individuals are deficient in UvrABC-mediated repair and, as a result, are highly sensitive to ultraviolet radiation: see e.g. XERODERMA PIGMENTOSUM.

In *Escherichia coli*, proteins encoded by genes *uvrA*, *uvrB* and *uvrC* constitute the UvrABC enzyme system, also called *ABC excinuclease*.

UvrABC-mediated repair appears to occur in the following way. Dimerized UvrA ($UvrA_2$) interacts with either one or two molecules of UvrB, thus forming a $UvrA_2B$ or $UvrA_2B_2$ complex. This complex appears to translocate along the helix (in an ATP-dependent manner) to the exact site of damage. UvrA then dissociates, leaving UvrB bound to DNA. UvrC now binds to the UvrB–DNA complex, and this new complex (specifically UvrC) mediates initial cleavage of the strand at a phosphodiester bond 3 or 4 nucleotides 3′ of the damaged site. [Studies on the first incision by UvrC: EMBO J (2005) 24(5):885–894.] The next cleavage of the strand occurs 7 or 8 nucleotides 5′ of the damaged site. The sequence of nucleotides containing the lesion, together with UvrC, is displaced by UvrD (= helicase II), and the single-stranded gap is filled by DNA polymerase I, which also displaces UvrB.

In *Escherichia coli* the excised region is only ∼10 nucleotides in length (hence the name 'short patch repair').

UvrD (*syn*. helicase II) See UVRABC-MEDIATED REPAIR.

UvsX enzyme (of phage T4) See RPA (sense 1).

V

V (1) A specific indicator of ambiguity in the recognition site of a RESTRICTION ENDONUCLEASE (or in any other sequence of nucleic acid); 'V' indicates A or C or G.
(2) L-Valine (alternative to Val).

V factor A growth requirement for certain species of HAEMOPHILUS: NAD^+ or $NADP^+$ (or nicotinamide mononucleotide). V factor is available in yeast extract, and is also available in heat-lysed erythrocytes (red blood cells).
(cf. X FACTOR.)

v-*onc* An unspecified viral ONCOGENE.

v-*onc*$^+$ The designation of a virus or viral genome containing an ONCOGENE.

v-*onc*$^-$ The designation of a virus or viral genome which lacks an ONCOGENE.

V5 epitope A fusion tag encoded by some types of expression vector; it consists of the following sequence of amino acids: Gly-Lys-Pro-Ile-Pro-Asn-Pro-Leu-Leu-Gly-Leu-Asp-Ser-Thr (which is recognized by an anti-V5 monoclonal antibody). It facilitates detection of the FUSION PRODUCT (i.e. the tagged, recombinant protein) by an anti-V5 monoclonal antibody.

The V5 tag and biotin were used in CHROMATIN IMMUNOPRECIPITATION studies as an alternative to the use of antibodies (raised against the target proteins) [BMC Mol Biol (2009) 10:6].

[Other uses: J Exp Clin Cancer Res (2009) 28(1):106; PLoS Pathog (2009) 5(6):1000493; Proc Natl Acad Sci USA (2009) 106(29):11989–11994.]

vaccine Any preparation administered, either orally or parenterally, in order to stimulate the recipient's protective immunity in respect of specific pathogen(s) and/or toxin(s). Until quite recently, the only vaccines available consisted of suspensions of killed or attenuated pathogens or of specific antigen(s), but it has now been established that protective immunity can be acquired from vaccines consisting primarily of the *DNA* that encodes the relevant antigen(s): see DNA VACCINE.

Recombinant forms of conventional types of vaccine can be made by genetically engineering (live) organisms – including attenuated bacterial pathogens; on infection, these organisms exhibit one or more specific types of protective antigen.

valproic acid 2-Propylpentanoic acid, an agent used e.g. as an inhibitor of histone deacetylases.
(See also HDAC.)

vancomycin A complex glycopeptide antibiotic, active mainly against Gram-positive bacteria, that inhibits synthesis of the cell-wall macromolecule peptidoglycan (by blocking transfer of a subunit from the cell membrane to the periplasm).

variable number of tandem repeats (VNTR) A phrase which refers to genomic loci consisting of tandem repeat sequences of nucleotides – the number of tandem repeats varying, at a given locus, in different individuals and/or differing from the number at other VNTR loci in the same genome.

For some authors, 'VNTRs' involve any type of tandemly repeated unit, such as MICROSATELLITE DNA or minisatellite DNA, but other authors equate the phrase specifically with a particular type of repeated unit.

VNTRs are associated with high rates of mutation. This is apparently due to the relative ease with which events such as SLIPPED-STRAND MISPAIRING and/or recombination may take place in these sequences. Studies on 28 VNTR loci within the genome of *Escherichia coli* strain O157:H7 reported single-locus rates of up to 7×10^{-4} mutations per generation, some of the observed mutations being consistent with a slipped-strand mispairing model [J Bacteriol (2006) 188(12):4253–4263].

A number of (separate) VNTRs in a particular genome may be exploited e.g. for TYPING isolates of microorganisms, and may also be used for distinguishing between DNA samples from eukaryotic/mammalian individuals; this form of typing is termed *multiple loci VNTR analysis* (see MLVA). This kind of approach to typing requires prior identification of suitable VNTRs in genomic DNA [see e.g. BMC Microbiol (2006) 6: 44].

VNTRs which consist of a *single type* of nucleotide (a so-called *single-nucleotide repeat*, or SNR) are associated with very high rates of mutation; they have been used for typing isolates of *Bacillus anthracis* (the causal agent of anthrax) [J Clin Microbiol (2006) 44(3):777–782].

variant antigenic type (of *Trypanosoma*) See VSG.

variant surface glycoprotein (of *Trypanosoma*) See VSG.

VariFlex™ The trade designation (Stratagene, La Jolla CA) for various products which are designed to facilitate studies on protein expression – including enhancement of solubility (see SOLUBILITY ENHANCEMENT TAG) and quantitation of soluble proteins (see Q-TAG, sense 1).

vector (*DNA technol.*) Typically, a PLASMID, construct or virus into which a gene (or nucleic acid fragment) is inserted prior to TRANSFECTION (sense 1); in general, a vector is a vehicle used for transporting nucleic acid into cells.

(See e.g. CLONING VECTOR, EXPRESSION VECTOR; see also DENDRIMER, GENE-DELIVERY SYSTEM, NANOPARTICLES and the TARGETRON.)

For other entries relating to the insertion of DNA into cells see e.g. ENDO-PORTER, LIPOFECTION, MELITTIN, MAGNETOFECTION, MICROINJECTION, NUCLEOFECTION, PROTAMINE SULFATE and SPINOCULATION.

The type of vector used in any given case depends at least partly on the size of the given gene or fragment to be transfected. Various types of vector are used for genes/fragments in the size range ~10 kb to ~1 Mb: see the entry CLONING for some examples.

In host cells, some types of vector replicate autonomously, i.e. as extrachomosomal elements (independent replicons that are not integrated into the host cell's chromosome(s)); this is the case with the typical cloning vector and some expression vectors. In other cases, a vector (with its *insert* – i.e. nucleic acid fragment or gene) integrates into the host's chromosome and is replicated as part of the chromosome.

Plasmid vectors are usually covalently closed (i.e. circular) molecules of DNA, the molecule remaining covalently closed after the gene or fragment has been inserted. However, *linear* vectors are used in some cases – e.g. in the BIGEASY LINEAR CLONING SYSTEM. (See also USER CLONING.)

[Plasmid vector design and production (a review): Gene Therapy (2009) 16:165–171.]

The Ti plasmid (see CROWN GALL) can be used as a vector for introducing DNA into plant cells (see AGROINFECTION).

Viral vectors incorporate insert DNA in their (recombinant) genomes; this is introduced into cells during viral infection. Commercial virus-based vectors include e.g. AAV HELPER-FREE SYSTEM, ADEASY XL ADENOVIRAL VECTOR SYSTEM and VIRAPORT RETROVIRAL GENE EXPRESSION SYSTEM.

(See RETROVIRUSES for comments on replicating and non-replicating viral vectors.)

Phages and phagemid particles can function as vectors in prokaryotic cells (see e.g. LAMBDA FIX II VECTOR). This kind of vector has also been used in some eukaryotic cells: see e.g. GENE THERAPY.

In some cases a viral vector is modified so that it can infect only particular type(s) of cell: see PSEUDOTYPING.

Bacteria have been used to transfer genes into mammalian cells: see section *Vectors in gene therapy* in the entry GENE THERAPY. Bacteria can also transfer genes to e.g. fungi: see the entry INSERTIONAL MUTAGENESIS.

In some cases a gene/nucleic acid fragment is inserted into a small intermediate vector (a *univector*) that is not *directly* used for transfection; the univector is inserted into a second vector that is inserted into the host cells [see BioTechniques (2005) 39(3):301–304].

(See also BINARY VECTOR SYSTEM.)

Some vectors do not contain the insert (gene/nucleic acid). For example, an *insertional targeting vector* may be used to introduce an insertion site (for subsequent use) into a specific sequence of genomic DNA. Such an engineered insertion site can include e.g. recognition sequences for a given recombinase – so that target DNA (in a second vector) can be inserted into the given site by RECOMBINASE-MEDIATED CASSETTE EXCHANGE [Proc Natl Acad Sci USA (2005) 102(18):6413–6418].

In order to carry out their function, some vectors require specialized (engineered) *host cells*. For example, cells with a *lacZ*Δ*M15* mutation are used for the blue–white screening of recombinant PBLUESCRIPT vectors. Moreover, some vectors are inherently unstable in certain types of host cell: see e.g. lentiviral vectors in the entry CLONING VECTOR.

Some vectors encode a peptide tag which forms a FUSION PRODUCT with the protein of interest – facilitating the subsequent detection or isolation of the expressed proteins (see e.g. PESC VECTORS and also SOLUBILITY ENHANCEMENT TAG).

(See also entries DESTINATION VECTOR, GAPPED VECTOR, GENE FUSION VECTOR, INSERTION VECTOR, REPLACEMENT VECTOR, SHUTTLE VECTOR and SUICIDE VECTOR.)

Some examples of destination vectors are given in the table in GATEWAY SITE-SPECIFIC RECOMBINATION SYSTEM.

Checking for the presence of an insert in a vector

The procedures used for inserting DNA into a vector molecule often involve insertion of the gene or fragment into a site created by cleavage of the vector with a particular restriction enzyme. In some of these procedures, vector molecules may simply re-circularize, without including an insert; accordingly, methods have been devised to check for the presence of an insert in the vector. In some methods, cells that receive vector molecules (e.g. by electroporation or lipofection) can signal the presence or absence of the insert when they form colonies on certain media: see e.g. *SUP4* gene in YEAST ARTIFICIAL CHROMOSOME. A different method involves so-called blue–white screening (see the entry PBLUESCRIPT).

Vector NTI Advance™ 9.1 Sequence analysis software: a tool for BIOINFORMATICS marketed by Invitrogen (Carlsbad CA, USA) which is compatible with Windows® PCs.

The software includes a number of modules that permit e.g. creation, mapping and analysis of sequences; sequence alignment of nucleic acids and proteins; the analysis of genomic sequences; and motif mapping of nucleic acids and proteins, as well as annotation.

There is also a facility for Gateway® cloning *in silico*.

vectorette PCR (bubble-linker PCR) A form of ANCHORED PCR which is used to copy an unknown sequence of nucleotides adjacent to a specific sequence.

