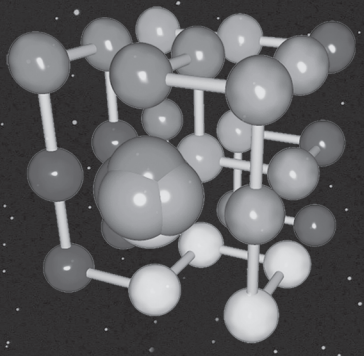

Giant
Molecules
Here, There, and Everywhere
Second Edition

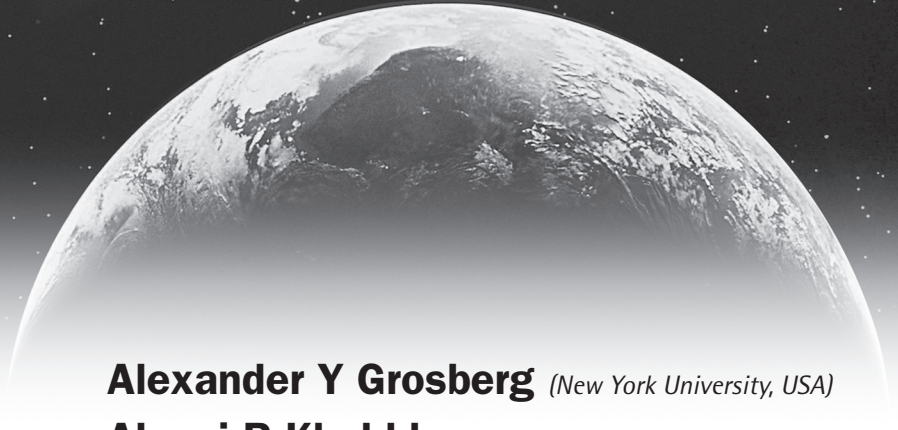
This page is intentionally left blank



Giant Molecules

Here, There, and Everywhere

Second Edition



Alexander Y Grosberg *(New York University, USA)*

Alexei R Khokhlov *(Moscow State University, Russia)*

Foreword by

Pierre-Gilles de Gennes

World Scientific

NEW JERSEY • LONDON • SINGAPORE • BEIJING • SHANGHAI • HONG KONG • TAIPEI • CHENNAI

Published by

World Scientific Publishing Co. Pte. Ltd.

5 Toh Tuck Link, Singapore 596224

USA office: 27 Warren Street, Suite 401-402, Hackensack, NJ 07601

UK office: 57 Shelton Street, Covent Garden, London WC2H 9HE

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library.

GIANT MOLECULES — 2nd Edition
Here, There, and Everywhere

ISBN-13 978-981-283-922-0
ISBN-10 981-283-922-4

Printed in Singapore.

To Vera and Natasha

This page is intentionally left blank

Foreword by P.G. de Gennes

The idea of atoms goes back to the Greeks: but for them it was really just a formal postulate, avoiding the intricacies of infinitely small objects. More than 2000 years were required to transform this into a reality. To show that the usual forms of matter around us are made with atoms, and clumps of atoms which we call molecules. The first determination of the size of a molecule is probably due to Benjamin Franklin: he knew (again from the Greeks) that a small amount of oil suppresses the waves on the sea. He then went to a pond in Clapham Common, choosing a day with a light wind, where the surface of the pond showed ripples. He then poured a spoonful of oil on the water, and measured the area upon which the ripples had disappeared: this area turned out to be huge. In our modern parlance, he had constructed a very thin monolayer of oil molecules. Dividing the volume (a spoonful) by the area, he could measure the size of a molecule (in this case, something like 2 nanometers).

Unfortunately, Franklin did not perform this calculation himself — it was done only a hundred years later by Lord Rayleigh (as explained in a beautiful book by C. Tanford[1]). But this experiment was a historical landmark: for the first time, molecules were not a figment of a philosopher's imagination. They became a physical object, with a well defined number, measuring their size!

A second step concerned the giant molecules which are the topic of this book. Many things around us (wood, cloth, food, our own body...) are made of macromolecules, or polymers — as we call them now. But the concept of macromolecules emerged very slowly. During the 19th century, many chemists synthesized new polymers and threw them down the sink! In these days, the chemical dogma was to make a new substance, to purify it as

[1] C. Tanford, "Ben Franklin stilled the waves", Duke University Press, 1980.

much as possible, and to test the purity by measuring a property such as the melting point: if the melting point was sharp, the product was considered as "good." But macromolecules, unfortunately, do not have a sharp melting point (for reasons which are, to a certain extent, explained in the present book). They were thus considered as "dirty," and rejected. Ultimately, around 1920, H. Staudinger proved conclusively the existence of long chain molecules to the community of chemists. Physicists then entered the game: Kuhn first, who understood the flexibility of many polymers, the role of entropy in these systems, and the resulting elasticity of rubber. Again here, we have a great book describing the story[2]. Then came P. Flory, who mastered most of the physical properties of polymers, using very simple, but deep, ideas. The next step was due to S.F. Edwards, who pointed out a profound similarity between the conformation of a chain and the trajectory of a quantum mechanical particle. This allowed for fifty years of theoretical know-how, accumulated in quantum physics, to be transposed to polymer science!

The present book describes the final state of this evolution. The two Russian authors have had the talent of writing it in a simple style — avoiding most of the heavy formalism which is beloved in countries of strong mathematical bias, such as Russia or France.

The final product is accessible for university students and to research engineers. I am convinced that it will play a very useful role in this context. Giant molecules are important in our everyday life. But, as pointed out by the authors, they are also associated with a culture. What Bach did with the harpsichord, Kuhn and Flory did with polymers. We owe a lot of thanks to those who now make this music accessible.

P.G. de Gennes
March 1996

[2]H. Morawetz, "Polymers: the Origins and Growth of a Science", John Wiley & Sons (USA), 1985.

From the Reviews of the First Edition

"*Giant Molecules* is a beautiful book on polymer science which is written by two of the leaders in the field who are also tremendously skilled at putting the science in both historical and scientific contexts. The book is actually a marvelous introduction to polymer physics ... which is scientifically accurate but can also be read as a wonderfully articulate and amusing history of the subject. The book must be on the shelf of all polymer scientists and will go a long way in explaining this sub-discipline to the broad public."

Philip Pincus
University of California
Santa Barbara
(from the review of the manuscript, 1996)

"Giant molecules is one of the hottest topics in science today. This book, written by two brilliant physicists, will guide readers through this new frontier of polymer science ... and the authors make the topic equally applicable to any curious reader. The authors are skilled story-tellers, which makes this scientifically relevant book entertaining as well as informative. *Giant Molecules* will be of use to all levels of science enthusiasts who are curious about the newest developments in polymer science. This book is not to be missed!"

Toyoichi Tanaka (1946-2000)
Massachusetts Institute of Technology
(from the review of the manuscript, 1996)

"Who would have thought a pair of theorists would produce a very readable and perceptive monograph of polymer physics? Yet this is exactly what

Alexander Y. Grosberg and Alexei R. Khokhlov have done in this attractive book. ... The explanations ... are about clearest I have read anywhere."

Edwin L. Thomas
Department of Materials Science and Engineering
Massachusetts Institute of Technology
(Nature, v. ***388****, p. 842, 1997)*

"... it might seem almost impossible to write a book on macromolecules ... using almost no maths. However, the authors have succeeded in writing an accurate and precise book. ... I never found a place where simplification led to scientifically questionable description ... This makes it a valuable book for both the scientifically interested layreader and non-expert student, as well as for the experienced scientist ...

I noted with pleasure the citations from classic literature at the start of each chapter which hint at some surprising parallels in thinking between scientists and the cited authors ..."

Kurt Kremer
Director of Max Planck Institute for Polymer Research
Mainz, Germany
(Physics World, December 1997, p. 49)

"The book reviews the fundamental concepts of polymer physics and discusses some of the modern frontiers of the subject, particularly in biology. The overall level is suitable for an advanced undergraduate in physics, chemistry or chemical engineering ... Practitioners will also find this book stimulating ..."

Thomas Halsey
Exxon Research and Engineering
Annandale, New Jersey
(Physics Today, February 1998, p. 73)

"... this is an easy read and readers with a desire to learn more about the biology and physics of polymers will find *Giant Molecules* friendly and welcoming."

Bernd Eggen
University of Sussex
(New Scientist, August 16, 1997, p. 41)

"As a scientific text it is without doubt one of the easiest to read I have ever encountered. Despite this, it remains, highly informative. ... The authors have managed, without compromising their scientific contents, to include a number of interesting anecdotes that place the science in its true context ... I would commend this book to anyone with an interest in polymer science, whether established experts or complete newcomers — it really is an excellent starting point for the subject."

Simon Biggs
The University of Leeds
*(Molecules, v. **3**, p. 142, 1998)*

"I am a physics professor working on semiconductor materials. Polymer is not my area ... I can't believe this great book hasn't been reviewed yet. Yes, it is written by two Russian scientists. But who said Russians can only write rigorous math books? This book is not a monograph ... it explains a lot of phenomena in a clear, concise and humorous language. There is a little math, not hard at all. Freshmen level calculus would be sufficient to understand the book."

A reader's review on Amazon.com web site

This page is intentionally left blank

(IMAGINARY) EDITOR (*sceptically*): Oh, not you again...

AUTHORS (*bashfully*): Well, you see, we've written a book on giant molecules...

EDITOR: What molecules?

AUTHORS: giant ones. (*Getting more excited*) Just listen to this bit here!

EDITOR (*impatiently*): Oh, no, I haven't time to listen to it. Anyway, you've already published a book on them[3], so what's the point?

AUTHORS: Yes, but that was for experts, while this one...

EDITOR (*losing his temper*): And this one is for housewives, presumably! Look, why don't you leave the preface with me, and I'll see what I can do.

AUTHORS: Here it is!

Preface

The very nature of the genre suggests the question that ought to be answered in the preface: For whom is this book intended?

We hope that this book may interest anyone with general curiosity about the world. And this is not just because we think too highly of ourselves! Rather, what really gives us hope is the unique position of this field. It is right at the crossroads of so very many paths of contemporary development and ardent interest. It is about all kinds of things, e.g. modern

[3] A.Y. Grosberg, A.R. Khokhlov, "Statistical Physics of Macromolecules", AIP Press, 1994.

materials (including really fascinating "smart" materials), and the famous DNA which is not just enthralling in its own right, but is already becoming a tool which is used, for example, in criminology and as a "computer in a glass of water". Polymer physics is also about modern medicines, and lots more. To sum up, many things that people talk about every day have their roots in our science.

That is why we decided that it was time to write a clear, comprehensible story about giant molecules.

A college or university student should be able to read our book from cover to cover and get a superficial but coherent idea of the subject. A scientist, whether a physicist, a chemist, a materials engineer or a molecular biologist, may be interested to see how we approach familiar topics avoiding the complexities of scientific language.

Very frequently, sophisticated science is treated with rather ambitious mathematics. And the experience indicates that this aspect is the most scary for many students. Indeed, mathematical methods become necessary when and if a student wants to become professional and to build new inroads into science. We keep the use of mathematics at bay, our mathematics is restricted to simple algebra and never goes beyond the typical high school curriculum. At the same time, our physics is at times quite sophisticated.

Last but not least, we hope that any reader may just browse through the book and find out what is meant by "molecular architecture", what will happen if you chop up a cauliflower, or who used to be called the queen of the world and her shadow.

Just one more thing. There is a well-known saying by Dostoevsky, "beauty will save the world". While one can interpret these words in different ways, there is no doubt that the intellectual beauty is one of the most astonishing features of science. Indeed, why does the most effective so frequently happen to be the most beautiful as well? We do not know, but it seems to be a fact! In this book we have tried to demonstrate the beauty of polymer and biopolymer science.

For the present edition, we have modified the text in many places and have written new chapters on polymer synthesis, protein folding, polymer knots and new sections on molecular motors, semi-flexible and worm-like polymers, and several others. We have included many new figures. Overall, about 50% of the book is new.

Previous edition included the CD ROM with computer simulations of polymers. We decided not to include it in this edition, because, as it turned

out, this part was getting obsolete too fast. We are working now on the ways to disseminate the corresponding material in a more efficient form.

All color figures in the book are grouped in three places: (i) pages 81–85, (ii) pages 215–221 and (iii) pages 277–281. References to them are labeled with letter "C", like Fig. C2.4 etc.

We have tried to make this book both interesting and useful. Whether we have succeeded or not is for our readers to decide.

The Authors

EDITOR (*murmuring to himself*): Well, if they are not lying, perhaps it *is* interesting after all... It sounds like, apart from the general reader, the book may interest people in *(counting on his fingers)* the APS, ACS, MRS, BPS ... I think we ought to publish it.

This page is intentionally left blank

Acknowledgments

This book underwent a considerable evolution. The first version was published in Russian in 1989[4], as a part of the so-called "Quantum bibliotheca" — a series of books widely read by high school students and professors alike. We are indebted for the invitation to contribute to this distinguished series.

The first version of the manuscript was carefully read by Drs. M.A. Livshitz and S.G. Starodubtsev. We are thankful for their useful comments.

We are indebted to Drs. T.A. Yurasova and C.J.B. Ford for their unlimited patience in the translation into English of the text that we originally wrote in Russian. Their work allowed for the first English edition of 1997[5].

This book was used as a text or a supplementary material in a number of Universities, and we thank all who shared with us their positive remarks or criticisms. Many readers informed us about mistakes and inaccuracies in various places in the book. We are particularly indebted to Dr. Byron K. Christmas, Center for Applied Polymer Science Research University of Houston-Downtown, for his correspondence. Dr. Nathan Moore of Winona State University gave us a long list of found mistakes and typos.

In preparation for the current edition we received a lot of help from Dr. Artem A. Aerov. He edited the whole text, again filtering out our lapses, and helped us in many other ways, in particular, he wrote the section on QWERTY in Chapter 14.

Professor Olga E. Philippova and Dr. Elena V. Chernikova greatly helped us editing the chemical parts of the book. Professors Vijay Pande, Rob Phillips, Eric Vanden-Eijnden, Alexander Vologodskii read various parts of the manuscript and provided valuable feedback.

[4] A.Y. Grosberg and A.R. Khokhlov, "Physics in the World of Polymers", Moscow, Nauka, 1989.

[5] A.Y. Grosberg and A.R. Khokhlov, "Giant Molecules: Here, There, and Everywhere ...", Academic Press, 1997.

Professors Andrey A. Askadskii, Sergei Buldyrev, Pavel G. Khalatur, Amit Meller, Vijay Pande and Jean-Louis Sikorav prepared several beautiful figures for this book. Table of knots was custom made for this book by Dr. Robert Scharein. Dr. Sergei B. Ryzhikov was instrumental in making photographs of several experimental devices. We gratefully acknowledge their help.

We owe many thanks to Sergei Buldyrev, Dmitry Cherny, Aleksandr K. Gladilin, Nicholas Hud, Mehran Kardar, Alexei Likhtman, Tom McLeisch, Amit Meller, Leonid Mirny, Jean-Louis Sikorav, Eugene Stanley, Peter Virnau, Alexander Vologodskii and Jakob Waterborg for their permissions and help in reproducing figures borrowed from their publications. Jean-Francois Joanny, Kurt Kremer and Michael Lomholt greatly helped us with quotations from the sources in their respective native languages.

Late Professor Piere-Gilles de Gennes (1932–2007) wrote an introduction to the first English edition of this book. He also very much encouraged our effort to present the material in a simple way, downplaying the technicalities, particularly the excessive mathematics, but conveying the aesthetic and cultural underpinnings of polymer science. We are grateful for his help and support.

Last but not least, we are deeply indebted to Professor Ilya M. Lifshitz (1917–1982); both of us were lucky enough to have him as a teacher. He was very good at creating a special atmosphere of ardent, inspiring interest in science. It is now up to you, the reader, to decide whether we succeeded, at least partially, in recapturing this atmosphere in our book.

Contents

Chapter 1

Introduction: Physics in the World of Giant Molecules

Molecules are supposed to be small, aren't they? Quite apart from anything else, even the very word *molecule* comes from a Latin phrase that literally means "a tiny mass of something". Nevertheless, what would you say about a molecule about 1 meter long? Or another one that weighs almost 1 kilogram? There are many molecular giants of the kind. They are called polymers; perhaps you have heard this word. Thus, our book is about polymers. The world of polymers.

The world of polymers... Are polymers really so diverse and numerous that they make up a whole world? Is this not an exaggeration?

Well, what are polymers? The first things that come to mind may be plastic bags, and other common plastics. You may also think of rubber and all its products. Then, synthetic fibres and fabrics, as well as natural ones, of course. In fact, the list is endless: for example, cellulose (which makes up both timber and paper), the shell of a space probes traveling to Venus or Mars, and artificial valves implanted into a human heart... Polymers are used for all sorts of purposes. Huge quantities of them are made these days throughout the world. In fact, the volume of polymers produced already exceeds that of metals (although metals still win by weight).

The applications alone are a good enough reason to study polymers. This is just the same as with semiconductors, for example. However, it is not only their applications that make polymers so fascinating. The greatest incentive to do polymer science is life itself. Even a schoolchild knows these days that our so called "genetic blueprint" (that is, what one is born to be, a dog or a cat, a boy or a girl, and what color of skin, hair, and eyes one is to have, etc.) is contained in molecules of a special polymer, DNA (deoxyribonucleic acid). Modern biology regards a living cell as a kind of factory, finely tuned, and controlled by DNA. Meanwhile, all the working

devices in this factory (be they chemical, electrical, mechanical, optical, or whatever) are based on another type of polymer called proteins. In addition to this, polymers make hooves and horns, hair, and lots more!

It is not just that polymers are found in abundance in nature, they actually play a crucial role. So M.D. Frank-Kamenetskii was not really joking when he called his popular book on DNA "The Most Important Molecule" (Ref. [45] in the list at the end of the book).

You may say, "All right, I believe you, polymers are important. Perhaps one can even talk about the world of polymers if one wants. But why physics?" Good question. We shall try to answer it in a minute, but before that let's make one more comment.

We would hate to sound like totally boring people who believe in doing only useful things. In fact, sometimes it is a good idea just to pursue whatever takes your fancy! At least, it works very well in scientific research. After all, it is seldom clear from the start what use you can make of a discovery or idea. What is fortunate is that good scientists usually have well developed "taste": what they like and want to do, tends to be also useful.

Well, let's go back to the question. Why study the physics of polymers? We can now give one good reason. It is merely very interesting! And it has a lot to offer. Beautiful effects, fundamental analogies with other areas, and clear physical principles explaining complex phenomena. These are just what we shall try to give a feel for in this little book. As for various applications, there are other people who can write a better story on those. Chemists could talk with confidence about synthetic polymers. And molecular biologists know a lot about biological polymers. However, even in these areas, physicists have no reasons to feel too much out of place. Without physics, one can hardly reach a proper understanding of polymer chemistry or molecular biology. This is why all polymer scientists know the physics of polymers, and all use it to some extent in their work. Quite often the combination proves very fruitful.

There was even a period, in the 1940s and 1950s, when polymer physics was developed mainly by professional chemists. The most notable among them was Paul Flory (1908–1982), an American physical chemist who went down in scientific history chiefly due to his pioneering work in polymer physics. He received a Nobel prize for this in 1974.

However, science tends to become more and more specialized. So it is not surprising that polymer physics has eventually grown into an independent field of research. This was helped by some eminent physicists,

such as I.M. Lifshitz in Russia, S.F. Edwards in England, and P.G. de Gennes in France, who in the middle of the 1960s turned towards the study of polymers. They revealed basic analogies between problems in polymer physics and some of the most burning and tantalizing questions of general physics. Polymers emerged on to the pages of the world's main physics journals and at major international conferences. Rather rapidly, a harmonious system of simple models and qualitative ideas formed about the basic physical properties of polymers at a molecular level. All these concepts have been used successfully both in physical chemistry and in molecular biology. This brought also some terminology simplification. For example, we shall frequently follow physics tradition and call the units of polymer chain "monomers," not the "monomer units," as chemists prefer.

If you know about the physics of polymers you will understand why they are so widely used in everyday life and in industry, as well as how they work in biology.

Chapter 2

What Does a Polymer Molecule Look Like?

> L'essentiel est invisible pour les yeux.
> (What is essential is invisible to the eye.)
>
> Antoine de Saint-Exupéry,
> *Le Petit Prince*

2.1 Polymers are Long Molecular Chains

There used to be a time when in scientific essays all substances were described just in terms of how human senses perceived them. Even now one may come across this way of presenting things in some textbooks; for example: "Water is a liquid which has no color, no taste, and no smell". These days such a description could also include information obtained from various measuring instruments, such as the spectrum or a material parameter. However, it would not be an exaggeration to say that modern scientists — be they physicists, chemists, or biologists — who study a substance should first of all have some image of a molecule of the substance.

This is why we shall start with what we can call portraits of polymer molecules. Polymers are substances consisting of long molecular chains, so-called macromolecules. A helpful image is some sort of long, entangled, three-dimensional thread, chain, rope or wire.

What could be the chemical structure of such a macromolecule? Figure 2.1 *a* shows schematically the structure of the simplest polymer chain, a polyethylene macromolecule. One can see that the macromolecule consists of indefinitely repeating identical CH_2 groups which are connected by covalent chemical bonds to form a chain. Other polymers (e.g., polystyrene

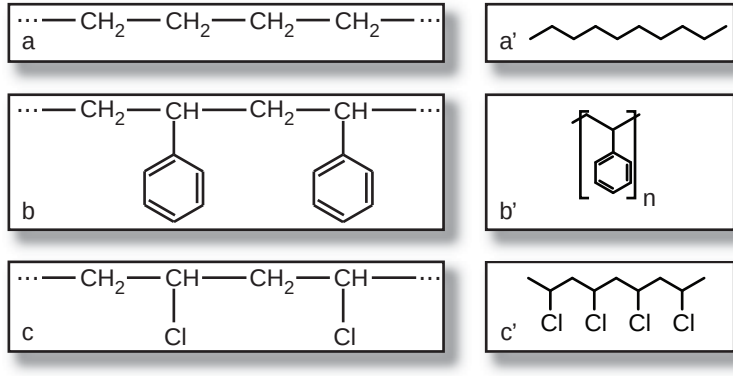


Fig. 2.1 Chemical structure of (*a*) polyethylene, (*b*) polystyrene, (*c*) polyvinyl chloride. To illustrate various ways to present such structures, panels (*a'*), (*b'*), and (*c'*) show the same polymers in various other notations. For instance, hydrogen atoms are usually not shown, carbon atoms may not be explicitly shown as well (*a'* and *c'*), and only one repeat unit can be shown instead of the chain (*b'*).

or polyvinyl chloride) are still organized into a chain of repeating units, although units themselves may have very different atomic structures (Figure 2.1 *b, c*). In this book, we will call the elementary units of polymer chains as monomer units, or simply monomers[1].

To be considered as a polymer, a molecule must consist of a great number of units, $N \gg 1$. Molecules of the types shown in Figure 2.1, if artificially synthesized in a chemical laboratory or industrial process, normally contain from hundreds up to tens of thousands units: $N \sim 10^2 \div 10^4$. Natural polymer chains can be even longer than such "synthetic" polymers. The longest known polymers are DNA molecules. The number of monomer units in DNA can reach a billion ($N \sim 10^9$) or even ten billion ($N \sim 10^{10}$). It is just because they can be so long that polymer molecules are called macromolecules ("macro" is the Greek for large).

[1]We must warn the reader of the terminological subtleties on this point. In chemical literature, the term monomers is frequently reserved for the relatively small molecules employed as the initial building blocks in purposeful making, or preparation, of polymer chains. In this language, the units of a polymer, or "links" of a polymer chain, are sometimes referred to as monomer residues (because monomers typically loose some chemical groups, such as OH, when combined into chains). We will discuss these issues in somewhat more details in Chapter 3, but mostly we will follow the tradition of physics literature and use the simple word monomer for the units of already prepared molecular chains.

The fact that polymer molecules consist of long chains of monomers was not originally realized. At the beginning of the 20th century it was finally proved that matter consists of atoms and molecules. Yet no one attempted to look at polymers from a molecular point of view, even though some natural polymers (such as rubber, cellulose, silk, and wool) were widely used. At that time, the predominant opinion about polymers was that they were a sort of complex colloid system. It was not until the early 1920s that seminal works by the German physical chemist Hermann Staudinger appeared. He suggested, after analyzing many experimental results, that polymer molecules are chains. The idea met with some scepticism at first, and even with a fair amount of mockery in scientific circles. Once, for instance, at a seminar, Staudinger was asked the question: "So what kind of length are your molecules after all — the size of a nail, or of a finger?" All those present thought it was very funny, and burst out into guffaws. Of course, from the modern point of view, there was nothing to joke about — DNA macromolecules, measured along the chain, can be as long as a few centimeters.

Although his hypothesis was not accepted at once, Staudinger stuck to it, and went on accumulating more and more experimental evidence. As a result, by the beginning of the 30s, the concept of the chain structure of macromolecules became generally established. It is sometimes reckoned that looking at the evolution of any scientific idea one can discern three different stages — at the beginning people say: "It's impossible!", then: "There may be something in it!", and eventually: "Oh well, but that's a well-known fact!" The concept that macromolecules are long molecular chains went through these three stages over a period of just ten years. Remarkably, Staudinger had to wait for about quarter of a century until eventually Nobel Prize in chemistry was awarded to him in 1953 ("for his discoveries in the field of macromolecular chemistry").

2.2 Flexibility of Polymer Chains

The work by Staudinger prepared the ground for physics to intrude into the "Polymer World" — it had become possible to explain physical properties of various polymers by taking into account the chain structure of their constituent molecules. But first, polymer scientists had to discern the specific shapes, or conformations, of molecular chains for different polymers.

For example, let's consider a polymer molecule diluted in some ordinary solvent (say, in water). What kind of shape does the molecule's chain have?

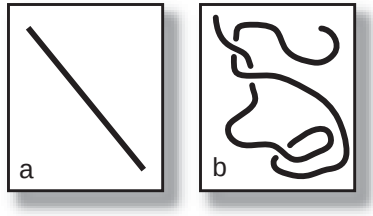


Fig. 2.2 (*a*) Rectilinear conformation of a polymer chain; (*b*) conformation of an entangled coil.

Judging from the linear structure of the polymer chains (Figure 2.1), at first glance it seems reasonable to assume that the chain looks vaguely like a straight line (Figure 2.2 *a*). But this is not true; as a matter of fact, it gets tangled up into a random loose three-dimensional coil (Figure 2.2 *b*). This is simply a result of the chain's flexibility. Let's emphasize: it is *not* the result of any particular specific *chemical* structure, it is the general *physical* consequence of the linear chain structure of the molecules.

Generally speaking, the idea of flexible polymer chains may appear rather surprising. At school one is taught that the atoms in a molecule are joined together by covalent bonds in some specific order. Therefore their positions in space with respect to each other must be fixed too — just following from the chemical formula for the structure. And if one looks at a small strand of the chain only, this argument will be quite correct.

For example, Figure 2.3 shows the spatial structure of a little segment of a polyethylene macromolecule. One can see that the main chain is a sequence of carbon atoms connected with covalent bonds, and that each carbon atom is also joined to two hydrogen atoms. So in complete agreement with the naive chemical concept, the atoms of each monomer unit as well as the atoms of neighboring units are located in a well determined way with respect to each other[2]. And although the main chain bonds form a zigzag pattern, Figure 2.3 seems to suggest that overall chain shape should be more or less like a straight rod, as in Figure 2.2 *a*.

There is even a separate branch of research called conformational analysis of polymers. It deals with the geometry of atoms' positions in reasonably short chain segments, for much more complex structures than polyethylene, of course. An example is depicted in Figure C2.4: a strand of a DNA double helix. (We shall talk about DNA structure in more detail in Sections 5.5

[2]For the moment, we ignore the fact that the conformation of a polyethylene segment shown in Figure 2.3 is not the only possible one. A few different conformations can be realized because there are several rotational isomers of the molecule (see later). By the way, this is the main reason for the flexibility of polyethylene chains.

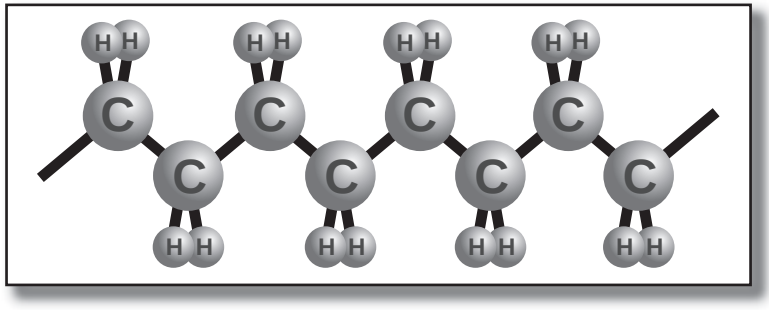


Fig. 2.3 Spatial structure of a polyethylene chain segment in the most energetically favorable configuration.

and 5.6.) The "portrait" of double helix is really one of the cultural icons of our time, it is found everywhere, from calendars and T-shirts to some architectural designs. The very fact that it is so easily recognizable indicates that each atom really does occupy a particular place.

At the same time, in reality, of course, the atoms of a molecule are not strictly fixed in their equilibrium positions. Indeed, if we think physically, the atoms may be pushed away from their equilibrium positions by a force or a kick, resulting, say, from thermal collisions between a given macromolecule and molecules of the solvent. Left alone after the kick, the atoms may also oscillate around these equilibrium positions. In a real system, changes in atoms' positions occur, firstly, because the bond angles (i.e. the angles between adjacent chemical bonds) can be deformed. Secondly, parts of the molecule can rotate with respect to each other, around the axes of single covalent bonds (but not around double ones). This rotation is sometimes expressed in terms of a molecule having a few different "rotational-isomeric forms". But the oscillations hardly ever alter the lengths of covalent bonds.

Thus, in many cases, you can regard a molecule as a construction of rigid rods, a bit like a miniature imitation of the Eiffel tower. The rods, representing covalent bonds, swing slightly from side to side, about the atoms, with angles between bonds changing. The amplitude of such bond-angle oscillations, as well as the probability of various rotational-isomeric forms, depends on the temperature. For example, at room temperature ($T \approx 300$ K) the oscillation amplitude of the bond angles ϕ for typical molecules normally varies from one to ten degrees: $(\Delta\phi)_{T\,=\,300\ \mathrm{K}} \sim 1^\circ \div 10^\circ$. Obviously, for an ordinary small molecule such oscillations would not appear too significant. Indeed, Eiffel tower also undulates a little bit, with

its top swinging by several meters in a windy weather, which does not seem like much when you either look at the tower from afar or sit in the famous restaurant on top. Similarly, in a short segment of a polymer chain only low-amplitude fluctuations occur. This is why the chain's flexibility is hardly noticeable at such a small scale, and short chain segments can indeed be depicted in the way shown in Figures 2.3 and C2.4.

At larger scales, however, all the small angle deformations add up along the chain and eventually result in the chaotic coiling of the polymer (Figure 2.2 *b*). Exactly how long the chain should be in order for the local fluctuations to result in global tangling up depends on the specifics of a particular chemical structure, but if the chain is long enough, then the coiling is inevitable.

2.3 Flexibility Mechanisms

As we have seen, any sufficiently long molecular chain does indeed have some flexibility, just because of its linear structure and considerable length. However, the nature of this flexibility may be different for different kinds of polymers. For example, the majority of the most commonly used synthetic polymers (including all those in Figure 2.1), as well as all protein molecules, have single C–C chemical bonds along their main chains. Such molecules appear flexible basically due to rotational isomerism, that is, because parts of a molecule may rotate around the single bonds. The main contribution to the discovery and study of this type of polymer flexibility was made by the physicist M.V. Volkenstein (1912–1992) and his group from St. Petersburg (at that time Leningrad).

A classic example of a polymer with a different flexibility mechanism is a DNA double helix (Figure C2.4). Since it consists of two entwined "threads", rotations in one of them are prevented by the other. So the only remaining way in which the chain can flex is by deformation of the angles between the bonds. Each bond angle gets distorted slightly, and so the flexibility is distributed fairly uniformly along the double helix. Nowhere may there be a kink or a right-angle bend, for example. DNA therefore looks like an elastic wormlike thread as shown in Figure 2.5 *a*. A model chain in Figure 2.5 is called a worm-like chain and is used to describe flexibility of this sort (sometimes it is also called Kratky–Porod model, after the researchers who introduced it in 1949). The fact that double helical DNA is well represented by a worm-like chain model is nearly obvious upon a

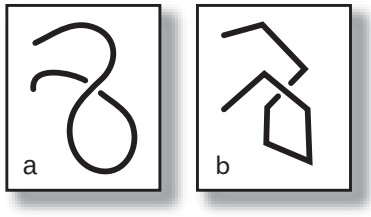


Fig. 2.5 A worm-like polymer chain (*a*) and a freely jointed polymer chain (*b*) — two simple and most common models of polymer chain flexibility.

single glance on Fig. C2.4; but for a while this fact seemed abstract, remote of any practical use. This view changed when in 1992, C. Bustamante and his co-workers at the University of Oregon performed a completely new type of experiment — they were able to measure elasticity of a single DNA molecule! Their results turned out possible to understand in terms of worm-like flexibility.

Yet another, and maybe the simplest, model for a polymer's flexibility is the so-called freely jointed chain. This is a sequence of rigid rods, each of length ℓ, joined together with freely rotating hinges as sketched in Figure 2.5 *b*. Such hinges hardly ever occur in a real polymer. However, as long as one is only interested in large-scale properties of a polymer coil, then the particular nature of the chain flexibility ceases to be important. (In Section 6.5, we are going to discuss why this independence of the details holds in most cases, and why it sometimes fails.) Therefore, for the sake of simplicity, we shall use the freely jointed model in this book to explain some concepts and results.