In this technique, the double-stranded linker (*vectorette*), which contains the anchor sequence, includes a central (i.e. non-terminal) region in which the bases in the two strands are *not* complementary; the lack of hybridization between these two regions of the vectorette results in a 'bubble' that consists of the two unpaired single-stranded regions.

As in the conventional form of anchored PCR, one of the two primers (primer A) has a binding site within the known sequence of the sample DNA.

The second primer (primer B) consists of a sequence which is *identical* (not complementary) to one of the two single-stranded bubble regions in the vectorette; this primer therefore cannot bind to *any* sequence in the vectorette and, at the start of PCR, has no function as a primer.

PCR starts when DNA synthesis is initiated from primer A; the resulting strand contains a sequence complementary to one of the two 'bubble' sequences in the vectorette – and *this* sequence can act as a binding site for primer B. PCR then continues as normal, with primers A and B.

With this arrangement, PCR products are formed only if amplification is initiated from primer A, i.e. the occurrence of amplification depends on the presence of a suitable binding site for primer A.

verocytotoxins Shiga-like toxins: see EHEC.

vertical transmission (of genes) Transmission of gene(s) from one individual or cell to its offspring/progeny.

(cf. HORIZONTAL TRANSMISSION.)

vesiculostomatitis virus See VSV.

Vibragen Omega A veterinary product (used for canine parvovirus infections) synthesized as a recombinant protein within a (live) silkworm host (see BACULOVIRUS EXPRESSION SYSTEMS).

Vibrio cholerae A species of Gram-negative bacteria; many of the (lysogenic) strains can cause cholera. The cholera toxin (genes *ctxA* and *ctxB*) is encoded by a bacteriophage (phage CTXΦ). (Genes for cholera toxin can also be acquired via the filamentous phage VGJφ through a form of specialized transduction [J Bacteriol (2003) 185:7231–7240].)

The *V. cholerae* genome consists of two dissimilar circular chromosomes that differ, slightly, in GC% (see also the table in the entry BACTERIA). The smaller of the two chromosomes is reported to begin replication later in the cell cycle than the larger chromosome, so that replication of both chromosomes is completed at about the same time [EMBO J (2007) 26(13): 3124–3131].

Vibrio fischeri (in quorum sensing) See QUORUM SENSING.

Vif protein (of HIV-1) See the entry RETROVIRUSES.

(See also VLP.)

vIL-6 A form of IL-6 (interleukin-6) encoded by HHV-8 (the human herpesvirus-8).

violacein A purple pigment: see QUORUM SENSING.

viral vectors See VECTOR.

ViraPort® retroviral gene expression system A commercial GENE-DELIVERY SYSTEM (Stratagene, La Jolla CA) suitable for expressing a given gene of interest in any of a wide range of target cells.

In the ViraPort® system the gene of interest is inserted into an expression vector which is based on elements of Moloney murine leukemia virus (MMLV). This gene-carrying vector is then inserted into suitable 'packaging' cells *together with* two other vectors: one vector carrying a retroviral *gag–pol* region (encoding the viral structural proteins and reverse transcriptase) and another containing the *env* region (encoding viral surface proteins). In these cells the transcript of the gene of interest is packaged by products encoded by the other two vectors – to form infectious (but replication-deficient) viral particles.

Virions formed in, and released from, the packaging cells are used for transfecting *target* cells. Within the target cells the transcript is reverse transcribed (using the reverse transcriptase synthesized in the packaging cells) and the DNA is integrated into the target cell's genome. The gene of interest can then be transcribed and expressed in the target cell. (As the target cells do not contain the viral structural genes, they cannot produce virions.)

By using vectors that encode different types of *env* protein, the virions which are assembled by this process can be made suitable for transfecting different types of target cell – e.g. human, mouse, rabbit – because the products encoded by this vector determine the type(s) of cell which can be infected by the virions.

As the genes for the viral components are dispersed among several vector molecules, and as there is little or no sequence overlap among the three vectors, there is minimal possibility of (undesirable) production of replication-competent virions through a recombination process.

[Uses (e.g): gain-of-function study [BMC Genomics (2009) 9:254]; study on the attachment site of a coronavirus [J Virol (2009) 83(2):1026–1035.]

(cf. the AAV HELPER-FREE SYSTEM and VIRAPOWER LENTIVIRAL EXPRESSION SYSTEM.)

ViraPower™ lentiviral expression system A system (Invitrogen, Carlsbad CA, USA) used for gene expression in a wide range of mammalian cells, including both dividing and non-dividing cells.

The plasmid vector (pLenti), into which the gene of interest is inserted, includes retroviral 5′ and 3′ LTR sequences (see RETROVIRUSES), a ψ (psi) site (which promotes packaging of viral RNA) and a gene encoding resistance to blasticidin or to Zeocin™.

The retroviral *gag–pol* and *rev* sequences, and a sequence encoding the G protein of the vesiculostomatitis virus (VSV) (see PSEUDOTYPING), are provided on separate plasmids (in the 'packaging mix').

The vector (containing the insert) and the plasmids of the packaging mix are then used jointly to transfect the cells of a producer cell line (293FT). In these cells, the transcript of the gene of interest is packaged by products encoded by the other plasmid vectors; this forms infectious, replication-deficient virions which can be stored or used to transfect target cells.

[Use (e.g.) studies on the effect of apoptotic human cells on granulocyte migration: J Clin Invest (2009) 119(1):20–32.]

(cf. the AAV HELPER-FREE SYSTEM and VIRAPORT RETROVIRAL GENE EXPRESSION SYSTEM.)

virion A single virus particle.

viroid Any of a variety of small, circular (covalently closed) molecules of single-stranded RNA, each of several hundred nucleotides in length, which are associated with symptomatic infection of plants; many types of viroid cause economically important diseases. Some examples of viroids include e.g. the *Citrus* exocortis viroid (CEV), coconut cadang-cadang viroid (CCCV), hop stunt viroid (HSV), potato spindle tuber viroid (PSTV), tomato bunchy top viroid (TBTV) and the tomato 'planta macho' viroid (TPMV).

Transmission of viroids may be mechanical (for example, it may occur during grafting procedures) or may occur via seed or pollen.

Viroid structure, apparently similar in all viroids, includes double-stranded regions (due to intramolecular base-pairing) and single-stranded (unpaired) loops.

A viroid may localize in the nucleus or in the chloroplast. In at least some cases, replication initially involves synthesis of complementary (minus) strands several times longer than the wild-type viroid.

Homology between viroids and group I introns lead to the suggestion that these agents may have derived from the (self-splicing) introns.

The viroid-pattern HAMMERHEAD RIBOZYME has been used

in various genetically engineered applications.

A database for viroids (and other forms of subviral RNA) is available at:
http://subviral.med.uottawa.ca

virtual RLGS See the entry RESTRICTION LANDMARK GENOMIC SCANNING.

virulent phage See BACTERIOPHAGE.

virulon In species of the Gram-negative bacterium *Yersinia*: a plasmid-encoded system which essentially comprises a range of virulence proteins and a transport channel for transferring these proteins into the cytoplasm of a eukaryotic target cell, following direct cell–cell contact.

Similar virulence systems have been found e.g. in species of *Salmonella* and *Shigella*.

virus-like particle See VLP.

visna virus See LENTIVIRINAE.

vitamin H *Syn*. BIOTIN.

vitamin K_3 2-Methyl-1,4-naphthoquinone: see MENADIONE.

Vitravene An oligonucleotide antisense therapeutic agent (ISIS Pharmaceuticals), approved by the FDA (USA) in 1998, that is used e.g. for treating cytomegalovirus (CMV) retinitis in AIDS patients; this 21-nucleotide product (administered by direct injection into the eye) binds to viral mRNA and blocks replication of the virus.

VLF-1 (in baculoviruses) Very late expression factor 1: see the entry BACULOVIRIDAE.

VLP Virus-like particle: any construct used as a surrogate for a virus – e.g. a VLP comprising the major capsid protein (L1) of a human papillomavirus (HPV) has been used in place of HPV in a vaccine.

[Protective immunity by a VLP vaccine of influenza virus: J Virol (2007) 81(7):3514–3524.]

A construct of twelve copies of a pentameric capsid protein from adenovirus has been used as a vehicle for increasing the uptake, by cells, of the antitumor agent bleomycin – the agent being tethered (linked) to this construct [PLoS ONE (2009) 4(5):e5569].

A VLP was used to study the inhibition of the viral protease inhibitor 3-*O*-(3′,3′-dimethylsuccinyl)betulinic acid by HIV-1 Vif protein [Virol J (2008) 5:162].

VNTR See VARIABLE NUMBER OF TANDEM REPEATS.

von Gierke's disease GLYCOGEN STORAGE DISORDER type Ib.

Voyager™ vectors Vectors (marketed by Invitrogen, Carlsbad CA) which, when expressed in mammalian cells, give rise to the target protein fused with the VP22 structural protein of the herpes simplex virus type 1; the fusion protein is then able to translocate to adjacent, non-transfected cells in a cell culture – within which it localizes in the nucleus. In this way, the recombinant protein can be delivered to ~100% of cells in a culture following the initial transfection procedure; thus, the protein can be delivered to those cells which had not been transfected by the Voyager vector.

The ability of the fusion protein to be translocated from the synthesizing cell to an adjacent, non-transfected cell is conferred by the VP22 fusion partner.

The target protein can be expressed as either a C-terminal or an N-terminal fusion to the VP22 partner.

All of these vectors incorporate e.g. a T7 promoter, a CMV promoter, selectable markers and a 6×His sequence (the latter allowing purification of the protein by a NICKEL-CHARGED AFFINITY RESIN).

Some of these vectors include a facility for topoisomerase-mediated insertion of a gene/fragment (see TOPOISOMERASE I CLONING).

VP22 A structural protein of herpes simplex virus 1 (HSV-1); it is able to translocate in a Golgi-independent, cytochalasin D-sensitive manner from a given mammalian cell to an *adjacent* cell within a cell culture. This property has been exploited by using the VP22 gene to partner the gene of a target protein in VOYAGER VECTORS; the fusion protein has the same ability as VP22 to translocate between mammalian cells.

Vpu protein (of HIV-1) See the entry RETROVIRUSES.

vRLGS See RESTRICTION LANDMARK GENOMIC SCANNING.

VSG (variant surface glycoprotein) In some species of the protozoan *Trypanosoma*: a component of the cell-surface layer that exhibits extensive ANTIGENIC VARIATION; such variation involves periodic switching from one *VSG* gene to another to produce a VSG of different antigenic specificity (thus helping the organism to evade the immune response of a *mammalian* host). An individual trypanosome may have ~1000 different VSG-specifying genes.

(Note that *T. brucei*, the causal agent of sleeping sickness, replaces its VSG layer with a PROCYCLIN surface layer when it is in the gut of the tsetse fly vector.)

Only one *VSG* is normally expressed at a given time (joint expression of two has been reported in a mutant strain). For expression, a *VSG* must be located within a specialized *expression site* situated within a telomere; there may be 20 or more of these sites within a genome. Different individuals expressing different *VSGs* are said to be different *variant antigenic types* (VATs).

The switching of *VSGs* appears to take place in at least two distinct ways. One involves transfer of transcription from one expression site to another (containing a different *VSG*).

The alternative mode involves recombination between one of the (many) genomic *VSGs* and a *VSG* in an expression site, possibly by homologous recombination and gene conversion.

Antigenic variation in African trypanosomes was reported to involve the histone methyltransferase DOT1B; deletion of this function caused e.g. persistent expression of two *VSGs*, simultaneously, for weeks [PLoS Biol (2008) 6(7):e161].

Reducing the number of *VSG* transcripts to 1–2% of their normal value by RNA INTERFERENCE was reported to induce cell cycle arrest, blocking cell division; it was suggested that this may protect the *trypanosome* – i.e. blocking cell division may prevent harmful 'dilution' of the (protective) VSG layer in the absence of VSG synthesis.

The mechanism of antigenic variation via the *var* genes of the malaria parasite, *Plasmodium falciparum*, appears to have certain differences in comparison with the antigenic variation

system in *Trypanosoma* [Eukaryotic Cell (2007) 6(9):1511–1520].

VSV Vesiculostomatitis virus: an enveloped ssRNA-containing virus of the family Rhabdoviridae; it can infect a wide range of animals, including mammals and birds, and it can replicate in many types of cultured cell.

The G (envelope) protein of VSV has been widely used e.g. for PSEUDOTYPING because it confers on a virion the ability to infect a wide range of types of cell.

VT1, VT2 Verocytotoxins 1 and 2: see EHEC.

VTEC Verocytotoxic *Escherichia coli*, an alternative name for strains of enterohemorrhagic *E. coli*: see EHEC.

W

W (1) A specific indicator of ambiguity in the recognition site of a RESTRICTION ENDONUCLEASE (or in any other sequence of nucleic acid); for example, in C↓CWWGG (enzyme StyI), the 'W' indicates A or T. (In RNA 'W' indicates A or U.)
(2) L-Tryptophan (alternative to Trp).

W-Beijing strain (also called Beijing/W) A strain of *Mycobacterium tuberculosis*, found particularly in Asia and in the former Soviet territories, associated e.g. with drug resistance and failure of anti-tuberculosis treatment. The genome of this strain is characterized by spoligotype (see SPOLIGOTYPING); the origin of replication was reported to contain a copy of the insertion sequence IS*6110*.

Rapid PCR-based detection of this strain has been reported in which a characteristic 641-bp PCR product was obtained by amplifying a sequence at the mycobacterial interspersed repetitive unit (MIRU) locus 26 within the *M. tuberculosis* genome [J Clin Microbiol (2006) 44(1):274–277].

[Other reports of PCR-based detection: J Clin Microbiol (2006) 44(2):302–306, and (2007) 45(3):1022–1023.]

W reactivation See WEIGLE REACTIVATION.

WAS See WISKOTT–ALDRICH SYNDROME.

WASP See WISKOTT–ALDRICH SYNDROME.

Watson–Crick model See B-DNA.

Weigle mutagenesis See WEIGLE REACTIVATION.

Weigle reactivation (W reactivation) A phenomenon in which certain phages (including phage λ) which have been damaged by ULTRAVIOLET RADIATION exhibit higher rates of survival in host cells which have been irradiated with UVR prior to infection with the phage (compared with non-irradiated host cells).

Growth of the phage in the irradiated host cells also results in higher rates of mutagenesis in the phage (*Weigle mutagenesis*).

The phenomenon can be partly explained by the presence of repair enzymes (including error-prone DNA polymerases) in the irradiated cells. However, other factor(s) appear to be involved because Weigle reactivation depends on irradiation of host cells even when they are mutant strains in which the SOS genes (including those encoding error-prone DNA polymerases) are constitutively active in the absence of UVR.

Werner syndrome A rare, human autosomal recessive disorder involving various features of age which develop prematurely (i.e. in the young), including graying/loss of hair but also e.g. cancer, cataracts etc. and/or other serious conditions. Werner syndrome results from the loss of a functional WRN protein, a RecQ-type DNA helicase encoded by gene *WRN* (chromosome location: 8p12–p11.2).

Various studies have reported on the activity of the Werner helicase [Nucleic Acids Res (2007) 35(17):5706–5716; PLoS ONE (2009) 4(3):e4673].

(See also RECQ PROTEIN.)

Western blot analysis *Syn*. IMMUNOBLOTTING.

Western blotting A procedure, analogous to SOUTHERN BLOTTING, in which *proteins* are transferred from a gel to a matrix using e.g. electrically driven transfer.

If the proteins on the matrix are probed by anti-protein *antibodies*, the procedure is generally referred to as *Western blot analysis* or IMMUNOBLOTTING (q.v.).

(See also NORTHERN BLOTTING.)

WGA (1) WHOLE-GENOME AMPLIFICATION.
(2) Wheat germ agglutinin.

WGD Whole-genome duplication.

Whipple's disease A malabsorption syndrome in which there is e.g. diarrhea, steatorrhea, lymphadenopathy and involvement of the central nervous system; typically there is an infiltration of the intestinal mucosa by macrophages which contain PAS-positive material (i.e. material that gives a positive test in the periodic acid–Schiff reaction).

Whipple's disease has been associated with infection by the Gram-positive pathogen TROPHERYMA WHIPPLEI.

whole-animal cloning See CLONING (whole-animal cloning).

whole-genome amplification (WGA) Any of certain methods that attempt to amplify genomic DNA, typically by copying randomly primed sequences; one such method is MULTIPLE DISPLACEMENT AMPLIFICATION.

Note. The phrase 'whole-genome amplification' is somewhat misleading. It does not mean, literally, that the entire genome is replicated (as *in vivo*). Rather, it refers to those techniques in which primers are used for the simultaneous amplification of different parts of sample DNA – (theoretically) the whole genome being available as a template.

WGA with bioinformatically optimized primers is reported to give better results than MDA with degraded genomic DNA [DNA Res (2006) 13(2):77–88].

WGA of bisulfite-treated DNA has been found to be a useful approach when studying cytosine methylation with small amounts of DNA [BioTechniques (2006) 41(5):603–607].

Wigglesworthia A genus of Gram-negative bacteria that occur as endosymbionts within specialized cells (*bacteriocytes*) in the anterior midgut of the tsetse fly, *Glossina*.

(cf. BUCHNERA, SODALIS and WOLBACHIA.)

wild-type strain (1) Any strain (of a given organism) whose characteristics are those typical of strains that are common in the natural environment.

The strains used in genetic engineering are often selected, or modified, to have specific deficiencies or other abnormal features (see e.g. COMPETENT CELLS).

(2) A strain which does *not* have specified mutant feature(s).

Wilms' tumor (oncomir overexpression) See ONCOMIR.

Wiskott–Aldrich syndrome (WAS) An (X-linked) immunodeficiency disorder that is associated with thrombocytopenia (i.e. low levels of platelets), eczema, and a susceptibility to recurrent infection. Hemorrhage is a major feature.

The WAS protein (WASP), a mutant cytoskeletal regulator, was reported to bind to the platelet CIB protein that normally binds Ca^{2+} and integrins.

wobble hypothesis An hypothesis (proposed by Crick) relating to the observed degeneracy of the third base of a CODON; the hypothesis provides an explanation of the way in which some tRNAs can recognize several codons that differ in their third position.

In the wobble hypothesis, base-pairing occurs normally between the first two bases in the codon (i.e. bases in positions 1 and 2, 5′-to-3′) and the corresponding bases in the anticodon, i.e. bases in positions 3 and 2, respectively (reading 5′-to-3′).

By contrast, the base in position 1 of the anticodon (i.e. the so-called 'wobble position') can base-pair in a 'non-Watson–Crick' manner with the base in position 3 of the codon; thus, a G in the anticodon's wobble position may pair with either C or U in the codon's position 3, and U (in the wobble position) may pair with either A or G.

Certain modified bases can occur in the wobble position of the anticodon. For example, A appears to occur invariably in a deaminated form (i.e. as inosine). In the wobble position, inosine can pair e.g. with C or U.

(See also QUEUOSINE.)

While the wobble hypothesis is generally applicable, other factors (such as the conformation of the tRNA's anticodon) may influence function.

wobble position See WOBBLE HYPOTHESIS.

Wolbachia A genus of Gram-negative bacteria which are found intracellularly or extracellularly in various invertebrate hosts (e.g. *W. pipientis* in certain mosquitoes; *W. persica* in ticks of the genera *Argas*, *Dermacentor* and *Rhipicephalus*; *W. melophagi* in the gut lumen of sheep keds (*Melophagus ovinus*); and certain strains in the reproductive tissues of the tsetse fly, *Glossina*).

Wolbachia is also an endosymbiont in the filarial nematode *Onchocerca volvulus* – causal agent of human onchocerciasis (manifested e.g. as 'river blindness'); chemotherapy with the tetracycline doxycycline depletes *Wolbachia* and, as a result, causes sterility in the nematode, thus blocking transmission of the disease.

Wolbachia has been considered as a 'gene drive system' in the development of genetically modified mosquitoes: see the entry GMMS.

(See also BUCHNERA, SODALIS and WIGGLESWORTHIA.)

***WRN* gene** See WERNER SYNDROME.

WRWYCR A peptide that inhibits junction-resolving enzymes at a HOLLIDAY JUNCTION (q.v.).

wyosine See Y BASES.

X

X chromatin *Syn.* BARR BODY.

X chromosome In human cells: one of the two sex chromosomes, the other being the Y chromosome; the normal female is characterized by an XX complement, while the normal male is characterized by an XY complement.

In females, in any given cell, one of the X chromosomes is inactivated; this reflects an event early in embryonic development: see X-INACTIVATION.

(See also BARR BODY.)

Genetically based disorders resulting from abnormal alleles on an X chromosome are called X-LINKED DISORDERS.

(See also DNA SEX DETERMINATION ASSAY and SEX SORTING (of spermatozoa).)

X factor A requirement for the aerobic growth of some species of HAEMOPHILUS. The requirement can be satisfied by hemin or protoporphyrin IX; the X factor is needed for the synthesis of cytochromes and catalase etc.

(cf. V FACTOR.)

X family (of DNA polymerases) A group of DNA-DEPENDENT DNA POLYMERASES that include prokaryotic, eukaryotic and viral enzymes – e.g. mammalian pol β, λ, and μ, and template-independent terminal deoxynucleotidyl transferase; *Sc*Pol4 and *Sp*Pol4 in yeasts; and bacterial enzyme *Bs*PolX.

The X family polymerases are involved in repair processes (e.g. in BASE EXCISION REPAIR and NON-HOMOLOGOUS DNA END-JOINING).

[Structural insight into biological functions of the X family polymerases: DNA Repair (Amst) (2007) 6(12):1709–1725.]

Note on the nomenclature of X family polymerases

There is little adherence to any basic format for naming the X family polymerases. Some use the approach shown above – e.g. an enzyme from bacterium *Bacillus subtilis* is designated *Bs*PolX. However, an X family enzyme from this species was designated PolX_{Bs} [Nucleic Acids Res (2008) 36(18):5736–5749], while one from *Thermus thermophilus* was designated ttPolX [Nucleic Acids Res (2009) 37(6):2037–2052].

(See also A FAMILY, B FAMILY and Y FAMILY.)

X-gal 5-bromo-4-chloro-3-indolyl-β-D-galactoside: a substrate that is hydrolysed by enzyme β-galactosidase to a blue-green product; bacteria forming this enzyme give rise to blue-green colonies when grown on an agar medium containing X-gal.

(See also BACTERIAL TWO-HYBRID SYSTEM, PBLUESCRIPT and REPORTER GENE.)

Bluo-gal is a related compound that produces a darker blue coloration; it has uses similar to those of X-gal.

Both X-gal and Bluo-gal are soluble in dimethylformamide.

X-inactivation A (normal) process that occurs in the development of embryos of female mammals in which one of the X chromosomes (i.e. from the maternal or the paternal source) is silenced (made transcriptionally inactive) in a proportion of the cells of the embryo; the other X chromosome (i.e. from the other parental source) is silenced in the remaining cells of the embryo.