2.4 A "Portrait" of a Polymer Chain

A typical conformation of a freely jointed chain consisting of a great number of units is shown in Figure 2.6. You can easily create a similar pattern yourself; if you have access to a personal computer, it is also a good exercise in programming! However, if you only have a sheet of paper we suggest the following routine. Draw a straight line of unit length, let's say, 1 centimeter. Then choose some random direction; you can do this, for example, by depicting a kind of a "wind-rose" (i.e. a diagram of the relative frequency of wind directions at a place) with six directions, numbering them in order from 1 to 6, and then tossing a die. (On a computer, instead of a die, you would simply use a random number generator.) Now, starting at the end of your straight line, draw a new one of the same length in the chosen direction,

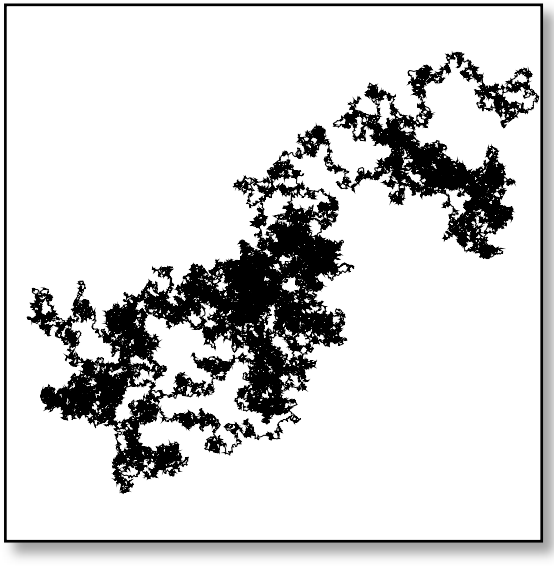


Fig. 2.6 A typical conformation of a polymer coil. The freely-jointed chain of 10^6 segments has been simulated computationally in three-dimensional space. Two-dimensional projection shown could have appeared as a Gaussian random walk on the plane, except the length of each step on the two-dimensional projection is not constrained to be unity. The figure is courtesy of S. Buldyrev.

and repeat this operation many times (i.e. choose a random direction again, independently from the previous one, add another straight line, etc). As a result, you get a "portrait" of a polymer chain, just like the one in Figure 2.6. Actually, this figure was indeed obtained by a very similar procedure (on a computer); the only difference is that the "wind-rose" had many more than six different directions, and it was situated in three-dimensional space rather than on a plane.

Looking at Figure 2.6 you might think that you have already seen something similar when studying molecular physics. You would not be wrong, although there is no chapter on polymers yet in most textbooks on molecular physics. However, Brownian motion is included in all of them. They often show a photograph, made with a microscope, of the random path of a tiny dust particle suspended in a fluid and buffeted chaotically by numerous molecules. Such a random walk and the polymer conformation in Figure 2.6 are as alike as two peas in a pod. Why should this be the case? We are going to find out in Chapter 6.

Figure 2.6 also makes it clearer how a polymer chain tangles up into a random coil due to its flexibility (as we have already discussed, see Figure 2.2 *b*). One can reproduce the same kind of pattern using any model for a chain's flexibility, it does not have to be a freely jointed one.

Having read all this you may be wondering why at the very beginning of our polymer story we started talking about such things as bending and the shape of polymer chains. In fact, bending of chains (or, in other words, their conformation) plays a key role in the properties of polymers. Nearly all of this book is a collection of examples of this, but here we shall give only one simple illustration. DNA molecules in human chromosomes are almost about a meter long (there is quite a lot to be recorded there, hence the considerable length!) If DNA chains were not flexible but rigid like spokes, how could they be packed and kept in a cell nucleus as small as one micron, or 10^{-6} m? As Figure C2.7 suggests, this is the problem even for a bacteria: once outer shell of bacteria is destroyed, DNA spills out; it must have been very dense inside given how much gets out!

2.5 Heteropolymers, Branched Polymers, and Charged Polymers

You now know that what is special about polymers is their chain structure, great length, and flexibility. These are common features of all polymers. They cannot explain everything though. One complication is that each monomer unit has a particular chemical structure; besides that, there are three major physical facts which make things more intricate, as we shall now discuss.

2.5.1 *Heteropolymers*

Simple polymer chains, such as the ones in Figure 2.1, consist entirely of identical monomer units and are sometimes referred to as homopolymers. However, some macromolecules are built of monomer units of a few different sorts. They are known as heteropolymers, or copolymers as chemists say (we shall use both terms interchangeably)[3]. Most interesting and important

[3]Once again, there is a terminological subtlety, largely due to historically different chemistry and physics cultural traditions. Chemists pay much attention to the fact that some polymers (including all examples of the Figure 2.1) have only carbon atoms in their main chains, while main chains of other polymers include the so-called hetero-atoms, that is, atoms other than carbon, such as nitrogen, oxygen, etc. Practically important examples include most plastics, cellulose, biopolymers of DNA and proteins, etc. There is special name for the latter type of compounds — heterochain polymers, but chemists sometimes also call them heteropolymers. As always, we in this book stick to the simplified terminology, which in this case also universally adopted not only in physics, but also in biophysics: by our definition, heteropolymers are the same as copolymers — chains of more than one type of monomers.

among them are biopolymers such as DNA (having four different types of monomer), and proteins (20 different types). The sequence of monomers along the chain forms the primary structure of this chain. One can compare the primary structure of a biopolymer to a sequence of letters in a long line of very interesting and informative book written in a language which we do not yet completely understand.

Some heteropolymers are not biological, but are artificially synthesized. Their primary structures, in the spirit of our previous comparison, resemble a book that a monkey would have created if it were allowed to use a typewriter. It would either be a totally random sequence of characters (i.e., a statistical copolymer) or a number of blocks of repeating identical letters, such as "$\mathcal{BBBBBZZZCCCC}$" (i.e., a block-copolymer) or maybe a simple periodic sequence, such as "$\mathcal{ABABABABABABABAB}$" (the latter can be also treated as a homopolymer whose repeating units, or monomers, are $\mathcal{AB}$ each). Lack of "sense" in their primary structures, by the way, does not prevent random and block-copolymers from having some very interesting physical properties, or from being widely used in applications.

2.5.2 *Branched Polymers*

Together with simple linear chains, polymer science also deals with branched macromolecules. They can have the shape of combs (Figure 2.8 *a*), stars (Figure 2.8 *b*), or an even more complicated structure (Figure 2.8 *c*). Another species of this kind is a macroscopic polymer network (Figure 2.8 *d*) which takes the idea of branching to its extreme. This huge molecule emerges when lots of entangled polymer chains are chemically connected,

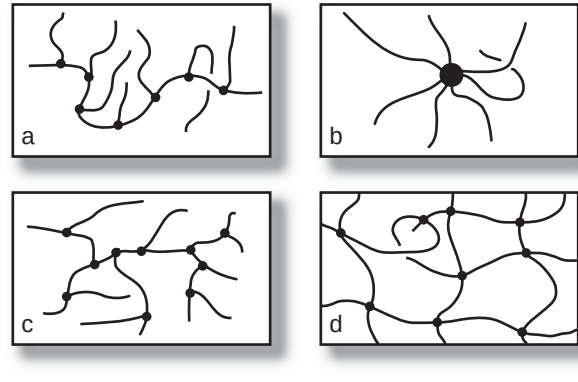


Fig. 2.8 Branched macromolecules: (*a*) a comb, (*b*) a star, (*c*) a randomly branched chain; (*d*) a polymer network.

or cross-linked, with each other (see Section 3.5). It can be many centimeters across. Scientists have a special word for it: a gel. Meanwhile, chefs, who may not even suspect that they are talking about polymer networks, use the same word in a slightly different form: jelly! So, when you are eating your favorite jell-o, you hold in your hands a single molecule: isn't it a *giant molecule*?! (Well, a rigorist would say that jell-o is not really a single molecule, for there are many small molecules of water and others inside... The rigorist is, as always, right, but we are right, too: at least one of the molecules in jell-o is big.)

2.5.3 *Charged Polymers*

None of the polymers depicted in Figure 2.1 contains electrically charged monomers. However, there are some polymers whose monomers may lose low molecular weight ions and become charged. Polymers of this sort are called polyelectrolytes, and the ions which break off are usually known as counterions.

The simplest of the polyelectrolytes are polyacrylic and polymethacrylic acids (Figure 2.9). When in solution in water, if an alkali is added, the monomers of these polymers dissociate and become negatively charged. Biopolymers, such as DNA and proteins, are also polyelectrolytes in their natural aqueous environment — DNA's chain has a large negative charge, whereas the monomers in proteins can be neutral or carry a positive or

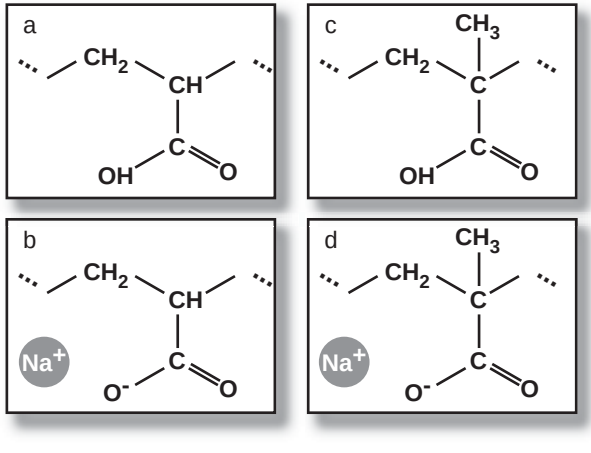


Fig. 2.9 A monomer unit of polyacrylic (*a*,*b*) and polymethacrylic (*c*,*d*) acids in the neutral (*a*,*c*) and charged (*b*,*d*) forms. The way a unit gets an electric charge is by dissociation in water solution if you add an alkali (e.g., NaOH; in this case the role of counterions for the charged units (*b*) and (*d*) is played by the Na^+ ions).

negative charge, depending on the type of monomer and composition of the solvent. By the way, such chains containing both positive and negative monomers also have special name, they are called polyampholytes.

2.6 Ring Macromolecules and Topological Effects

Some polymer molecules can have the shape of a loop or a ring (Figure 2.10 *a*). Studying these, it is important to remember that parts of such closed chains cannot go through each other (Figure 2.11) in the way that ghosts, or phantoms, would do. In other words, as they sometimes say in scientific literature, the chains are "not phantom". Hence, the number of conformations in which a ring molecule can appear in its thermal motion is restricted. Anything that one can obtain from the original shape by various movements and deformations is allowed, but not the passing of the chain through itself. The mathematical properties of such objects are studied in a course on topology and are therefore called topological properties.

However, we do not even need to know topology to understand that a ring molecule can be tied into a knot of some sort (Figure 2.10 *b*). A few rings can form various entanglements with each other (Figure 2.10 *c*). A

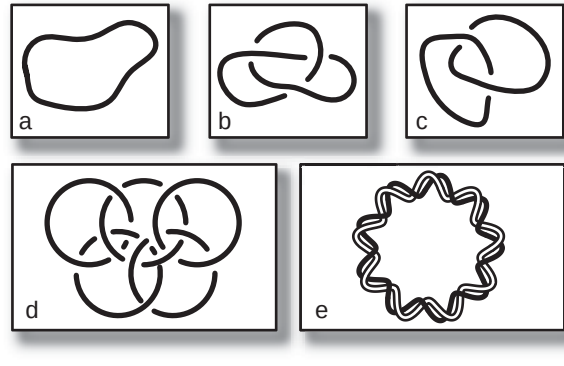


Fig. 2.10 An unknotted (*a*) and knotted (*b*) ring macromolecule. The link of two ring macromolecules (*c*). An Olympic gel (*d*). The tangling of two complementary strands into a double helix (*e*).

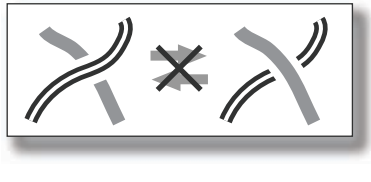


Fig. 2.11 An impossible type of motion: two chains or two segments of the same chain cannot go through each other.

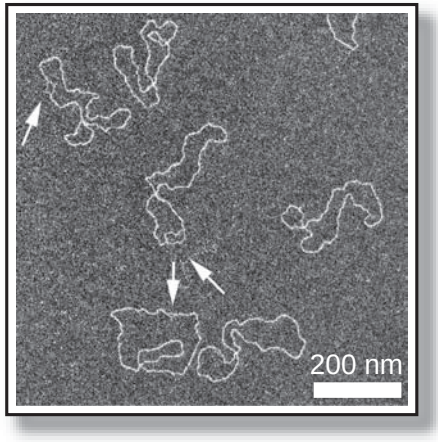


Fig. 2.12 Rings of DNA about 3,000 base pairs long. The sample was prepared at very low concentration of salt in water, this led to very strong electrostatic repulsion between negatively charged phosphate groups on the opposite DNA strands, which is why double helix was unwound in several places indicated by the arrows. Electron microscopy image is courtesy of D. Cherny. Another image of a DNA ring, Figure 11.3, having a knot, will be discussed later in Chapter 11.

peculiar thing about Figure 2.10 *c* is that the molecules are not connected with chemical bonds, yet cannot be easily separated. Even such a thing as the so-called Olympic gel (Figure 2.10 *d*) can in principle exist. It looks like a kind of molecular chainmail, and obviously acquired its name due to its resemblance to the coupled rings of the Olympic emblem. Of course, there are the same sort of topological constraints in polymer networks too (See Figure 2.8 *d*).

We will discuss more about polymer knots in Chapter 11, but cannot postpone mentioning one of the reasons why topological effects are of special interest: natural DNA molecules normally, and perhaps even always, have a ring shape (Figure 2.12). The two strands of the double helix form a link of a very high order as shown in Figure 2.10 *e*. You may get some idea of how important the topology is from the following fact. Living cells have "provided" themselves with special topological enzymes which can do rather intricate jobs. They can, for instance, break one of the strands of a ring-shaped DNA molecule, then use some energy to "rearrange" the double helix by twisting it a particular extra number of times, and finally "heal" the break. Obviously, this is not just accidental, but is done for some good reason.

We should also mention in passing that some organisms, called kinetoplasts, have their genomes in the form of Olympic gel type construct of many DNA rings.

Linear polymer chains (of an open rather than a closed shape) are certainly not topologically constrained in the same sense. They can always come together or move apart. On the other hand, you have probably sometimes had to wrestle with a bundle of entangled ropes or cables. We all know how time-consuming this is. And the knowledge that, in theory, ropes can be separated does not really help! So, based on this mundane experience, we may expect that systems of densely entangled linear chains should exhibit rather interesting and unusual dynamic behavior. We shall talk about this in Chapter 12.

Chapter 3

How are Polymers Made?

> Mein Märchen ist aus, dort lauft eine Maus, wer sie fängt, darf sich eine große Pelzkappe daraus machen.
> (My tale is done, there runs a mouse, whosoever catches it, may make himself a big fur cap out of it.)
>
> Grimm Brothers,
> *Hänsel and Grethel*

We have talked about different types of polymer molecules. Now it seems the right time to ask how all these various types are actually made, ranging from the simplest linear polymer chain to a polymer network of a complex, densely entangled structure.

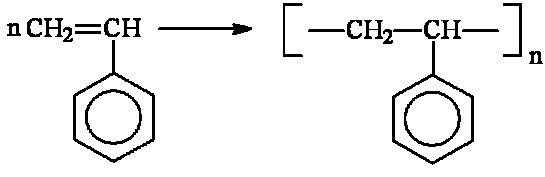
In a living cell, chains of biopolymers (DNA, RNA, proteins, polysaccharides) are built by special systems in an enzyme-mediated process called biosynthesis. This is a very robust process. Suffices it to say that if we fully stretch all DNA macromolecules synthesized in the human body during the life period the total length turns out to be of astronomical scale: two light-years! And all the macromolecules forming this way are practically identical.

The synthesis of artificial polymers is much less robust. This is a major task of polymer chemistry. This book, however, is meant to concentrate on physics, so we shall not discuss this question in any great detail. Nevertheless, it might help to have some general idea of the methods of polymer synthesis. It would allow us to understand physical properties of polymers better and more profoundly.

Long polymer chains are synthesized from low molecular weight compounds that are monomers. There are two main methods of synthesis: polymerization and polycondensation.

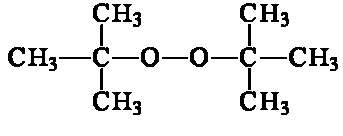
3.1 Polymerization

During polymerization, monomers are joined successively (one-by-one) to the main chain, according to the rule $A_N + A \rightarrow A_{N+1}$. For example, polystyrene (Figure 2.1 *b*) is obtained through polymerization of styrene:



(3.1)

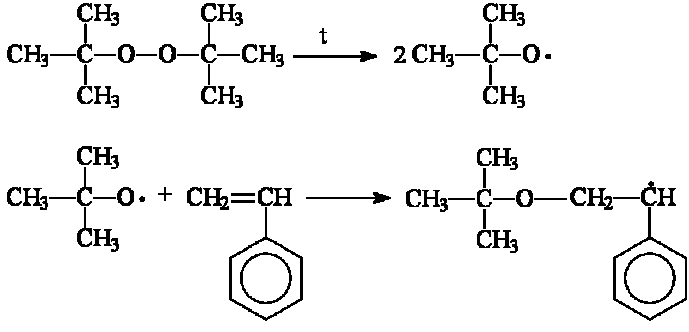
It would be natural to ask here what conditions are needed for a chain to start growing. And how does the process stop? A reaction like (3.1) cannot begin of its own accord. To start such a reaction the active center (it may be a free radical, cation or anion) should be produced first, for this purpose chemists are normally using the so-called initiators — special substances which can generate active species. In a simple example the initiators easily decompose and form free radicals, i.e. molecules containing unpaired electrons; the reaction initiated this way is called free-radical polymerization. Typical initiators for free radical polymerization are compounds with a labile bond, e.g. peroxide −O−O−; hydrogen peroxide is the most well-known example, but most widely used in the reactions of the type (3.1) are organic peroxides, e.g. di-*tert*-butyl peroxide:



(3.2)

The free radicals are usually highly reactive, because of unpaired electrons. In particular, they can react with the double bond in the compounds like styrene. In such bond one electron pair is held securely between the two carbon atoms (σ-bond). The other is more loosely held (π-bond). When

the free radical is approaching to styrene molecule, another stable σ-bond is formed instead of a π-bond, but the unpaired electron is transferred to monomer unit, as it is shown in the following reaction of di-*tert*-butyl peroxide with styrene:



(3.3)

The resulting free radical can react with another styrene monomer; the location of the radical is transferred to it, etc. In this way polymerization continues by itself with no outside help, making the chain longer and longer. The free radical is always located at the growing chain end. This process is called chain propagation.

From this example we can discern the main features of the polymerization process. First, to enable this kind of synthesis, a monomer molecule has to have a double (or triple) chemical bond. Second, the whole process is merely a rearrangement of chemical bonds between the molecules (e.g., a double bond transforms into two single ones). This is why no byproducts are normally created during polymerization, and the growing molecule in most cases consists of exactly the same atoms as the initial compounds. Third, from each free radical one polymer chain can emerge, therefore to get longer chains one should normally decrease the concentration of initiator and increase the concentration of a monomer.

If an active center at the end of a chain ceases to exist (say, unpaired electron of a free radical becomes passivated), then the chain stops growing. It is said that in this case polymerization terminates. The termination process can happen spontaneously (if, for example, the ends of two independently growing chains meet together and react forming a combined chain), but it can also be deliberately stimulated by special substances called inhibitors. An active center can also be transferred from one macromolecule to another (so called transfer reaction); in this case the macromolecule loses

its capability to propagate and stops to grow, but simultaneously the new active chain appears and its propagation starts. Obviously, the chain will also stop growing upon the exhaustion of monomer supply. In one way or another, the chain propagates and eventually grows into a macromolecule, for instance, like polystyrene shown in Figure 2.1 *b*.

Polymerization is used to obtain most common polymers shown in 2.1. Polymerization of styrene (3.1) goes on at rather mild conditions (around 100°C, normal atmospheric pressure, benzene or toluene used as a solvent). However, for polymerization of ethylene more extreme conditions are needed (300°C and 2,000 atm).

Heteropolymers can also be synthesized in this way, but, of course, there should be different types of monomers in the mixture.

3.2 Polycondensation

Polycondensation is rather different than polymerization: segments of a polymer chain, with functional groups at the ends, gradually join on to each other: $\mathrm{a{-}A_N{-}b} + \mathrm{a{-}A_M{-}b} \rightarrow \mathrm{a{-}A_{N+M}{-}b} + \mathrm{a{-}b}$. Polycondensation can be visualized if one thinks of a group of people holding hands to form a human chain: each person has two hands, and they correspond to two reactive functional groups in polymer language.

An example of such a process is the reaction between terephthalic acid ($\mathrm{HOOC{-}C_6H_4{-}COOH}$) and ethylene glycol ($\mathrm{HO{-}C_2H_4{-}OH}$). At certain conditions COOH and OH groups of these molecules can react, forming ester COO bond:

$$\begin{aligned} \mathrm{HOOC{-}C_6H_4{-}COOH} &+ \mathrm{HO{-}C_2H_4{-}OH} \\ &\rightarrow \mathrm{HOOC{-}C_6H_4{-}COO{-}C_2H_4{-}OH} \\ &+ \mathrm{H_2O}\ . \end{aligned} \tag{3.4}$$

The resulting compound will grow further, because there are potentially reactive functional groups COOH and OH at the ends of it. It is clear that in this process the chain segments of different lengths will be formed, and these segments can further join each other with the formation of a new ester bond and a longer united chain having the following structure of repeat monomer unit: $\mathrm{{-}CO{-}C_6H_4{-}COO{-}C_2H_4{-}O{-}}$. This polymer is called polyethylene terephthalate. Everyone knows this polymer very well from ordinary life, since plastic containers for cold beverages are made of it. Also, when you buy clothes and want to check the fraction of synthetic

fibers in the fabric, this is normally designated "polyester $x\%$". In this case polyester is just another word for polyethylene terephthalate (this is a class of polymers to which this macromolecule belongs).

During polycondensation low molecular weight substances are normally produced as a by-product. (For instance, in the reaction (3.4) it happens to be water.) This is why, in contrast to polymerization, the composition of a growing molecule changes compared to that of the initial compounds. Another special feature of polycondensation is that long chains can form only if most of functional groups (like COOH and OH groups in the example (3.4)) have been involved in reaction. However to achieve this result the monomers should be taken in equal amounts. Otherwise we will end up with short fragments (oligomers) of polyethylene terephthalate chains. In the analogy with human chain, the situation with slight excess of one of the reactants corresponds to a chain of men and women in which every man can only hold hands with a woman, and vice versa. It is clear that in this case the excess of either men or women will lead to finite chains, and the more is the excess the smaller is the average chain length.

Sometimes it is necessary to deliberately ensure lower length of the obtained polymer (for example, for better processibility). For the reactions similar to (3.4) one of the methods to reach this aim is to add one of the reactants in slight excess. The polycondensation will then proceed up to a point when one reactant is completely used up and all the chain ends possess the same functional group (either COOH or OH). Another method to achieve desired chain length at polycondensation is the addition of a small amount of a monomer with only one functional group (chain terminator).

It is interesting to mention that the first truly synthetic (not based on natural products) polymer material was bakelite obtained in 1907 via polycondensation of phenol and formaldehyde. This material had good dielectric properties and was used mainly as an electrical insulator. The most famous polycondensation polymer is probably nylon belonging to the class of polyamides. Other common classes of polycondensation polymers are polyesters (like polyethylene terephthalate), polysiloxanes, polycarbonates, polysulfides, polyethers and polyimides.

3.3 Catalysts for Polymer Synthesis

An important aspect of polymer synthesis is the use of different catalysts to facilitate polymerization and polycondensation. For example, in the

so-called coordination polymerization monomer can add to a growing chain end only if it forms a complex with the organometallic active center. The most famous catalysts for coordination polymerization were discovered by Karl Ziegler (1898–1973) and Giulio Natta (1903–1979) in the 1950s (Nobel Prize in 1963), they are based on titanium tetrachloride and methylaluminoxane. By the use of these catalysts it was possible, for example, to obtain polypropylene macromolecules which are much longer, not branched and stereoregular (i.e. not atactic — see Section 4.2 for more details). Another important class of catalysts for polymerization are metallocenes; the reader wishing to learn more about those has to consult the books on polymer chemistry.

We would like to add here only one remark on the catalysis for polymerization processes in living nature. For example synthesis of DNA is catalyzed by the protein called DNA polymerase. Contrary to the systems mentioned above, this proceeds at room temperatures, normal pressures, and does not involve metals. This is another testimony to the fact that Nature in the course of molecular evolution has done a great job and scientists have still a long way before them to improve their methods.

3.4 Polydispersity, Living Polymerization

It would seem that polymer chains constructed from monomers as a result of random chemical reactions should have a rather wide distribution in their lengths. This is indeed true, and the name for this phenomenon when chains of various lengths coexist in a polymer substance is polydispersity. Polydispersity has to be borne in mind when analyzing polymer properties. In practice, there are some ways to reduce polydispersity by separating chains with different length.

An interesting method to obtain polymers with relatively narrow polydispersity without additional separation processes is based on the so-called living polymerization. This is a polymerization for which the ability of a growing chain end to terminate or to transfer its active species to another molecule has been removed. To reduce polydispersity the rate of chain initiation should be much larger than the rate of chain propagation. As a result practically all the chains grow simultaneously at a more or less constant rate until the monomer is exhausted. Living polymerization is a popular method for synthesizing block copolymers (see Sections 2.5 above and 4.7 below), since growing chain ends remain active after the process

with initial monomer is completed (no chain termination). Therefore, at this stage it is possible to add another monomer, and a new chain block will start to grow.

First example of living polymerization was discovered by Michael Szwarc (1909–2000) in 1956 in the anionic polymerization of styrene in special catalytic system. In this type of polymerization active chain end is negatively charged which prevents most of termination processes.

Anionic and later cationic polymerization gave most of examples of living polymerization systems until recently, when more sophisticated methods of manipulation with free-radical polymerization processes become available. These methods are based on the use of the compounds which reversibly react with propagating radical and convert it to the so-called "dormant species". When the equilibrium between the active and dormant species is regulated by special catalysts based on a transition metal, this process is called atom transfer radical polymerization (ATRP). If this equilibrium is provided by stable radicals such as nitroxides, the process is called stable free-radical polymerization (SFRP). In the case when dormant species are formed via a chain transfer rather than reversible termination reactions, this process is referred to as reversible addition fragmentation chain transfer (RAFT) polymerization. All these techniques allow to produce macromolecules of desired architecture and molecular masses.

3.5 Branched Polymers

Let's now talk briefly about branched polymers. If, say, polycondensation is going on, and initial monomers have only two functional groups each, then we shall end up with linear polymer chains (with a small proportion of loops). However, if the monomers have three or more functional groups, a branched macromolecule can be synthesized (see Figure 2.8 *c*). Given plenty of "multifunctional" monomer units at the start, one can even obtain a polymer network (Figure 2.8 *d*).

Branched macromolecules and polymer networks can also be formed by the cross-linking of linear chains. There are various chemical ways of cross-linking. Sometimes chemically active cross-linking agents are used; they establish covalent bonds between different chain strands. Alternatively, ionization in a polymer system can be stimulated by radiation, etc. The simplest everyday life example of cross-linking is vulcanization — during

this process viscous natural rubber becomes a highly elastic polymer network (see Chapter 7 for more details).

Chapter 4

What Kinds of Polymer Substances are There?

> . . . they knew how to weave stuffs of the most beautiful colors and elaborate patterns . . .
>
> Hans Christian Andersen,
> *The Emperor's New Clothes*

4.1 "Traditional" States of Matter and Polymers

We now can easily visualize polymer molecules — they are long chains tangled up into coils. However, even knowing the molecular structure of a substance, it may still be hard to predict for sure all its properties. For example, water, consisting always of the same well-known molecules H_2O, depending on the conditions can be a liquid, a solid (ice), or a gas (steam). So what about polymer substances? How do they look, and what states can they exist in?

Everyone is familiar with the three simplest ordinary states of matter: solid (crystal), liquid and gaseous. There is also a fourth one, a plasma. Normally it emerges at extremely high temperatures when thermal motion is so intense that it leads to the ionization of the atoms. If a polymer is heated up to such a temperature, its molecular chains will merely fall apart. So the substance will not be polymeric any longer, or, in other words, the destruction of the polymer will occur. Thus, the state of a high-temperature plasma is not possible for polymers.

It seems we are left with the three "traditional" states of matter after all. This sounds too small a number though, if we try to imagine all the diversity of polymer substances in everyday life: there are plastics and

rubber, fibers and fabrics, timber and paper, polymer films and varnishes, dyes and paints, not to mention the various polymers found in nature! You would be right to suspect therefore that the common concept of the three states of matter is not quite applicable to polymers, especially as two of the three states, a gas and a crystal, are not really typical for polymers.

Indeed, if one wanted to create a polymer gas, one would have to make long heavy molecules (like the ones in Figure 2.1) "fly" around. This would only be possible if there were no gravity, and also if you could maintain a low pressure in the container (i.e. you would have to provide a high vacuum there). Obviously, such exotic conditions are very hard to achieve; this is exactly why polymer gases have not been heard of so far.

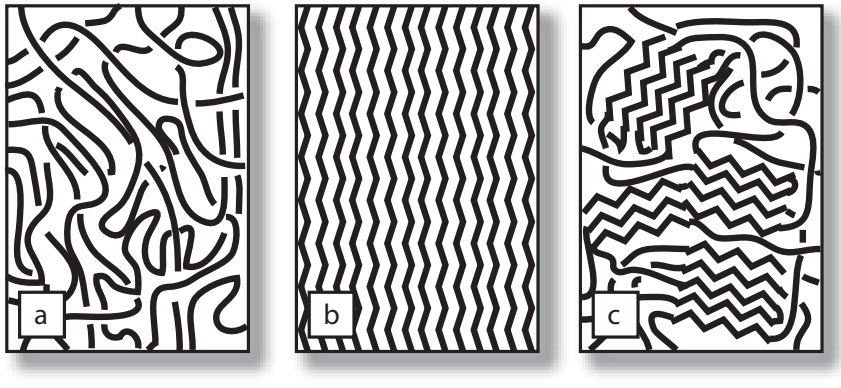


Fig. 4.1 A cartoon of molecular arrangements in typical polymeric substances: (*a*) a liquid; (*b*) a perfect crystal; (*c*) a partially crystallized polymer.

Perfect single crystals (see Figure 4.1) cannot be obtained from polymers for another reason. Let's experiment with a liquid of polymer molecules (Figure 4.1 *a*). If we cool it down to below the crystallization temperatures, then the perfect crystal (Figure 4.1 *b*) will be energetically the most favorable state. It cannot be formed straightaway, though. Crystallization goes on totally independently in different parts of the system. So what appears at the start is a number of crystalline "nuclei" randomly orientated with respect to each other. Clearly, when the nuclei grow big enough the entire structure becomes somewhat "frozen". (This is because, in the crystalline phase, in order to move with respect to each other polymer chains have to overcome enormous energetic barriers.) Hence, further evolution towards the perfect structure of Figure 4.1 *b* appears hardly possible. This

is why crystallizing polymers normally form a semi-crystalline phase so that crystalline regions are separated by amorphous layers (Figure 4.1 *c*). Sometimes perfect single crystals of a polymer can still be obtained by special techniques, but they have not found any broad practical use.

4.2 Possible States of Polymer Substances

Do we have to class all polymers as liquids, now we know that they can be neither gases nor, except rarely, crystals?! In the broad sense, we would — if we only regard a liquid as a dense substance that has no long-scale order in the atoms' positions. However, this definition would not be terribly informative. This is why there is another, more fruitful way to classify polymers' phases. A distinction is normally made between a semi-crystalline state, a polymer glass, an elastic, and a viscous polymer. Which of the four phases occurs depends on the kind and strength of interactions between the monomers.

We have already talked about semi-crystalline polymers. Let's now describe in brief the other three states. A polymer in a viscous state is purely a liquid of macromolecules as sketched in Figure 4.1 *a*. Long chains all mingle together, but, in thermal motion, they can rather easily move with respect to each other. If an external stress is applied, some overall motion of the molecules occurs, i.e. the polymer starts to flow. The flow develops quite slowly, due to a great number of entanglements. This explains why the viscosity of polymeric liquids is normally rather high. Naturally, this state of a polymer is called viscous; another name for it is a polymer melt.

Let's now see what will happen if molecular chains of a polymer melt are joined together with covalent chemical bonds (cross-links) to form a network (see Figure 2.8 *d*). (We talked about different techniques of how to synthesize a polymer network in Chapter 3.) Clearly, the chains will no longer be able to move long distances relative to each other (simply because they will all be tied together into a network). Thus it becomes impossible for the polymer to flow. Meanwhile, on a smaller scale (i.e. shorter than an average distance between two neighboring cross-links) the mobility of the chains will not be constrained by the cross-links. This is why, if you apply tension to a polymer network, its chains, which were originally coiled up (Figure 2.8 *d*), stretch quite considerably, resulting in exceptionally large elastic reversible deformations. This state of a polymer is called elastic. Rubber is, obviously, a well-known example of it.

Cross-linking of chains in an elastic polymer does not necessarily have to be caused by covalent bonds between neighboring molecules, however. The role of effective cross-links can be performed by nuclei of a crystalline phase (Figure 4.1 *c*), or by topological entanglements (Figure 2.10 *c* and *d*). It can also be played by some small regions where, due to particular local conformations of the chains, there are comparatively high potential barriers for the chains to move with respect to each other ("glassy" or "frozen" regions). Thus, an elastic polymer substance can in principle be produced without chemical cross-linking.

If the temperature decreases, many polymers tend to change from a melt to a semi-crystalline state. However, far from all polymers crystallize when they are cooled. The crystal formation begins when little crystalline seeds start developing. This happens when, on the one hand, the crystalline phase is thermodynamically favorable but, on the other hand, the thermal motion is enough to enable the rearrangement of the polymer chains to form seeds. If the cooling is fast enough, we can easily avoid that stage, and so a crystal does not form. This statement is also true for substances with low molecular weights. The new thing for polymers is that the "fast" cooling does not necessarily have to be very fast in the usual sense of the word. As you can see from Figure 4.1 *c*, it takes much more time to form a crystalline seed for heavily entangled chains than for atoms or small molecules which are not chained together.

Moreover, some polymers cannot be crystallized even in principle. Indeed, crystallization may only appear if there is long-scale order in the molecules' positions (as in Figure 4.1 *b*). However, say, for a statistical copolymer whose chains consist of two types of units, $\mathcal{A}$ and $\mathcal{B}$, long-scale order is impossible. (This is simply because the sequences of $\mathcal{A}$ and $\mathcal{B}$ along the chains are totally random.) Such copolymers can never crystallize on cooling.

The same effect is observed for homopolymers whose monomers, although chemically identical, may appear in a few different spatial configurations, with high barrier for the inter-conversion between these states. As an example, Figure 4.2 shows two possible configurations of the repeat unit of propylene. If the synthesis is carried out under usual conditions, these two configurations will be present in roughly equal proportions, and will alternate randomly along the chains. This kind of polypropylene is called atactic; obviously, it cannot crystallize. Yet there is a special technique for synthesizing the so-called isotactic polypropylene instead, whose monomers

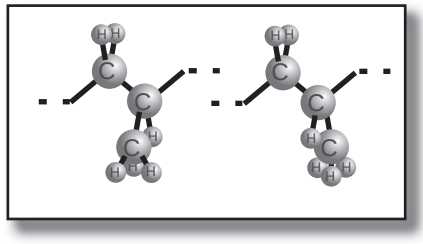


Fig. 4.2 Two possible configurations of the monomer unit of propylene.

are all arranged in only one of the two possible configurations. Crystallization is then quite straightforward.

There are many other polymers that, like polypropylene, are atactic (i.e. unable to crystallize) if synthesized under normal conditions. This includes, for instance, polystyrene, polymethyl methacrylate (perspex), and polyvinyl chloride (PVC), to mention but a few from everyday life.

So crystallization does not happen, but what kind of processes do happen in the "noncrystallizable" polymers if the temperature is reduced? Since thermal motion becomes less, energy barriers for relative motion of the molecules grow effectively higher and higher. Gradually, this motion becomes "frozen out", first of all at the largest scale, and then at increasingly smaller ones. In the end, any thermal motion at any scale larger than the size of a monomer ceases to exist. Polymers in such a "frozen" state are known as polymeric glasses, and the process that we have just described is called a glass transition. It normally occurs in a rather narrow temperature range around the "glass transition temperature" T_g.

Thus, polymers which are unable to crystallize tend to become a glass at low temperatures. You may be quite familiar with those more or less transparent glasses made from the atactic polymers already mentioned — polystyrene, perspex and PVC. (For the first two of them $T_g \approx 100°\text{C}$, whereas for the last one $T_g \approx 80°\text{C}$.) However, ordinary silicate glasses (used for windows, for example) are not polymeric; they are made of low molecular weight compounds such as silicon dioxide, oxides of boron, sodium and calcium. Such a mixture forms a glass on cooling, roughly in the same way as polymers.

The very term "glass" leaves no doubt that the substances are mainly used as transparent partitions. You may wonder why the majority of polymeric glasses are actually transparent, whereas semi-crystalline polymers are normally not. The reason is that the structure shown in

Figure 4.1 *c* has a great number interfaces between the crystalline and amorphous phases. Light gets reflected by the interfaces many, many times, and eventually gets totally "lost" between them. As a result, the sample does not let light out, hence the lack of transparency. Meanwhile, so-called acrylic glasses, say, have a much more uniform structure, or at least the scale of inhomogeneities is much smaller than the wavelengths of visible light. Therefore, light can penetrate through such substances without being scattered — which is why some nice tableware is made of this material.

4.3 Plastics

All four states of polymers are very important from the practical point of view. Rubber substances are used in their elastic state, for example; we shall talk about them in detail in Chapter 7.

In this section we are going to look at plastics, the materials that we know very well from everyday life. Over the last few decades they have become widely used in industry too. The worldwide yearly production of plastics exceeds now 150 million tons. We hardly imagine life without these materials: what else can we use to make pens or water bottles? At the same time, plastics are endangering the environment: about 10 million tons of them are now floating in the Ocean alone, and no one knows how to deal with that. Concentrate on usefulness or on dangers — one thing is clear: we have to understand plastics. Thus, what are they?

By definition, plastics are those polymers which whilst being processed are either elastic or viscous, whereas the materials in the actual use have to be either glasses or semi-crystalline substances. How can a polymer be transformed from one state into another? In many cases, it is done by changing the temperature: A polymer can be processed at elevated temperatures, where it is a viscous liquid, and afterwards cooled to become glassy or semi-crystalline. Materials produced in such a way are called thermosoftening plastics. However, some polymers tend to show the opposite behavior — they become solid with increasing temperature. For instance, epoxy resin mixed up with a hardener very quickly becomes solid if heated up. This is simply because cross-links are formed more rapidly at higher temperatures. Such materials are sometimes called thermosetting plastics.

Thus, all four states of polymers share "responsibility" for the properties of plastics. Some of them are involved at the production stage, and others

come into play when the plastic is put to practical use. We should also point out in what way the properties of glassy and semi-crystalline plastics differ from each other. Semi-crystalline thermosoftening materials (such as polyethylene, terylene, nylon, and teflon) are much more deformable and elastic, and much less fragile, than polymeric glasses. Normally they are not transparent either, but, in contrast to rubber, they tend to retain their shape under moderate deformation.

The extent to which solid materials can be deformed is described in physics by Young's modulus, E. It is defined in the following way. Let's imagine that we are stretching a cylindrical rod of length ℓ and cross-sectional area S, applying a force f along the axis. As the English scientist Robert Hooke noticed as early as 1660, the deformation $\Delta\ell$ of the rod (i.e. variation of its length) is proportional to the force (provided that $\Delta\ell$ is not too big),

$$\sigma = \frac{f}{S} = E\frac{\Delta\ell}{\ell} \ . \tag{4.1}$$

In this formula σ is the stress, i.e. it is the force per unit cross-sectional area, and E is Young's modulus. The value of E depends on the material of the rod, but not on its shape or size.

Let's now look at some materials that we are going to talk about later in this chapter, and see what sort of values of E they have at room temperature. As a point of reference, it makes sense to choose the hardest inorganic substances, such as steel, cermet alloys, etc. Their Young's moduli range from 10^{11}–10^{12} Pa[1] Inorganic glasses (as used in windows) have E in the range of 10^{10}–10^{11} Pa. Meanwhile, for polymeric glasses typical values are $E \sim 10^9$–10^{10} Pa which means that their deformability is two orders of magnitude higher than that of steel. As we have already said, semi-crystalline plastics are even more easily deformed; indeed, they have $E \sim 10^8$–10^9 Pa. As for various sorts of rubber, as well as other polymers which are normally used in their elastic state, their Young's moduli tend to be exceptionally low: $E < 10^6$ Pa.

How can we account for such a great difference in values of E for different polymeric materials? Thermal motion in an *elastic polymer* is intense enough to enable the chains to move freely with respect to each other. However, long-distance movements of the chains (i.e. flow) are much harder to

[1] As formula (4.1) indicates, E has units of pressure — force per unit area. This is why we express it in Pascals. Let's remind that the unit Pa is defined as one Newton per m^2, and that normal atmospheric pressure is very close to 10^5 Pa. Since we are going to understand the molecular world, it is also useful to realize that one MPa (megaPascal) can be thought of as $1\frac{\mathrm{pN}}{\mathrm{nm}^2}$, where pico Newton is 1 pN $= 10^{-12}$ N and 1 nm $= 10^{-9}$ m.

perform because of the cross-links. Under an external tension, the chains can be easily stretched — this explains very low values of Young's modulus.

In contrast, in *polymeric glasses*, relative motion of the chains is hardly possible even on scales as small as the size of a monomer. This is why their Young's moduli are significantly higher. A detail to notice here is that for polymeric glasses room temperature values of E are still an order of magnitude lower than those for inorganic glasses. It shows that at room temperature motion is not as "frozen" in polymers as in silicate glasses; there is still some freedom for the chains to rearrange their conformations locally — this increases deformability and reduces the value of E.

Finally, in a *semi-crystalline polymer* there are amorphous partitions between regions of crystalline phase (Figure 4.1 *c*). For many materials these partitions are not completely in the glassy state at room temperature. So what we really have is a mixture of solid crystal "islets" separated by a kind of "grout" made from a rubber-like polymer. Clearly, such material should be less fragile, with lower values of Young's modulus, than a polymeric glass.

4.4 Polymeric Fibers

We have described possible states of polymers, and have discussed how the most commonly used materials, plastics and rubber, come into the picture. However, there is another class of polymeric materials that we have missed so far, namely, fibers. Fibers are by no means any less important, one good reason being that nearly all our clothes are produced from fibers. So what are they, and what state of matter do they represent?

First of all, we should note that polymeric fibers can be either of natural origin, or produced in a chemical laboratory or a factory. Cellulose, for instance, is the most widespread natural fiber. Molecular chains of cellulose form the walls of biological cells in most plants; they are also the chief constituent of timber. Natural cellulose fibers are obtained from flax, cotton, hemp, etc. Other well-known kinds of natural fiber are wool and silk. They, of course, have "animal" origin: Wool is given to us by sheep, goats, and camels, whereas silk is produced by a caterpillar (silkworm) with a Latin name *Bombyx mori*. (It is amazing that just a single thread of fiber made by a silkworm is about a kilometer long!) Chemically, silk and wool consist of polymer chains of particular proteins, called keratin (wool) and fibroin (silk).

The fibers obtained in a factory or a chemical laboratory are called chemical fibers. There are two types: artificial fibers which are made from natural threads, but modified in order to improve their properties, and synthetic fibers which are synthesized from some simple chemical compounds.

Artificial fibers are mainly obtained from various kinds of natural cellulose. For example, timber cellulose is used to make viscous rayon, and also the so-called acetate and triacetate fibers can be produced from cotton cellulose.

Among the most interesting synthetic fibers are nylon and its various brands (they are called polyamide fibers as they are built from polyamide molecular chains), terylene (polyether fibers), and lastly orlon and acrylon (polyacryl nitrile fibers).

The particular physical state of polymeric fibers in which they are actually used, depends on the purpose they serve. Obviously, the fibers have to be reasonably tough and should not stretch significantly under the influence of longitudinal forces which occur in the fiber during its use[2]. This immediately rules out the viscous and the elastic states. As for the other two, the fibers can be either semi-crystals (cellulose fibers, nylon, and terylene) or polymeric glasses (orlon). There is, however, something special about the structure of the fibers in these states.

Are semi-crystalline fibers arranged in the same way as in Figure 4.1 *c*, and polymeric glasses as in Figure 4.1 *a*? In fact, if they were, such materials would be of rather poor quality. They would not be strong enough, but would be quite easily stretched (you may see this if you just compare the values of the elastic modulus given above, for typical polymers in the two states). Experiments on semi-crystalline natural fibers have revealed, however, that their crystalline regions are not orientated randomly (as in Figure 4.1 *c*), but are mainly parallel to the axis of the fiber (Figure 4.3 *a*). This is just the kind of structure, with the chains predominantly parallel to the axis of the fiber, that is aimed at when chemical fibers are produced (both semi-crystals and polymeric glasses, Figure 4.3 *b*). Thus, polymer fibers are always anisotropic because polymer chains have a preferential orientation along the axis of the fiber. The higher is the anisotropy, the

[2]By the way, for the fibers used in the making of clothes, the limiting factor is not the stress caused by the wearing of the clothes. When making cloth from fibers, the fibers are first spun into thread, and then the thread is woven or knitted to form the final product. In mass fabrication, stresses occur during the spinning and weaving or knitting stages. These stresses are normally much higher than those that arise during the use of the finished fabric.

greater is Young's modulus for longitudinal deformations, and the stronger is the fiber.

How can we explain this? Why should a fiber become stronger when its anisotropy increases? In order to answer this question, let's look once again at the perfect polymer crystal shown in Figure 4.1 *b*. Let's see what happens if we start stretching the sample along the direction of the polymer chains' orientation. The stretching will be hindered by the covalent bonds that hold the monomers together in the long chains. Say, the chains have carbon "backbones" (which is true for many important polymers, e.g. polyethylene and polyvinyl chloride). In this case the covalent bonds which are in charge of building up the chains are obviously C – C bonds. We can estimate their deformability if we remember that diamond whose crystalline structure is also formed by C – C bonds has Young's modulus $E \sim 10^{12}$ Pa. Naturally, for the crystal shown in Figure 4.1 *b* the order of magnitude of E should be the same. The breaking strength for the two materials should be reasonably similar too (but only, of course, if the tension is applied to the sample in Figure 4.1 *b* along the polymer chains).

On the other hand, we know that for disordered semi-crystalline polymers $E \sim 10^8$–10^9 Pa because of the amorphous layers between the crystalline areas. If the crystalline areas start getting ordered in a particular direction (in other words, if the structure in Figure 4.1 *c* starts transforming into the structure in Figure 4.3 *a*), then more and more chains appear to be stretched along the axis of the fiber. Such chains take most of the strain arising from the deformation; they make the fibers considerably stronger, and the Young's modulus increases. Certainly, one cannot achieve the strength of diamond in such a way. Yet it is quite possible to improve the mechanical properties of the material by 1.5 to 2 orders of magnitude,

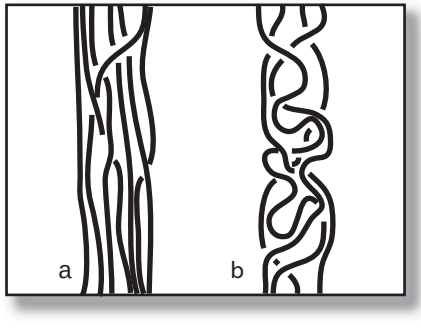


Fig. 4.3 Cartoon representation of the structures of orientated semi-crystalline (*a*), and amorphous (*b*) fibers.

due to an increase in the anisotropy of the fiber. Exactly the same idea is used to strengthen glassy fibers — pushing towards the transition from the structure shown in Figure 4.1 *c* to the ones presented in Figure 4.3 *b* and then preferably in the direction of Figure 4.3 *a*.

So far we have talked about the final structure of fibers when ready for use. In the fibers obtained from natural products (e.g. from cotton or wool), nature itself has provided the right structure. Indeed, these fibers, though semi-crystalline, are highly anisotropic (as in Figure 4.3 *a*). But what about chemical fibers? You may ask, first of all, how they are made, and, secondly, how the necessary degree of anisotropy is provided.

The usual strategy is as follows. Take a polymer from which you want to make a fiber, and convert it to a viscous state. This can be done by heating it up (if the polymer has a reasonably low melting point, like nylon or terylene). Otherwise, for refractory polymers in particular, polymers can be diluted in a good solvent; as a result, one gets a concentrated fluid solution of the polymer, called a spinning solution. Then the fiber is made by squeezing the solution (or the melt) through little holes, spinnerets, into a medium which helps to solidify the polymer in the shape of thin fibers. If the fibers are formed from a melt, the medium can just be cold air. On the other hand, if a solution of a polymer is used, the solvent has to be removed from the fibers after passing through the spinnerets. The solvent can be evaporated by placing the fibers in a jet of hot air. Alternatively, the threads can be treated in a so-called precipitating bath, which contains a special medium that makes it energetically favorable for the polymer to shrink and squeeze out the solvent.

However, the fiber obtained in such a way is not yet sufficiently aligned. Its structure looks like that in Figure 4.1 *a* or *c*. In order to make structures like the one in Figure 4.3 *b*, or to push it even towards Figure 4.3 *a*, one has to stretch the solid fiber at a temperature which is high enough for the polymer not to form a glass. This process, called orientational stretching, causes the polymer chains become more aligned.

Figure 4.4 shows a typical dependence of the stress σ on the relative elongation $\Delta\ell/\ell$ for a semi-crystalline polymer material. If σ is not too big, Hooke's law (Equation 4.1) is valid; the deformation in this case is elastic (reversible) — the fiber starts going back to its initial state after the force has stopped acting. However, when the stress becomes as high as σ_0 (see Figure 4.4) the situation changes dramatically. The deformation starts increasing of its own accord whereas the stress remains the same or even decreases slightly. During this process a sort of a "neck" develops in

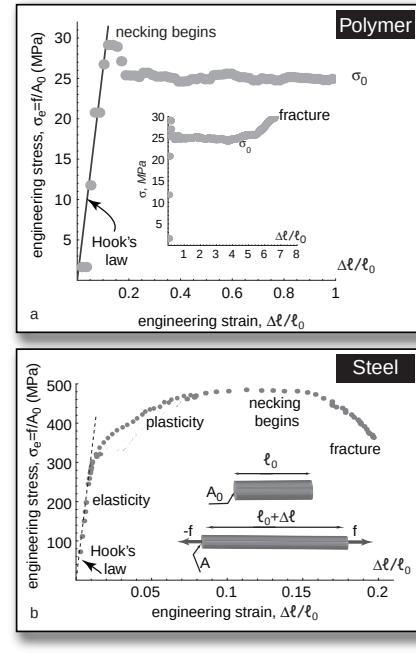


Fig. 4.4 Stress-strain diagrams of a typical polymeric material (*a*) and, for comparison, for steel (*b*). The scheme of the experiment is shown in the inset in the lower figure: one takes a sample of given dimensions and pulls on it with a measured force f. The stress σ is defined as the ratio of force to the *initial, undeformed* cross-sectional area of the sample, A_0: $\sigma = f/A_0$. This is called *engineering stress*, it is easier to measure in practice, because to determine the true stress f/A, one needs to measure both f and A. Stress has dimension of pressure and here it is presented in the units of megaPascal (a useful reminder for the book on molecules: 1 MPa$= 1\frac{\text{pN}}{\text{nm}^2}$). Strain is presented as unitless relative elongation of the sample $\Delta\ell/\ell_0$. Panel (*a*) presents data for an isotropic semi-crystalline polymeric material, specifically — low density polyethylene film. The inset presents a wider interval of strains. Notice that at low strain the deformations for both polymeric material and steel obey the usual Hook's law, i.e., stress is linear in strain, as shown by the straight lines. At larger $\Delta\ell/\ell_0$ deformation for these materials usually becomes irreversible. The curve for *engineering stress* in many materials goes through a maximum when neck develops in the sample (see Figure 4.5); interestingly true stress keeps increasing because actual cross-sectional area becomes substatially smaller than A_0. The reader should also realize that while overall shape of the stress-strain curve is somewhat similar for steel and polymeric film, the relevant scales are vastly different: at more than an order of magnitude lower strains steel develops more than an order of magnitude higher stress. At the same time the reader should realize that while these plots are quite typical, specific materials may differ from one another quite substantially. Measurements for this figure were performed at room temperature, film thickness 60 microns, the strip width 2 mm and length 10 mm. The plots are based on the data courtesy of A. Askadskii.

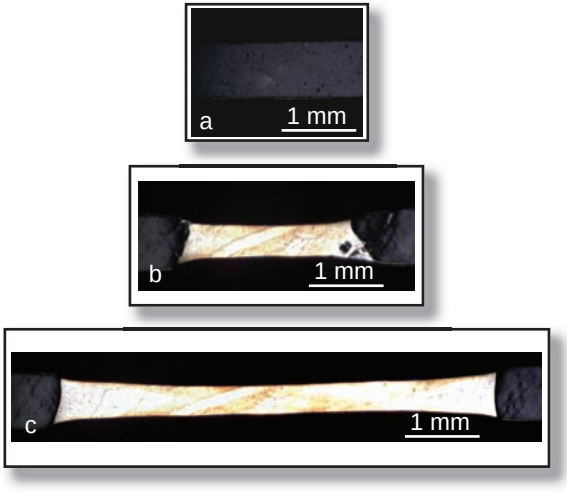


Fig. 4.5 The development of a "neck" during the stretching of a stripe of polymer film. Panels (*a*), (*b*), and (*c*) represent three successive stages of neck development. The photographs were taken using polarized light. The images are courtesy of A. Askadskii.

the fiber; it lengthens, and eventually runs the whole length of the sample (Figure 4.5). As a result, the fiber may stretch by a factor of about two or even more. Naturally, the deformations occurring in the fiber after the "neck" has been formed are irreversible.

Here is a very simple experiment which you can easily do yourself to show that the processes described really do take place in polymer fibers. Cut carefully a strip of polyethylene film, 10 cm long and 1 cm wide (a piece of a plastic bag would work OK). Stretch it, and you will see that, after a certain amount of deformation, a very elongated zone appears in the middle of the strip. It spreads along the whole strip with further deformation. This zone is similar to the "neck" developing on a stretched polymer fiber — as soon as it emerges the deformation becomes irreversible.

At a molecular level, the development of the "neck" means that the stress applied to the fiber is apparently rather high; therefore it leads to the break up of the semi-crystalline structure with random orientation of the crystallites (i.e. the one shown in Figure 4.1 *c*). In the "neck" zone the structure rearranges since polymer chains align on stretching. As a result, the chains are eventually orientated as shown in Figure 4.3 *a*. So the fiber becomes anisotropic and therefore stronger.

The strengthening is also helped by one more thing. When Figure 4.1 *c* is transformed into Figure 4.3 *a*, the degree of crystallization increases (quantitatively, there in an increase in the volume fraction of crystalline regions in the semi-crystalline fiber). This happens because some preliminary

orientation of the chains paves the way for further crystallization. The latter proceeds much more smoothly than in an isotropic sample because all the crystalline domains are already orientated roughly along the axis of the fiber. Such orientation helps the growth of the domains, and smaller ones can more easily join up with each other. It explains the higher degree of crystallization of the fiber and, therefore, its better mechanical properties.

In the end, one obtains some truly fantastic materials. For instance, the airbags which allow for soft landing of spacecraft on Mars are made of liquid crystalline polymeric fiber called vectran. This must be a truly special material!

4.5 Polymeric Liquid Crystals and Super-Strong Fibers

We have now described how chemical fibers are produced. In many respects, such man-made fibers are no worse than natural ones. They are often used these days in the textile industry. Nevertheless, typical values of the breaking strength or the Young's modulus of polymeric fibers are one or two orders of magnitude lower than those of steel. So the question arises: Could we use the same physical idea as we have just described, but take it further to try to remove the difference? Is it possible to produce polymeric fibers of nearly the same strength as steel? Of course, the problem of creating such super-strong fibers is of great importance. There are many applications where light but strong materials are needed.

You can indeed make a super-strong fiber, even stronger than steel, from a polymer, but the polymer must be converted into a special liquid-crystalline state which is really a variety of the viscous state. If you think of a viscous polymer as of some "polymeric liquid", then a liquid-crystalline polymer can be regarded as an "anisotropic polymeric liquid". The anisotropy occurs spontaneously, with no help from outside (such as orientating fields, mechanical stresses or whatever).

Let's look at the simplest example (Figure 4.6) to see how this spontaneous orientation may appear. Just throw a bunch of randomly oriented matches onto a surface (Figure 4.6 *a*). Now start reducing the area covered by the matches, but make sure that they are still orientated in the same random way. We gradually come to the situation in Figure 4.6 *b*. At this stage it becomes impossible to decrease the area any further while retaining the orientational disorder. Does this mean that we have already reached a close-packed arrangement of the matches? Certainly not, they

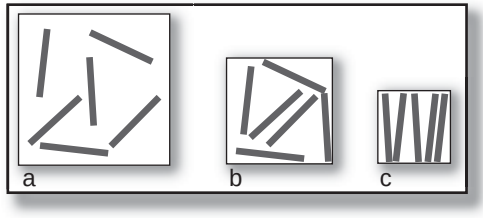


Fig. 4.6 Experimenting with matches on a surface. If the surface area available is large, matches do not interfere with each other, and distribute in a random isotropic fashion (*a*). If, on the other hand, the available surface is small, then matches have no choice but to order their orientations (*c*).

can be crammed into a much smaller square (Figure 4.6 *c*). However, their orientations would no longer be random; all the matches would be facing in the same direction. Hence, we conclude that the system of matches can be confined to a smaller area than in Figure 4.6 *b*, but the system would then be anisotropic.

Now let's imagine that, instead of matches, we have a system (a solution) of molecules with an elongated shape. What will happen if we gradually raise the concentration of the solution? At lower concentrations we shall observe the pattern shown in Figure 4.6 *a* — the distribution of the molecules' orientations will be isotropic. Then, while the concentration grows, we shall eventually reach the threshold regime as in Figure 4.6 *b*. Obviously, at higher concentrations the solution can only be anisotropic. This anisotropy occurs for no external reason, but spontaneously, just because a dense enough system of elongated particles cannot possibly be arranged in any isotropic way.

This is exactly what a liquid-crystalline state is — an anisotropic state that spontaneously develops in a solution of elongated molecules at higher concentrations. Starting from some degree of asymmetry of the molecules (i.e. the ratio of length to diameter), the liquid-crystalline state can appear in a melt as well.

The name "liquid crystal" reflects the duality of such materials; according to their properties they could be placed somewhere in between ordinary liquids and crystalline solids. Like liquids, liquid crystals lack long-range order in the positions of their molecules; most liquid crystals are indeed fluid. At the same time, just like solid crystals, liquid crystals are anisotropic as their molecules are orientated in an anisotropic way.

It is clear from Figure 4.6 that the liquid-crystalline state should be more typical for substances whose molecules have an elongated shape. Moreover, the greater the asymmetry of the molecules, the lower the critical concentration of the solution at which the molecules start to align spontaneously.

This suggests that solutions of stiff polymer chains should become liquid crystals rather easily, in quite a broad range of concentrations. Indeed, on larger scales such molecules are tangled up into coils. Therefore their asymmetry is determined by the ratio of the longest chain segment ℓ which can still be regarded as approximately straight (i.e. the Kuhn segment; see Section 6.5) to the characteristic diameter d of the chain. For rod-like polymer chains the ratio ℓ/d can be rather high; in the case of aromatic polyamides, for instance, it can even reach a few hundred. This means that even when the volume fraction of an aromatic polyamide is only a few percent, such a solution should still be liquid-crystalline; in other words, its molecular chains should be aligned predominantly along the axis of spontaneous orientation.

Now let's go back to the problem of how to produce super-strong fibers. A natural strategy is to use the inherent anisotropy of a liquid-crystalline solution, and to form the fiber directly from this solution. If we do this, we shall end up with a highly oriented fiber immediately after the moulding, extra stretching, and expulsion of the solvent. The orientational order will be much higher than you normally achieve by orientational stretching alone. In practice, this technique allows liquid-crystalline solutions of aromatic polyamides to be converted into really amazing fibers whose strength and Young's modulus are of the same order of magnitude as those of steel. Such fibers were first created about 1970s, and now they are widely used in various areas of industry.

4.6 Polymer Solutions

So far when talking about various states of polymers, we have usually meant substances consisting purely of polymer molecules, although when discussing polymer fibers, we did mention that they are often formed from solutions of polymers.

Polymer solutions are, obviously, liquid mixtures of long polymer chains and small, light solvent molecules. They play a very important role in polymer physics; this is why it makes sense to give here a brief description of them. We shall discuss two qualitatively different uniform states of polymer solutions.

These states are illustrated in Figure 4.7; polymer chains are shown with solid lines, and small molecules of a solvent are not depicted at all. Figure 4.7 *a* corresponds to a dilute polymer solution; macromolecules are

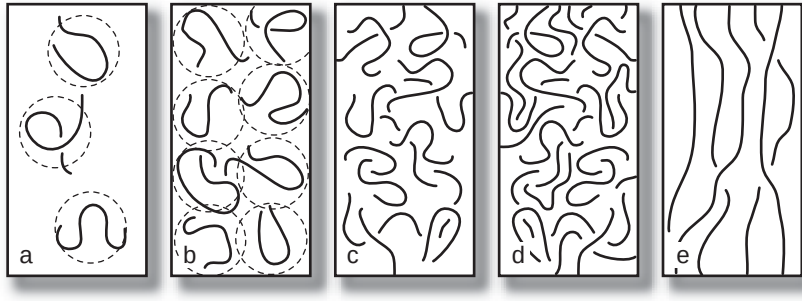


Fig. 4.7 A sketch of concentration regimes of a polymer solution: (*a*) dilute polymer solution — coils do not overlap, $c \ll c^\star$; (*b*) cross-over between a dilute and a semi-dilute solution — coils (depicted by dashed lines to guide the eye) are on the brink of overlapping, $c \simeq c^\star$; (*c*) semi-dilute solution — coils strongly overlap, but concentration is still low, $c^\star \ll c \ll c_{\mathrm{melt}}$; (*d*) concentrated solution $c \approx c_{\mathrm{melt}}$; (*e*) liquid-crystalline solution.

separated by large distances and hardly interact with each other at all. The properties of such solutions are governed merely by the properties of the individual macromolecules. For instance, from light scattering or viscosity measurements, we can judge the shape and size of the polymer coils. So a dilute polymer solution is, in a way, the most basic polymer system, because as we study it we actually learn about the properties of the individual macromolecules. In this sense, it is similar to a low-density gas of ordinary small molecules. Commonly, in more complex polymer systems the chains are highly entangled, and strongly interact with each other; therefore, it is much harder to discern the contributions of individual macromolecules. To find out about the individual chains, you would have to look at data for dilute solutions.

With increasing concentration, the polymer coils sooner or later start to overlap; then we eventually get to the picture of densely entangled coils as shown in Figure 4.7 *c*. Obviously, the intermediate regime between Figures 4.7*a* and *c* will be when the coils do not yet overlap, but just touch each other (as in Figure 4.7 *b*). This means that the critical concentration $c^\star$ corresponding to the intermediate regime is the same order of magnitude as the concentration of monomers in each coil. It is useful to know how to calculate this value, or at least to estimate it. It would help us to understand which concentration regimes are realistic or typical for a polymer solution under different conditions. In Section 6.6 we shall come back to this question and find an approximation for $c^\star$.

4.7 Polymer Blends and Block-Copolymers

Polymer chains can, of course, mix not only with solvent molecules but also with other polymers. This is how polymer blends are formed. However, there are not that many pairs of different polymers that can blend in any proportion. A mixture of two polymers $\mathcal{A}$ and $\mathcal{B}$ typically will tend to separate into nearly pure phases of $\mathcal{A}$ and $\mathcal{B}$. This happens even if the repulsion between the monomers of $\mathcal{A}$ and $\mathcal{B}$ is so weak that they would be able to mix if they were not linked into a chain.

An interesting thing occurs when immiscible polymers $\mathcal{A}$ and $\mathcal{B}$ form one chain (Figure 4.8 *a*). This is what we call a block-copolymer (see

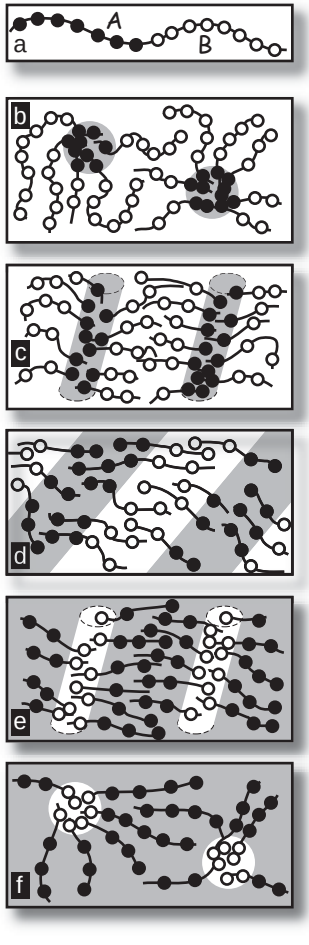


Fig. 4.8 Panel (*a*): A block-copolymer chain, consisting of two blocks, $\mathcal{A}$ (black) and $\mathcal{B}$ (white); panels (*b* — *f*): several types of microstructure in block-copolymer melts. Gray refers to the regions where black monomers dominate, white monomers dominate in white regions.

Section 2.5). In this case the blocks $\mathcal{A}$ and $\mathcal{B}$ try to separate from each other. However, a macroscopic phase separation is impossible because the $\mathcal{A}$ and $\mathcal{B}$ blocks are tightly linked to each other within the chains. As a result, we get a pattern of micro-domains which contain mainly $\mathcal{A}$ blocks or $\mathcal{B}$ blocks, separated by fairly thin interphase regions (Figures 4.8 and C4.9). This effect is known as microphase separation in block-copolymers, and the structure that emerges is called a micro-domain structure.

Depending on the relative lengths of $\mathcal{A}$ and $\mathcal{B}$ blocks, micro-domains can take different shapes. In particular, they can look like little spheres (micelles) of $\mathcal{A}$, fitted onto a regular three-dimensional lattice submerged into the "sea" of $\mathcal{B}$ (Figure 4.8 *b)*. This happens if $\mathcal{A}$ blocks are much shorter than $\mathcal{B}$ ones. Increasing the length of $\mathcal{A}$ blocks changes this picture, and we end up with cylindrical $\mathcal{A}$ domains in a "sea" of $\mathcal{B}$ units (Figure 4.8 *c*). When the two blocks have roughly the same lengths, there are alternating $\mathcal{A}$ and $\mathcal{B}$ layers (Figure 4.8 *d*). Finally, for shorter $\mathcal{B}$ blocks, cylindrical (Figure 4.8 *e*) and spherical (Figure 4.8 *f*) $\mathcal{B}$ "islands" appear in the "sea" of $\mathcal{A}$ units.

Spherical micelles, cylindrical micelles, and lamellas (layers) are not the only possible microphase segregated structures. For instance, the characteristic feature of micelles is this: imagine a small molecule which can diffuse through one of the phases (say, $\mathcal{A}$), but not soluble in, and cannot enter the other phase ($\mathcal{B}$). Such molecule can travel arbitrarily far in the inverted micelles case (4.8 *e* or *f*), but will be locked in direct micelles (4.8 *b* or *c*). By the way of analogy, we can imagine a city and two different persons: one can walk in the streets, but cannot enter any building, while the other is inside a building and cannot exit from it. Obviously, the former person can walk very far, while the latter cannot. In the city, and in general in two dimensions, one cannot imagine two separate sets of non-intersecting streets. Interestingly, in three dimensions such a situation is perfectly possible, and the corresponding phase is called bicontinuous. Along with other microphase segregated states it is illustrated in Figure C4.9. It does not show polymer chains explicitly (unlike Figure 4.8), and the reader should imagine them going from the region of one color to the other.