An X chromosome destined to be silenced expresses a non-coding RNA, called Xist, that coats the chromosome and also recruits other factors; this blocks transcriptional activity of genes on that chromosome. (Compare IMPRINTING CONTROL REGION.) (See also XIC.)

Following X-inactivation, the inactivated X chromosome (designated Xi) remains inactivated in successive rounds of cell division.

The active (transcribed) X chromosome is designated Xa.

[High-resolution analysis of epigenetic changes associated with X-inactivation (in murine cells): Genome Res (2009) 19:1361–1373.]

X-linked disorder (sex-linked disorder) Any genetic disorder resulting from an abnormal allele on an X CHROMOSOME; an example is hemophilia.

An X-linked disorder may be dominant or recessive.

Males inheriting an X-linked disorder which is recessive in the female parent exhibit the abnormal trait because (having only one X chromosome) they lack the corresponding normal X-linked allele. Females who have a single normal allele may be asymptomatic, but will be carriers of the condition.

Mating between an affected male and an unaffected female will not transmit the trait to a son: male-to-male transmission does not occur because a son inherits only the Y chromosome from his father.

X4 strains (of HIV-1) See HIV-1.

Xa The designation of the *active* X chromosome in a cell of a female mammal (see X-INACTIVATION).

Xaa The designation sometimes used to indicate an unknown amino acid residue in a polypeptide.

xanthine The product of the oxidative deamination of guanine.

xenotransplantation Transplantation (of tissues or organs etc.) across the species barrier.

xeroderma pigmentosum (XP) A genetically based, autosomal recessive disorder which is characterized by extreme photosensitivity and a tendency to develop skin cancers. Most XP patients are deficient in UVRABC-MEDIATED REPAIR, while a minority are deficient in other repair functions.

One approach to the diagnosis of XP involves assessment of DNA repair capability in cells that have been pre-exposed to DNA-damaging UV radiation (usually at 254 nm); control cells (from individuals who do not suffer from XP) exhibit a given level of DNA repair activity, while the cells from XP sufferers typically show a marked reduction in repair activity. After UV challenge, the cells are incubated for a fixed period of time in a medium which includes a labeled nucleoside or a nucleoside analog (such as BRDU); the cells are then assayed for the amount of label incorporated in their DNA. The assay measures *unscheduled DNA synthesis* (UDS); this refers to the DNA synthesis required to repair the UV-induced damage – that is, it excludes DNA synthesis associated with normal (S-phase) DNA replication. In order to measure *only* UDS, (i.e. to exclude the uptake of label by normal DNA

replication), non-S-phase cells may be selected under a microscope; this is facilitated e.g. by maintaining the cells on coverslips throughout the diagnostic procedure.

In one type of XP diagnostic test, ^{3}H-labeled thymidine is used to assess DNA repair synthesis by a *liquid scintillation* procedure – involving generation of light when a scintillant material is exposed to a radioactive source.

One (accurate) form of XP test involves the measurement of ^{3}H-thymidine incorporation using autoradiography of cells on coverslips. A less-accurate, but less time-consuming, form of the test measures incorporation of the nucleoside analog BrdU; here, measurement of incorporated BrdU requires the denaturation of DNA and the use of an anti-BrdU antibody to detect the incorporated analog.

A method that is quicker and more sensitive than the BrdU approach exploits 5-ethynyl-2′-deoxyuridine (EdU – a nucleoside analog) to assess DNA repair activity. EdU incorporated in DNA is detected by an azide-conjugated dye, Alexa Fluor® 488, which binds covalently to the EdU. (This approach thus avoids the need for denaturation of DNA and the requirement for antibodies.) Using this technique, a standard UDS assay was reported to be finalized within half a day [Nucleic Acids Res (2009) 37(4):e31].

(See also COCKAYNE SYNDROME and MELANOMA.)

Xgal See entry X-GAL (under 'X-', above).

Xi The *inactive* X chromosome (see X-INACTIVATION).

Xic X-inactivation center: a regulatory region on an X chromosome which is involved in X-INACTIVATION. It includes the gene for Xist.

***xis* gene** (phage λ) See the entry PHAGE LAMBDA.

Xist A non-coding RNA which is involved in silencing one of the two X chromosomes during X-INACTIVATION.

XL1-Blue A general cloning strain of *Escherichia coli* (Stratagene, La Jolla CA, USA) designed to permit e.g. blue–white screening and single-strand rescue of phagemid DNA; a *lacI*^{q} gene (enhancing synthesis of the repressor protein of the LAC OPERON) allows tight control of *lac*-controlled gene(s).

XL1-Blue MRF′ See the entry RESTRICTION-MINUS CELLS.

XL1-Red A MUTATOR STRAIN of *Escherichia coli* – see entry for details of genotype and examples of use.

XP A common abbreviation for XERODERMA PIGMENTOSUM.

Xpress™ A fusion tag (Invitrogen, Carlsbad CA), encoded by certain expression vectors, which facilitates detection of the fusion product; the tag, Asp-Leu-Tyr-Asp-Asp-Asp-Asp-Lys, is detected by the anti-Xpress™ monoclonal antibody.

[Example of use: RNA (2009):15(1):33–43.]

XSCID (SCIDXI) X-linked SCID (severe combined immunodeficiency): see SCID in GENETIC DISEASE (table).

***xseA* gene** See EXONUCLEASE VII.

***xthA* gene** See EXONUCLEASE III.

Y

Y (1) A specific indicator of ambiguity in the recognition site of a RESTRICTION ENDONUCLEASE (or in any other nucleic acid sequence); for example, in GTY↓RAC the 'Y' indicates C or T. (In RNA 'Y' indicates C or U.)

(2) L-Tyrosine (alternative to Tyr).

Y bases Certain modified forms of guanine found in molecules of tRNA; in these bases an additional ring is fused to the purine skeleton. One example is *wyosine*. Y bases occur e.g. in the tRNA for phenylalanine in bacteria and yeasts.

(cf. Q BASES.)

Y-box-binding protein See YB-1 PROTEIN.

Y chromosome In human cells: one of the two sex chromosomes (found in males), the other being the X CHROMOSOME (q.v.).

(See also entries AMELOGENIN, AMPLICON (sense 5), DNA SEX DETERMINATION ASSAY, HETEROSOME, SEX SORTING.)

Y family (of Alu sequences) See ALU SEQUENCES.

Y family (of DNA polymerases) DNA polymerases which can continue to synthesize DNA through lesions (they are said to 'bypass' lesions), incorporating incorrect nucleotides opposite those lesions. These 'error-prone' or 'translesion' polymerases occur in all three domains of life; they include, for example, the *Escherichia coli* DNA polymerases IV and V (which are active in the SOS response to DNA damage); human DNA polymerases pol η (eta), pol ι (iota), and pol κ (kappa); and archaeal DNA polymerases Dpo4 (in *Sulfolobus solfataricus*) and Dbh (in *Sulfolobus acidocaldarius*). (See also 8-OXOG.)

The translesion DNA polymerases are also found in other categories – e.g. DNA polymerase ζ (zeta) of *Saccharomyces cerevisiae* (an enzyme of the B FAMILY).

(See also DNA-DEPENDENT DNA POLYMERASE and TRANSLESION SYNTHESIS.)

(cf. A FAMILY and X FAMILY.)

Y family DNA polymerases have been reported to enhance survival in *Escherichia coli* when DNA replication is inhibited by nucleotide starvation brought about by treatment with hydroxyurea; this is apparently a role distinct from the ability of the Y family polymerases to bypass the effects of DNA damage [EMBO J (2006) 25(4):868–879].

Thermostable polymerases of the Y family were reported to be useful for PCR amplification of damaged or ancient DNA (and may have applications e.g. in forensic science) [Nucleic Acids Res (2006) 34(4):1102–1111].

(See also FORENSIC APPLICATIONS.)

Y family polymerases were suggested to mediate mutagenesis by incorporating oxidized nucleotides into nucleic acids [J Bacteriol (2006) 188(13):4992–4995].

Translesion DNA polymerases act in combination to bypass lesions, the enzymes involved in a given case depending on the nature of the DNA damage [EMBO J (2009) 28(4):313–314.]

(See also REVERSE GYRASE.)

Y-linked meiotic drive system (gene drive system) See GMMS.

Y-STR See FORENSIC APPLICATIONS (*Types of investigation*).

Y2H See YEAST TWO-HYDRID SYSTEM.

Y3H See YEAST THREE-HYBRID SYSTEM.

YAC YEAST ARTIFICIAL CHROMOSOME.

YB-1 protein Y-box-binding protein 1: a protein that is multifunctional and is characterized by its interactions with nucleic acids. It is present in informosomes (see entry MRNP) and is involved in the translation of mRNA. In the nucleus, it acts as a Y-box-binding transcription factor which can activate genes associated with e.g. growth, differentiation and apoptosis.

YCp YEAST CENTROMERE PLASMID.

yeast When used without qualification, 'yeast' generally refers specifically to the species *Saccharomyces cerevisiae* (a very common experimental organism) – cf. entry YEASTS.

(See also entry SACCHAROMYCES.)

yeast artificial chromosome (YAC) An *in vitro* construct that can carry a large target gene/insert (up to ∼1 Mbp) and which can replicate within yeast cells (commonly *Saccharomyces cerevisiae*) in a way that generally resembles the behavior of a small chromosome (including e.g. division at mitosis and meiosis). YACs have been used for CLONING genes and other sequences of DNA which are too large for cloning in other kinds of vector; they have also been used e.g. for expressing certain genes in mice (see later under subheading).

(See also the entries BACTERIAL ARTIFICIAL CHROMOSOME and HUMAN ARTIFICIAL CHROMOSOME.)

The following account refers primarily to the preparation of a YAC for use in *yeast* cells, in particular the strain AB1380 of *S. cerevisiae* (see SACCHAROMYCES for details).

The minimal structural requirements for a YAC are: (i) an autonomous replicating sequence (see ARS) to allow ongoing autonomous intracellular replication; (ii) a centromere sequence for mediating separation of chromatids etc. during the stages of division; and (iii) TELOMERE sequences, at each end of the (linear) construct, to maintain the integrity of the YAC during successive rounds of cell division.

A YAC may be constructed as a circular molecule which is subsequently cut, with RESTRICTION ENDONUCLEASES, into two separate linear pieces: the so-called left and right arms.

Each arm has a telomere sequence at *one* end. The gene of interest/insert is ligated between a left arm and a right arm – leaving the telomere sequences in terminal positions. A YAC – excluding the insert – is only a few hundred base-pairs in length.

For cloning, the circular molecule is cut, at several sites, to form the left and right arms. During this process, a section of the circular molecule (the 'stuffer') is excised – leaving a telomeric sequence at *one* end in both the left and right arms. A mixture – containing the left and right arms of the YAC, together with copies of the gene/insert to be cloned – is then incubated with a ligase.

The product required from the ligation step is a construct

consisting of the gene/insert of interest flanked on one side by a YAC left arm and on the other side by a YAC right arm. (Following incubation, the reaction mixture will also include unwanted constructs – e.g. YACs containing no insert and inserts flanked by two left arms or two right arms; unwanted constructs are automatically de-selected – see below.)

Products from the ligation step are inserted into SPHEROPLASTS of *S. cerevisiae* by transformation. (Yeast cells with intact cell walls are not used.)