Thus, there is a rich variety of very interesting structures even for the simple diblock-copolymer. And we have a simple tool for controlling the microstructure of a block-copolymer melt. All you need to do is to vary the length ratio between $\mathcal{A}$ and $\mathcal{B}$ blocks. Even richer variety of structures exist in three-block copolymers — it is actually too rich to discuss in this book.

4.8 Ionomers and Associating Polymers

A rich variety of nano-scale structures and inhomogeneities is very typical of polymers. Fully uniform states are the exception rather than the rule. Now we shall explore yet another peculiarity of polymer structure, by looking at the so-called ionomers.

We have already discussed polyelectrolytes in Section 2.5. They are formed when small ions, called counterions, break off from the chain. They leave behind monomer units of the opposite charge. If a counterion escapes and sets out on a "journey" on its own, the whole chain acquires an electrical charge, and becomes a polyelectrolyte (Figure 4.10 *a*). However, this is not the only scenario. Thermal motion may not be strong enough for the counterion to tear itself away from the ionized monomer. Instead, the two form an "ion pair". The counterion stays in the vicinity of the charged monomer (at an average distance a), the two charges making a dipole (Figure 4.10 *b*). If all the counterions tend to stay in such pairs, the chain is called an ionomer.

Can we tell exactly when each of these two cases, a polyelectrolyte (Figure 4.10 *a*) and an ionomer (Figure 4.10 *b*), would occur? Assume that the charges of the dissociated monomer and the counterion are the same in magnitude, and equal to the electronic charge e. Suppose also the dielectric constant of the medium is ε. Then the energy of the Coulomb interaction[3] of the ions in a pair is $e^2/\varepsilon a$. If this energy is much less than the characteristic energy of thermal motion k_BT, where k_B is Boltzmann's constant, and T is the absolute temperature (see Section 7.6 below for a more detailed discussion of characteristic thermal energy), i.e.

$$\frac{e^2}{\varepsilon a k_B T} \ll 1 \ , \tag{4.2}$$

then counterions break off the chain. Thus we get the polyelectrolyte regime. Conversely, if

$$\frac{e^2}{\varepsilon a k_B T} \gg 1 \ , \tag{4.3}$$

[3]Throughout this book, we use the so-called Gauss system of units as far as electrical and magnetic quantities are concerned. These units are in fact the most convenient ones in all respects, except they do not agree with the tradition accepted in electrical engineering, such as the unit of Ampere for the current. Since we will not deal with any technical aspects anyway, we stick to the Gauss units, in which, for instance, the expression for Coulomb energy does not have the annoying coefficient $1/4\pi\varepsilon_0$. If you, the reader, feel more comfortable with some unit system of your choice, we encourage you to repeat all our simple calculations using your preferred units and see for yourself that the results stay unchanged.

Here, we have to digress and explain the signs $\ll$ and $\gg$ and their usage. Formally, they mean "much less" and "much greater", respectively. Of course, one may ask — how much is "much"? In other words, if $x \gg y$, does it mean x should be twice larger than y, or ten times larger, or what? The general answer has to do not so much with the specific numerical values of x and y, but rather with *dependencies* of things on x and y when they *change*. Let's illustrate this rather abstract point with the specific example of Equations (4.2) and (4.3). Formula (4.2) says that we have polyelectrolyte regime if the ionization energy is *much smaller* than thermal energy $k_B T$; the meaning of it is as follows: the smaller the value of dimensionless parameter $e^2/(\varepsilon a k_B T)$, the more complete ionization, and the more accurate polyelectrolyte concept. Similarly, formula (4.3) states that ionomer regime occurs if the ionization energy is *much larger* than thermal energy $k_B T$, which means: the larger the ratio of energies $e^2/(\varepsilon a k_B T)$, the more accurate the ionomer model. Qualitatively then, portraying the situation in imprecise impressionists strokes, we can say we deal with either a polyelectrolyte or an ionomer regime.

Quite similarly, we discussed in Section 4.6 the regimes of dilute and semi-dilute solutions, realized at concentrations c such that $c \ll c^*$ (Figure 4.7 *a*) and $c \gg c^*$ (Figure 4.7 *c*), respectively.

Of course, such a description leaves a "gray zone", an intermediate regime, or a cross-over, when parameter is neither big nor small. However, when parameters, such as concentration or ionization energy, change by many orders of magnitude, the "gray zone" is in many cases insignificant: yes, it is possible that a system is neither really a polyelectrolyte nor an ionomer, neither dilute nor semi-dilute, but something in between, but if we understand well both limits, we are usually not scared by the intermediates, too. This is why, in this book, as it is customary everywhere in modern physics, we will consider the limiting regimes, delineated by "strong inequalities", $\ll$ and $\gg$; moreover, we shall frequently simply write $<$ instead of $\ll$ and $>$ instead of $\gg$, pretending silently that the cross-over "gray zone" is rather narrow and not interesting to us. With this in mind, let's return to ionomers and polyelectrolytes.

then the thermal motion cannot break up the ion pairs, and the chain is an ionomer.

We have already said that proteins, DNA, and polyacrylic and polymethacrylic acids are all polyelectrolytes when dissolved in water (see Section 2.5). We emphasize that the solvent should be water. Why is this so essential? The dielectric constant of water is extremely high ($\varepsilon \approx 80$). Therefore, the ratio $e^2/(\varepsilon a k_B T)$ is relatively small, and inequality (4.2) holds. However, if you use other solvents, with much smaller values of ε (usually between 2 and 20), then inequality (4.3) holds instead, and the polymer is an ionomer rather than a polyelectrolyte. This is why polyelectrolyte regime is typical for polymers dissolved in water, including the biopolymers, while ionomer regime is typical for polymers dissolved in organic solvents.

Ion-containing polymer chains in a melt (in the absence of a solvent) are also typically in the ionomer regime. This is because the dielectric constant of a pure polymer tends to be rather low. What is the structure of such an "ionomer" melt? Ionomer chains contain some (usually small) proportion of monomers in the form of ion pairs (see Figure 4.10 *b*). They interact strongly with each other, since they are electric dipoles. The other monomers have no electrical charge. Dipoles always arrange themselves in such a way that the interaction between them is attractive (see inset in the Figure 4.10). This is why ion pairs are strongly attracted to each other.

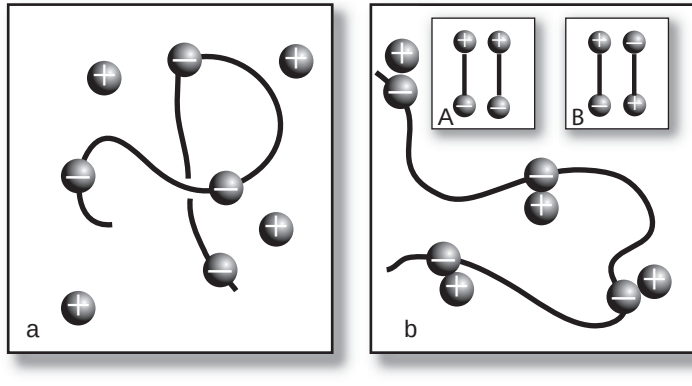


Fig. 4.10 A polyelectrolyte (*a*): all the counterions are free and not attached to the polymer chain; an ionomer (*b*): counterions are "condensed" on the charges of the chain and form ion pairs. *Inset:* Dipoles (ion pairs) are free to choose any orientation with respect to each other. The one they prefer is (*B*), since it gives the lower energy than (*A*). It corresponds to attraction.

Fig. 4.11 (*a*): A cartoon of multiplet structure in a melt of ionomers. Strongly interacting ion pairs are shown explicitly. Multiplets are circled and shaded. (*b*): A sketch of a solution of an associating polymer. Strongly interacting monomers are depicted as big empty circles. Associates are circled and shaded. In contrast to Figure (*a*), the space between associates is mainly taken up by the solvent molecules — the volume fraction occupied by monomer units of the chains is rather small.

However, they are part of a polymer chain, so they cannot be separated into a distinct phase. As a result, small islands emerge in a sea of neutral monomers (Figure 4.11 *a*). Such aggregates are called ionomer multiplets.

Compare Figure 4.11 *a* and Figure 4.8 *b*. You may notice that the multiplets in a polymer melt somewhat resemble the spherical domains in a block-copolymer melt that form when one of the blocks is much longer than the other. This similarity is not surprising. The structure in Figure 4.11 *a* can be obtained from Figure 4.8 *b* by letting the shorter block tend to one monomer, and increasing the attractive force between the monomers. As a matter of fact, it is quite a common pattern in polymer structure when small spherical multiplets ("associates") are formed by strongly attracting monomers, whatever the nature of their attraction. Such polymers are called associating.

Associating polymers have many practical uses. Let's give the simplest example. Suppose we wanted to increase the viscosity of a liquid substantially. Can we do it by adding just a little bit of some other substance? If yes, then how to choose the substance? As you might have guessed, an associating polymer consisting of two different types of monomers (Figure 4.11 *b*) can play this role. Let's see how it works. The greater part of the

monomers easily dissolve in the liquid. However, a small fraction of the monomers try to avoid the solvent. Any contact with the liquid molecules is extremely energetically unfavorable for them. Such strongly associating monomers join together to form aggregates (multiplets). The structure of the resulting solution is shown in Figure 4.11 *b*. It looks like a polymer network (a gel). The special thing about it is that the role of covalent cross-links is played by associates of strongly interacting monomers. This is not a real network (like the ones described in Section 2.5). The associates can dissociate from time to time, as well as appear in a new place. Therefore, if we apply shear stress to such liquid, it will start flowing, together with the associating polymer that is dissolved in it. Nevertheless, the viscosity of the liquid is substantially increased because dissociation of such associates is relatively rare. Even a very small amount of an associating polymer is enough to increase the viscosity of the liquid substantially.

4.9 Conductive Polymers

Our story about polymer materials would not be complete if we did not mention here that some of them may be electric conductors. In Section 3.1 we touched upon the difference between covalent bonds of σ- and π-types. Within the context of electrical conductivity, further difference is that σ-electrons (i.e., participating in a σ-bond) are strongly localized between the two connected atoms, they cannot move, and thus do not contribute to the electrical conductivity of the material. On the other hand, in the case of π-bonds the electrons are much more delocalized and therefore may exhibit much higher ability to move from place to place.

Still, if a polymer material is pure (undoped), its electrical conductivity is very low — normally between 10^{-10} to 10^{-8} S/cm even if it is rich with π-bonds (for comparison, conductivity of sea water is close to 0.02 S/cm)[4]. The reason is that π-electrons are present essentially on every bond, therefore, any particular one of them cannot move anywhere because of Pauli exclusion principle — all states around are occupied by other π-electrons. The situation changes after doping. For polymer systems doping normally means oxidation, i.e. removal of some of the delocalized π-electrons, this

[4]S/cm (Siemens over centimeter) is the commonly used unit of electric conductivity. Siemens is the unit of conductance (inverse resistance), it is equal ohm^{-1}. Conductance of a sample of length L and cross-sectional area A equals $\sigma A/L$, therefore, conductivity σ, which is the property of a material, must have the dimension of conductance divided by length.

process is analogous to the formation of "holes" in semiconductors. In this case, even at low values of doping ($< 1\%$) electrical conductivity can increase several orders of magnitude, up to values around 0.1 S/cm. Highest conductivity in polymers reported up to now is 8×10^4 S/cm for highly doped polyacetylene. These materials were obtained in 1980s; in 2000, Nobel Prize in Chemistry was awarded to Alan Heeger, Alan Mac Diarmid and Hideki Shirakawa "for the discovery and development of conductive polymers".

Main classes of conductive polymers include polyacetylenes, polypyrroles, polythiophenes and polyanilines. All these polymers are rich with π-electron bonds. There are numerous attempts to use these polymers to design organic solar cells, organic light-emitting diodes, electrochromic materials, electroluminescent materials, super-capacitors etc. The biggest advantage to use polymers instead of inorganic materials is normally their processability, good mechanical properties, and low cost.

With this we conclude our brief review of the kinds of states in which the simplest polymers can exist. Of course, we did not cover the whole variety of polymer systems. A few more examples will be found further on in this book, as well as in some other popular books listed in the Suggested Further Readings.

Chapter 5

Polymers in Nature

> La verdad adelgaza y no quiebra, y siempre nada sobre la mentira como el aceite sobre el agua.
> (Truth will rise above falsehood as oil above water.)
>
> Miguel de Cervantes,
> *Don Quixote*

> In a bath, in a tub, in a shower,
> In a stream, in a brook, in the sea,
> Here and there, and everywhere —
> Glory to water forever be!
>
> K. Chukovskiy,
> *Wash-into-holes* (*Russian children's poem*)

A great many fascinating biological objects consist of polymers. For example, the shell of a tortoise or the stiff back of a beetle are "built" from a polymer called chitin whose chains are held together by proteins (which are polymers too!) Then there are viruses, little boxes made from protein chains, with a nucleic acid chain inside each. There are far too many examples to tell of them all! We shall therefore have to stick to three, ones which we, as physicists, believe are the most interesting and fundamental.

However, before we start our story, there is one more thing to say: The main biopolymers function in the medium of water. A human body consists of 60% water by mass; some animals carry around even more water in their bodies. Water reservoirs are a source of life (as we shall discuss in more detail in Chapter 14). Therefore, it might be helpful to learn a bit about

the molecular structure of water, before we plunge into the discussion of biopolymers.

5.1 A Few Words about Water and the Love or Fear of it

A molecule of water, H_2O, is triangular in shape (Figure 5.1 *a*). The electron cloud tends to be shifted away from the hydrogen nuclei towards the oxygen nucleus by, on average, 0.02 nm $= 2 \cdot 10^{-11}$ m.

As a result, the positive charge of the hydrogen nuclei is not quite compensated. Similarly, there is an uncompensated negative charge around the oxygen nucleus. This peculiarity of the structure may not seem of great significance at first sight. However, it is the real cause of all the special properties of water which make it play such an important role in living organisms. What are these properties?

First, a water molecule has a considerable dipole moment, $p = 0.6 \cdot 10^{-29}\mathrm{C} \cdot \mathrm{m}$, i.e. the water is polar (this is what we call substances whose molecules have a non-zero dipole moment). This means that in an external electric field water molecules can be regarded as little "dipoles", each carrying two charges, $+e$ and $-e$, separated by a distance a (e is the charge of proton, $e = 1.6 \cdot 10^{-19}$ C); then $p = ea$. Given the value of p mentioned above, we can calculate $a = p/e = 0.04$ nm $= 4 \cdot 10^{-11}$ m. Such little dipoles have no difficulty in becoming aligned in an external electric field; this explains why the dielectric permeability of water is much higher than for all other common liquids: $\varepsilon \approx 80$.

In Section 4.8 we decided that such a high value of dielectric constant means that many monomers dissociate in water solutions. In other words, the corresponding polymers are polyelectrolytes. In particular, the polyelectrolyte nature of the main biopolymers, DNA and proteins, is crucial for their biological functioning.

Second, water molecules appear to be able to form so-called hydrogen bonds between each other. A hydrogen bond is a kind of saturable, attractive interaction between a couple of atoms, say O, C, N, etc. One of the two atoms should be joined to a hydrogen atom by a covalent bond. For instance:

$$\mathrm{O{-}H}\ldots\mathrm{O}\ ,$$

where the dots mark the hydrogen bond, and the solid line the covalent bond. Roughly speaking, the attraction occurs because the hydrogen atom's electron is shifted towards the oxygen atom along the covalent

bond. As a result, there is some extra positive charge near the H nucleus as well as some extra negative charge around the O nucleus. Thus an H nucleus can be attracted to an O nucleus of another molecule, linking the two molecules together. The binding energy of a hydrogen bond is of order 0.1 eV $= 1.6 \cdot 10^{-20}$J; this is one or two orders of magnitude smaller than a covalent bond's energy (which is about 1 to 10 eV), but somewhat larger than the thermal energy at room temperature (300°K): $k_BT \sim 0.03$eV where $k_B \approx 1.38 \cdot 10^{-23}$ J/K is Boltzmann's constant. The comparison of these energies is very telling. On the one hand, random molecular motions with energy about k_BT are practically unable to break covalent bonds, which, therefore, act like reliable locks at room temperature. On the other hand, hydrogen bonds at room temperature are far less reliable; their energy is still few times k_BT, so they are connected most of the time, but do break every now and then due to random molecular hits. Therefore, the molecular structure of water at any instant just looks like a three-dimensional network of hydrogen bonds, but, in contrast to a gel, every piece of this network gets torn apart and stuck together in a new manner over and over again, due to the thermal motion.

The network of hydrogen bonds is a key concept clarifying many properties of water, e.g. water's high heat capacity. Indeed, in order to increase the temperature of water you have to expend a fair bit of energy to break the hydrogen bonds.

What we have said about water also explains its special features as a solvent. Nonpolar substances (i.e. substances whose molecules have no dipole moment, for example the simplest organic compounds — fats and oils) are barely soluble in water, whereas the solubility of polar substances is normally much greater. This can be explained in the following way. If a polar molecule is placed into water it experiences a strong attraction to the water molecules. This is due to the interaction between the little dipoles, which have the ability to line up antiparallel to each other (compare the inset in Figure 4.10 and Section 4.8). For a low-molecular weight molecule, the energy of such attraction is usually around 0.1 eV, and quite often this is enough to provide significant solubility. In contrast, if there is a nonpolar molecule in the water, there will be no attraction, in fact just the opposite will occur — the water's molecular structure will be distorted as some hydrogen bonds will be broken. Obviously, this is not energetically favorable, and so the water molecules will try to "push" the alien molecule out. Such molecules have practically zero solubility.

Polar and nonpolar substances are also known respectively as hydrophilic and hydrophobic. These names start making sense when translated from Greek: *hydro* ($\upsilon\delta\rho o$) of course means water, *philos* ($\varphi\iota\lambda o\varsigma$) means friend, and *phobos* ($\varphi o\beta o\varsigma$) fear.

The concept of hydrophilic and hydrophobic behavior is very important in molecular biology.

5.2 Head-and-Tail Molecules

Add some hydrophilic substance to a glass of water, and it will merely mix with the water, just like sugar. In other words, it will be dissolved. On the other hand, a hydrophobic substance cannot be dissolved, it will separate out from the water, just like oil. However, there is a more complex "amphiphilic" kind of molecule; each molecule contains both a hydrophilic and a hydrophobic part. What happens to them in water?

Each of us must make such experiment many times a day, observing the interaction between amphiphilic substances and water, since even ordinary soap consists of amphiphilic molecules. (How could we avoid mentioning soap having chosen an epigraph from the book *Wash-into-holes*?) Besides, amphiphilic molecules are often encountered in biological systems. Most often such molecules consist of a polar atomic group, the "head" (Figure 5.1 *b*), and a hydrophobic "tail" which is attached to the head. The tail is a carbohydrate chain $(-CH_2-)_n$ of moderate length; normally n varies in

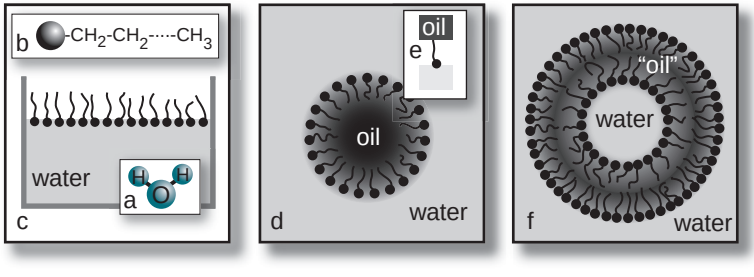


Fig. 5.1 The behavior of amphiphilic molecules in water. (*a*): One molecules of water; (*b*): A schematic diagram of a typical amphiphilic molecule, consisting of a hydrophilic head (a ball) and a hydrophobic hydrocarbon tail; (*c*): If there are not too many amphiphilic molecules in contact with water, they prefer to locate on the surface; (*d*): amphiphilic molecules can surround a drop of oil in water; (*e*): in this sense a soap molecule can be said to connect water and oil; (*f*): A liposome contains a small amount of water inside, separate from the bulk of water by an oily shell.

the range from 5 to 20. The whole molecule looks very much like a tadpole. Strictly speaking, "tadpole" molecules like the ones in Figure 5.1 *b* are not quite polymers since the number n is not high enough. Nevertheless, it is just because they have their flexible tails that such molecules exhibit some rather special and interesting properties.

Well, so what will happen if you try to dissolve molecules like those in Figure 5.1 *b* in water? A straightforward guess is that while there are not too many of them they will stay on the surface, immersing their heads in the water and sticking their tails out of it (Figure 5.1 *c*). By the way, this makes it clear how soap actually works. Oil, fat, and other nonpolar organic compounds cannot be easily washed off by water because they just do not dissolve in it. However, there is a great difference as soon as amphiphilic molecules of soap come along. Their hydrophobic tails will cling to the oil. The reason is that it is energetically favorable for hydrophobic particles in the water to come together. For them, getting together is simply a way of defending each other from being too close to the water. As a result, the water molecules form a kind of "coating" around the drops of oil (Figure 5.1 *d*). The whole surface of such "coated" particles consists of the hydrophilic heads of the soap molecules; therefore they are soluble and easily washable by water. Thus, in a sense, the amphiphilic soap molecules stick the oil to the water (Figure 5.1 *e*).

The next question to ask is: What if there are too many "tadpoles" and they cannot all be accommodated on the water's surface? Each of us has his or her own experience with soap and detergents, and we all know one good way to accommodate more "tadpole" molecules — to increase the surface area considerably by forming a lot of bubbles, or a foam. (This is why the substances formed of interface-loving molecules, such as "tadpoles", are even referred to as surface-active). However, if there is no way to increase the surface area any more, then the approximate scenario shown in Figure C5.2 gradually develops with the increase in the "tadpole" concentration.

The first stage is that the "tadpoles" get together to form spherical particles, called micelles. Each micelle's outer surface is made up of hydrophilic heads, and is in direct contact with water, whereas the hydrophobic tails are hidden inside (Figure C5.2 *a*). On the one hand, obviously, such micelles dissolve easily in water, because to water they seem purely hydrophilic! On the other hand, they behave like highly stable, almost indestructible units.

If there are even more amphiphilic molecules in the solution, the spherical micelles start to feel rather cramped. The "tadpoles" reorganize to form a system of parallel cylindrical micelles (Figure C5.2 *b*). Now let's imagine

that the number of "tadpoles" continues to grow. You can regard it as just having less water — no longer enough to fill in the gaps between the cylindrical micelles. So the amphiphilic molecules are forced to rearrange themselves once again, this time forming parallel layers known as lamellae (Figure C5.2 *c*). If there is even less water then inverted cylindrical micelles (Figure C5.2 *d*) and, later, inverted spherical ones (Figure C5.2 *e*) gradually develop. We end up with lots of different and very beautiful structures!

Substances with such structures have quite unusual properties. They are fluid, but in the cases *b* and *d* they tend to flow differently parallel and perpendicular to the cylinders. Meanwhile, for the lamellar structure *c* there is only one possible direction of flow — the layers can only slide parallel to each other. Certainly, light too propagates differently along and across the cylindrical micelles or the lamellae; hence birefringence is typical.

To tell the truth, you cannot usually get all five successive stages (Figure C5.2 *a*–*e*) with the same substance. Either the tails are too thick to form spherical or even cylindrical micelles, or, on the contrary, they may be too thin to construct inverted micelles. Normally one substance can exhibit only two or three of the structures in Figure C5.2.

Let's compare Figure C5.2 and 4.8. You can easily spot the similarity between the structures that "tadpoles" form in water and those appearing in block-copolymer melts. This is not a coincidence. Indeed, block-copolymers are also amphiphilic molecules, just like the "tadpoles". The only difference is that, instead of a tail and a head, they simply have two tails connected with each other.

The structures like the ones in Figure C5.2 are often used in a special kind of polymerization process called emulsion polymerization. In Section 3.1 we described how polymerization occurs. We discussed that the chain growth often "terminates" because its two growing ends come together. If we knew how to make the ends less likely to encounter each other, we would be able to produce longer polymers. One of the solutions is to carry out polymerization in a system such as that shown in Figure C5.2 *a*. Indeed, assume that both the initiator and the monomers are insoluble in water, but that they do not mind the hydrophobic tails of the "tadpoles". Then, if you dissolve them both in the system shown in Figure C5.2 *a*, they will be mostly absorbed by hydrophobic micelles. Polymerization will start in the micelles where the initiator molecules ended up. You can adjust the concentration so that there is no more than one molecule of the initiator in most micelles. Then the chain growth will be safe from terminating. The polymerization will go on until there are no monomers

left in the micelle. Thus, we create longer polymers than if we use one of the usual methods. Moreover, the chain tends to grow faster, because monomers are trapped in a special "microreactor", a micelle. Since this "microreactor" has a microscopic size, it helps to solve another problem, the one of taking away the heat that is given out during the reaction.

This is all fairly interesting, but you may start wondering: what does it have to do with biology? Here is the answer: Molecules of phospholipids have the shape of a "tadpole", although normally with two, or sometimes even three, tails. They are the chief constituent of membranes that separate biological cells from the outside world and divide the cells into compartments. The considerable thickness of the double tails prevents phospholipids from clumping into micelles, hence they form into layered walls.

Phospholipids can even be used as a material to make a model of a real cell. All you have to do is to take a suspension of phospholipids and to give it a good "shake" with an ultrasound signal of the appropriate wavelength. This forms "liposomes" which are comparable in structure to that shown in Figure 5.1 *f* Liposomes are used, for example, to study how different drugs may penetrate into a cell through the cell membrane.

However, the phospholipid layer is not the only part of a membrane. There are also some proteins "floating" in the lipid medium, as shown in the cartoon in Figure 5.1 (the image was created by Mariana Ruiz Villareal and is available in public domain `http://commons.wikimedia.org/break wiki/File: Cell_membrane_detailed_diagram_en.svg`). Moreover, the membrane (and thus the whole cell) is held in shape by the so-called cytoskeleton. It consists of proteins and polysacharides (which are polymers too!) The strange name comes from the Greek for cells, *cytos* ($\kappa\upsilon\tau\tau\alpha\varsigma$).

A few years ago, Vincent Noireaux (now at the University of Minnesota) and Albert Libchaber (of Rockefeller University in New York) published an article[1] with a telling title "A vesicle bioreactor as a step toward an artificial cell assembly"; they reported data of a series of experiments in which they used a vesicle like the one in Figure 5.1 *f* and tried to equip it with at least the elements of bare necessity for a simplest biological cell. They succeeded in implanting some of the proteins into the artificial membrane, placing some DNA inside the cell, and making a few more steps towards artificial cell.

The study of cell membranes is one of the most rapidly developing branches of modern biology; it even has its own name, *membranology.*

[1] Proceedings of the National Academy of Sciences of USA, v. **101**, n. 51, pp. 17669–17674, 2004.

There is a great variety of interesting phenomena in this area, and quite a few of them are related to polymers. Unfortunately, there is no way to describe them all in this book, but we cannot help giving one particular example. In principle, the "fatty" layer of a membrane can exist in two different states. One of the states is nearly solid, with the hydrophobic "fatty" tails lying parallel to each other. In contrast, the other state is liquid, and the tails are randomly entwined. There is a possibility of a phase transition between the two states. (To be precise, the cytoskeleton is actually involved in this transition too.) At different stages of a normal cell's life cycle, its membrane appears in the two different states, sometimes changing from one to the other. (A cancer tumor cell, however, which is prone to uncontrollable division, is incapable of such transitions, so its membrane remains liquid all the time. This may turn out to be important in understanding cancer, although as yet we do not know why.)

Thus, some rather complicated structures can be built from phospholipid molecules in nature. Later we shall see even more interesting "architecture" when discussing proteins and nucleic acids. However, we first ought to explain what made us mention architecture.

5.3 Molecular Biology and Molecular Architecture

In books on the history of architecture you may come across an interesting theory. People who are keen on scientific explanations may find it quite attractive. In our own words, it is the following.

How could you work out, if you wanted to, what kind of architectural style was typical of some period in history? It turns out that you do not really have to study the aesthetic views of that epoch. All you need is just some knowledge of the mechanical properties, the elasticity and strength, of the building materials used at that time.

To make this clearer, we give some examples. They, of course, will be a digression, but they are interesting, and will help us to get the idea. A structure made of huge unattached stone blocks appears tremendously strong in compression, but very weak in shear; bending (torsion) may only be supported by individual blocks. The Egyptian pyramids in which the pharaohs were buried are an extreme example of this method of construction. They have the most pointed tops that one can create using material that cannot withstand shear stresses.

By contrast, for a building without a pointed top, columns (which work in compression) are used to support continuous beams (working in bend). This is exactly how the glorious Parthenon in Athens was designed and constructed.

Layers of bricks or a system of small stone blocks glued to each other would still be very strong in compression, and now satisfactory in shear. However, they still cannot tolerate tension or bend at all. Vaults of a ceiling made of such a material have to arch upwards. This kind of design can be seen in the Hall of Facets in the Moscow Kremlin, and in white stone churches all over Russia. It also occurs in the Gothic cathedrals of Western Europe. In fact, in the latter case, numerous little towers and buttresses are designed to save every part of the walls from any tensile stress, even on the expense of introducing more compression.

Wooden logs are a different example — they are strong in compression perpendicular to their length and in tension along it. Hence, apparently, were built the wooden temples in the North of Russia.

And finally, reinforced concrete can withstand beautifully all types of stresses, and this explains giant vertical and horizontal surfaces in modern constructions, such as the sky scrappers or several hundred meters tall TV towers.

Of course, whatever style and material, there may be all sorts of architecture — some buildings show no spark of talent, whereas some others are real works of genius. But that is a very separate discussion indeed!

Back to biology: If we think of a chemical substance as of some kind of architectural construction, and of the molecules as of a building material, we shall get quite a similar situation. In the previous chapter we looked at how properties of various polymer chains are determined, in the end, by the chain structure of the individual molecules. However, no architect in his or her worst nightmare would ever dream of becoming a polymer technologist. The trouble is that a polymer scientist is unable to pile up the molecules, one by one, in specially determined places, as if they were bricks or logs. Instead, indirect methods must be used, such as, for instance, heating and cooling, or dilution and sedimentation etc., in order to encourage the molecules to arrange themselves in a way at least vaguely similar to what is desired. (Just imagine an architect trying to build something sensible out of a pile of bricks, by merely shaking it, or floating the bricks in water and pouring them out!) This is why the order of molecular segments and the geometrical structure of synthetic materials (particularly polymers) always has many faults (or defects) and can never be perfect.

(For example, if entwining polymer chains in a fiber could be arranged to the perfection of a tidy little girl's hair plait, then we would have fibers of amazing strength — as much as an order of magnitude stronger than that for the best known ones made from a liquid crystalline solution.) However, this is only the case with artificial materials. (Architecture will be mentioned again in this book — see Figure C13.2.)

The situation in biology is totally different. Molecular biology does have something in common with molecular architecture. First, a number of molecules in biological systems are such that they manage to keep, naturally, very high order among themselves (we have seen it with lipids). Second, there are some special systems in a living cell that are capable of arranging molecules in a given structure; an example is a ribosome, which is responsible for protein synthesis. This is why a peculiar novel language comes into use when a physicist starts talking about proteins and nucleic acids – rather unusual and different from all the rest.

5.4 Molecular Machines: Proteins, RNA, and DNA

The special thing about biological macromolecules (proteins, RNA and DNA) is that they have biological functions to fulfill. You could say that proteins, or RNA, or DNA are not only molecules of a particular substance, but each of the molecules is also a device or a machine to do particular operations. In this sense it is more straightforward to talk about such polymers in the language used to describe robots.

In particular, as we have already said, a strictly fixed sequence of different monomer units in a chain of a biopolymer can naturally be compared to a text written with the appropriate molecular "alphabet". Since such a sequence determines chemically the individuality of, say, a protein, then in the spirit of this analogy we can say that a "protein text" lists or codes the function of the protein, it should be compared to a blueprint of a machine. The sequence of monomer units in DNA, as everyone knows, contains genetic information, and it codes the "texts" of proteins by means of the so-called genetic code. This is the cybernetic terminology which is commonly used in molecular biology.

This cybernetic analogy is beautiful and comprehensive, but it does not tell us anything about the way in which the processes actually occur. Why can one protein with a given piece of text detect photons of light in the retina of an eye, and another with a different piece of text cause the physical effort of a muscle, whereas a third controls the immune system,

and a fourth...? (One cannot easily count all the functions of proteins: catalysis of strictly specific reactions, including the biosynthesis of proteins and DNA; transportation of molecules through membranes; tangling and disentangling of knots on DNA, etc. — as a matter of fact, all the processes in a cell are carried out by proteins.) So how is the DNA text read, and how, according to the instructions that it contains, is a protein built? Certainly all these and similar questions are connected with the physics of biopolymers. We already know a lot, but there is still far to go to reach a complete understanding. Perhaps you, however young or old you may be today, will still have many interesting open questions left for your curiosity if you decide someday to take part in these studies. For now, we shall just describe in brief what is already known.

5.5 The Chemical Structure of Proteins, DNA and RNA

5.5.1 *Proteins*

First, a few words about the chemistry of the subject. Monomer units of a protein chain are residues of the so-called amino acids, and have a structure of the sort $-\mathrm{CO}-\mathrm{CH}\mathcal{R}-\mathrm{NH}-$. They are called residues, because amino acids have extra OH group on the left end and extra H atom on the right end; when amino acids combine to form a peptide, (almost) each of them looses water ($\mathrm{H_2O}$), and what enters the chain is a residue. The exceptions are two monomers at the ends: the one which keeps the OH group is called the C-terminal, while the other one which keeps the H is called N-terminal of the chain.