YACs are designed to facilitate the selection of those yeast cells that have been transformed with the required construct (i.e. insert flanked by left and right arms). For this purpose, use is made of specific, modified strains of *S. cerevisiae*. One such strain is AB1380 (see the entry SACCHAROMYCES for details). A YAC designed for use in strain AB1380 contains a functional (wild-type) *TRP1* gene in one arm and a functional *URA3* gene in the other arm. Hence, only those cells which have received constructs containing *both* arms of the YAC will be selected (i.e. will grow) on media lacking tryptophan and uracil. However, although selecting for both arms of the YAC, this procedure does not select positively for those cells in which the construct includes an insert.

To select for the presence of an insert, the YAC is designed to include a functional *SUP4* suppressor gene which converts the mutant (red) cell (see strain AB1380 in entry SACCHAROMYCES) to a (wild-type) colorless cell. This gene contains an EcoRI restriction site (a cloning site) within which the insert can be ligated. An insert in this site inactivates gene *SUP4*; hence, a cell which contains an insert – as well as both arms – will grow with the (mutant) red coloration, while a cell with an insert-free YAC (i.e. containing a functional *SUP4* gene) will form (wild-type) colorless colonies.

YAC transgenes in mice

Many studies have shown that genes present in YACs can be expressed in (transgenic) mice. In at least some cases, a transgene is expressed in a way similar or identical to the way in which it is expressed in the original host species. Thus, for example, an early study [Proc Natl Acad Sci USA (1993) 90:7593–7597] reported that the human β-globin gene, when encoded by a 248-kb YAC, was expressed appropriately in transgenic mice. A YAC mouse model was also used e.g. for studies on Huntington's disease [Neurobiol Dis (2008) 31(1):80–88].

yeast centromere plasmid (YCp) A (low-copy-number) ARS- and centromere-containing expression vector used in yeasts.

yeast episomal plasmid (YEp) A (high-copy-number) plasmid based on the TWO-MICRON PLASMID.

Yeast Gene Order Browser (YGOB) An online tool used for comparative genomic studies on hemiascomycetous yeasts.

[Syntenic relationships in hemiascomycetes using YGOB: Nucleic Acids Res (2006) 34(Database issue):D452–D455.]

yeast genetic marker See YEAST MARKER.

yeast integrating plasmid (YIp) A vector which integrates into yeast chromosomal DNA.

yeast marker (yeast genetic marker) Any chromosomal gene of a yeast which can act as a selectable marker when present in a construct (e.g. a YEAST ARTIFICIAL CHROMOSOME) that is used for transforming cells.

Marker genes in common use include e.g. those encoding essential products involved in amino acid biosynthesis (e.g. *HIS3*, *LEU2*) and in the biosynthesis of uracil (e.g. *URA3*). Thus, e.g. insertion of a YAC containing a functional (wild-type) *HIS3* into an auxotrophic strain of yeast – in which the corresponding gene is inactivated – will permit growth (and, hence, selection on appropriate minimal media) of those cells which have taken up the YAC.

yeast three-hybrid system (tribrid system; Y3H) An experimental system, based on the YEAST TWO-HYBRID SYSTEM, that can monitor *in vivo* interactions involving three proteins.

In this system, one of the vectors encodes the extra protein – whose expression is regulated by controlling the conditions of the experiment.

When expressed, the third protein may interact with prey and/or bait proteins in various ways. Thus, for example, it may stabilize prey–bait binding or it may modify one or both proteins (e.g. by phosphorylation) to active or inactive forms, thereby promoting or inhibiting interaction.

As in the two-hybrid system, suitable reporter genes signal the occurrence of specific interactions.

Expression of the third protein is made inducible in order to be able to distinguish between duplex reactions (between the bait and prey proteins) and triplex reactions (that involve all three proteins). One inducible system involves a methionine-activated promoter; however, in this case, the pre-induction conditions (depletion of methionine) may have an inhibitory effect on the growth of the test organism (which could affect the results of the three-hybrid assay). One approach to avoiding this problem is to use tetracycline-regulated expression of the third protein (which has no obvious effect on the yeast's phenotype, or on the gene expression pattern) [Nucleic Acids Res (2009) 37(2):e11].

yeast two-hybrid system (Y2H) A technique for the detection of protein–protein interactions *in vivo* (within yeast cells). (See also BACTERIAL TWO-HYBRID SYSTEM and CYTOTRAP.)

This technique exploits the fact that some transcription activators can be physically split into two inactive domains: a DNA-binding domain (DBD) and an activation domain (AD) – both of which are necessary in the functional transcription activator.

The DBD has a binding site for a promoter-related region, while AD is required for initiation of transcription, from that promoter, by RNA polymerase II.

A functional transcription activator can also be formed if a fusion protein (i.e. *hybrid* protein – see GENE FUSION) which *includes* DBD binds to a fusion protein that *includes* AD. In this technique, combination (i.e. binding) occurs between the pair of *fusion partners* – i.e. between the protein that is fused to DBD and the protein that is fused to AD; these two fusion partner proteins, whose interaction is under study, are called the 'bait' and 'prey' proteins respectively.

Given bait–prey binding, transcription is initiated from the corresponding promoter, which regulates a REPORTER GENE; transcription of the reporter gene indicates that a functional transcription activator is present – that is, that the two fusion partner proteins have interacted physically.

The fusion proteins are encoded on vector molecules that are introduced into cells; fusion proteins are expressed within the cells.

Various types of reporter gene can be used.

The occurrence or otherwise of interaction between the bait and prey proteins can be indicated by either positive selection or negative selection.

For positive selection, cells of *Saccharomyces cerevisiae* can be engineered so that their growth depends, absolutely, on the occurrence of the bait–prey interaction. For example, the endogenous *HIS3* gene can be inactivated; thus, without the corresponding gene product (HIS3) the cells cannot grow on a particular selective medium. A copy of the *HIS3* gene, inserted into the cells (as a reporter gene) in the two-hybrid system, will therefore allow cells to grow on the selective medium *if* bait–prey interaction occurs – i.e. if the reporter gene (*HIS3*) is transcribed and expressed.

For negative selection (counterselection) in the cells of *S. cerevisiae*, the endogenous *URA3* gene can be deleted and replaced by *URA3* on a reporter cassette – expression of *URA3* then being governed by the bait–prey interaction. Bait–prey interaction promotes expression of *URA3* (which encodes orotidine-5′-phosphate decarboxylase) – with consequent activation of the pathway for uracil biosynthesis. Under these conditions, the addition of 5-fluoro-orotic acid (5-FOA) to the medium results in the formation of a suicide substrate, so that no growth occurs. (This type of approach may be useful e.g. for selecting proteins that are *defective* in interaction.)

One potential problem is *self-activation*: transcription of the reporter gene without initial bait–prey interaction (yielding a false-positive result). This problem can be addressed e.g. by testing for self-activating proteins before assay, or examining putative positives following removal of the prey-encoding plasmid by a selective process.

[Resolving the network of cell signaling pathways using the evolving yeast two-hybrid system (review): BioTechniques (2008) 44(5):655–662.]

The yeast two-hybrid system was found to permit quantitative detection of protein interaction over a greater range than the bacterial system while the latter had less self-activation (= autoactivation). [Combined yeast/bacteria two-hybrid system: Mol Cell Proteom (2005) 4:819–826.]

[Examples of use of yeast two-hybrid systems: PLoS ONE (2009) 4(4):e5117; J Biol Chem (2009) 284(14):9489–9497; Plant Physiol (2009) 149(3):1478–1492.]

[Y2H in system biology (a review): Int J Mol Sci (2009) 10(6):2763–2788]

yeasts A convenient, but non-taxonomic, category of fungi that includes organisms which are *typically* unicellular, fermentative and saprotrophic, and which reproduce asexually by the process of budding. *Saccharomyces cerevisiae* is an example of a typical yeast.

Many organisms that are included in the category *yeasts* do not exhibit *all* the typical features. For example, the species *Candida albicans* can grow in a mycelial form (as well as in a unicellular form – i.e. it is a *dimorphic* fungus); moreover, this species can also be pathogenic.

The species *Schizosaccharomyces pombe* is a 'fission yeast' – i.e. the cells divide by fission (rather than by budding).

Some yeasts (e.g. *Hansenula canadensis*, *Lipomyces* spp) are non-fermentative.

Some otherwise 'typical' yeasts (e.g. *Pichia pastoris*) can carry out a type of metabolism (methylotrophy) which is not common in this category of organisms.

The category *yeasts* is based on morphological and physiological criteria rather than on taxonomic criteria.

yellow fluorescent protein (YFP) A fluorescent protein which is widely used e.g. as a partner in fusion proteins, acting as a reporter system. (See GENE FUSION.) YFP is also used e.g. in BIMOLECULAR FLUORESCENCE COMPLEMENTATION.

An enhanced form of YFP (eYFP, or EYFP) – analogous to enhanced green fluorescent protein (see EGFP) – is also used [e.g. Biochem J (2009) 418(3):529–540; PLoS ONE (2009) 4(1):e4286].

(Some related entries: GREEN FLUORESCENT PROTEIN, RED FLUORESCENT PROTEINS.)

YEp YEAST EPISOMAL PLASMID.

YFP YELLOW FLUORESCENT PROTEIN.

YGOB YEAST GENE ORDER BROWSER.

YIp YEAST INTEGRATING PLASMID.

YOYO-1® A dimeric cyanine fluorophore (from Molecular Probes/Invitrogen, Carlsbad CA, USA) used e.g. for DNA STAINING; it is a bis-intercalator which is reported to intercalate between GC pairs preferentially. On excitation (e.g. with an argon-ion laser at 488 nm), the dye fluoresces at 509 nm.

YOYO-1® has been reported to promote the formation of TRIPLEX DNA [PLoS ONE (2007) 2(3):e305].

YOYO-1® has been used to stain erythrocytes (red blood cells) in order to detect infection by *Plasmodium* (the malaria parasite) – such staining detecting *plasmodial* DNA (because RBCs have no DNA) [PLoS Pathog (2009) 5(4):e1000366].

YOYO-1® has been used for imaging Rad51 on individual molecules of DNA [Proc Natl Acad Sci USA (2009) 106(2): 361–368].

YTH system YEAST TWO-HYBRID SYSTEM *or* YEAST THREE-HYBRID SYSTEM.

YycFG system See TWO-COMPONENT REGULATORY SYSTEM.

Z

Z-DNA Left-handed helical dsDNA, with ∼12 bases/turn, first observed in sequences such as . . .GCGCGCGC. . . under some *in vitro* conditions (such as high concentrations of salt); the sugar–phosphate backbone forms an irregular zig-zag.

Some proteins (e.g. the dsRNA adenosine deaminase) have a site, designated Zα, which can interact specifically with the Z conformation of nucleic acids, including hybrid DNA/RNA duplexes as well as dsDNA and dsRNA.

In vivo, Z-DNA has a role e.g. in the transcription process. One study demonstrated that the dinucleotide repeat sequence $d(CG)_9$ – located upstream of the TATA box – can enhance transcription by forming a Z-DNA secondary structure in a promoter-specific way, the enhancement being influenced by the distance between $d(CG)_9$ and TATA box [Proc Natl Acad Sci USA (2007) 104(7):2229–2234].

Studies on Zα-bound Z-DNA are generally carried out with CG repeats. However, information from a study on three non-CG-repeat DNA sequences has suggested that conformation, rather than sequence, is an important factor in the recognition of the Z-DNAs by Zα [Nucleic Acids Res (2009) 37(2):629–637].

Z ring A ring-shaped structure, composed of molecules of the FtsZ protein, which develops mid-way along the length of a dividing bacterial cell prior to the formation of the septum; it marks the plane of the forthcoming septum.