In the above chemical formula for amino acid residue, $\mathcal{R}$ stands for a radical which can be of 20 possible types. In the simplest case it is just a hydrogen atom (-H), and the corresponding amino acid residue is called glycine (Gly). For the remaining 19 amino acids the radical $\mathcal{R}$ has a more complex structure, such as $-\mathrm{CH_3}$ (alanine, Ala), $-\mathrm{CH_2}-\mathrm{OH}$ (serine, Ser), $-\mathrm{CH_2}-\mathrm{CH_2}-\mathrm{S}-\mathrm{CH_3}$ (methionine, Met), $-\mathrm{CH_2}-\mathrm{CO}-\mathrm{NH_2}$ (asparagine, Asn), $-\mathrm{CH_2}-\mathrm{COO^-}$ (aspartic acid, Asp), and $-(\mathrm{CH_2})_4-\mathrm{N^+H_3}$ (lysine, Lys). The latter two examples show that protein chains may contain monomer units which carry a positive or negative electrical charge. As usual, the system as a whole must be electrically neutral, so the amino acid residue might be charged when and if it dissociates, which means it releases its counterion into the surrounding water, or it receives an ion from dissociated water molecule. The sequence of amino acid residues in the chain

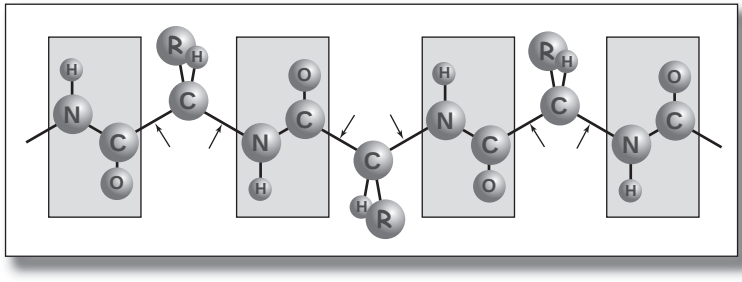


Fig. 5.4 The chemical structure of a protein chain. In the picture, flat amide groups are enclosed in rectangles. They all contain $-\mathrm{N}-\mathrm{C}-$ peptide bonds which form the main chain backbone when the chain is finally complete. The arrows show the bonds about which the chain may rotate. (The corresponding angles of rotation are called Ramachandran's φ and ψ angles.) $\mathcal{R}$ symbolizes side groups of various amino acid residues.

is different for different proteins; you can regard this sequence as a kind of "text" written in a 20-character protein "alphabet". The important point, corroborating the analogy with architecture, is that all individual molecules of any particular protein type have exactly the same sequence of residues; for instance, all hemoglobin molecules in your body have identical sequences. The number of monomer units N in each molecule varies from protein to protein, usually in the range of a few tens up to a few hundreds.

The spatial arrangement of atoms in a short piece of a protein chain is sketched in Figure 5.4. The $-\mathrm{CO}-\mathrm{NH}-$ bond links together $-\mathrm{CH}\mathcal{R}-$ groups which are specific to each unit. It is called a peptide bond; that is why the whole protein molecule is often referred to as a peptide (or polypeptide) chain.

5.2.2 *Nucleic Acids*

The chemical structure of DNA strands is illustrated in Figure 5.5. Each strand is made up of alternating sugar (deoxyribose) and phosphate groups. A nitric base is attached to each sugar group. There are four possible bases: adenine ($\mathcal{A}$), cytosine ($\mathcal{C}$), guanine ($\mathcal{G}$), and thymine ($\mathcal{T}$). They are all shown in Figure 5.5. RNA strands have a similar structure, only with a different type of sugar in the main chain, and the base uracil ($\mathcal{U}$) replacing thymine.

As for the three-dimensional structure, you probably know that in a living cell, DNA molecules consist of two strands (like the one in Figure 5.5) which form a double helix (see Figure C5.10 a little further on). It is

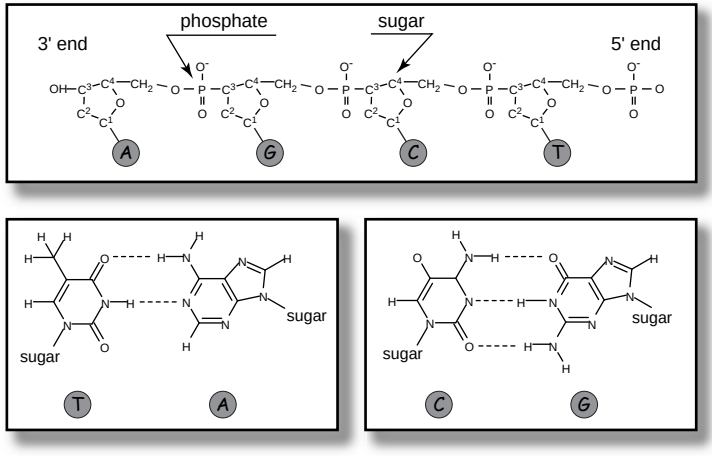


Fig. 5.5 The chemical structure of a single strand of DNA and chemical structures of complementary base pairs.

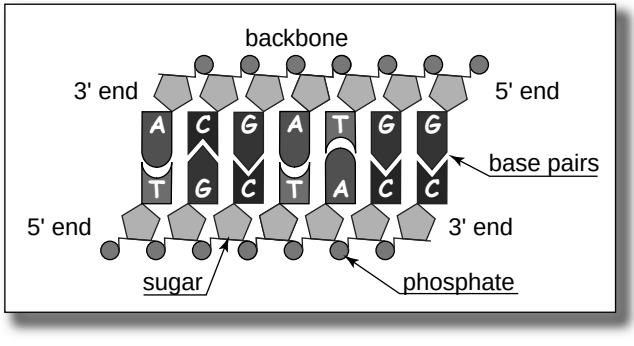


Fig. 5.6 A cartoon explanation of the mutual complementarity of the two strands of the DNA's double helix.

essential that these two strands are mutually complementary. This means that, say, adenine in one of the strands always corresponds to thymine in the other, whereas guanine always corresponds to cytosine. Physically, the reason for this is that the nitric bases are located in the very core of the double helix, where only the pairs $\mathcal{A}-\mathcal{T}$ and $\mathcal{G}-\mathcal{C}$ can fit perfectly without distorting the shape of the double helix. Figure 5.6 explains in a somewhat simplified way why this is so. Hence, the second strand of the double helix contains no extra information, but merely helps to reproduce the information and to make multiple copies of it.

Phosphate groups are negatively charged at normal conditions. That means, of course, that their respective positive counterions dissociate from them and wander in the surrounding water. As a rule, both in the living cell and in the laboratory, there are many other ions, apart from DNA counterions. For instance, there are usually salts dissolved in water, such as regular table salt NaCl, or KCl, or $MgCl_2$, etc. These molecules also dissociate, and they are present in water as ions, such as Na^+ (or K^+, or Mg^{2+}) and Cl^- etc[2]. Tight control of delicate balance of ions is one of the most stringent requirements of survival for every cell, both single cell organism or a cell from, say, anywhere in our body. Not going to any details, we only point out that the most important and the simplest condition is overall electroneutrality: positive ions in any macroscopic volume are almost exactly compensated by the negative ones. However, when we are talking DNA, on the molecular scale, we are dealing with a negatively charged molecule. And the charge of DNA is pretty large. If all of the phosphate groups were to be dissociated, the linear charge density would be $-2e/a \approx -5.9e/\text{nm}$, where a is the distance between neighboring phosphate groups, about 0.34 nm; in other words, surface charge density on the double helix would be $-2e/\pi da \approx -e/\text{nm}^2$, where $d \approx 2$ nm is the diameter of the double helix. This is a huge charge density! It is so large, that in fact only about 25% of all counterions truly depart from DNA into the surrounding medium, the other 75% remain in close vicinity to the double helix (so called Onsager–Manning condensation). But 25% is also a lot — it is in fact practically maximal possible charge density for any molecular system, natural or artificial. So, in a word, DNA is a very strongly charged thing!

For the readers knowing some electrostatics it would be no surprise that charge phosphate groups are located on the outer surface of the double helix. This is because the dielectric constant of water is very large, it is about $\varepsilon \approx 80$ (because water molecules have dipole moments, they are polar), while the interior of the double helix is not polar and only weakly

[2]Under such circumstances, it makes little sense to ask which specific positive ions are "attributed" to DNA as its "own" counterions — protons H^+ or some of the present positive metal ions. This may lead to the question whether DNA and RNA deserve the character A in their abbreviated names, which stands for "acid", or, maybe, it is more productive to think of DNA as a salt of the corresponding acid? This question seems purely terminological to a physicist, but a chemist may have a different opinion. On this topic we can only cite the very first sentence from the famous very first paper by J. Watson and F. Crick, where they announce their discovery of the double helix: "We wish to suggest a structure for the salt of deoxyribose nucleic acid (DNA)".

polarizable, so energy of the charged groups is dramatically reduced when they are in contact with water.

5.6 Primary, Secondary, and Tertiary Structures of Biopolymers

Going into the physics of biopolymers, the first thing to understand is the hierarchy of their structure. As you must have guessed from the title of this Section, there is a primary, a secondary, and a tertiary structure, and sometimes even a quaternary one.

5.6.1 *Primary Structures: Sequences*

The primary structure, as we have already mentioned, is the sequence of units in the chain, that is, the "text". It is created during the biosynthesis of each molecule and is "memorized"; in other words, it is not distorted by the thermal motion unless the whole molecule is destroyed. For a book, it is obvious that thermal motion of atoms in the page is far insufficient to swap letters in a word (making "no" from "on" etc.); for a molecular book, it is much less obvious — but still true, because the sequence of letters in either a DNA or a protein is fixed by strong covalent bonds which are practically never broken by thermal fluctuations under normal conditions.

Biochemists have learned how to "read out" the primary structure of a protein molecule. Although more difficult, methods have been developed to sequence natural DNA's as well. In 2003, an amazing project was completed — the sequencing of (almost) entire human genome: the sequence of about 3×10^9 nucleotides is now known and is stored in a computer memory. Genomes of other organisms, ranging from viruses to mammals, are now being deciphered with an ever increasing speed, and there is a huge, constantly increasing international database of the corresponding data[3]. Unfortunately, early methods of DNA sequencing were rather laborious and, therefore, expensive: the first sequences were decoded at the cost close to \$1 per nucleotide, which meant billions of dollars per genome. But the hopes for a better health care are now largely revolving around the knowledge of DNA sequences — the knowledge of genomes of pathogen viruses or bacteria, the knowledge of abnormal human genes, etc. Therefore, we

[3]Incidentally, the widely known use of DNA in criminal investigations is also based on sequencing, but does not involve genes; it looks at so-called short tandem repeats in non-coding parts of DNA.

— the humankind — desperately need to reduce the sequencing price from its original level by as much as at least six (!) orders of magnitude, that is, a million fold, to make individual sequencing accessible as a diagnostic tool. The work in this direction continues, the sequencing price tag rapidly diminishes, but we need more.

One promising avenue to make DNA sequencing fast and cheap must be mentioned in this book, because of its distinct polymer physics flavor. The idea is as follows (see a sketch in Figure C5.7). Imagine taking a volume of salty water and divide it into two halves by a membrane. Imagine now that there is one little hole in the membrane. With two electrodes in the two halves of the volume, we can use a battery to drive an electric current carried through the hole by the ions (presented in water because it is salty; in a typical case, when salt is the regular table salt, the positive ions are sodium and negative ions are chloride). This part of experiment is very easy, and everyone can make it at home. Now the not so easy part comes. First, let's add some DNA into water on one side of the membrane. Second, let's make the membrane hole so small that only one DNA chain can squeeze there. This is of course not only not easy — this is actually very difficult, but scientists did master it. Moreover, there are two ways to do it. First, one can use a special protein, for example, the one called α-hemolysin, which self-assembles to form a nice pore in a lipid membrane; second, one can use semi-conductor technology to make a little hole in a solid membrane. Another serious difficulty is that when membrane hole is only of a molecular size, the current of ions through it is really tiny small, only about 10^{-9} amps (to put it in prospective, a typical wristwatch consumes from its battery tens of thousands times larger current); it is difficult to measure such current — but scientists did master that, too. Now, when DNA chain squeezes itself into the hole and crawls through it like a snake — it blocks the ions from penetrating through the hole and reduces the current even further. Moreover, the chances of an ion to get through depend to some extent on which particular nucleotide is presently in the pore — therefore, optimists say, if we monitor the tiny fluctuations of ion current in real time as DNA crawls through the hole, we can read out the DNA sequence from this electric signal. Sounds beautiful! Is it possible to realize, is it practicable? Unlike science questions, to which definitive answers are established, this is for now an open matter of opinion - and various opinions, including the diametrically opposite ones, do exist; we recommend a nice review of this topic in the article [51]. But what we can say for sure — this is a truly

exciting area of work, of course, as well as some other propositions for sequencing.

Sequencing is akin to reading. But when a young child masters reading, he or she also learns to write at about the same time; are there then analogs of writing for biopolymer sequences? Yes, there definitely are! First, there is recombinant technology — one can add the desired piece of DNA into an existing organismal genome, such as the plasmid of bacteria. This way one can force the bacteria to produce some protein which is entirely foreign to this bacteria, for instance, human insulin. Second, making a short DNA strand — up to a couple of thousand nucleotides — with any desirable sequence is now a routine procedure: you can write down a desired sequence on the proper web site, pay a modest fee using your credit card ... — and then receive the DNA sample in the mail. The procedure behind this "DNA-made-to-order" is quite complex, but its major component is the so-called polymerase chain reaction, universally called by an abbreviated name PCR. (Using abbreviations instead of words does not add to the language elegance, but the fact of the matter is that some abbreviations — DNA being the most striking example — do become universally accepted technical terms.) PCR uses the DNA polymerase enzyme to produce copies of DNA initially presented by very few molecules. As the reaction develops, the generated copies of DNA automatically become templates for yet new copies, which is why this is the *chain* reaction, it develops exponentially and large number of copies can be quickly produced[4]. The beauty of PCR is that natural DNA polymerase enzyme is employed essentially as an efficient technological device.

It is also possible now to "write" protein sequences: the approach, called protein engineering, was pioneered in 1980s by Sir Alan Fersht in the University of Cambridge in Great Britain. His idea was to use a natural system of biosynthesis that is at work in every living cell; in this sense, it is the entire cell that is now a technological device. Given our ability to prepare any DNA sequence, in principle, we could get a cell to produce protein molecules with any primary structure we wish. However, even if the technical side were no problem, the trouble is that no one really knows what to wish for. When the total number of units $N \sim 10^2$, and there are 20 "candidates" for

[4]The word "chain" in PCR has nothing to do with molecular chains we discuss in this book. In this case, it is the chain reaction because the product of one step in the reaction is immediately used as the reagent in the next step; this type of chemical reactions is particularly common in such areas as combustion and explosion.

each unit, we end up with approximately 20^{100} different possible sequences for a chain of length 100. How can we choose some sensible ones out of such a tremendous number? Of course, one can merely "duplicate" some genuine proteins just as they exist in nature. However, this would be quite an extravagant pastime; it would be only worth trying if we managed to "improve" on nature. It is a bit like a publisher who is unable to read: he is free to publish any book, but how would he choose an exciting and informative text that would become a best-seller? As for biopolymers, particularly proteins and RNA, it only becomes clearer whether their texts are "exciting and informative" once the secondary and tertiary structures are formed.

5.6.2 *DNA Methylation*

We should mention here important complication. Under certain circumstances, the chemical structure of DNA can be modified by the addition of a methyl group for example, to the number 5 carbon of the cytosine pyrimidine ring (see Figure 5.5). This usually plays a regulatory role, for instance, it may have the specific effect of reducing gene expression. DNA methylation was observed in both the cells of adult somatic tissues, and in the embryonic stem cells. There are suggestions that long term memory storage in humans may be regulated by DNA methylation, and that methylation aspect of the sequences can be inheritable.

5.6.3 *Secondary Structures*

Secondary and tertiary structure are the short-scale and long-scale order in the monomers' positions, respectively.

The main secondary structures of proteins were discovered in the 1940–1950s by a chemist Linus Pauling (1901–1994) at California Institute of Technology in Pasadena near Los Angeles (these studies were an important part of the work for which Pauling was awarded the Nobel prize in Chemistry in 1954; by the way, he also won the Nobel Peace prize). They are called α and β structures. They are made stable by the hydrogen bonds. In fact, the reason why the loops of the helix and the β-folds are formed is simply that this is the arrangement that achieves the maximum saturation of the hydrogen bonds.

Both α- and β-structures of polypeptides are quite universal, their structure is only marginally dependent on the sequence of aminoacids (this

dependence, however small, might still be important, for instance, some aminoacids make α-helix more likely, others promote β-strands, and yet others prefer to be in the loops between secondary structure elements). The cartoon of α-helix is shown in the Figure C5.8, while similar cartoon of the β-sheet is presented in Figure C5.9.

The most common secondary structure of DNA was discovered in 1953 by Francis Crick and James Watson at the University of Cambridge in England using the experimental data of Rosalind Franklin at King's College in London — it is the famous Watson and Crick double helix. The discovery of the double helix has won a firm and very fair reputation as one of the major achievements in the history of science. The striking beauty of the double-helix model of DNA is the way it explains, with brilliant ease, one of the real marvels of nature — the ability of all living things to reproduce themselves. Indeed, as soon as the two complementary strands move apart, they form something like a pair of "printing plates" or templates which are ready to make two identical copies. This is exactly how biological inheritance occurs at the molecular level. The canonical Watson–Crick double helical DNA is shown in Figure C5.10.

We should also mention in passing that some DNA sections, with particular types of primary structure, may form very unusual — called also non-canonical — secondary structures under certain conditions. These structures are different from the familiar Watson-and-Crick right-handed double helix (which in this context is called the B-form). In particular, if the DNA is torsionally stressed (e.g., by improperly forming a ring or by magnetic tweezers in the lab), its parts can form a left-handed double helix (the Z-form). There is a possibility to make a triple helix (the H-form), and so on. As another example, you can encounter palindromes in the primary structure of segments of DNA. These are sentences which read the same in both directions; a few funny examples in English are "A man, a plan, a canal — Panama", "Draw pupil's lip upward!", "And DNA", etc. Palindromic bits of DNA often take the shape of a cross (Figure 5.11).

There can be some unfavorable conditions when secondary structures do not develop in biopolymer chains. This can be seen upon a close look at the Figure 2.12: it shows electron micrograph of some DNAs prepared under the conditions of very low salt concentration in solution. Under such conditions, negatively charged phosphate groups of the opposite DNA strands repel strongly and this leads to the unwinding of the double helix at least in some places, shown by arrows in the Figure 2.12. More generally, if the temperature is increased, or some low molecular weight substances

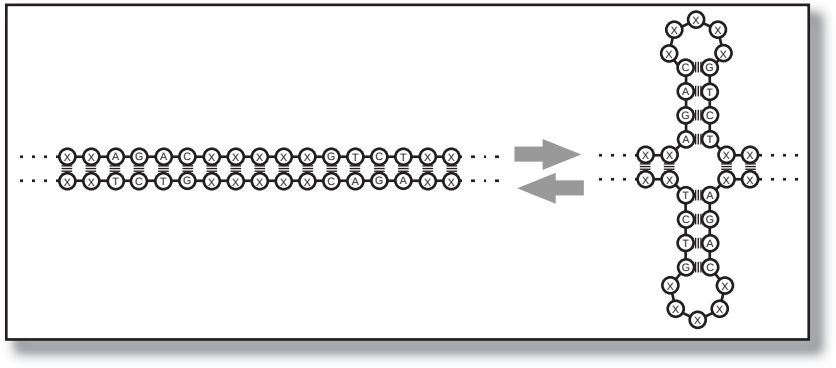


Fig. 5.11 The cross-like structure of a palindromic strand of DNA. To make this structure possible, the sequence must be symmetric, or be a palindrome. The second string's palindrome is not identical, but complementary to the first string's one.

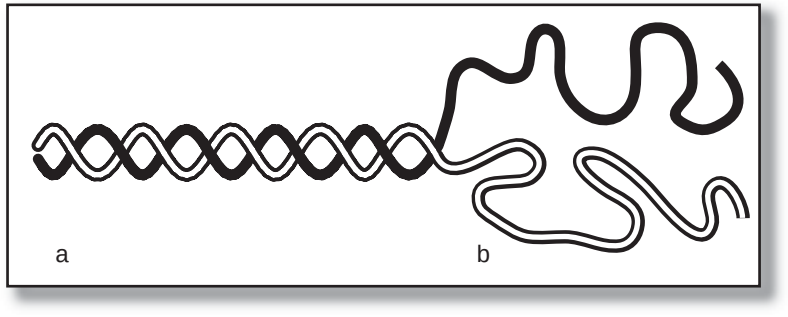


Fig. 5.12 DNA strands: (*a*) in a double helix, and (*b*) in the molten state.

(such as salt) are added to the solution (or removed from it), it can cause untwisting of the helices, called a helix-coil transition. This transition has this name because a spiral-shaped polymer chain is rather rigid, whereas a non-spiral one is a relatively flexible coil (Figure 5.12). Therefore the helix-coil transition is also called the melting of the helix. The physics of helix-coil transition is of great interest.

Helix-coil transition is beautifully used in PCR. Indeed, once new copies of DNA strands are produced tightly wound with their templates of the previous generation. To unwind them and to make them serve as templates once again, experimenter raises temperature, causing helices to undergo helix-coil transition and complementary strands to diffuse away from each

other, after which temperature can be lowered again to repeat the whole process. In fact, historically, the decisive step in implementing PCR as a routine reliable procedure was to employ the DNA polymerase enzyme (so-called Taq polymerase) isolated from a thermo-stable organism (typically the bacterium Thermus aquaticus) which can withstand multiple heating-cooling cycles.

The analogy of helix-coil transition with the ordinary melting of solids is particularly appropriate since both transitions occur very sharply with increasing temperature, over a narrow temperature range. As in melting, helix-coil transition is accompanied by considerable absorption of heat which is used to break the hydrogen bonds forming the helix. Furthermore, just as a crystal melts in big pieces, not atom by atom, in a similar way in helix-coil transition, not only are individual turns of the helix destroyed (this would be similar to the melting of individual cells of a crystalline lattice), but whole chunks of the helix break down. This is a cooperative effect, i.e. the loss of one turn helps the neighboring one to fall apart. Thus, the analogy between ordinary melting and the helix-coil "melting" is rather pervasive, but... However good is the analogy, in some ways the helix-coil transition is different from ordinary melting. The main difference is that spiral and non-spiral strands do not separate out (as, say, regions of ice and water do in a winter river or in your glass of a drink with chunks of ice), but they are mixed along the chain. In the language of theoretical physics, we can say that phase segregation does not occur in the helix-coil transition, and, therefore, strictly speaking, this is not a phase transition.

The theoretical interpretation of all this is rather interesting. Apparently, the helix-coil transition is indeed a real melting process, although not of a three-dimensional crystal, but of a one-dimensional one. In the one-dimensional world, melting is a rather rapid process, but it does not lead to phase separation. This fact is known to physicists as Landau's theorem — because it is briefly mentioned in one of the volumes of the 10-volume "Course of Theoretical Physics" by Lev D. Landau and Evgenii M. Lifshitz.

It is also interesting that a real heteropolymer with a non-uniform primary structure does not melt as sharply as a specially prepared homopolymer. Figure C5.13 explains why. It compares polymers with different primary structures showing their melting curves. These are dependencies of the helical fraction ϑ (sometimes also called helicity; it is the fraction of helical units in the chain) on inverse temperature T^{-1}. Two uniform homopolymers, say $\mathcal{A}-\mathcal{A}-\ldots-\mathcal{A}$ and $\mathcal{B}-\mathcal{B}-\ldots-\mathcal{B}$, both melt rather sharply, but at different temperatures. (For example, the difference in

melting point for DNA molecules which consist only of $A-T$ pairs or only of $G-C$ pairs is as big as 40°C) — see Figure C5.13 *a*. Obviously, a copolymer $\mathcal{A}-\mathcal{A}-\ldots-\mathcal{A}-\mathcal{B}-\mathcal{B}-\ldots-\mathcal{B}$ would melt in two stages (see Figure C5.13 *b*). Hence it is not surprising that the melting of a real heteropolymer with a complex sequence of monomers $\mathcal{A}$ and $\mathcal{B}$ is a somewhat gradual process (Figure C5.13 *c*).

These differences in behavior become even more obvious if we look at the so-called differential melting curves – dependencies of the derivative $\partial\vartheta/\partial T$ on the inverted temperature, T^{-1} (lower graphs in Figure C5.13). Differential melting curves characterize the slope of the $\vartheta(T^{-1})$ dependencies (upper graphs in Figure C5.13). Typical differential melting curve looks like a set of peaks, and Figure C5.13 explains why: if a peak is observed at some particular temperature T_0, then it suggests melting of a particular piece of the helix at or around T_0 — namely the piece whose primary structure happens to mix "stronger" and "weaker" base pairs in such a proportion as to melt at T_0.

5.6.4 *Tertiary Structures*

Thus, the secondary structure of a protein has the geometry of an α-helix or a β-fold with an elementary unit (i.e. a turn of a spiral, or a "hairpin") which includes some three to ten monomers of the chain.

Meanwhile, the tertiary structure is the way the chain is laid out as a whole, i.e. the geometry in which the pieces of the secondary structure are brought together. A tertiary structure is intrinsically different from the secondary one. When the secondary structure is formed, only the monomers which are close to each other along the chain are brought together. On the other hand, the formation of tertiary structure may bring close to each other any parts of the chain, even those separated by very long strands of the chain.

An example of globular tertiary structure of a protein is presented in Figure C5.14. This particular protein is called aspartic protease endothiapepsin, but its name is not really important for us at the moment. Let's look first at the image C5.14 *a*. There, the spirals show the strands of α-spirals (there happens to be only a few of them in this particular protein), and the flat arrows represent the pieces which have β-structure. This way of depicting the protein structure was suggested by the biophysicist Jane Richardson of Duke University. It is good for its clarity; if you tried to draw tertiary structure in more detail, the picture would appear too

complex and hard to understand. The image C5.14 *b* presents the same structure in a different, space-filling way. This representation is useful to show that the globule is actually very compact, the groups of atoms are packed quite densely — but, of course, this way we do not see the elements of secondary structure inside.

The protein in Figure C5.14 is a globular one. This means that its tertiary structure fold into a dense, compact bundle called a globule. We shall talk in detail about polymer globules in Chapters 9, 10, and 14. Among other things, we will discuss how these globules are formed and how they can be destroyed (or denatured, as it is more commonly called), whether this is similar to melting or not, whether it is phase transition or not, and why this issue — dubbed as protein folding problem — is considered one of the central problems in modern biological physics. There, we will also discuss globular states of other biopolymers, particularly DNA (see Section 9.12).

5.7 Globular Protein Enzymes

Quite a lot of proteins have a globular structure. Above all, these include enzymes which catalyze all kinds of chemical reactions in a living cell, in particular biosynthesis of new proteins and DNA. Remember that a catalyst is a substance that speeds up a chemical (or some other) reaction, but is not itself affected by the reaction. A light-hearted example is a subway escalator. Its function is to take passengers up and down. Let's think of these two operations as of two "reactions" going in opposite directions:

$$[\text{person in subway}] + [\text{electrical energy}] \qquad [\text{person above ground}]$$

From the point of view of energy conservation, it is all very simple and straightforward. If a person is in the subway and there is enough electrical energy available, then he or she *can* be moved to the surface (the direct "reaction"). Alternatively, if the person was on the surface and is going down, then his or her potential energy *can* be transformed into electrical energy (the reverse "reaction"). The escalator itself would not be affected by taking passengers up and down. This is exactly what a typical catalyst does. You can come up with many more such examples for yourself. In fact, any kind of machine tool acts as a catalyst. By the way, biological enzymes behave more like man-made machines, rather than like ordinary chemical catalysts (such as, say, a platinum powder which speeds up the

oxidation of sulphur dioxide to sulphur trioxide by nearly 1,000 times, and is used in the industrial production of sulphuric acid).

There are two fundamental similarities between enzymes and machines. First of all, the acceleration of the reaction is extremely high. Usually the reaction does not even occur without the enzyme, in the same way as a rod does not spontaneously shape itself into a bolt without a lathe! The second point in common is the extreme selectivity. An enzyme may work with one substance, but would not work with another, even a very similar one. It is like a cutting tool which cuts right-handed bolts of a certain diameter, but would not make left-handed ones or the ones of even slightly different diameter.

So how do these molecular machines, the enzyme globules, actually work? Figure 5.15 shows the mechanism schematically. A "starter" molecule is to undergo the treatment. It dives into a special cavity, or a pocket, on the surface of the globule which is called an active center. Inside the pocket, the molecule might press a kind of "button", and sometimes the name of active center is reserved for that. Whatever terminology we use, the essence of the matter is that the electron shells of the active center are set into fast motion; then other parts of the globule start moving (although not as fast). They squeeze the "starter" molecule as if with a pair of pincers and pull it, snap it, wring it, and so forth, to make it into the desired shape. In a similar fashion, other proteins fight the "invaders" of our body, such as bacteria and viruses; these proteins are capable of highly specific recognition of other molecules. In fact, the particular protein illustrated in Figure C5.14 is a "recognition molecular machine", and the figure depicts how the recognition is achieved in this case.

Certainly, our description of enzymes and immunoglobulines is rather approximate. On the other hand, a detailed theory of how these proteins really work has not yet been completed. This study forms a subject called enzymology. In any case, what seems apparent at this stage is that since each tool or machine is not just a random pile of bits and pieces, similarly an

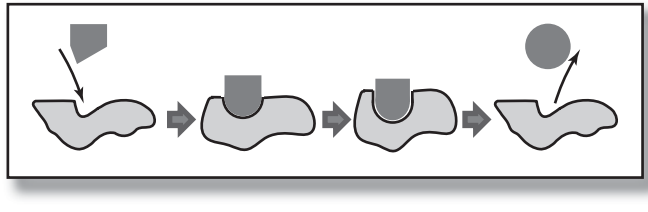


Fig. 5.15 A sketch of different stages of catalysis by an enzyme.

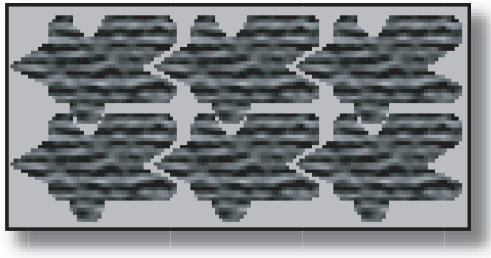


Fig. 5.16 A sketch of a protein crystal.

enzyme's globule should be an organized structure, with all the monomers in well defined places.

This conclusion can be easily tested by experiment. If it is true, then all the molecules of a particular protein should be globules of a strictly identical shape. Therefore, if they are fairly concentrated they would line up in a regular periodic lattice (as in Figure 5.16), a protein crystal. This is exactly what happens. If you extract a protein from a cell and make a concentrated solution, then after some skilful work (aimed at precise determination of properly conducive conditions, such as temperature, salinity, pH — which frequently have to fit into very narrow "windows"), you end up with protein crystals. They are so perfect that they can provide sharp diffraction patterns when illuminated with X-rays. It is worth to emphasize that not only globules form a regular lattice, but every internal element of every particular globule is positioned inside the globule in exactly the same way as its counterparts in all other globules. Studying such diffraction patterns is a good way to find out the spatial structure of the globules. This is just what the scientists do; as a matter of fact, several thousands of tertiary protein structures have already been "decoded" by now.

So what are the forces that hold a protein in the shape of a globule? The globule must be very dense since all the monomers have to occupy fixed positions. It turns out that the shrinking of the protein molecule to a globule is mainly caused by the hydrophobic effect which we have already discussed. About a half of all the twenty amino acid residues are hydrophobic, so they are crammed into the inside of the globule, letting the hydrophilic ones take up positions on the surface. This arrangement reminds us of spherical micelles (Figures 5.1 *d* and C5.2 *a*). The only difference is that now we have only one long chain strung back and forth through the globule. Thus the structure is not quite spherical, but much more complex.

We can now understand better some of the difficulties of protein engineering. The aim is to design the primary structure so that the chain will coil up into a globule with the required tertiary structure, and will act in the desired way. It is no easier than to write a decent book on polymer physics in a language that seems like gibberish, just by manipulating letters and words randomly!

5.8 Molecular Motors

We mentioned that protein enzymes are akin to human-made machines, albeit on a nano-scale. We cannot resist mentioning that some of these machines are correctly called motors, because what they do is they perform mechanical motion against friction on the expense of chemical energy. For instance, some of the neurons (neural cells) in your body have their cell soma (main body) somewhere in your back close to your waist, while their axons go as far as the tip of your toe — perhaps the distance of the order one meter. Some proteins are synthesized in the cell soma and have to be delivered to the cell end. How can they be transported? The cell cannot afford waiting for this to happen via simple diffusion (with a diffusion coefficient around $10^{-9}\frac{\mathrm{m}^2}{\mathrm{s}}$, this would require about 10^9 s, or 30 years). Instead, special motor proteins, called kinesin, deliver the cargo by traveling along the microtubules — by the way, yet another polymer-like object. Another motor proteins, called dynein, deliver cargo in the opposite direction.

Perhaps the most widely known motor protein is myosin, responsible for the work of our muscles. Yet another motor proteins push DNA to make the virus. And the list of examples is easy to continue.

From chemical point of view, what these motor proteins do is they catalyze the reaction of hydrolysis of ATP; each act of hydrolysis yields energy about 14 k_BT. Motor proteins use this energy to perform *directed walk* in one particular chosen direction (instead of random walk in random direction in thermal equilibrium).

We have to emphasize that molecular motors are *not* classical heat engines familiar from physics and thermodynamics classes. It is not because molecular motors are small. It is because they use chemical energy directly transforming it into mechanical work, while human built heat engines transform chemical energy of a fuel into heat and then transform part of the heat into mechanical work. Only part of the heat can be transformed (only free

energy, not energy, is available to do work), which is why the efficiency of heat engines is always smaller than 100% (this statement is known to physicists as Carnot theorem). By contrast, molecular motors do not dissipate chemical energy into heat before doing mechanical work, which is why their efficiency is not restricted by any fundamental physics law, and, depending on how one defines it, may approach arbitrarily close to 100%.

The fact that molecular motors are *not* heat engines has also other very direct consequence: as mechanical devices, they are reversible. What does it mean? For instance, a molecule of kinesin, under physiological conditions, burns its fuel (hydrolyzes ATP) and spends the obtained energy to walk in a particular direction. But what would happen if we intentionally starve the kinesin on ATP, provide the excess of ADP and inorganic phosphate (products of hydrolysis), and forcefully pull kinesin molecule along its track? Well, you may guess, the machine is reversible — it will start catalyzing the reverse chemical reaction, producing ATP out of ADP and phosphate on the expense of the external source performing mechanical work to pull the kinesin forward. At the first glance this may sound a bit like pulling a car by its bumper in the hope that it will start producing gasoline — but the car motor is a heat engine, which is why it is irreversible and can never be run in the opposite (gasoline producing) direction, while kinesin molecule is a mechanical system and it can be reversed. In this sense, kinesin is more like an electric motor, for which the reversion — turning the shaft and getting the electricity out — is a very common lecture demonstration in physics.

It is very tempting to tell you more about these amazing motors and how they work, because it is very interesting, but we must restrict ourselves.

5.9 Physics and Biology

There is just one more question we would like to broach to conclude this chapter. We have been using the words: primary, secondary, tertiary. And what comes next? Sometimes the name quaternary structure is introduced when a few protein globules are stuck together, or when one protein chain forms a number of little globules. Clearly, there is the whole hierarchy of structures: There are complexes of chains, these complexes form parts of cells, the cells make up tissues, and so on.

Where does the dividing line between physics and biology go, in the face of such a variety of structures and systems? Physicists and chemists

study atoms and molecules. Biologists dig their way towards them, from the other end, by looking at organs, tissues, and cells... Do their paths ever meet, or even overlap, or are they separated by an impermeable stone wall? This is a crucial question. If there is a wall, then we will never succeed in understanding life. Biopolymers, with their three levels of structure (i.e., primary, secondary and tertiary) seem to have all the chance to connect the territories of physics and biology. On the one hand, a biopolymer chain is just a molecule. In this sense, it should be the job of physics to explore its properties; moreover, physics and physicists are well equipped to achieve that. On the other hand, a biopolymer molecule has aspects which could be called "specifically biological", specifically, it carries information which was created by evolution, and it is tailored to perform a certain function.

Indeed, what do all living things, from an elephant to a microbe, have in common? One of the main features is some kind of "design" or "construction" which the creature holds on to from birth to death. But this is just what we could say about a biopolymer chain too (although its design is much simpler). A biopolymer keeps its structure unchanged, from the moment when it is synthesized, chemically or in the cell, until it is destroyed. And the biopolymer molecule is apparently the simplest of all systems possessing the property of having a "construction" or "design". That is why we are so enthusiastic about biopolymers serving as a real bridge between the realms of physics and biology.

This idea is so attractive, and looks extremely simple. It seems obvious indeed! In fact, it seems the property of really good ideas to become obvious once they are formulated! And it is also usually quite difficult to identify who was the first to formulate such idea. The authors of the present book know that this deep view of biopolymers as physical objects possessing the property of design was clearly formulated in 1968 by our teacher, the Russian physicist Ilya M. Lifshitz of the Institute for Physics Problems in Moscow. He coined the term "linear memory" of biopolymers — as if they always "remember" the linear structure they were given when synthesized. The invention of the term does not seem particularly successful, it is rarely (if at all) used at present. But the idea itself is more than successful. Was Lifshitz the first? He was definitely *among* the first, and his thinking was perhaps the most clear and the most advanced at the time, but similar ideas were slowly coming into the existence through the works of quite a few people worldwide, and by the mid 1980s the idea had become a matter of course (you may want to remember our comments on how Staudinger discovered the chain structure of polymers).

Thus, it turns out that if you start to look into the physics of systems with linear memory (to use our teacher's word), you may hope one day to come face to face with the mysteries of biology. We shall talk more about this in the chapters that follow.

This page is intentionally left blank

Color Figures for Chapters 2–5

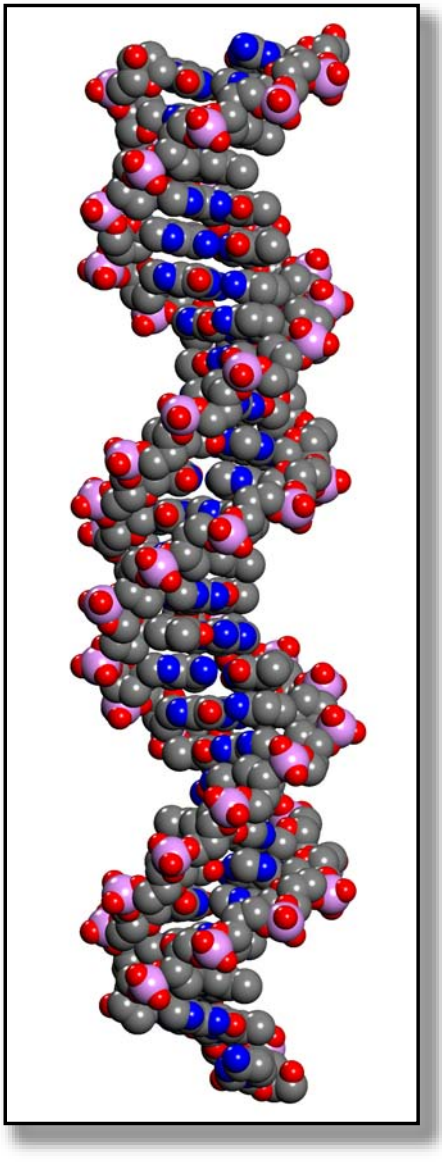


Fig. C2.4 Spatial structure of a strand of a DNA double helix. Atoms of different types are shown with spheres of different shades.

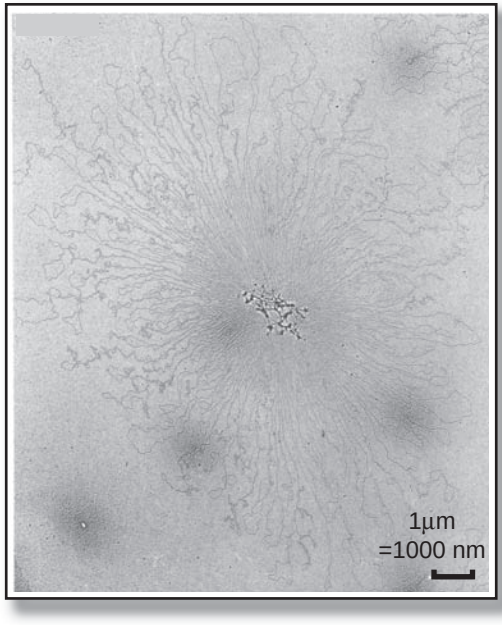


Fig. C2.7 Electron microscope picture of bacterial DNA partially released from the cell (bacteria *E. coli*) upon gentle breaking of the cell outer membrane (see Sections 2.4, 6.3 and 9.12). The figure illustrates how dense is the DNA packing under native conditions: we see that even spilled out DNA is rather dense and tangled, we can therefore imagine how dense it was while still inside the cell. The scale bar in the image corresponds to 1 mm, or 1,000 nm — about the size of the cell. The figure is reproduced with permission from the classical paper: Ruth Kavenoff and B.C. Bowen, "Electron Microscopy of Membrane-Free Folded Chromosomes from *Escherichia Coli*", Chromosoma, v. **59**, n. 2, pp. 89–101, 1976.

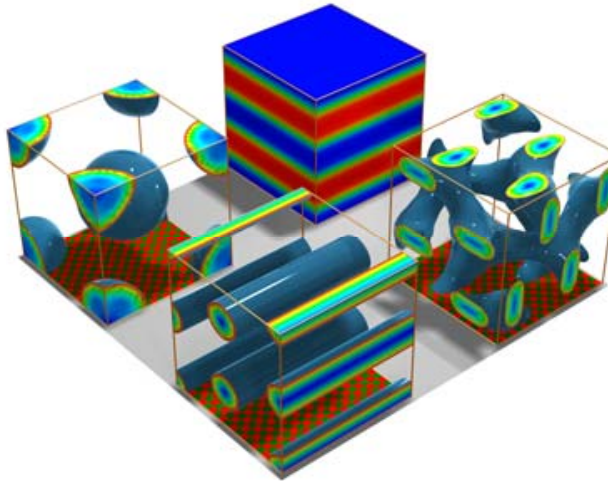


Fig. C4.9 Four different microphase segregated structures which can exist in diblock-copolymer melt. Layers (lamellas), spherical micelles, cylindrical micelles, and bicontinuous phase are shown. The figure is courtesy of P.G. Khalatur.

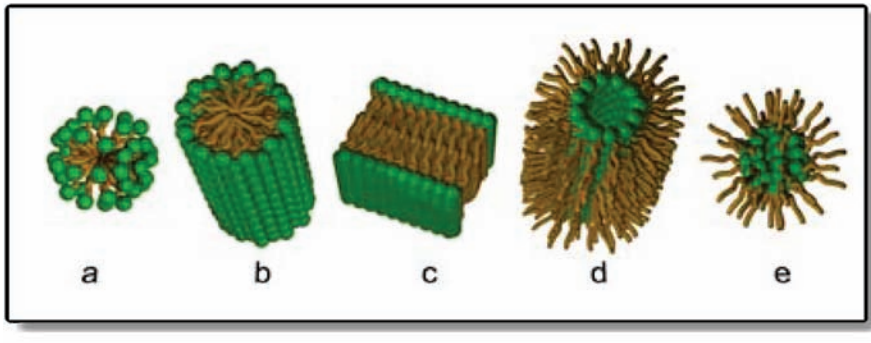


Fig. C5.2 Possible structures formed in solution of amphiphilic molecules: spherical micelles (*a*), cylindrical micelles (*b*), lamellas, or layers (*c*), inverted cylindrical micelles (*d*), inverted spherical micelles (*e*). In three-dimensional space these structures will organize pretty much in the same way as shown in Figure C4.9. The figure is courtesy of P.G. Khalatur.

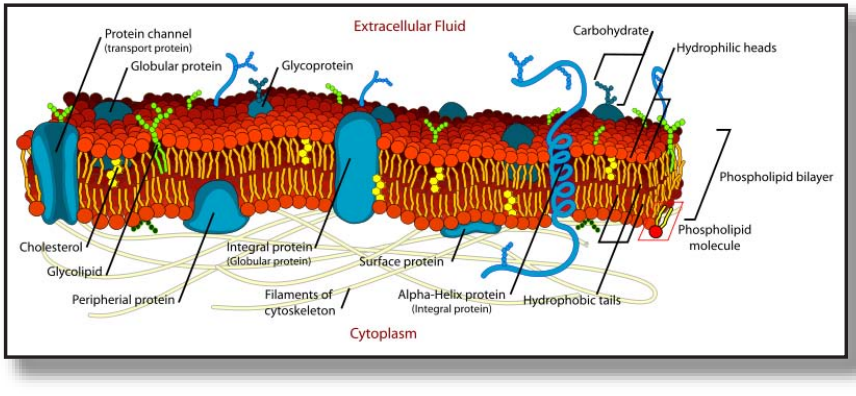


Fig. C5.3 A cartoon of a cellular membrane. The various parts of it are labeled in the figure. The image was created by Mariana Ruiz Villareal and is available in public domain.

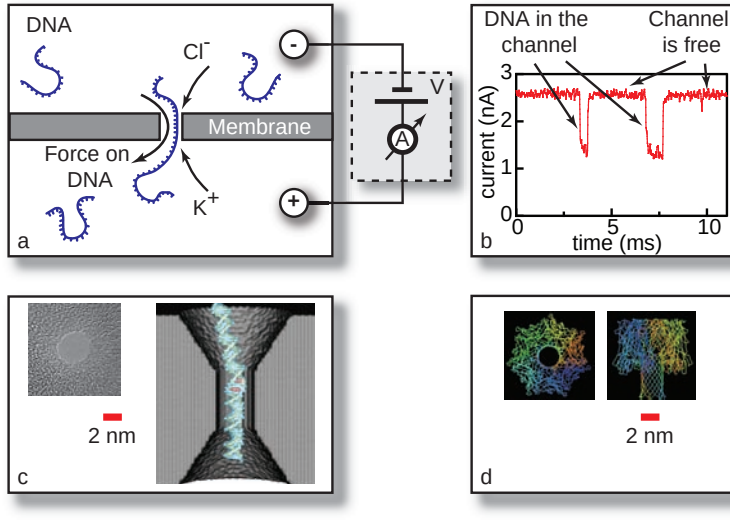


Fig. C5.7 An illustration of experiment on DNA translocation through a nanopore. Panel (*a*) is a cartoon of the overall experimental setup: a thin membrane with a single narrow channel separates two volumes of salty water. Once electrical voltage is applied by the battery, the current is induced through the pore, carried by the salt ions, typically K^+ and Cl^-. Since DNA is negatively charged, it is also the subject of force and every now and then DNA gets thread into the pore. Since the pore is so narrow, the DNA obscures the passage of ions, and thus blocks the current. Panel (*b*) shows typical current trace indicating two clearly visible events of DNA passage through the channel. Both the dwell time (needed for DNA to pass through the channel) and capture time (needed for the next DNA to arrive to the pore) can be reliably measured from such current traces. Panels (*c*) and (*d*) illustrate the two possible ways to make proper nanopores. The (*c*) is the solid state nanopore: left is TEM (transmission electron microscope) image of a 4 nm pore fabricated in 20 nm thick silicon nitride film; right is the reconstructed image of double stranded DNA threaded through a 4 nm pore. The (*d*) are top view and side view of the α-hemolysin protein — better to say, self-assembled system of 7 protein molecules, which is stable when its "stem" is submerged into a lipid membrane. Protein channel is narrow and is suitable for single stranded DNA translocation only (double helix does not fit into the hole). Notice that the scale in panels (*c*) and (*d*) is the same; protein nanopore is significantly smaller indeed. The figure is courtesy of Amit Meller. Parts of the figure are reprinted with permission from M. Wanunu, M. Sutin and A. Meller, "DNA Profiling Using Solid-State Nanopores: Detection of DNA-Binding Molecules", Nano Letters, 2009. Copyright 2009, American Chemical Society.

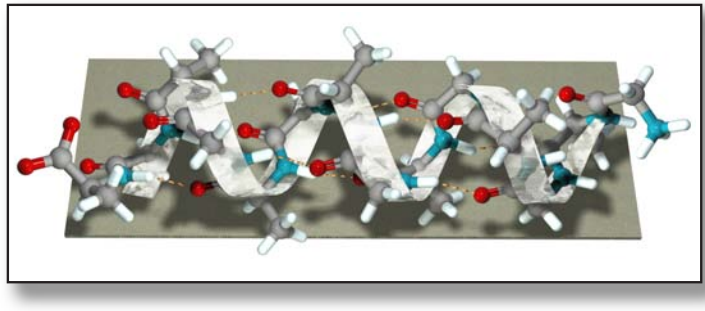


Fig. C5.8 One of the two most common secondary structures of protein: α-helix. Notice that while main chain is of helical shape, the side groups of aminoacid residues extend outward from the helix, which is why the geometry of the backbone in α-helix is relatively universal, it is only weakly dependent on the sequence. In this illustration, for simplicity, only the least bulky side groups (glycine, side group H, and alanine, side group CH3) are used. The geometry of α-helix is such that the advance per one amino acid residue along the helix axes is 0.15 nm, while the pitch (or advance along the axes per one full turn of the helix) is 0.54 nm. That means, the turn around the axes per one monomer equals $360^\circ \times 0.15/0.54 \approx 100^\circ$; in other words, there are $360^\circ/100^\circ \approx 3.6$ monomers per one helical turn. Accordingly, hydrogen bonds $\mathrm{CO}\ldots\mathrm{HN}$ (shown in the figure as dotted lines) connect residues k and $k+3$ for every k.

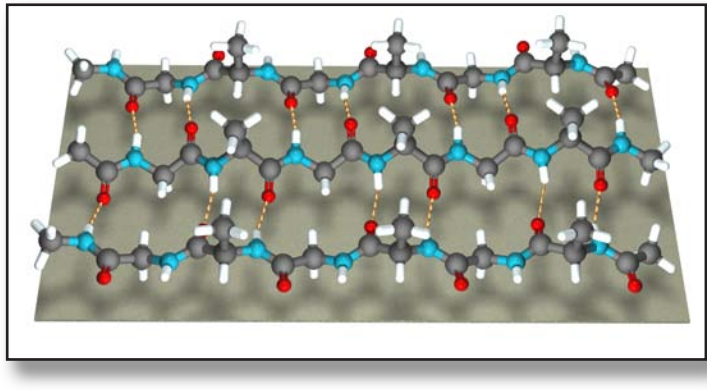


Fig. C5.9 One of the two most common secondary structures of protein: β-sheet. It is stabilized by hydrogen bonds between atoms of the main polypeptide chain, not involving the aminoacid residues side groups, which extend above and below the sheet. As in case of α-helix, only the least bulky amino acid residues, glycine and alanine, are used for this illustration.

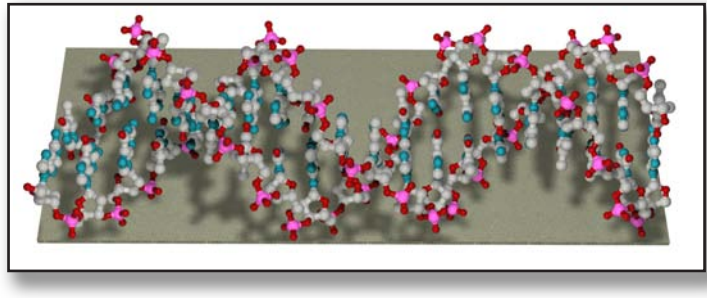


Fig. C5.10 The most common secondary structure of DNA is a double helix. Under normal physiological conditions (of temperature and ionic strength) it has about 10.4 base pairs per helical turn, with the distance between base pairs along the helix about 0.34 nm.

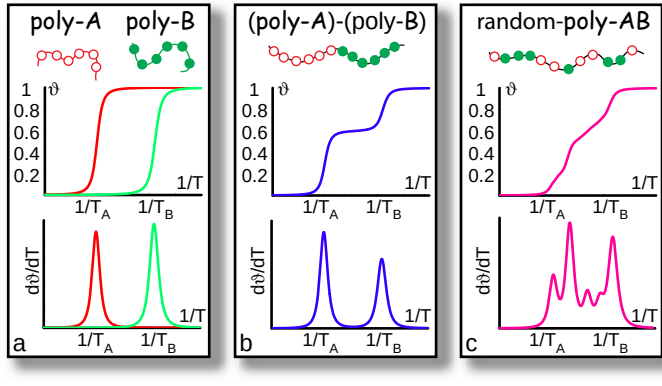


Fig. C5.13 Melting curves, $\vartheta(T^{-1})$, upper plots, and differential melting curves, $d\vartheta/dT$, lower plots, for different sequences, presented at the top: (*a*) poly-$\mathcal{A}$ and poly-$\mathcal{B}$ homopolymers; (*b*) block-copolymer with 60% $\mathcal{A}$ and 40% $\mathcal{B}$; (*c*) heteropolymer with random alternation of the monomer units.

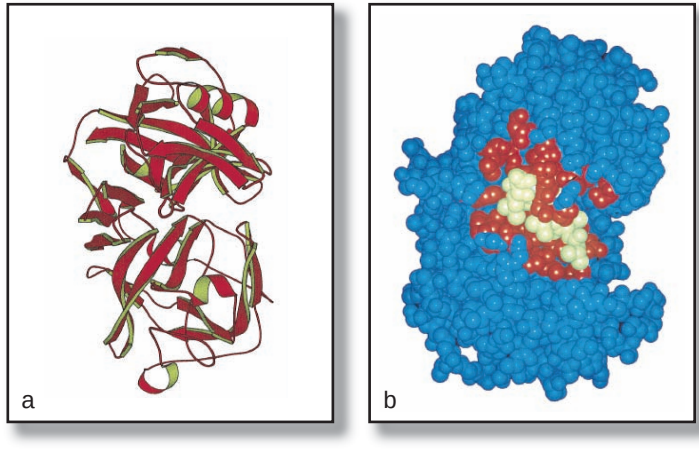


Fig. C5.14 A realistic image of molecular recognition by a protein called aspartic protease endothiapepsin. Figure (*a*) represents a ribbon diagram of three-dimensional structure of the globule in the form in which it was crystallized. It is seen that the globule consists of two domains; they have, respectively, 175 and 153 residues. The binding pocket is situated in the gap between two domains. In figure (*b*), the same protein is shown in space-filling representation; the main body of the protein globule is shown in blue, while yellow is the "target" molecule to be recognized. Red is the part of the protein that forms an active center. Reproduced with permission from: J. Gomez and E. Freire, "Thermodynamic Mapping of the Inhibitor Site of the Aspartic Protease Endothiapepsin", Journal of Molecular Biology, v. **252**, n. 3, pp. 337–350, 1995.

Chapter 6

The Mathematics of a Simple Polymer Coil

> It was customary for wealthy pirates to start with a proper wardrobe.
>
> Mark Twain,
> *The Adventures of Tom Sawyer*

6.1 Mathematics in Physics

In the two previous chapters we looked at the properties of real polymeric substances. We have come across both artificial polymers, which are used in industry or in everyday life, and natural polymers, the building bricks of life. We only used words to describe them, without any mathematics. However, it was more like a story than a science, so our description was rather superficial. In order to understand polymers better, as always happens in physics, one has to move on from words to mathematics. This is because "those who have mastered at least the principles of mathematics give the impression of people with one more sense than other mortals" (Charles Darwin[1]). Moreover, "mathematics is the language in which the gods talk to people" (Plato).

However, mathematical descriptions have their own "game-plan". Real systems are so extraordinarily complex that if you wanted to describe them fully, you would have to take into account an incredible number of different factors. This would be a hopeless task. The way out is to simplify reality, grasping the main features and ignoring all the less important. Fortunately, constructing a theory with even a very simple model usually pays off.

[1] Some other sources cite a slightly different formulation: "Mathematics seems to endow one with something like a new sense".

When you get a deep feel for the properties of the simple model, it opens your eyes to the behavior of the real system too.

In this chapter, we shall discuss different mathematical descriptions of the simplest model of a polymer, the "ideal polymer coil" (the reason for this name will become clear in Chapter 7).

To tell the truth, the authors are physicists, not mathematicians. So we can fully appreciate Goethe's joke when he said: "Mathematicians are like Frenchmen: whatever you say to them they translate into their own language, and forthwith it is something entirely different". Goethe's sarcasm was directed perhaps against the invading armies of Napoleon, rather than against the poor mathematicians. However, his attitude is echoed by John Ziman, the English theoretical physicist: "Nothing is more repellent to normal human beings than the clinical succession of definitions, axioms, and theorems generated by the labours of pure mathematicians". Therefore, we shall try to spice up the following chapters with some history and various physical analogies. Occasionally we may even wander off into some "non-polymer" physics. In any case, we are not going to do math just for the sake of it. The physical sense and meaning of mathematical formulae will be our main concern.

6.2 Analogy Between a Polymer Chain and Brownian Motion

Imagine you are in a thick forest. You have picked enough mushrooms and berries (or whatever you were gathering there), the weather has become bad, and all you want now is to get out of the wretched place. But how? The trees and bushes hinder your view and make it hard to walk. You cannot see the sun behind the clouds... It seems quite certain that you are facing a hard time — unless you have got a compass. (Well, in theory they say that an experienced person can tell directions by looking at how moss and lichen grow on tree trunks, and where ant hills are, etc. — but that is not really our subject.) But with a compass — would it be of any use without a map? You would not know which direction you need to take... Well, it appears a compass would still be extremely useful. Soon we shall see why.

We are telling you this story for a good reason. It will help us to comprehend a deep mathematical concept which has been very fruitful when explaining the behavior of polymers as well as many other things, in the

fields ranging from biology to economics. Historically, this concept was first developed when Brownian motion was studied.

Brownian motion itself, as its name reveals, was discovered in 1827 by the English botanist Robert Brown (1773–1858). Looking through a microscope at little particles of pollen suspended in water, he was fascinated by their random "dances". The particles were moving by themselves, apparently with no external encouragement. So a lot of people decided that there must have been some "living power" causing the motion (because the flowers were animate!) They reckoned this had proved that there was some mysterious "substance" which made the animate different from the inanimate.

The question was debated for a long time. Everyone was free to think what they wanted. Then there was a dramatic boom in interest at the end of the 19th century. Brownian motion was regarded as a kind of perpetual motion, so it tantalized those who were puzzled by the general problems of science. These included the nature of irreversibility (i.e. the distinction between past and future) as well as the difference between Darwinian biological evolution leading to perfection of species, and the thermodynamic evolution described by Clausius, Thompson, and Boltzmann, which leads to dissipation, or, as it was then called, "thermal death."

Eventually, the answer was found by Albert Einstein[2] and the Polish physicist Marian Smoluchowski (1872–1917), then a professor at the University of Lviv. The title of one of Einstein's papers on the theory of Brownian motion is rather telling: "On the motion of particles suspended in resting water which is required by the molecular-kinetic theory of heat". Einstein and Smoluchowski considered chaotic thermal motion of molecules and showed that it explains it all: a Brownian particle is "fidgeting" because it is pushed by a crowd of molecules in random directions. In other words, you can say that Brownian particles are themselves engaged in chaotic thermal motion. Nowadays, science does not make much distinction between the phrases "Brownian motion" and "thermal motion" — the only difference lies back in history. The Einstein–Smoluchowski theory was confirmed by beautiful and subtle experiments by Jean Perrin (1870–1942)[3]. This was a long awaited, clear and straightforward proof that all substances are made of atoms and molecules[4].

[2]By the way, Einstein presented his theory of relativity and the concept that light consists of photons in exactly the same year, 1905.

[3]You can read about Perrin's experiments in a very interesting book [55].

[4]The atomic hypothesis was suggested long ago by the ancient Greeks, but it had to wait for more than two thousand years to be proved!

We will skip further details of this adventure story. We just need to emphasize one more thing before we get back to polymers. Since a Brownian particle moves due to collisions with molecules, its path breaks into a sequence of many very short flights and turns. In this sense, a Brownian trajectory is pretty similar to the shape of the polymer chains which we saw in Section 2.4 (Figure 2.6). Another obvious example of this sort is of a man who is lost in a forest, with no compass, and has no choice but to wander at random.

Certainly, no microscope would let you see the twists and turns of an individual molecule's path. However, the Einstein–Smoluchowski theory tells us how to spot the difference between a "fuzzy" line which consists of a great number of tiny random kinks, and an ordinary smooth curve, even though we cannot discern the individual kinks. (We do not always need to see everything, e.g. we can happily tell water from alcohol even though the individual molecules are invisible!) In the same way, a polymer chain looks nothing like a shape stretched in a certain direction. And the path of a man in a forest would depend quite noticeably on whether he is equipped with a compass or not!

So what is the difference between a smooth and a "kinky" path?

$$\text{For motion in a straight line:} \qquad R = v\,(t_2 - t_1) \tag{6.1}$$

$$\text{For a Brownian particle:} \qquad R = \ell^{1/2}\,[v(t_2 - t_1)]^{1/2} \tag{6.2}$$

The notation here is as follows: in formula (6.1), R is the displacement, i.e. the distance $R = |\mathbf{R}_1 - \mathbf{R}_2|$ between the initial ($\mathbf{R}_1$ at time t_1) and final ($\mathbf{R}_2$ at time t_2) points of the motion ($t_1 < t_2$), v is the average velocity of the motion. In formula (6.2), R also characterizes the distance between initial and final positions, but since the motion is random, R should be understood as an average; more specifically, it is the root-mean-square displacement: $R = \left\langle (\mathbf{R}_2 - \mathbf{R}_1)^2 \right\rangle^{1/2}$, where the angle brackets indicate that the average is taken over a number of different Brownian paths. Apart from this technical detail of the definition of R, Equation (6.2) is fundamentally different from (6.1) because R is proportional to the square root of elapsed time instead of time itself. The price for that is the appearance of a new parameter ℓ which has the dimension of length and whose physical meaning we will have to discuss and explain[5]).

[5]The Einstein–Smoluchowski theory leads to the value $\ell = (mk_BT)^{1/2}/(3\pi\eta r)$ for spherical Brownian particles of radius r and mass m moving in a liquid of viscosity η at a temperature T.

What is the polymer analogue of the Einstein–Smoluchowski equation (6.2)? Let L be the contour length of a polymer chain. It is bound to be proportional to the number of monomers in the chain, given that the chemical structure does not change. The chain length L plays the same role for a polymer as the value $v(t_2 - t_1)$ for a Brownian particle, that is the total distance traveled by the particle along the path. Since the chain wiggles around a lot, the root-mean-square distance between its ends, $R = \left\langle (\mathbf{R}_2 - \mathbf{R}_1)^2 \right\rangle^{1/2}$, is totally different and not even proportional to the contour length L. You can easily find R from Equation (6.2) if you replace $v(t_2 - t_1)$ by L:

$$R = \ell^{1/2} L^{1/2} = (\ell L)^{1/2} \ . \tag{6.3}$$

6.3 The Size of a Polymer Coil

Let's look again at formulas (6.1) and (6.2). How different are they really? And how similar are the end-to-end distance R and the contour length L of a polymer coil? The answer is that the differences are very significant, and we shall try to explain why.

The main distinction between (6.1) and (6.2) can be spotted at the first glance — the power law for R depending on the time interval $\Delta t = t_2 - t_1$ is not the same! In case you think it is not important, or not even worth mentioning, we will use words and numbers rather than formulas to make it clearer.

Let's start with numbers, in other words, with estimating the order of magnitude of things. As we have seen, the rambler lost in a forest is very similar to what happens to a Brownian particle. Say, the rambler spends 10 h per day walking (i.e. $\Delta t = t_2 - t_1 = 10$ h) at a speed of $v = 3$ km/h. (It is hard to move much faster in the forest.) If he uses a compass or some other means of judging direction, his path will look more or less like a straight line. The displacement will be given by (6.1); in our case, it is 30 km. Having strolled that far, the well-prepared hiker is quite likely to find the way out, or at least to reach a major road or path. However, if he wanders randomly with no guidance, the situation becomes much more serious. Let's just take for granted for a moment that $\ell \approx 300$ m; then we can use (6.2) and discover that the poor chap will move no more than 3 km from where he started... Hence, good advice for stray hikers in a forest is: Do not rush about! Just keep going in a fixed direction. It does not

matter where, as long as you use some landmarks to guide you in a straight line[6].

By the way, lost travelers are often known to feel that they are going in circles. All kinds of absurd causes have been suggested to account for this. Some reckon it is because one's two legs may differ in length or strength, others blame Coriolis forces, or even giddiness due to the rotation of the Earth. However, the proper scientific explanation immediately follows from Figure 2.6. A meandering, entangled path tends to cross itself a number of times, so it is not surprising that a lost traveler may occasionally find himself in a place where he has been before. In Chapter 8, we shall estimate how many self-crossings, on average, there are along a random path.

Now, what does a "self-crossing" mean in the case of a polymer? The entangled shape brings some monomers, which are separated by a long piece of the chain, closer together. They may even collide. Interactions between such monomers during their collisions are known as volume interactions. They occur in the bulk of a polymer coil, or inside its "volume" — in contrast to "linear" interactions that hold together the neighboring monomers along the chain. A little later, we shall discuss in detail how various properties of polymers are affected by volume interactions.

Let's go back to formulas (6.1) and (6.2), and look at another example directly connected with polymers. You may remember that the contour length of a DNA double helix from a human or an animal cell can be as big as one meter. The question is, how small a coil would the double helix form, if it meandered randomly, in a Brownian-like way? The value of ℓ for DNA has been measured in experiments; to high accuracy, $\ell = 100$ nm $= 10^{-7}$ m. Thus, Equation (6.3) leads to:

$$R \approx 3 \cdot 10^{-4} \text{ m} = 0.03 \text{ cm} . \tag{6.4}$$

This is certainly much less than the one-meter contour length, yet far too big to squeeze into a cell nucleus which is about 10^{-6} m in diameter! Thus, the fact that DNA coils up is important to understand how DNA is arranged inside the cell, although it is not sufficient. Of course, if you think about it, you can realize that it could not possibly be sufficient, because DNA's

[6]Even if the forest is huge, and you know that one of its borders is much closer than the opposite one, you still should not wander around, but advance in a straight line. Only if the lapse of time suggests that you must be straying away from the closest border, should you turn abruptly (e.g., through 120°) and try another direction. This sounds like a lovely problem for Maths enthusiasts — to work out the best strategy for a rambler who starts to make his way out of a forest at a given distance from the closest straight-line border.

shape cannot be merely random, there has to be some regularity in it, otherwise it would be impossible to find its different strands and "read" the information from them. Scientists are still a bit vague on how exactly nature has managed to pack the double helix into a cell (see Figure C2.7). There are some ideas, however, which we are going to discuss in Chapter 9. For now we shall only say that the main part is played by the volume interactions mentioned above. They help DNA to fold in such a way that its size becomes proportional to the number of monomers N (i.e. the contour length L) not raised to the power one (as for a stick), nor even to the power $1/2$ (as for a randomly entangled coil described by (6.2) or (6.3)), but to the even smaller power $1/3$.

6.4 Derivation of the "Square Root" Law

Now we know that the main peculiarity of a randomly entangled polymer coil is that its size is proportional to the square root of the chain's contour length: $R \sim L^{1/2}$ (see (6.3), for example). This is why a polymer molecule appears much smaller in size than it would be if entirely stretched (given that it is reasonably long). Indeed, the ratio $R/L \sim L^{-1/2}$ decreases with L, and tends to zero when $L \to \infty$ for any ℓ. However, if we wanted to find the size R of a particular molecule with a given contour length L, we would need to know the value of ℓ. In any case, it would help if we knew more about the physical meaning of ℓ. Also, since we realize the importance of the "square root" law, it might be a good idea to see where it comes from.

To tackle both questions, we shall abandon the Brownian analogy and return to the purely polymer world, as it will be easier that way. Let's imagine a freely-jointed polymer chain consisting of N units (Figure 2.5 *b*). The end-to-end vector $\mathbf{R}_N$ is given by a simple formula which is obvious

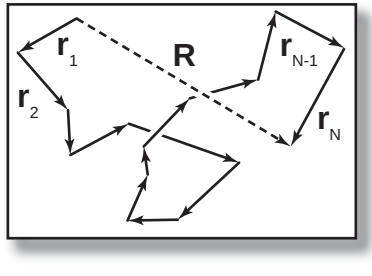


Fig. 6.1 Illustration for the formula (6.5): random walk or a polymer chain can be presented as a succession of vectors $\mathbf{r}_i$, and their vector sum is the end-to-end vector $\mathbf{R}$.

from Figure 6.1:

$$\mathbf{R}_N = \sum_{i=1}^{N} \mathbf{r}_i \ . \tag{6.5}$$

Here i labels chain segments, N is the total number of segments in the chain, and $\mathbf{r}_i$ is an end-to-end vector for the ith segment. The moduli of all the vectors $\mathbf{r}_i$ are the same, they are equal to the length of one segment: $|\mathbf{r}_i| = \ell$ (we have deliberately chosen this notation for the length of a segment). Meanwhile, the directions of the vectors $\mathbf{r}_i$ are completely random and independent of each other.