(See also GENE FUSION (uses).)

The polymerization of FtsZ proteins is inhibited in the SOS response to damaged DNA (see SOS SYSTEM).

Zα A site, on a protein, that interacts with Z-DNA.

zalcitabine See NUCLEOSIDE REVERSE TRANSCRIPTASE INHIBITORS.

zanamivir A sialic acid analog used as an ANTIVIRAL AGENT; it inhibits viral NEURAMINIDASE in types A and B influenza virus.

(See also *swine flu* in the entry INFLUENZAVIRUS.)

ZAP Express® vector A vector (Stratagene, La Jolla CA) that incorporates features of both phage and plasmid, and which is suitable for expression or cloning in bacteria and eukaryotes.

This (linear) vector can carry inserts of up to 12 kb.

Essentially, the ZAP Express® vector consists of a 4.5-kb PHAGEMID (pBK-CMV) bracketed by two PHAGE LAMBDA sequences, each having a terminal single-stranded *cos* region. One of the lambda sequences contains phage genes A→J that encode head, tail and morphogenesis. *Immediately* flanking the phagemid (i.e. between the phagemid and each end-piece) are two short sequences: the *initiator* sequence on one side of the phagemid, and the *terminator* sequence on the other side of the phagemid; these two sequences can be recognized by a helper phage (see later).

Phagemid pBK-CMV includes the ColE1 (plasmid) origin of replication, and the f1 (phage) origin of replication. It also includes a *lac* promoter (from the LAC OPERON), inducible by IPTG and used for gene expression in bacteria, and the CMV promoter for gene expression in eukaryotes.

The multiple cloning site (MCS) occurs, in the phagemid, within a sequence encoding part of the *lacZ* gene – enabling blue–white screening (see PBLUESCRIPT). The twelve cloning sites in the MCS allow directional insertion.

The MCS is bracketed by T3 and T7 promoters, in opposite orientation, allowing *in vitro* transcription of the insert.

Selection of (prokaryotic and eukaryotic) vector-containing cells is facilitated by antibiotic-resistance genes (neomycin or G418 sulfate).

Using the vector

The gene of interest/cDNA etc. is inserted into the MCS (in the phagemid).

Zap Express® vectors, containing inserts, are packaged e.g. in GIGAPACK components to form complete, infective phage particles. The phages are used to infect an appropriate strain of *Escherichia coli* – within which these vectors replicate as plasmids.

If required, the insert can be expressed in *E. coli* by using the agent IPTG to activate transcription from the *lac* promoter.

Instead, when the Zap Express® vector is intracellular, the phagemid (pBK-CMV) portion can be excised from the Zap Express® vector. This is achieved by infecting the cells with a (filamentous) helper phage. A product of this helper phage recognizes the initiator and terminator sequences (see above), which flank the phagemid, and a new strand of DNA (which corresponds to the phagemid, and its insert) is synthesized; this strand displaces the existing strand in the phagemid. The displaced strand is initially circularized and is then packaged in protein components encoded by the helper phage. Phagemid particles, containing single-stranded DNA (that includes the insert), are then secreted from the living cells in the form of filamentous phages.

If required, the single-stranded copy of the phagemid vector pBK-CMV (in the phagemid particles) can be converted to a double-stranded plasmid form by using the particles to infect a strain of *E. coli;* the strain of *E. coli* used is one containing an F plasmid because, in such strains, the cells have a PILUS (q.v) which is needed for the process of infection. Conversion to the plasmid form occurs within the cells.

If a *library* has been inserted into Zap Express® vectors, a common requirement is to screen for a given product that can be expressed in the *E. coli* host cells. This can be done e.g. by using the library-containing phage particles to infect a LAWN PLATE of *E. coli* in the presence of the inducing agent IPTG. Each PLAQUE (one of many) that is formed on the lawn plate will contain (i) progeny phage particles, and (ii) the protein product encoded by the insert in that specific clone of phages. The plaques on a given plate are 'blotted' onto a nitrocellulose membrane and then examined by a labeled probe (such as a specific, labeled antibody). If a positive plaque is found, the phages in that plaque can be collected and amplified (i.e. by replication in *E. coli*); the insert in this clone of phages may

then be isolated and sequenced.

The ZAP Express vector has been used e.g. for studies on biosynthesis of anandamide [Proc Natl Acad Sci USA (2006) 103(36):13345–13350]; on *Humulus lupulus* (the hop plant) [BMC Genomics (2008) 9:445]; on the frog (*Xenopus laevis*) [Proc Natl Acad Sci USA (2008) 105(7):2469–2474]; studies on a protein from *Plasmodium falciparum* (malaria parasite) [Parasitol Res (2009) 104(6):1389–1397]; and for studies on a predictor of recurrence-free survival in breast cancer [PLoS Med (2009) 6(5):e1000068].

ZBTB4 protein See KAISO PROTEIN.

ZBTB28 protein See KAISO PROTEIN.

Zea mays Maize (= corn). For a database on maize genetics and genomic see MAIZEGDB.

ZEBRA (*syn*. Zta) An activator of replication of the EPSTEIN–BARR VIRUS (EBV): a protein, encoded by (immediate-early) viral gene *bzlf-1*, which initiates replication of EBV within latently infected cells. In latently infected cells, ZEBRA can be induced by treating the cells with e.g. corticosteroids, anti-immunoglobulin, or the compound TPA (a phorbol ester: 12-*O*-tetradecanolyphorbol-13-acetate).

A co-activator of the EBV lytic cycle in latently infected cells is the transcription factor Rta (encoded by the viral gene *brlf-1*).

Studies on activation of the lytic cycle, using inhibitors of DNA methyltransferases and inhibitors of HDACs, reported that hyperacetylation of histones in the vicinity of the *bzlf-1* and *brlf-1* promoters was not, in itself, sufficient to activate the expression of the viral lytic-cycle genes [J Virol (2008) 82(10):4706–4719].

The switch from latent to lytic infection of EBV in cells is reported to be associated with sensitization of those cells to killing by NK cells [J Virol (2007) 81(2):474–482].

zebrafish A small teleost (bony) fish (*Danio rerio*) with a short generation time; it is a common experimental animal.

The formation of microRNAs (miRNAs) in (living) cells of zebrafish was studied by dual-fluorescence reporter/sensor plasmids [BioTechniques (2006) 41(6):727–732]. (See also microRNA-9 in the entry MICRORNAS.)

MORPHOLINO ANTISENSE OLIGOMERS are frequently used for KNOCKDOWN studies on the zebrafish.

In vivo gene-targeted mutations were introduced into zebrafish genomic DNA by the use of zinc-finger nucleases [PLoS ONE (2009) 4(2):e4348].

Temporally controlled site-specific recombination in zebrafish has been carried out by ligand-dependent expression of the Cre recombinase [PLoS ONE (2009) 4(2):e4640].

(See also MITOCHONDRIAL DNA.)

Zenon® antibody-labeling reagents Products from Invitrogen (Carlsbad CA, USA) used for the rapid (~10-minute) labeling of IgG antibodies. These reagents are labeled Fab fragments – a Fab fragment being one of the two arms of the (Y-shaped) immunoglobulin molecule; they bind to the Fc part of an IgG antibody, the Fc region being the stem of the Y-shaped molecule. Fab fragments may be labeled e.g. with a fluorophore or with BIOTIN, or with an enzyme such as HORSERADISH PEROXIDASE.

[Uses (e.g.): Mol Biol Cell (2006) 17(6):2722–2734; Proc Natl Acad Sci USA (2007) 104(9):3460–3465; J Clin Invest (2007) 117(1):258–269; J Biol Chem (2009) 284(10):6495–6506; PLoS ONE (2009) 4(3):e4827.]

Zeocin™ An antibiotic (Invitrogen, Carlsbad CA, USA) of the bleomycin/phleomycin group which cleaves DNA. Zeocin™ is an effective agent against eukaryotic cells (including yeast and mammalian cells) and also against at least some types of bacteria.

Resistance to Zeocin is conferred by a gene whose product blocks the binding of Zeocin to DNA.

[Examples of use of Zeocin™ as a selective agent: Cancer Res (2009) 69:369–378 (selection of B16 murine melanoma cells with 500 μg/mL); Neoplasia (2009) 11(1):77–86 (selection of human glioblastoma cells with 200 μg/mL).]

Zero Blunt® A cloning kit (Invitrogen, Carlsbad CA) used for cloning blunt-ended PCR products which have been prepared by amplification with a proofreading DNA polymerase. The Zero Blunt kit includes a control plasmid, a linearized 3.5 kb cloning vector, and T4 DNA ligase.

The vector includes an antibiotic-resistance gene, encoding resistance to either kanamycin or Zeocin™; two recognition sites for EcoRI (flanking the insertion site) for excision of the insert; and primer sites for sequencing. For positive selection, it also contains a copy of gene *ccdB* (see CCD MECHANISM) fused downstream of a *lacZ*α sequence, the cloning site being within the *lacZ*α sequence.

A TOPO® version of the kit is also available. In this kit, the blunt-ended fragments are inserted into the vector by the activity of topoisomerase I (see the entry TOPOISOMERASE I CLONING for basic details of topo cloning).

Zero Blunt® has been used (e.g.) in studies on a murine splice variant [PLoS ONE (2009) 4(3):e4986]; on molecular divergence in *Anopheles cruzii* populations [Malaria J (2009) 8:60]; and on a recombinant vaccinia virus [Mol Med (2009) 15(5–6):144–151].

ZFN (*DNA technol*.) ZINC-FINGER NUCLEASE.

zidovudine (AZT) 3′-azido-3′-deoxythymidine, a NUCLEOSIDE REVERSE TRANSCRIPTASE INHIBITOR used e.g. in the treatment of AIDS; its mode of action is similar to that of other NRTIs.

AZT undergoes glucuronidation in the liver; drugs which inhibit glucuronidation may influence plasma levels of AZT.

Side-effects from the drug include certain types of anemia.

Ziehl–Neelsen stain A stain used for detection of ACID-FAST BACILLI in e.g. smears of sputum. The dye used in this staining method (carbolfuchsin) is red.

ZiFDB Zinc finger database – see the entry ZINC FINGER.

ZIFT Zygote intrafallopian transfer: see the entry IVF.

zinc finger A type of motif which is common in DNA-binding proteins (e.g. certain transcription factors); it contains a zinc-binding site. Different families of zinc-finger proteins differ e.g. in structure and in the type(s) of zinc-binding amino acid

residue; cysteine residues are commonly involved, sometimes in conjunction with histidine residues. Characteristically, an α-helix component of a zinc-finger protein binds to the major groove in DNA.

(See also METHYLATION, TAGM, ZINC-FINGER NUCLEASE.)

[Zinc finger database (ZiFDB): a repository for information on C2H2 zinc fingers, and on engineered zinc-finger arrays: Nucleic Acids Res (2009) 37(Database issue):D279–D283.]

Type II myotonic dystrophy is caused by repeat expansion of MICROSATELLITE DNA (the CCTG sequence) in intron 1 of the zinc-finger gene *ZNF9*.

zinc-finger nuclease (ZFN) A composite, engineered nuclease used for generating gene-specific KNOCK-OUT MUTATIONS *in vivo* (i.e. within living cells). A ZFN is a dimer in which each monomer consists of a zinc-finger array (often three/four zinc fingers) fused to the (non-specific) cleavage domain of a type IIS restriction endonuclease, FokI (see the table in the entry RESTRICTION ENDONUCLEASE); thus, the binding specificity, for a given target site, is provided by the ZINC FINGER moiety while the cleavage domain of FokI contributes the nuclease (i.e. cutting) function.