Thus, the value of $\mathbf{R}_N$ that describes the size of the coil can be written as the sum of a large number of independent random terms, as in (6.5). Fortunately, mathematical properties of such sums have been very well studied, since they often emerge in areas of maths, physics, engineering, and biology.

We will pick just a couple of examples. The stress experienced by an airplane, for instance, depends on the total weight of the passengers, the fat and the skinny altogether, and so the total weight is the sum of random contributions. Similarly, in light scattering, the electromagnetic field of the scattered wave is the superposition (i.e. the sum) of the fields created by individual atoms. These contributions are totally random due to thermal motion of the atoms; when they interfere they can either enhance or diminish each other. In both examples, as well as in lots of others, the answer about possible deviations from the average can be found from the "square root" law. Let's derive this law from formula (6.5).

Together with $\mathbf{R}_N$, we need to introduce $\mathbf{R}_{N-1}$ which is a similar end-to-end vector, but for the the first $N-1$ units of the chain. We can write down:

$$\mathbf{R}_{N-1} = \sum_{i=1}^{N-1} \mathbf{r}_i \ , \qquad \mathbf{R}_N = \sum_{i=1}^{N} \mathbf{r}_i$$

$$\mathbf{R}_N = \mathbf{R}_{N-1} + \mathbf{r}_N \tag{6.6}$$

(Recursion relations of this kind are often useful when one tries to sort out random values.)

Now we can start thinking how to work out the average end-to-end distance. But, first of all, we need to decide what sort of average to look at. The point is that the average value of the vector $\mathbf{R}_N$ itself, as well as of all its components, is zero, i.e. $\langle \mathbf{R}_N \rangle = 0$. This is simply because the end-to-end vector can be equal to $\mathbf{R}_N$ and to $-\mathbf{R}_N$ with the same probability.

Therefore, it is the average length of the vector, $\langle|\mathbf{R}_N|\rangle$, that is meaningful, and which gives us an idea of the size of the coil. However, it is handier to calculate the value

$$R_N^2 \equiv \left\langle \mathbf{R}_N^2 \right\rangle = \langle \mathbf{R}_N \cdot \ \mathbf{R}_N \rangle = \left\langle |\mathbf{R}_N|^2 \right\rangle \ , \tag{6.7}$$

which also describes the coil's size. The definition (6.7) agrees with the way we defined R before.

According to (6.6), R_N^2 is given by:

$$\mathbf{R}_N^2 = \mathbf{R}_{N-1}^2 + 2\mathbf{R}_{N-1}\mathbf{r}_N + \mathbf{r}_N^2 = \mathbf{R}_{N-1}^2 + 2|\mathbf{R}_{N-1}|\ell\cos\gamma_N + \ell^2 \ , \tag{6.8}$$

where γ_N is the angle between vectors $\mathbf{R}_{N-1}$ and $\mathbf{r}_N$, whereas ℓ, as you may remember, is the modulus of $\mathbf{r}_N$, i.e. $\ell = |\mathbf{r}_N|$. In the case of a freely-jointed polymer, the direction of $\mathbf{r}_N$ does not depend on the shape of the rest of the chain. This is why the angle γ_N is equally likely to have any value from 0° to 180°, which means that $\cos\gamma_N$ is equally likely to be positive (when γ_N lies between 0° and 90°) and negative (for γ_N between 90° to 180°). Therefore, the average value of the cosine is zero, $\langle\cos\gamma_N\rangle = 0$. This should help us to find the average value of $\mathbf{R}_N^2$ straightaway from (6.8). Indeed, since the average of the second term on the right-hand side is zero,

$$\left\langle \mathbf{R}_N^2 \right\rangle = \left\langle \mathbf{R}_{N-1}^2 \right\rangle + \ell^2 \ . \tag{6.9}$$

So, if an extra segment is added to the chain, $\left\langle \mathbf{R}^2 \right\rangle$ will increase by ℓ^2. Now, applying induction, one can easily prove that:

$$\left\langle \mathbf{R}_N^2 \right\rangle = N\ell^2 = L\ell \ . \tag{6.10}$$

Now, at last, we can use Equation (6.7) to find the size of a polymer molecule consisting of N units:

$$R_N = \ \left\langle \mathbf{R}_N^2 \right\rangle^{1/2} = N^{1/2}\ell = L^{1/2}\ell^{1/2} \ . \tag{6.11}$$

Thus, the "$L^{1/2}$ rule" is proved.

6.5 Persistence Length and Kuhn Segment

We have proved Equation (6.11) only for a particular model of polymer, with independent, freely-jointed segments. Is the formula valid for other models (including random walks)? We would need a special investigation to find out. The investigation, however, can be reduced to a very simple argument.

As we know, the flexibility of a polymer chain is not very noticeable at smaller scales, but it starts showing up as the scale increases. This means

that there has to be some critical length for each polymer, ℓ_{eff}. Any segment shorter than ℓ_{eff} can be regarded as rigid; that is, its end-to-end distance is roughly the same as its contour length. At the same time, different segments of length ℓ_{eff} behave as nearly independent. Such segments of length ℓ_{eff} are called effective segments, or Kuhn segments, after the Swiss physical chemist Werner Kuhn (1899–1963) who was the first to use the statistical mechanics to understand polymers and, in particular, he suggested the idea of effective segment. Obviously, a molecule of contour length L contains $N_{\text{eff}} = L/\ell_{\text{eff}}$ Kuhn segments. Since Kuhn segments are nearly independent, we can imagine that they are freely jointed, and use Equation (6.11):

$$R^2 = \left\langle \mathbf{R}^2 \right\rangle = N_{\text{eff}} \ell_{\text{eff}}^2 = \left(\frac{L}{\ell_{\text{eff}}} \right) \ell_{\text{eff}}^2 = L \ell_{\text{eff}} \ . \tag{6.12}$$

Equation (6.12) gives, in fact, the definition of effective Kuhn length; that is $\ell_{\text{eff}} = R^2/L$. Comparing (6.12) with (6.3), we immediately discover that the value ℓ which appears in (6.3) and (6.2) is exactly the Kuhn length. It gives the length scale on which the polymer chain (or the path of a random walker) remains roughly a straight line. (This is precisely where our estimate for a lost hiker in a forest, $\ell \leq 300$ m, comes from.)

There is quite a large range of Kuhn lengths for real polymers. Rather modest values of about 1 nm are typical for simple synthetic chains, whereas DNA's effective segment stretches 100 nm. (This is a huge number in molecular world, considering that an atom's size is of the order of 0.1 nm!)

Why is there such a difference? In each case, ℓ_{eff} is determined by the flexibility of the chain. It might be interesting to see how exactly the two quantities — flexibility and Kuhn length — are linked together. Let's set out on a journey along the chain (Figure 6.2), assuming that the direction of the first segment is fixed. At first, the change in direction would be very smooth, and hardly noticeable. It feels as if the chain keeps a sort of "memory" of the initial direction. Farther on, this memory starts fading, and eventually completely disappears. To describe this quantitatively, we choose two points on the chain, separated by a contour length s (Figure 6.2). Since the chain flexes, its directions at the two points are different; let's say the angle between them is $\theta(s)$. This angle varies due to fluctuations (i.e. due to thermal motion). You could probably guess that a meaningful value is the average $\cos\theta(s)$. It turns out that, if s is reasonably large, $\langle \cos\theta(s) \rangle$ decays exponentially with s,

$$\langle \cos\theta(s) \rangle = \exp\left(-\frac{s}{l} \right) \ . \tag{6.13}$$

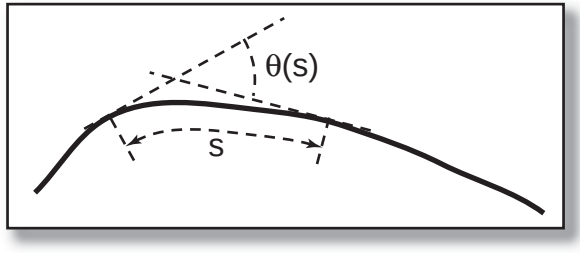


Fig. 6.2 Diagram explaining the concept of persistence length: angle θ between tangents of two points, contour distance s apart, depends on s such that at small s this angle is very small, it growth with increasing s and eventually becomes uniformly distributed between 0 and π (or 180°) — see Equation (6.13).

This formula is actually an exact definition of a very important quantity, l, which is known as the persistence length of a polymer chain.

What is the physical meaning of the relationship (6.13)? To answer this, let's first look at a segment which is shorter than l. When $s \ll l$, Equation (6.13) leads to $\langle \cos\theta(s) \rangle \approx 1$. Hence the angle $\theta(s)$ fluctuates around zero. This simply means that chain segments that are close compared to l have nearly the same direction. For the opposite case, $l \gg s$, Equation (6.13) results in $\langle \cos\theta(s) \rangle \approx 0$. Clearly, this indicates that $\theta(s)$ can be anything from 0° to 180° with equal probability. So the chain direction gets totally "forgotten" at lengths greater than l.

To summarize, persistence length is a parameter describing polymer chains quantitatively. Its physical meaning is the following: Memory of chain direction is retained on length scales shorter than l, but lost once l is exceeded. In other words, the name "persistence length" is very telling: the chain *persists* to have unchanged direction up to the length l.

Since the memory stretches in both directions, the Kuhn segment of length ℓ_{eff} must be roughly twice as big as the persistence length l. This is indeed true. Moreover, the relationship $\ell_{\text{eff}} = 2l$ is exact for a worm-like polymer chain (see Section 2.3); it is also valid for other models, although only approximately.

In principle, the persistence length should vary with temperature. The higher the temperature of a chain, the more it bends, and hence, the shorter its persistence length and Kuhn segment. However, in most cases this dependence is not important, since the range of temperatures where polymers may even exist is not that wide.

In the following, we will drop the index eff for brevity and denote effective Kuhn segment simply ℓ without index: $\ell = \ell_{\text{eff}}$.

6.6 The Density of a Polymer Coil and Concentration Ranges of a Polymer Solution

The phrase "polymer coil" may perhaps remind you of a ball of thread for knitting. In some sense, string or thread is indeed rather similar to a polymer chain. Despite this, a ball of string and a polymer coil have nothing in common whatsoever. A ball is wound tightly, with no gaps, whereas a polymer chain is arranged in a very loose manner, as Figure 2.6 showed. Yet structures as tight as a ball *are* known in the polymer world — they are called globules. We shall leave them until Chapter 9, and instead look for a theory to explain such low densities of polymer coils.

As we have seen, the size of a molecule made of N effective segments, of length ℓ each, is equal to $R = \ell N^{1/2}$ (see (6.12)). The volume of this coil can be estimated as[7] $V \sim (4\pi/3)R^3 \sim R^3 \sim \ell^3 N^{3/2}$.

Knowing the volume V and the number of segments N, we can now find the average concentration (number of segments per unit volume), $c^\star$ of the segments in the coil:

$$c^\star = \frac{N}{V} \sim \frac{N}{\ell^3 N^{3/2}} = \ell^{-3} N^{-1/2} \ . \tag{6.14}$$

Our main achievement so far is that we have found the actual dependence of $c^\star$ on N. In particular, estimate (6.14) shows that, if a chain is sufficiently long ($N \gg 1$), the segment concentration $c^\star$ becomes extremely low.

You might remember that we have already come across the concentration $c^\star$ in Section 4.6, when we talked about polymer solutions. There we said that if the overall concentration c of a polymer in the solution is less than $c^\star$ (i.e. $c < c^\star$) then individual coils hardly ever overlap, and the solution looks like a low pressure gas of coils (Figure 4.7 *a*). If, on the other hand, $c > c^\star$ then the coils penetrate deeply into each other, and the chains are entangled (Figure 4.7 *c*). The value $c^\star$ corresponds to the threshold regime (Figure 4.7 *b*). Now we have seen that the value $c^\star$ is extremely small for long polymer chains. It means that a polymer solution can be made up of separate, non-overlapping coils only at very low concentrations.

This conclusion becomes even clearer if we replace c by another quantity, the volume fraction ϕ of a polymer in the solution. Let v be the volume of a single segment, then, with c such segments per unit volume, the fraction of the whole volume occupied by the segments is $\phi = cv$. The advantage of ϕ (compared to c) is that it has no dimensions. Formulas become even

[7] The sign $\sim$ means "of the same order of magnitude". When we make rough, order of magnitude, estimates, such factors as $4\pi/3$ are not important, so we leave them out.

easier, e.g. in the case of a polymer melt (when there is no solvent at all) $\phi = 1$.

Let d be a characteristic thickness of a polymer chain. Then, thinking of an effective segment as a cylinder of diameter d and height ℓ, we can estimate its volume: $v \sim \ell d^2$ (we leave out as usual an unimportant factor of $\pi/4$). The threshold volume fraction $\phi^\star$ that marks the cross-over between Figure 4.7 *a* and Figure 4.7 *c* will be approximately the following:

$$\phi^\star \sim c^\star v \sim \left(\frac{d}{\ell}\right)^2 N^{-1/2} \ . \tag{6.15}$$

In practice, the ratio ℓ/d ranges from numbers about 2 or 3 for flexible synthetic polymers to 50 for a DNA double helix. Using (6.15), we can look at particular numbers. For example, if there are 10^4 units in a flexible chain, then the coils already start overlapping when the volume fraction of the polymer is 10^{-2}, i.e. when the polymer takes as little as 1% of the whole volume of the solution.

Therefore, if $N \gg 1$, there has to be a rather broad range of concentrations $\phi^\star < \phi < 1$ ($c^\star < c < 1/v$) in which the coils are heavily entangled ($\phi^\star < \phi$), yet there is still little polymer in the solution ($\phi < 1$). This type of polymer solution is known as semi-dilute, and is shown in Figure 4.7 *c*. Really concentrated solution corresponds to $\phi \sim 1$, when the volume fractions of the polymer and the solvent are comparable (Figure 4.7 *d*).

By the way, the intermediate, semi-dilute region is only possible because polymer chains are so extremely long ($N \gg 1$). Indeed, if $N \sim 1$ (i.e. if there were small molecules in place of a polymer in the solvent), the two inequalities $\phi^\star \simeq N^{-1/2} < \phi < 1$ would not work together. Therefore, the solution has to be either dilute ($\phi < 1$), in which case the individual molecules hardly interact with each other, or concentrated ($\phi \sim 1$), with strongly interacting molecules.

Figures 4.7 *a* – *d* depict isotropic polymer solutions in which there is no preferential orientation of polymer chains. However, as we have already mentioned, spontaneous ordering can occur in concentrated solutions of rigid ($\ell/d \gg 1$) polymers, and they become liquid crystalline (Figure 4.7 *e*). In most cases, such an anisotropic state is established at concentrations $c > c_{\rm cr}$, where $c_{\rm cr} \sim \ell^{-2} d^{-1} > c^\star$, so $c_{\rm cr}$ corresponds to a semi-dilute solution.

6.7 The Gaussian Distribution

There is another bunch of problems that we still need to tackle when portraying an isolated polymer coil. We might as well start with a very simple question: What does it mean to say that a coil size is proportional to the square root of the chain length, i.e. $L^{1/2}$ or $N^{1/2}$? Can the chain, accidentally, stretch out into a straight line? In the same spirit you may ask, can one day all the passengers of an airplane turn up very fat, and the plane fail to take off? Can the dipoles of an object that scatters light, by coincidence, all line up in the same direction? Can all molecules in the room gather in one corner?

You would probably agree that, in principle, all these things could happen, although they are extremely unlikely. In fact, the probability that the chain would stretch out into a straight line is just the same as the probability of any other particular conformation (assuming — plausibly — that all conformations have the same energy). But the whole point is that there are many many curled up and entangled shapes, whereas there is only one straight line, and no more. That is why a polymer, left on its own, is most likely to coil up into one of a myriad of conformations of a size $R \sim \ell N^{1/2}$. Hence, this is just what the average size of a polymer is, within an order of magnitude. The chances that due to fluctuations a polymer may expand up to $R \sim \ell N$ are exceedingly slim.

To bring some maths into play, we need to count up all the stretched and all the coiled conformations of a polymer chain. More precisely, we would like to know how many different conformations of the chain have the same end-to-end vector, $\mathbf{R}$. How much is this as a fraction of all the possible conformations? (In other words, what is the probability that a polymer chain picked at random has an end-to-end vector $\mathbf{R}$?) It is not a usual kind of task for elementary maths. The question really is in how many different ways can you choose the terms of a sequence, so that their sum stays the same.

We cannot afford to derive it here, so we shall merely tell you the answer:

$$P_N(\mathbf{R}) = Q \exp\left[-\frac{3\mathbf{R}^2}{2N\ell^2}\right] . \tag{6.16}$$

Here ℓ is the effective, or Kuhn, length, N is the number of Kuhn segments in the chain, and Q is a constant factor, which does not depend on R. The value of Q depends on what exactly we mean by $P_N(\mathbf{R})$. It can be two things, either the total number of conformations with the given $\mathbf{R}$ (in some

infinitesimal volume element, of course), or the probability of a conformation with this $\mathbf{R}$ (i.e. the ratio of the number of such conformations to the total number of all the possible conformations). In particular, if P_N is the probability, then

$$Q = \left[\frac{3}{2\pi N\ell^2}\right]^{3/2} . \tag{6.17}$$

Equation (6.16) is written for a vector $\vec{R}$. We may also want to look at its components R_x, R_y, and R_z (which show how far apart the two ends of the chain are shifted from one another along the x, y, and z axes). For any of the three components:

$$P_N(R_\alpha) = \left[\frac{3}{2\pi N\ell^2}\right]^{1/2} \exp\left[-\frac{3R_\alpha^2}{2N\ell^2}\right] . \tag{6.18}$$

Formula (6.16) can be obtained by multiplying together expressions of type (6.18) for all the three components $\alpha = x$, $\alpha = y$, $\alpha = z$; we leave it for you to check and explain.) The function $P_N(R_x)$ is plotted in Figure 6.3. You can see that all the values of R_x from zero up to about $\ell N^{1/2}$ have roughly the same probability. However, if R becomes larger, the probability decays very dramatically. We could phrase it like this: If the first unit of the chain is fixed at the origin, then there is a roughly equal chance that the last unit will be at any point inside a sphere of radius $\ell N^{1/2}$. On the other hand, the likelihood of finding the last unit outside such sphere is negligibly small.

Equations (6.16) and (6.18) for the probability of different values of $\mathbf{R}$ or its components, are known as Gaussian distributions, because famous mathematician K.F. Gauss (1777–1855) was the first to come up with this kind of formula, albeit in a different context. The diversity of situations in which one encounters Gaussian distribution is amazing. Apart from polymers and airplane passengers, consider, for example, some measurements. As you probably know, in order to obtain more accurate data and to reduce the impact of inevitable random inaccuracies in measurements, experimentalists must repeat the same measurement again and again. Let's say they do it N times. Then they need to find the average. To do this, they first have to add all the measurements together. Thus, the errors get added up too. However, some of the errors are positive and some are negative, at random. (It seems we can never get away from sums of random values!) This is why the error in the sum is proportional to $N^{1/2}$, rather than to N. Finally, to take the average, the sum is divided by N. That is why

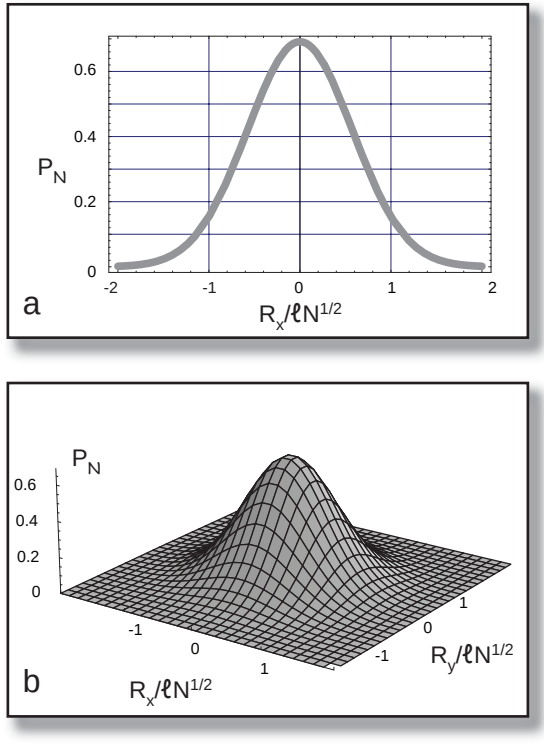


Fig. 6.3 Gaussian distribution. (*a*): The dependence of P_N on $R_\alpha/(\ell N^{1/2})$ ($\alpha = x$ to be specific) given by Equation (6.18). (*b*): Two-dimensional distribution P_N as it depends on re-scaled coordinates x and y, on $R_x/(\ell N^{1/2})$ and $R_y/(\ell N^{1/2})$. Three-dimensional distribution P_N depends on x, y, and z; it looks quite similar — but we cannot draw it.

the error in the average will behave as $N^{-1/2}$, hence it will indeed decrease with the number of measurements.

In general, we can say that the sum of a great number of random values is controlled by a Gaussian distribution of probabilities, like (6.16). This is one of the key ideas in probability theory. Due to its great importance, it was given a posh name, the central limit theorem (CLT).

Why "limit"? In fact, the Gaussian distribution (6.16) is valid exactly only in the limit of a very large number of terms, $N \gg 1$. For a finite number of terms, it is only approximate. However, even in the case of a moderate number of terms (or unit segments, N), a Gaussian distribution provides acceptable accuracy[8].

[8]If you have a liking for elegant mathematical trifles, you may be interested in the following example. It is about so-called "lucky" bus tickets. Russian bus tickets used to have six-digit numbers on them, and students liked to believe — seriously or not — that a ticket is "lucky" if the sum of the first three digits is the same as the sum of the last three digits. It is possible to prove that there are 55, 252 "lucky" tickets out of the total

Quite naturally, a free single polymer coil is often called a Gaussian coil, after the distribution (6.16).

quantity of $1,000,000$ six-digit numbers. So the probability of a lucky ticket is 5.5252%. On the other hand, the sum of three digits of a number is actually a sum of random terms. Although there are only three terms in this case, you could still try to use the CLT to estimate the probability of a lucky ticket. You would then get an approximate answer of 5%, which is surprisingly close to the accurate value. Thus, even if N is as small as 3, the CLT works reasonably.

Chapter 7

The Physics of High Elasticity

> As a matter of fact, he was as not normal as it is possible to be.
>
> J.K. Rowling,
> *Harry Potter*

7.1 Columbus Discovered . . . Natural Rubber

In some popular books, they say that natural rubber was the first polymer encountered by our ancestors. If you think about it for a minute, your reaction might be: "What rubbish! People had always known things like wood and timber. Not to mention that both prehistoric and contemporary human beings themselves are made of polymers." However, the fact of the matter is that in many materials the polymeric nature, although important, only shows up in rather subtle ways. For instance, there are low molecular weight compounds that look similar to polymers in the semi-crystalline, viscous, or glassy state. At the same time, there is a property which is both very noticeable and purely polymeric, that is, impossible for small molecules: high elasticity. And the first highly elastic substance that the Europeans came across was indeed natural rubber.

When they reached America, the first European explorers and immigrants were, of course, overwhelmed by novelties. They found potatoes, tobacco, sweet corn and tomatoes — to name a few. They met strange local people there who had an unusual way of life. It is not surprising that books on the discovery of America are so enthralling. Unfortunately, some of the newcomers were greedy and aggressive and these features sometimes dominated over the natural curiosity of discoverers. This did not only made

the Native Americans to suffer, but also caused irreversible damage to their culture.

The culture of making rubber was a lucky exception. People from the very first expedition of Columbus were amazed by the balls that the Native Americans played with. At that time in Europe, balls were typically made from bull's bladders encased in a leather shell (these are also polymers, by the way!) In contrast, the American balls were solid, heavy, and surprisingly bouncy. They were made from a substance which the Native Americans called "caoutchouc". Only ten years later, when the Europeans reached as far as Brazil, they found out that "caoutchouc" is a milky resin excreted by the tree called "heve", extracted by sloping cuts on the trunk. At that time, the invaders, in their bellicose excitement, forgot about the rubber tree for a while. More than two hundred years passed before "heve" was properly scientifically described by botanists (in 1738). Now it is fairly well-known as *Hevea brasiliensis*, and "caoutchouc" is still the word for natural rubber in some European languages.

7.2 High Elasticity

Rubber has very unusual properties. Under certain conditions, it remains solid (non-fluid), yet is extremely elastic. A fairly low stress is able to deform a piece of rubber quite significantly (much more than if it were an ordinary solid). The deformation is reversible (elastic), i.e. when the stress is released, the sample regains its original, un-deformed shape.

To appreciate the elasticity of rubber, let's see how it differs from ordinary materials, including the non-polymeric ones. Take steel or plastic, for example. Figure 7.1 compares the dependence of the stress σ on the strain $\Delta\ell/\ell$ for a steel rod (Figure 7.1 *a*), for a plastic string (Figure 7.1 *b*) and for a piece of rubber (Figure 7.1 *c*). The three graphs have some things in common. They all start off as a linear relationship between σ and $\Delta\ell/\ell$, i.e. they follow Hooke's law (4.1) for sufficiently small deformations. In this range, deformations are almost completely reversible. The linearity extends roughly to about point A in all three pictures. Then, between A and B, the lines bend. Here the deformations become significantly non-linear, but still remain reversible. Interesting behavior starts above B. This is where the reversibility is finally lost: the sample, as they say, starts to flow. Even if the stress is released, the material retains a certain residual (plastic)

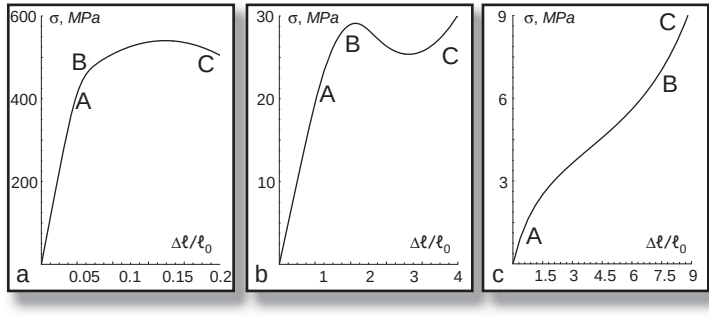


Fig. 7.1 A sketch of (engineering) stress-strain diagrams for steel (*a*), solid (partially crystalline) polymeric material (*b*) and highly elastic (rubbery) polymeric material (*c*). See caption of Figure 4.4 for the explanation of engineering stress. The basic points are indicated in each graph: up to the point *A*, the elastic response (stress) remains linear in deformation, i.e., material obeys Hook's law; simultaneously, deformation remains fully reversible. Between points *A* and *B* deformation is still reversible, the sample restores its size and shape upon switching off the load, but stress-strain relation is no longer linear. After point *B* the deformation becomes plastic, that is, irreversible. At point *C* the material fails (breaks). Although all of these features are common to all shown materials, they are exhibited at vastly different values of stresses and strains.

deformation, and never quite goes back to its original shape. Eventually, at point C, the sample gets torn apart.

Although all three graphs have points A, B, and C, they are positioned differently. There is also a huge difference in numbers. For instance, to break steel (i.e. to reach point C) you only need to stretch it by a few percent. And reversible deformations are never higher than 1%. In contrast, a rubber strip can happily extend up to eight times its original length (by 700%). Moreover, this is still reversible! The actual values of stress at breaking point are also beyond comparison. They are up to about $2,000$ MPa for steel, and 30 MPa for rubber. So we are talking about totally different scales both for stress σ and for strain $\Delta\ell/\ell$. This is why there is a disparity in the Young's modulus E as well. Indeed, according to Equation (4.1), Young modulus is determined by the slope of the linear part of a curve such as those in Figure 7.1. Hence, we get $E \approx 2 \cdot 10^5$ MPa for steel, and $E \leq 1$ MPa for rubber. This gap is enormous, more than five orders of magnitude. (We have already mentioned this in Chapter 4.) One more difference is that rubber has a wide range where deformations are non-linear yet reversible (between A and B), whereas for steel this area is almost missing. On the other hand, the curve $\sigma(\Delta\ell/\ell)$ for steel has a comparatively

broad region of plastic deformations (between B and C), whereas rubber almost immediately snaps as soon as it starts to flow (Figure 7.1 *c*).

Now you know what people mean when they talk about the high elasticity of rubber. In brief, high elasticity means that the material is prone to very high, non-linear yet reversible deformations as a result of rather moderate stress.

7.3 The Discovery of Vulcanization

Besides balls, the Native Americans used to make lots of other handy things from rubber, such as water-proof shawls, a kind of Wellington boots, flasks, etc. They had not quite reached perfection though. All the stuff was rather sticky and short-lasting. Even worse, on a hot day it would melt altogether! The Europeans took over, and kept on trying to find a better use for such an unusual substance. But it was not that easy.

The problem is that natural rubber is not really a solid. Any external force (such as gravity) makes it flow, albeit slowly. Strictly speaking, it is a liquid polymer melt in the viscous state. Therefore, any natural rubber product keeps changing shape. Exactly how much it "wobbles" or "oozes", depends strongly on the temperature. High above room temperature, natural rubber is more of a liquid. At low temperatures, the oozing nearly stops. But the high elasticity disappears as well, and the rubber hardens.

An amusing true-life story about the rise and fall of a man named Charles Mackintosh, from Glasgow, Scotland, illustrates the problem. The clever fellow decided to use rubber in the production of raincoats. A thin layer of rubber was placed between two layers of fabric. It worked out very nicely, and the raincoats (called macintoshes) became very popular in the notoriously wet Britain. Mackintosh rapidly became rich, shortly after he started his business in the winter of 1820. However, when summer came along, the temperature rose, and all the rubber flowed out of the macintoshes. The poor inventor went bankrupt, and the whole idea of padding coats with rubber was abandoned for many years.

Not for too long though. A breakthrough occurred in 1839 when the American C. Goodyear suggested the process for vulcanizing rubber. At the molecular level, rubber consists of polymer chains with frequent double bonds (Figure 7.2). The vulcanization involves adding sulphur atoms to the rubber. They form covalent bonds between the chains (Figure 7.2 *b*), so the chains become linked together by sulphur bridges. You get a

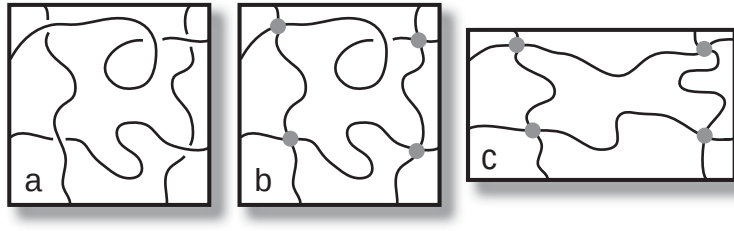


Fig. 7.2 A cartoon illustrating vulcanization and elasticity of vulcanized network. Panel (*a*) depicts a system of polymer chains prior to vulcanization. Panel (*b*) shows the same polymers after vulcanization: chains are cross-linked and form now a network (cross-linkers are shown as grey balls). Panel (*c*) gives an idea of network behavior upon deformation. In this figure, the sample is stretched in horizontal direction and shrinks somewhat in vertical direction to maintain practically unchanged volume (unchanged area in the figure). The deformation is achieved by chains uncoiling to a necessary extent, by releasing some of their loops and wiggles, along the direction of stretching; chains going in perpendicular direction become somewhat more crumpled. The main point is that lengths of chains practically do not change, only their shapes rearrange.

polymer network. It is not fluid, even at relatively high temperatures, when a normal polymer melt of unattached chains would start flowing (due to intense thermal motion, making the chains move with respect to each other). At the same time, there is nothing to stop such a network from expanding. Under strain all the chains would stretch as in Figure 7.2 *c*, so it is still highly elastic. Of course, Goodyear had no idea that rubber was a polymer. (This was discovered almost a hundred years later.) He did not even dream of explaining the vulcanization in the way we have just done. But his invention started the era of commercial use of new, vulcanized rubber.

The story of the discovery is very interesting in its own right. Charles Goodyear (1800–1860) was not a scientist in the modern sense of the word. His education was not very deep, and his aspirations were mainly directed toward business. Once he happened to buy a lifebuoy of india (natural) rubber. The unusual material captured the inventor's imagination. He became literally obsessed with the idea of making rubber strong and pliable. There was hardly anything Goodyear did not try! He mixed rubber with turpentine, soot, and oil. He burnt it in the oven, as the Native Americans were said to have made some progress by keeping rubber in the bright sun. There were times when Goodyear thought he had succeeded. Then he would persuade investors to support the enterprize, and immediately set up production on a really American scale. Alas, every time the rubber would

start to run. The products would ooze away, sometimes giving out such a horrible smell that they had to be buried in the ground! The debts were left unpaid, and Goodyear's numerous children had to live in poverty, and for a while he was even imprisoned for debt. But nothing could stop him.

To be fair, we need to say that Goodyear was not the only person to work on improving india rubber. There was even a kind of "rubber fever" in the 1820s and early 1830s in North America. However, it was none other than Goodyear who eventually made the breakthrough.

It was entirely accidental. One day he was mixing rubber with sulphur and various other ingredients when he dropped some on top of a hot stove. The next morning the stove had cooled, and one side of the rubber lump, which was next to the sulphur, had become unrecognizable. It looked like the normal rubber we now use every day. At this stage, Goodyear did something most important, which makes him really deserve the fame, not just credit for being a lucky guy who got the answer by chance. He noticed what had happened, realized its significance, and drew the right conclusion. The recipe for success had to do both, mixing rubber with sulphur and then heating it.

This is what Goodyear wrote about his discovery: "I was encouraged in my efforts by the reflection that what is hidden and unknown and cannot be discovered by scientific research, will most likely be discovered by accident, if at all, by the man who applies himself most perseveringly to the subject, and is most observing of everything related thereto[1]."