In a given ZFN, the zinc-finger array is engineered to bind to a specific target sequence in genomic DNA, allowing the FokI cleavage domain to make a double-strand break (DSB) in the DNA.

In mammalian cells, the repair of DSBs can occur either by a mechanism involving (homology-based) copying from the sister chromatid (to restore the normal sequence), or by the process of non-homologous end-joining (NHEJ). As its name indicates, NHEJ is a template-independent process that joins the two ends of a DSB – but it does so in such a way that the original sequence is generally changed to a mutant sequence corresponding to a knock-out mutation. In at least some types of mammalian cell, the repair of DSBs seems to occur more frequently by NHEJ than by homology-based repair; NHEJ can accordingly be exploited for the production of targeted knock-out mutations by engineering the intracellular activity of gene-specific ZFNs.

(See also NON-HOMOLOGOUS DNA END-JOINING.)

ZFNs are used as *hetero*dimers. During the development of this technology it was found that homodimerization of ZFNs could lead to off-target, genome-wide cleavage of DNA. This problem was addressed by improving the ZFN architecture so that (specific) cleavage of DNA by two variant types of ZFN occurs efficiently only when these two ZFNs form a heterodimer [Nature Biotechnol (2007) 25(7):778–785].

A modular approach has been described for the facilitated preparation of ZFNs. The methodology was tested by targeting a range of sites in the human *CCR5* gene, and a number of site-specific mutations were recorded [Genome Res (2009) 19:1279–1288].

Studies have been carried out to reduce the cytotoxicity of ZFNs [PLoS Genet (2009) 5(2):e1000376].

ZFN-based knock-out of the gene encoding dihydrofolate reductase (DHFR) has been demonstrated in Chinese hamster ovary (CHO) cells. ZFN-encoding expression vectors were introduced into CHO cells (with Lipofectamine™ 2000), and ZFNs were transiently expressed in the cells; the disruption of *DHFR* was reported to be complete within 2–3 days of the transient expression of ZFNs. Bi-allelic disruption of *DHFR* occurred, and three (genetically distinct) $DHFR^{-/-}$ cell lines were reported [Proc Natl Acad Sci USA (2008) 105(15):5809–5814].

This technology has also been used for the rapid mutation of genes in the zebrafish [PLoS ONE (2009) 4(2):e4348].

[Genetic analysis of zinc-finger nuclease gene targeting in *Drosophila*: Genetics (2009) 182:641–651.]

[Zinc finger database: a repository of information on C2H2 zinc fingers and engineered zinc finger arrays: Nucleic Acids Res (2009) 37:D279–D283 (doi: 10.1093/nar/gkn606).]

In previous approaches to targeted cleavage of DNA, the restriction endonuclease was linked e.g. to a triplex-forming oligonucleotide (see PROGRAMABLE ENDONUCLEASE).

ZIP (1) A LEUCINE ZIPPER.

(2) A sequence that binds a ZINC FINGER.

Zip nucleic acid (ZNA) A type of oligonucleotide which has been modified with a cation moiety to lower the electrostatic repulsion experienced by the oligonucleotide when it binds to a complementary strand of nucleic acid; such modification thus increases the T_m value of the hybridized oligonucleotide.

ZNAs can anneal to their target sequences at relatively high temperatures (permitting e.g. more satisfactory amplification, by PCR, of AT-rich sequences). As primers, they are reported to be ideal in terms of both specificity and affinity for target sites [Nucleic Acids Res (2009) doi: 10.1093/nar/gkp661].

(See also LOCKED NUCLEIC ACIDS.)

ZipChute™ probe A free (i.e. non-immobilized) probe which incorporates a sequence that can hybridize to a *unique* 'ZipCode' sequence in another, immobilized probe; each type of ZipChute™ probe can therefore hybridize only to the specific probe that carries the corresponding ZipCode; the *availability* of any given (immobilized) probe for hybridization indicates the particular feature being assayed by that probe.

Each ZipChute™ probe also incorporates an individualized 'mobility modifier'. When the ZipChute™ probes are eluted from their (immobilized) target probes they undergo capillary electrophoresis; the mobility modifiers cause different ZipChute™ probes to migrate at different rates, thus facilitating their differentiation and identification.

zipcode Within an mRNA molecule: a sequence of nucleotides which contributes to *targeting* the molecule to an appropriate location in the cell; correct targeting is essential to ensure that the (encoded) polypeptide/protein is synthesized in the right subcellular location. Zipcodes, which, in some cases, are only several nucleotides in length, typically (although not always) occur in the 3′ UTR region of the mRNA. (Interestingly, the ability of a zipcode to direct translocation of the mRNA to a particular intracellular location is independent of the flanking RNA; thus, if an unrelated sequence of RNA is fused to an *isolated* zipcode, that sequence will be directed to the same

subcellular location as the original mRNA.)

Other factors, including *trans*-acting proteins, are involved in targeting mRNAs; such proteins may be required e.g. for linking an mRNA to transport machinery and/or for tethering it to the required subcellular location.

[*Cis*-acting determinants of asymmetric cytoplasmic RNA transport: RNA (2007) 13(5):625–642.]

zipcode array An approach that avoids some of the problems inherent in a conventional MICROARRAY (q.v.). Each of the (immobilized) probes in a zipcode array incorporates its own *unique* sequence of nucleotides (the zipcode), and each target sequence being examined by the array has a complementary copy of *one* of these zipcodes; hence, a given probe can bind only the specific target sequence carrying the complementary copy of its own zipcode.

All the zipcodes in the array are designed to bind optimally to their complementary sequences at a similar temperature – so that (temperature-dependent) stringency is similar for all the probe–target pairs; this is in contrast to the situation in a conventional microarray, in which the operating temperature cannot be optimal for all the probe–target pairs.

A zipcode array has been used e.g. for detecting mutations within a given region of genomic DNA. This region of DNA was initially amplified by PCR. The resulting amplicons were then used as *templates* for SINGLE-BASE EXTENSION assays in which a primer's 3′ nucleotide was designed to be complementary to a *specific* mutation site in the amplicon, and its 5′ end carried a sequence complementary to a particular unique zipcode. A primer that was extended by a (BIOTIN-labeled) ddNTP was subsequently hybridized to a specific probe (via zipcode interaction); the identity of that probe (and thus the type of mutation detected) could then be determined by using a STREPTAVIDIN-linked label which bound to the biotin.

The zipcode approach is particularly useful in that an established zipcode array can be used for successive sets of different target molecules.

An analogous system, termed *barcoding*, is used e.g. in the technique MOLECULAR INVERSION PROBE GENOTYPING.

(See also PADLOCK PROBE.)

zipper (molecular) See MOLECULAR ZIPPER.

zippering (1) Ongoing (progressive and rapid) hybridization of two complementary strands of nucleic acid after initial binding between pairs of bases. It was proposed that, in the cases examined, an initial 'nucleation' (local duplex formation) is likely to be followed by a further nucleation event before the balance of hybridization proceeds by a zippering mechanism [Nucleic Acids Res (2008) 36(9):2981–2989].

(2) The mode in which (certain) invasive bacteria attach to their (mammalian) host cells prior to internalization; zippering involves e.g. formation of multiple points of attachment between pathogen and host cell. This form of invasion occurs e.g. with *Listeria monocytogenes* and *Yersinia enterocolitica*.

ZNA See ZIP NUCLEIC ACID.

ZNF217 An oncogene whose product was reported to bind to a large number of genomic sites – see CHROMATIN IMMUNOPRECIPITATION.

zoo blot A membrane (prepared e.g. by SOUTHERN BLOTTING) containing fragments of genomic DNA from a wide range of species. Probing a zoo blot with a labeled fragment may give useful information about that fragment. Thus, for example, a fragment that hybridizes to DNA from each of a number of widely differing species may be expected to include a highly conserved coding sequence because, in general, non-coding DNA tends to be poorly conserved. In other cases, a fragment may exhibit positive hybridization to DNA of only closely related species.

ZsYellow A fluorescent reporter protein used e.g. in testing for antibiotic resistance in *Mycobacterium tuberculosis* – see the entry ANTIBIOTIC RESISTANCE TESTING.

Zta (*bzlf-1* gene product) See ZEBRA.

zwoegerziekte virus See LENTIVIRINAE.

zygospore A sexually derived, thick-walled resting spore which is characteristic of fungi of the subdivision Zygomycotina.

zygote A diploid cell resulting from the fusion of two (haploid) gametes.

zygote intrafallopian transfer (ZIFT) See the entry IVF.

zygotic meiosis The occurrence of meiosis in a zygote prior to the formation of haploid vegetative cells in those life cycles in which a haploid phase predominates.

zymogen (proenzyme) An (inactive) precursor of an enzyme; it may be converted to the active form of the enzyme by proteolytic cleavage. (See e.g. caspases in the entry APOPTOSIS.)

zymolyase A preparation that has enzymic activity against most (1→3)-β-glucans (which are found e.g. in yeast cell walls). It is often used e.g. in the preparation of yeast SPHEROPLASTS. (cf. DRISELASE.)

[Uses (e.g.): studies on polymorphisms in *Saccharomyces cerevisiae* [PLoS Genet (2009) 5(6):1000502]; investigation of protein transport in *Saccharomyces cerevisiae* [Mol Biol Cell (2009) 20(5):1592–1604.]

Appendix
Alphabetical list of genera

(read *across* the columns)

(A = archaean; B = bacterium; Bd = bird; C = crustacean; F = fungus (yeasts and other species); Fsh = fish; I = insect; M = mammal; N = nematode; P = plant (includes algae and higher plants); Pz = protozoan)

Abies (P)	*Acanthamoeba* (Pz)	*Acer* (P)
Acetabularia (P)	*Achlya* (F)	*Acinetobacter* (B)
Actinobacillus (B)	*Aedes* (I)	*Aeromonas* (B)
Agalliopsis (I)	*Agaricus* (F)	*Agrobacterium* (B)
Agropyron (P)	*Albugo* (F)	*Alcaligenes* (B)
Allomyces (F)	*Alnus* (P)	*Alternaria* (F)
Alteromonas (B)	*Amanita* (F)	*Anabaena* (B)
Anguina (N)	*Anopheles* (I)	*Aphelenchoides* (N)
Apis (I)	*Arabidopsis* (P)	*Arcella* (Pz)
Arcellinida (Pz)	*Armillaria* (F)	*Arthrobacter* (B)
Aschersonia (F)	*Ascosphaera* (F)	*Aspergillus* (F)
Astacus (C)	*Ateles* (M)	*Avena* (P)
Babesia (Pz)	*Bacillus* (B)	*Bacteroides* (B)
Bartonella (B)	*Bdellovibrio* (B)	*Beauveria* (F)
Beggiatoa (B)	*Belononlaimus* (N)	*Berberis* (P)
Beta (P)	*Bifidobacterium* (B)	*Bodo* (Pz)
Bodonina (Pz)	*Boletus* (F)	*Bombyx* (I)
Bordetella (B)	*Borrelia* (B)	*Bos* (M)
Botrytis (F)	*Brassica* (P)	*Bremia* (F)
Brevoortia (Fsh)	*Brochothrix* (B)	*Brucella* (B)
Bryopsis (P)	*Buchnera* (B)	*Bullera* (F)
Burkholderia (B)	*Byssochlamys* (F)	*Caedibacter* (B)
Caenorhabditis (N)	*Callithrix* (M)	*Camellia* (P)
Campylobacter (B)	*Candida* (F)	*Cantharellus* (F)
Carchesium (Pz)	*Cardiobacterium* (B)	*Carnobacterium* (B)
Caryophanon (B)	*Caseobacter* (B)	*Castanea* (P)
Caulerpa (P)	*Caulobacter* (B)	*Cellulomonas* (B)
Cephalosporium (F)	*Ceratocystis* (F)	*Cercopithecus* (M)
Chaetomium (F)	*Chenopodium* (P)	*Chionoecetes* (C)
Chlamydia (B)	*Chlamydomonas* (P)	*Chlorella* (P)

Appendix: alphabetical list of genera

Chondrus (P)	*Chromobacterium* (B)	*Cicadulina* (I)
Citellus (M)	*Citeromyces* (F)	*Citrobacter* (B)
Cladophora (P)	*Cladosporium* (F)	*Claviceps* (F)
Clostridium (B)	*Cocos* (P)	*Coffea* (P)
Colletotrichum (F)	*Colpidium* (Pz)	*Colpoda* (Pz)
Coriolus (F)	*Corynebacterium* (B)	*Costia* (Pz)
Coxiella (B)	*Cricetus* (M)	*Criconemoides* (N)
Cristispira (B)	*Cryptococcus* (F)	*Cryptosporidium* (Pz)
Culicoides (I)	*Culiseta* (I)	*Curtobacterium* (B)
Cyathus (F)	*Cyclidium* (Pz)	*Cymbidium* (P)
Cyniclomyces (F)	*Cyprinus* (Fsh)	*Dacrymyces* (F)
Dactylis (P)	*Daedalea* (F)	*Daldinia* (F)
Danio (Fsh)	*Darna* (I)	*Deinococcus* (B)
Dekkera (F)	*Dendrolimus* (I)	*Desulfobacter* (B)
Desulfobulbus (B)	*Desulfococcus* (B)	*Desulfomonas* (B)
Desulfosarcina (B)	*Desulfovibrio* (B)	*Desulfurococcus* (A)
Desulfuromonas (B)	*Dictyostelium* (F)	*Dientamoeba* (Pz)
Difflugia (Pz)	*Diplocarpon* (F)	*Ditylenchus* (N)
Dolichodorus (N)	*Drosophila* (I)	*Dunaliella* (P)
Duttonella (Pz)	*Ehrlichia* (B)	*Eikenella* (B)
Eimeria (Pz)	*Elsinoë* (F)	*Emericella* (F)
Endamoeba (Pz)	*Endothia* (F)	*Ensifer* (B)
Entamoeba (Pz)	*Enterobacter* (B)	*Enterococcus* (B)
Epichloë (F)	*Erwinia* (B)	*Erysipelothrix* (B)
Erysiphe (F)	*Escherichia* (B)	*Euglena* (P)
Euplotes (Pz)	*Fagus* (P)	*Festuca* (P)
Filosea (Pz)	*Fistulina* (F)	*Fomes* (F)
Frankia (B)	*Frateuria* (B)	*Fusarium* (F)
Fusobacterium (B)	*Gaeumannomyces* (F)	*Gaffkya* (B)
Gallionella (B)	*Gallus* (Bd)	*Ganoderma* (F)
Gardnerella (B)	*Gelidium* (P)	*Giardia* (Pz)
Gigartina (P)	*Gilpinia* (I)	*Gloeobacter* (B)
Glossina (I)	*Gossypium* (P)	*Gracilaria* (P)
Graminella (I)	*Gregarina* (Pz)	*Haemophilus* (B)
Hafnia (B)	*Halimeda* (P)	*Halobacterium* (A)
Halococcus (A)	*Halomonas* (B)	*Hansenula* (F)
Helicobacter (B)	*Helicotylenchus* (N)	*Heliothis* (I)

Helvella (F)	*Hematopinus* (I)	*Hemicycliophora* (N)
Heterodera (N)	*Hexamita* (Pz)	*Histomonas* (Pz)
Histoplasma (F)	*Homo* (M)	*Hoplolaimus* (N)
Hormoconis (F)	*Humicola* (F)	*Humulus* (P)
Hydnum (F)	*Hyphomonas* (B)	*Hyposoter* (I)
Ichthyophthirius (Pz)	*Ictalurus* (Fsh)	*Inocybe* (F)
Inonotus (F)	*Ipomea* (P)	*Isospora* (Pz)
Issatchenkia (F)	*Janthinobacterium* (B)	*Juglans* (P)
Kentrophoros (Pz)	*Klebsiella* (B)	*Kloeckera* (F)
Kluyvera (B)	*Kluyveromyces* (F)	*Koserella* (B)
Kurthia (B)	*Lachnospira* (B)	*Lactobacillus* (B)
Lactococcus (B)	*Laminaria* (P)	*Legionella* (B)
Leishmania (Pz)	*Lentinula* (F)	*Lentinus* (F)
Lepista (F)	*Leptinotarsa* (I)	*Leptomis* (Fsh)
Leptomonas (Pz)	*Leptonema* (B)	*Leptospira* (B)
Leuconostoc (B)	*Limulus* (C)	*Lipomyces* (F)
Listeria (B)	*Lolium* (P)	*Longidorus* (N)
Loxodes (Pz)	*Lupinus* (P)	*Lycopersicon* (P)
Lymantria (I)	*Lyngbya* (B)	*Macaca* (M)
Malassezia (F)	*Mallomonas* (P)	*Malpighamoeba* (Pz)
Malus (P)	*Marmota* (M)	*Megasphaera* (B)
Megatrypanum (Pz)	*Meles* (M)	*Melilotus* (P)
Meloidogyne (N)	*Melopsittacus* (Bd)	*Methanobacterium* (A)
Methanococcoides (A)	*Methanococcus* (A)	*Methanosarcina* (A)
Methanospirillum (A)	*Methanothermus* (A)	*Minchinia* (Pz)
Methanothrix (A)	*Mobiluncus* (B)	*Monocystis* (Pz)
Monostroma (P)	*Moraxella* (B)	*Morchella* (F)
Morganella (B)	*Mucor* (F)	*Mus* (M)
Mustela (M)	*Mycena* (F)	*Mycobacterium* (B)
Mycoplasma (B)	*Nacobbus* (N)	*Nadsonia* (F)
Naegleria (Pz)	*Nannomonas* (Pz)	*Nectria* (F)
Neisseria (B)	*Neodiprion* (I)	*Nesoclutha* (I)
Neurospora (F)	*Nicotiana* (P)	*Nitrobacter* (B)
Nitrosococcus (B)	*Nitrospina* (B)	*Nocardia* (B)
Nosema (Pz)	*Nostoc* (B)	*Nudaurelia* (I)
Olpidium (F)	*Opalina* (Pz)	*Oryctolagus* (M)
Oryza (P)	*Oscillatoria* (B)	*Ostracoblabe* (F)

Appendix: alphabetical list of genera

Oxalobacter (B)	*Pacifastacus* (C)	*Paramecium* (Pz)
Paratylenchus (N)	*Paspalum* (P)	*Pasteurella* (B)
Pelodictyon (B)	*Penicillium* (F)	*Penaeus* (C)
Peptostreptococcus (B)	*Peranema* (Pz)	*Perilla* (P)
Perkinsiella (I)	*Peronospora* (F)	*Petunia* (P)
Peziza (F)	*Phaffia* (F)	*Phaseolus* (P)
Phellinus (F)	*Phlebotomus* (I)	*Phleum* (P)
Phoma (F)	*Photobacterium* (B)	*Phycomyces* (F)
Phycopeltis (P)	*Physarum* (F)	*Phytolacca* (P)
Phytophthora (F)	*Pichia* (F)	*Pilobolus* (F)
Pinus (P)	*Piptoporus* (F)	*Pisum* (P)
Planktothrix (B)	*Planobispora* (B)	*Planococcus* (B)
Plantago (P)	*Plasmodiophora* (F)	*Plasmodium* (Pz)
Pleurotus (F)	*Pneumocystis* (F)	*Polyporus* (F)
Popillia (I)	*Populus* (P)	*Poria* (F)
Prevotella (B)	*Procyon* (M)	*Propionibacterium* (B)
Prorodon (Pz)	*Proteus* (B)	*Prunus* (P)
Pseudallescheria (F)	*Pseudomonas* (B)	*Psoroptes* (I)
Puccinia (F)	*Pulex* (I)	*Pyricularia* (F)
Pyrococcus (A)	*Pyrodictium* (A)	*Pyronema* (F)
Pyrus (P)	*Pythium* (F)	*Quercus* (P)
Radopholus (N)	*Rattus* (M)	*Rhipicephalus* (I)
Rhizobium (B)	*Rhizoctonia* (F)	*Rhizomucor* (F)
Rhizopus (F)	*Rhodococcus* (B)	*Rhodotorula* (F)
Rickettsia (B)	*Robinia* (P)	*Rubus* (P)
Saccharomyces (F)	*Saccharum* (P)	*Saimiri* (M)
Salix (P)	*Salmo* (Fsh)	*Salmonella* (B)
Saprolegnia (F)	*Sarcina* (B)	*Scenedesmus* (P)
Schizosaccharomyces (F)	*Sciara* (I)	*Scleroderma* (F)
Sclerotinia (F)	*Sclerotium* (F)	*Scolytus* (I)
Scutellinia (F)	*Scutellonema* (N)	*Scytonema* (B)
Selenomonas (B)	*Septoria* (F)	*Serpula* (F)
Serratia (B)	*Shigella* (B)	*Sordaria* (F)
Sorghum (P)	*Sparassis* (F)	*Spartina* (P)
Spermophilus (M)	*Sphaerotilus* (B)	*Spirillum* (B)
Spiroplasma (B)	*Spirulina* (B)	*Sporobolomyces* (F)
Sporothrix (F)	*Staphylococcus* (B)	*Stentor* (Pz)

Streptococcus (B)	*Streptomyces* (B)	*Sulfolobus* (A)
Sylvilagus (M)	*Synchytrium* (F)	*Synechococcus* (B)
Synechocystis (B)	*Tetrahymena* (Pz)	*Thea* (P)
Theileria (Pz)	*Theobroma* (P)	*Thermoplasma* (A)
Thermus (B)	*Thiobacillus* (B)	*Thomomys* (M)
Tilia (P)	*Tilletia* (F)	*Torula* (F)
Torulopsis (F)	*Toxoplasma* (Pz)	*Toxothrix* (B)
Trachelomonas (Pz)	*Trametes* (F)	*Trebouxia* (P)
Tremella (F)	*Trentepohlia* (P)	*Treponema* (B)
Trichomonas (Pz)	*Trichophyton* (F)	*Trichoplusia* (I)
Trichosporon (F)	*Triticum* (P)	*Tropheryma* (B)
Trypanosoma (Pz)	*Typhula* (F)	*Tyzzeria* (Pz)
Ulmus (P)	*Urticaria* (P)	*Ustilago* (F)
Venturia (F)	*Verpa* (F)	*Verticillium* (F)
Vibrio (B)	*Vicia* (P)	*Vigna* (P)
Volvariella (F)	*Volvox* (P)	*Vorticella* (Pz)
Wickerhamia (F)	*Wickerhamiella* (F)	*Wolbachia* (B)
Wolinella (B)	*Xanthomonas* (B)	*Xenopsylla* (I)
Xenorhabdus (B)	*Xylaria* (F)	*Yersinia* (B)
Zea (P)	*Zoogloea* (B)	*Zythia* (F)