Goodyear died almost as poor as he had been in his youth. Nevertheless, his invention became widely popular even during his lifetime. The method of vulcanization that he designed has survived till now with hardly any changes. Furthermore, many of Goodyear's ideas on how to obtain different sorts of rubber with particular features are now successfully exploited. For instance, incorporating an inert filler (such as carbon black), results in a very hard and robust rubber that is especially good for tires. (The way it works is that little particles of soot fill in the mesh of the network. This makes it harder to squash.) To obtain the opposite effect, a plasticizer (e.g. some oil that would help the particles of filler move along the network) is added. This gives rubber that is easily worn away, like that used to make erasers and the like.

Thus, since the second half of the 19th century, the rubber industry has developed very rapidly. The latex of *Hevea brasiliensis*, growing in the

[1]Quoted from the book [53], p. 124.

wild, had long remained the only raw material for the industry. However, in 1870 the English smuggled about 100,000 *Hevea* seeds from Brazil.

Then young trees were cultivated from the seeds in British botanical gardens. They gave birth to vast plantations of rubber trees in the colonies, mainly in Malaysia, Indonesia and Ceylon (Sri Lanka). By the First World War, only a negligible part of the world production of rubber was from Brazil.

7.4 Synthetic Rubber

In countries that had no access to the tropical plantations of rubber trees, especially in Russia and Germany, scientists tried to work out how to make synthetic rubber from available raw materials. The studies were a success. They found a way of making synthetic rubber from butadiene and in the 1930s production started. It worked out well, synthetic rubber was satisfactory, and looked very similar to natural rubber.

All other industrial countries remained content with natural rubber. Its qualities were still better overall. However, all was to change during the Second World War, when almost all the rubber plantations in South-Eastern Asia were occupied by the Japanese. This encouraged the search for new methods of synthesizing rubber, especially in the United States and Canada. Soon, the world production of artificial rubber caught up with and, by the 1960s, had surpassed the production of natural rubber.

In contrast to vulcanization, this time it was all worked out scientifically. The studies on rubber synthesis kept up with the new idea that polymers were made of long molecular chains. (You may remember that H. Staudinger pioneered this theory, and gave credence to it by many experiments in the 1920s and 1930s.) The endeavor with synthetic rubber not only brought in new products, it also had scientific value. Those studies confirmed that high elasticity was not a unique feature of natural rubber, but should be typical of any polymer network or gel (as long as there is no glass transition or crystallization under certain circumstances, otherwise the motion of the chains would be constrained).

7.5 High Elasticity and Stretching of an Individual Polymer Chain

We have said that high elasticity is a common property of polymer networks. However, this sounds a bit too general. Let's zoom in, and examine what

it implies for particular molecules. Figure 7.2 *b* portrays a typical polymer network. You can see a set of long molecular chains, bridged together with cross-linkers (covalent chemical bonds). It looks like a kind of framework in three-dimension. What would you regard as an elementary "brick" of such a structure; what is the smallest piece we have to consider if we want to understand network deformation? The answer is clear from Figures 7.2 *b* and *c*: It is a strand of the chain between two neighboring cross-links (bridges). We shall introduce a new word, *subchain*, for such strands. Because polymer chains are so flexible and randomly coiled, it is usually sufficient to uncoil slightly some of the chains to achieve a rather significant deformation of the network. Thus, if a polymer network is stretched, many subchains will be somewhat uncoiled, while some of them will be actually somewhat compressed (Figure 7.2 *c*). Therefore, the elasticity of the whole polymer is a sum of the elasticities of all the individual subchains. This is why it makes sense to explore elastic properties of a single subchain first, before looking at the whole network.

Thus, let's see what happens if a single chain is pulled upon by an external force $\mathbf{f}$, as shown in (Figure 7.3). Actually, such a single chain stretching experiment can be done — which is a marvelous experimental achievement. It was first done by C. Bustamante and his co-workers in the University of Oregon in 1992, and then repeated in many laboratories around the globe. There are several versions of the apparatus by which experimenters can manipulate single molecules, such as optical tweezers, magnetic tweezers, and some others; we cannot talk about it here, but encourage the reader to read on this subject, for instance, in the article [52]. It is actually somewhat ironic that the most convenient polymer for performing such single chain stretching experiment is nothing lesser than double helical DNA. Needless saying, experiments with DNA are exciting because they shed light on the properties of this most important of all molecules — and we will return to this point (see Section 7.12). But for now let's just use the example of DNA to sort out the basics of polymer physics.We imagine that one chain end is attached to an immobile support at the origin (we always can choose origin as we like), then the position of the other end, where the force is applied, is described by the familiar end-to-end vector $\mathbf{R}$ (see Figure 7.3). We want to find the average value of $\mathbf{R}$ as it depends on the applied force $\mathbf{f}$.

It turns out useful to think first about the opposite problem. Suppose we want to maintain the chain's dangling end at the given position $\mathbf{R}$; what kind of force $\mathbf{f}$, on average, should we apply in order to retain the desired

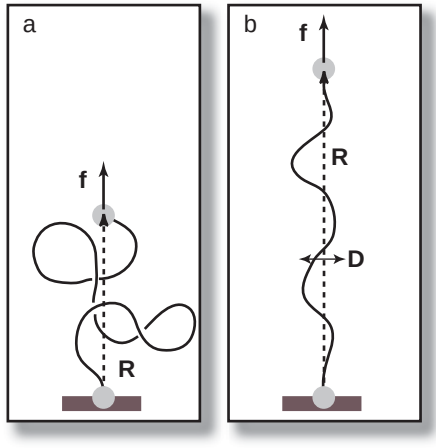


Fig. 7.3 In order to keep a given end-to-end vector $\mathbf{R}$ for a single polymer chain, we need to apply an external force $\mathbf{f}$. Similar figure in the first edition of this book was considered a rather abstract theoretical concept; right now this very experiment is performed rather routinely in many laboratories using DNA as a polymer to test, and optical or magnetic tweezers to apply the force. Our cartoon illustration is unrealistic only in one aspect, namely, the size of the bead (to which the force is applied) in these experiments is an order of magnitude larger than the DNA coil size. Panel (*a*) shows the situation with weak force, when chain makes numerous loops and remains randomly coiled. Panel (*b*) depicts strong force regime, when chain makes only limited excursions in perpendicular direction, characterized by the length scale D; this situation is more fully considered below in Section 7.12.

value of $\mathbf{R}$? You may be a little surprised by this last question. Is a force really needed to keep $\mathbf{R}$ unchanged? As we learned in Section 6.7, even with no force at all the end-to-end vector may have any possible value, including the value $\mathbf{R}$ we wish. There is no doubt about that. However, without the force, if $\mathbf{f} = 0$, the dangling end will not stay at the desired point $\mathbf{R}$ for any length of time. It will go on fluctuating. All directions of $\mathbf{R}$ will be equally likely. This is why on average the end-to-end vector will be equal to zero, just as you expect from the symmetry of the distribution $P_N(\mathbf{R})$ (6.16). Hence, in order to keep $\mathbf{R}$ fixed, you need to use a force $\mathbf{f}$. You can guess, based on a simple symmetry argument, that $\mathbf{f}$ should point in the same direction as $\mathbf{R}$ (Figure 7.3); indeed, what other direction might it point to? It is an external force, i.e. a force from some external object, acting on the chain. And what about the force that the chain in its turn exerts on the external object? Newton's third law tells us that it should point in the opposite direction, i.e. towards the origin of coordinates. So it is a restoring, elastic force. Thus, we have naturally come to the conclusion that a polymer chain resists being stretched. In other words, it exhibits elasticity.

We encourage the reader to re-think the above paragraph to realize that all our arguments are perfectly applicable even to the simplest model of a chain — a freely-jointed polymer chain made of a large number $N \gg 1$ of

elementary segments, each of the same length ℓ and, moreover, with all volume interactions between segments completely neglected (see Section 6.3). (The latter approximation is known as an "ideal polymer chain"; we will come back to it in Section 8.1, but for now we can think of it as just sticks of practically zero thickness).

Our arguments may not satisfy you completely. Indeed, what is the physics of polymer chain elasticity if it exists even for inextensible segments with no interactions? To answer this, let's think of an ordinary solid crystal. What makes it elastic? When a crystal is stretched, the atoms are pulled further apart (Figure 7.4). Thus, the elastic force in this case would be a result of many interatomic interactions. Sometimes it helps to describe the same thing in terms of energy. The undeformed crystal is in equilibrium; that is, the potential energy of interatomic interactions is a minimum (Figure 7.4). An external deforming force pulls the atoms up the slope from the bottom of the potential well. Suppose an external force f causes an elongation of the crystal, Δx. The work it does, $f\Delta x$, is used to increase the internal potential energy ΔU of the crystal, which is the total energy of interatomic interactions: $f\Delta x = \Delta U$, or:

$$f = \frac{\Delta U}{\Delta x} \ . \tag{7.1}$$

Equation (7.1) gives you a recipe for finding the stretching force f (and the elastic force of the crystal, which is opposite to it). All you need to know

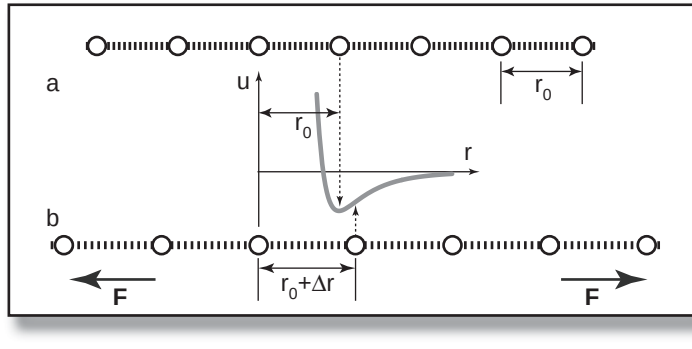


Fig. 7.4 An illustration of the elasticity of a crystal. (*a*): Initially, in an non-deformed crystal, the distance, r_0, between any two neighboring atoms corresponds to the minimum of the interactional potential energy $U(r)$. (*b*): To stretch the sample, we have to increase distances between atoms, to make each of them $r_0 + \Delta r$, and the displacement Δr determines the shift away from the minimum of potential energy: we have to increase the potential energy, which is why the crystal develops the force of elastic response.

is how the crystal is constructed. Then you should be able to work out its internal energy ΔU.

Now let's go back to a polymer chain. This time, we are not talking about a stretched array of atoms, as in the case of a crystal. What we really see when a polymer chain is stretched is an increase in the end-to-end distance (Figure 7.3). The only way this can happen is, of course, if some wiggled bits of the chain straighten up and disentangle. This is particularly striking and easy to follow if we think of a freely-jointed (Figure 2.5 *b*) ideal chain. (Once again, ideal chain approximation assumes that the only interaction between the neighboring monomers comes from their being joined together into a chain. All other monomers do not interact at all, just like the molecules in an ideal gas. You will be getting used to this approximation.)

Suppose we apply a force $\mathbf{f}$, and change the end-to-end vector $\mathbf{R}$ by some value $\Delta\mathbf{R}$. Hence, we did an amount of work $\mathbf{f}\Delta\mathbf{R}$. Where did the energy go? Previously, when we looked at crystals, we managed to find the answer quite easily. Unfortunately, the way we did it would not work in this case. In an ideal system, interactional potential energy is zero, both before and after the deformation. Thus, Equation (7.1) is of no use. As for kinetic energy of molecules, it is determined by the temperature. If the temperature is constant during the deformation, the kinetic energy will not change either... Are we lost?

Before giving up, let's think of another analogy. Strange as it may seem, help comes from an ideal gas. In a sense, an ideal gas has "elasticity the other way round". Suppose you wanted to hold the gas under a piston (Figure 7.5) in a vessel of a certain volume (it is just like holding a polymer chain to retain a non-zero $\mathbf{R}$). You would have to apply a squeezing force $f = pA$, where p is the gas pressure, and A is the area of the piston. In order to reduce the volume, you would always need to do some work.

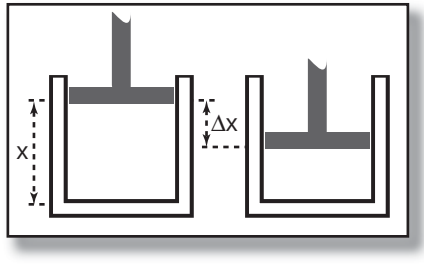


Fig. 7.5 Ideal gas in a vessel with a piston before (left) and after (right) the compression by the amount Δx.

Nevertheless, both the potential and kinetic energy of the gas molecules stay the same, given that the compression is isothermal. So where on earth does the work go? (As you see, we end up with the same question once again.)

Of course, in the long run, all the work transforms into heat which is dissipated in the surroundings. How do we know? Well, if there were no surrounding medium to take the heat away, both a gas when compressed and a polymer when stretched would get warmer (see also below Section 7.11). Does that mean elasticity of a polymer chain depends on the environment which absorbs the heat? Well, we know that the pressure of an ideal gas does *not* depend on the type of the environment, so maybe there is something similar for a polymer?

Also, is there any hope we can still manage with simple energy arguments? Or do we need to tackle the problem by means of mechanics, tracing all the molecules? It would not be hard for an ideal gas. Its pressure is just an overall result of all the individual hits by the molecules on the piston. This concept immediately leads to the ideal gas equation of state (we encourage the reader to reproduce this derivation, it is very beautiful). However, there is no such simple picture for a polymer. Even in the case of an ideal chain, the motion of segments is extremely complicated, due to knots and entanglements.

But let's think. We know that the surrounding medium plays no other role but to maintain the constant temperature T. Our proof of polymer elasticity was based on a very general idea. Indeed, we showed that when a polymer chain is stretched, it is pulled from a more probable to a less probable state. Hang on a minute! Is there perhaps some universal way of finding the energy cost of lowering the probabilities at a constant temperature, without getting bogged down in the mechanics of molecular collisions?

There is indeed a very general rule, known as the Boltzmann principle. It states the following. Suppose there are Ω ways in which molecules can occupy a certain state. (In our case, this number is proportional to the probability $P(\mathbf{R})$ — see Equation (6.16)). Then we need to find the quantity

$$S = k_B \ln \Omega \ , \tag{7.2}$$

where k_B is Boltzmann's constant. The energy equivalent of probability we are seeking is the change in the value:

$$U_{\text{eff}} = -TS \ , \tag{7.3}$$

where T is the absolute temperature. In the case of a polymer chain, according to (6.16),

$$S(\mathbf{R}) = -k_B \frac{3\mathbf{R}^2}{2N\ell^2} + \text{const} \ , \tag{7.4}$$

or

$$U_{\text{eff}}(\mathbf{R}) = k_B T \frac{3\mathbf{R}^2}{2N\ell^2} + \text{const} \ , \tag{7.5}$$

where const is a quantity independent of $\mathbf{R}$ (which arises because Ω is *proportional* to $P(\mathbf{R})$, not equal to it). Using Equation (7.5), we can easily find the elastic force from (7.1). We shall do so a little later, in Section 7.9. Thus, the Boltzmann equation $S = k_B \ln \Omega$ has rescued us when we had nearly lost hope. It is not surprising that this formula was engraved on the tombstone of the author, Ludwig Boltzmann (1844–1906)! But what does it mean, and where does it come from? It is an interesting question in its own right. We shall devote the next two sections to it, and then come back to the discussion of high elasticity.

7.6 Entropy

Rapidly developing science gives rise to a new vocabulary. This gives us an excuse to reflect on how human languages evolve. It is fascinating to be able to trace this process, spanning from the dawn of mankind to the modern day. For example, officers of the Russian army, after entering Paris in 1814, used to spur on French waiters in Russian. The Russian word *bistro* (meaning "quickly") soon became absorbed into French and then to other western languages, everybody now understands that "Bistro" is a modestly set small restaurant offering inexpensive simple meals. Much more recently, we witnessed how the English language acquired such strangers as "pogrom", "sputnik", "perestroika" from the Russian, French language picked up "airbags" from English, whereas the Russians borrowed words like "computer" and even some English abbreviations became words in Russian such as PR (public relations). Novel words usually enrich the language, as they represent new things and ideas. For instance, the word "computer" is literally absorbed into Russian to distinguish the modern universal device from a "machine for calculations"; likewise, the word *sputnik* in English does not mean "satellite" in general (which would have been the correct translation), but rather refers to the first Russian satellite, and so evokes memories and mood of that period of time.

Speaking more specifically of the scientific words, many people are afraid of them, and indeed some of the newly invented ones may sound pretty horrible (like "uniformitarianism" or "compartmentalization"). Such words have a very narrow use, and clutter up the language. We honestly think that their authors must have lacked a sense of moderation! Here is a telling statement by Samuel Goudsmit, an editor of one of the leading scientific journals, *Physical Review*: "We find that [neologisms] are often ungrammatical, frequently ugly, sometimes chauvinistic, likely to be obscure, and usually unnecessary". Nevertheless, there have been some really valuable scientific contributions to the world's vocabularies, and the word "entropy" is among them; moreover, it certainly deserves a place near the top of the list.

Together with energy, time, and so on, entropy is one of the most crucial concepts of physics, and of science in general. Unfortunately, ever since the idea of entropy appeared, it has always been surrounded by a halo of mystery. For instance, the following definition is attributed to the well-known physical chemist Wilhelm Ostwald (1853–1932; he was born in Riga, educated in Tartu, worked most of his life in Leipzig, and awarded Nobel prize in 1909): "Energy is the queen of the world, and entropy is her shadow!" Such an attitude is not without reason. How do people hear about entropy the first time? Quite often it gets mentioned in the context of the most global and tantalizing problems, such as the origin of life, or the future of the Universe. Perhaps, this explains why there is usually no room for entropy in the school curriculum. However, it is quite a straightforward thing. To get to know it in the first instance, you do not need to dive into obscure philosophical matters. Moreover, it is hard to manage without entropy, if you are aiming to describe atomic properties of matter. It would be a bit like trying to explain the rules of football without mentioning the ball! This is especially true for polymers. Now you understand why we need to digress from the main theme, and talk about entropy in more detail.

Let's think of energy, for a start. How would you define it? Of course, you can split it up into various forms, e.g. potential energy, kinetic energy, etc, and describe them separately. However, the real meaning of energy is revealed by the conservation law. Consider a complex system. Suppose we know that somewhere in this system a certain form of energy has decreased. This means that the energy of the other parts must have increased (given that the system is isolated). Thus, we are able to draw the right conclusion straightaway. The great thing is that we don't need to know anything about the way the system functions or what it is made of.

Now back to entropy. Equation (7.2) can be regarded as its definition. As we have already said, entropy is the energy equivalent of probability. In other words, if you look at how much the value $(-TS)$ has changed, it will tell you exactly how much work has been done to transfer the system from a more probable to a less probable state. In this case yet again, just like with energy, you need not worry about the details, e.g. what did the work (a piston, or an electric field, etc.), how the molecules collided with the object doing the work, and so forth.

What exactly does the Boltzmann principle (7.2) mean? Its main idea is that the quantity $U_{\text{eff}} = -TS$ defined by (7.3) and (7.2) can be regarded as some sort of potential energy. Indeed, if the system is left to itself, it is most likely to drop down into the most likely state (sorry for this tautology!) According to (7.2) and (7.3), this would mean an increase in entropy, and hence a decrease in U_{eff}, which is just what the principle of minimum potential energy predicts.

Figure 7.6 sketches the function $U_{\text{eff}}(\mathbf{R})$ for an ideal polymer chain, in accordance with (7.5). The graph has the shape of a potential well. However, you cannot say that "sitting" at the bottom of the well corresponds to the equilibrium. We are talking about non-zero temperatures here. Suppose you have a little ball at temperature T, and you put it into a proper potential well (not U_{eff}, but merely U). What will it do? It will go jittering around the equilibrium position, in a random Brownian way. The typical size of the swings will be such that the potential energy increases by about $k_B T$. (By the way, this is just how physicists estimate the amplitudes of thermal oscillations of atoms in a crystal.) A similar thing can be said about U_{eff}. As you can see from Equation (7.5), the condition $U_{\text{eff}}(\mathbf{R}) - U_{\text{eff}}(0) \sim k_B T$ leads to the result that the distance $\mathbf{R} \sim N^{1/2}\ell$ (as usually, we dropped numerical factors of order unity). This result is

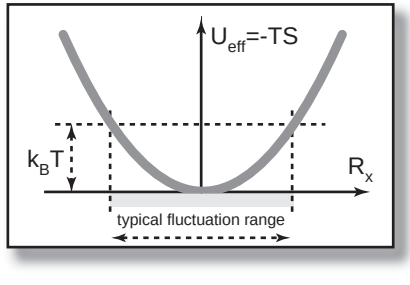


Fig. 7.6 The dependence of the effective potential energy $U_{\text{eff}} = -\text{TS}$ of a polymer on the x-component of its end-to-end vector $\mathbf{R}$. This picture shows how the amplitude of the fluctuations in R can be found from the condition that U_{eff} reaches up to about $k_B T$ above the minimum.

exactly what we want — the most probable end-to-end distance for a single polymer coil (see Section 6.7).

Now the Boltzmann equation has become a little clearer, because we have sorted out U_{eff}, and agreed that it is something like potential energy. Yet it is so tempting to actually try to derive the equation! Let's do it for an ideal gas, in the next section.

7.7 Entropic Elasticity of an Ideal Gas

Assume there is an ideal gas in a vessel with a piston. The ideal gas equation of state (also called sometimes Mendeleev–Clapeyron equation) gives us the pressure of the gas, p:

$$pV = Nk_BT \ , \tag{7.6}$$

where N is the total number of molecules, and V is the volume of the vessel. The force acting on the piston will be

$$f = pA = \frac{Nk_BTA}{V} \ , \tag{7.7}$$

where A is the surface area of the piston.

Suppose we have pushed the piston down by Δx thus slightly compressing the gas. The volume of the vessel has obviously decreased by $\Delta V = A\Delta x$ (Figure 7.5). If Δx is very small, we can neglect the tiny variations of the force and pressure while the piston is moving. Hence the work done will simply be given by

$$f\Delta x = pA\Delta x = p\Delta V = \frac{Nk_BT\Delta V}{V} \ . \tag{7.8}$$

At this stage, it would help if we remembered that $\Delta V/V$ can be approximated as $\Delta(\ln V)$, in the limit of the small V. Indeed, from calculus,

$$\Delta(\ln V) \approx \frac{\partial(\ln V)}{\partial V}\Delta V = \frac{\Delta V}{V} \ . \tag{7.9}$$

Since N is constant,

$$N\frac{\Delta V}{V} = N\Delta(\ln V) = \Delta(N \ln V) = \Delta(\ln V^N) \ . \tag{7.10}$$

Hence, we get:

$$f\Delta x = kT\Delta \ln V^N \ , \tag{7.11}$$

or:

$$f = -\frac{\Delta U_{\text{eff}}}{\Delta x} \ , \tag{7.12}$$

where

$$\Delta U_{\text{eff}} = -k_B T \Delta \ln V^N \ . \tag{7.13}$$

So we have the force in the form (7.12), which looks similar to (7.1). Now we need to specify what U_{eff} is. We are going to link it with the number of ways in which a molecule can be positioned in the vessel.

Clearly, a gas cannot decrease in volume of its own accord. Yet it can expand with no extra help. This inequality of rights has to do with the difference in probabilities. Apparently, a rarified state of gas is more probable than a denser state. The reason is that it can be realized in more different ways. But how do we know? Can we really count all the possible ways for each state? The answer is yes, and the simplest procedure is the following. Let's divide the whole volume of the gas into little cubic cells, of volume a^3 each (Figure 7.7). To make it simpler, let's assume that each molecule can only be located in the centers of the cells. Hence, we are bringing in a discrete distribution of the molecules' positions, instead of a continuous one. There will be V/a^3 different ways of accommodating each molecule in the volume V. Suppose there are N molecules all together. Then you will be able to arrange them in $(V/a^3)^N$ ways in the volume V. This is because the molecules of an ideal gas do not interact with each other; so you can spread them around the cells totally independently. If the piston was originally at a height x_1, the initial volume was $V_1 = Ax_1$, whereas the final volume is $V_2 = Ax_2 = A(x_1 - \Delta x)$. The number of ways, Ω_1 and Ω_2, in which the two states can be realized is given by:

$$\Omega_1 = \ (V_1/a^3)^N \qquad \Omega_2 = (V_2/a^3)^N \ . \tag{7.14}$$

Obviously, $\Omega_1 > \Omega_2$ (because $V_1 > V_2$), so indeed the final, more compressed state is the less probable one. However, you may feel a bit suspicious

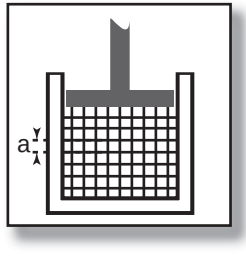


Fig. 7.7 Counting the number of states of an ideal gas.

about the method of calculation. We jumped rather carelessly from the continuous set of the molecules' positions to a discrete one. In fact, it is quite all right to do so. You will see a little later that the size of a cell a will cancel out from all the final formulas for physical quantities. This means that it does not matter after all how you break up the volume and count the states.

How much do the quantities Ω_1 and Ω_2 really differ? Let's look at the ratio Ω_1/Ω_2. Equation (7.14) leads to:

$$\frac{\Omega_1}{\Omega_2} = \left(\frac{V_1}{V_2}\right)^N \quad . \tag{7.15}$$

The ratio V_1/V_2 is greater than one; the power N is a large number, much greater than one. Therefore, $\Omega_1/\Omega_2 \gg 1$. Let's make a simple estimate. Say, for example, $V_1/V_2 = \ 1.01$, and $N = 6 \ \cdot \ 10^{23}$ (this is roughly the number of molecules in one mole of a gas). Then

$$\frac{\Omega_1}{\Omega_2} = \ (1.01)^{6\cdot 10^{23}} \approx 10^{2.6\cdot 10^{21}} \ , \tag{7.16}$$

which is an incredibly enormous number; it is even greater than the number of atoms in the whole universe! Thus, according to (7.15), the final state is incomparably less likely than the initial one.

It seems obvious that spontaneous evolution of a system can only go in one direction, from a less probable to a more probable state (especially when there is such an immense gap between the two probabilities). Now you understand why a gas can spontaneously expand (at a constant temperature), but is unable to shrink of its own accord. In order to compress the gas, you have to push the piston with some force. The counter reaction against your effort is the elastic force of the ideal gas.

Using (7.15), we can rewrite expression (7.11) in terms of Ω:

$$\Delta\left(\ln V^N\right) = \ln V_1^N - \ln V_2^N = \ln\left(\frac{V_1}{V_2}\right)^N = \ln\left(\frac{\Omega_1}{\Omega_2}\right) = \Delta\left(\ln\Omega\right) \ . \tag{7.17}$$

Comparing this with (7.13), we can conclude that, for an ideal gas,

$$U_{\text{eff}} = k_B T \ \ln\Omega \tag{7.18}$$

which is the same as the Boltzmann equation (7.2).

7.8 Free Energy

There is one more tricky bit that we need to sort out. What if we have a complex structure where both the potential energy U and the probability

Ω of the state (or the entropy S as in (7.2)) are changing at the same time? In fact, this is exactly what happens in practice, in real physical systems, including polymers.

The answer is obvious. During any isothermal process, an external force has to do work on both things at a time, i.e. altering the energy U as well as the number of states Ω. The work done is determined by the change in the total $U + U_{\text{eff}}$. This quantity is known as the free energy F, and is normally written as

$$F = U + U_{\text{eff}} = U - TS \ . \tag{7.19}$$

This leads us right to the very basic principle of minimum free energy. Any system, left on its own at fixed temperature, always behaves in such a manner that its free energy goes down. The minimum of the free energy corresponds to the equilibrium state. However, the equilibrium is only defined in a statistical sense — the system never stops its random thermal jittering around the equilibrium position. We say that it fluctuates.

The two terms in the free energy (7.19) are often known as the "energy part" and the "entropy part". Using this, we can make some interesting generalizations. An ideal gas and an ideal polymer both appear to have a zero energy part of the free energy. On the other hand, an ideal crystal has a zero entropy part.

Later on, we shall have a few more chances to explore free energies of polymers. But for now, let's try to solve the question which you might have had for some time. Why "free"? What a strange name! In fact, the concept of free energy (as well as of entropy) belongs to thermodynamics, which is known as "the child of the age of steam". It was once a very applied area, concentrating on the problems of heat-engine design. Even now, if you look in the wrong textbook, you might get the impression that thermodynamics is a strange, out-of-date study of steam engines. This is certainly far from true. Thermodynamics is probably a unique example of a science that originated from rather narrow practical problems and gradually formed into a very general field of knowledge, spanning from cosmology to biology.

It is most amazing that scientists such as Carnot and Clausius, who laid the foundations of thermodynamics, still believed in a very naive caloric theory of heat, which held that heat is a form of fluid. The main practical question they faced was the following. Suppose there is some hot steam coming from a boiler. How much of its energy can it give away to produce useful work? (Presumably, it cannot give up all its heat!) In other words,

what fraction of the energy is free to be converted into work? The answer is: the free energy (7.19)! Hence the name.

7.9 Entropic Elasticity of a Polymer Chain

Let's now go back to the high elasticity. As we have just seen, the internal energy of an ideal polymer does not change: $\Delta U = 0$. So there is no energy contribution to the elasticity; the elasticity is explained in terms of entropy alone. Indeed, when a chain is stretched, we move from a more probable state (realized in more different ways) to a less probable one (realized in fewer ways). The chain starts getting uncoiled, and loses some freedom. In the extreme case, a chain stretched out in a straight line has no freedom at all ($\Omega = 1$, $S = 0$).

We have already found the entropy of an ideal chain (7.4). Now we can use formula (7.1) to find the elasticity:

$$f = -T\,\frac{\partial S}{\partial R} = \left(\frac{3k_BT}{N\ell^2}\right) R\,. \tag{7.20}$$

The vectors $\mathbf{f}$ and $\mathbf{R}$ are parallel (as you can see in Figure 7.3). This is why we can rewrite (7.20) in the vector form:

$$\mathbf{f} = \left(\frac{3k_BT}{N\ell^2}\right)\,\mathbf{R}\,. \tag{7.21}$$

Thus, the force $\mathbf{f}$ has turned out to be proportional to the "displacement" $\mathbf{R}$. We can say that an ideal chain obeys the well-known Hooke's law. However, perhaps we need to be a bit more cautious. Compare (7.20) with an ordinary form of the law (4.1). The main discrepancy is that the average value of $\mathbf{R}$ in a non-deformed chain equals zero. Therefore, we cannot bring in anything like the relative deformation $\Delta\ell/\ell$ which appears in the usual form of Hooke's law.

Still, we could think of an "elastic constant" of a polymer chain: it would be the coefficient of the linear relation between the force $\mathbf{f}$ and the deformation $\mathbf{R}$. According to (7.20), it happens to be $3k_BT/N\ell^2$. First, notice that it is proportional to $1/N$, which makes it a very small quantity if the chains are fairly long. This means that polymer chains are very susceptible to external forces; this is exactly what accounts for the high elasticity of rubber and other similar polymers. The second thing we can notice is that the elastic constant is proportional to the temperature T. This is because the elastic forces are due to entropy, as you can see from (7.3).

7.10 Entropic Elasticity of a Polymer Network

We have explored what happens when an individual polymer chain is stretched. This was not just an exercise. We have shown that the elasticity of a network is built up from the elasticities of all the subchains (Figure 7.2), so we can make use of what we have found. There is one tricky question though. Let's imagine a highly elastic solid body, say, a rubber ball. The macromolecules are rather closely packed in it and interact strongly with each other. So can we really treat each subchain as an ideal polymer, with no volume interactions at all?

The answer is that we can. Of course, in such a dense structure, the thermal motion of the molecules will be nothing like that of an ideal single chain. Atomic groups within one monomer will oscillate and rotate in a totally different fashion. However, the density of the surroundings will make no difference to the entangled shape of the macromolecules (i.e. the size of a coil will still be proportional to the square root of the chain length). The Gaussian distribution (6.16) will not be affected either. In general, the large-scale properties of chains are the same for both ideal and highly-elastic polymers. This idea was voiced clearly for the first time by P. Flory in 1949; thus it is often called the Flory theorem. You can explain it qualitatively in this way. In a uniform, amorphous substance all the conformations of a certain chain are equally likely (in the sense that they correspond to the same energy of interaction with the other chains). This is because the surroundings of each unit are roughly the same. But this is the only assumption we actually made when deriving the elasticity of an ideal polymer.

Now, as we are convinced we are on the right track, let's investigate the stretching of a polymer network (see Figure 7.2). We shall treat it as a set of ideal subchains. Suppose each subchain consists of N freely-jointed segments, each of length ℓ. (To make it simpler, we neglect the polydispersity of the polymer.) When the network is stretched, all the subchains are also stretched on average. Their entropy (7.4) decreases (as the end-to-end distance R grows). This causes an "entropic" elastic force. It does not explain the high elasticity yet. The high elasticity is the capability of bearing huge reversible strains at rather moderate stresses. It occurs because the "elastic modulus" of each chain is fairly small (see (7.20)).

Imagine a polymer network in the shape of rectangular parallelepiped. Let's draw the x-, y-, and z-axes along its sides. Suppose we have elongated

the network by factors λ_x, λ_y, and λ_z along these axes (respectively). Then, if the initial length of the network along the x-axis was a_{0x}, it will now be $\lambda_x a_{0x}$, etc. Now we need to make some assumption about how the network is deformed. The simplest is to assume what is called affinity (where the cross-links and the whole network deform in the same way). Say, the end-to-end distance of a certain subchain was initially $\mathbf{R}_0$, with components R_{0x}, R_{0y}, and R_{0z}. After the deformation, the vector becomes $\mathbf{R}$ such that its components are $R_{0x}\lambda_x$, $R_{0y}\lambda_y$, and $R_{0z}\lambda_z$. According to (7.4), the change in entropy of the subchain is

$$\begin{aligned}\Delta S(\mathbf{R}) &= S(\mathbf{R}) - S(\mathbf{R}_0) \\ &= -\frac{3k_B}{2N\ell^2}\left[\left(R_x^2 - R_{0x}^2\right) + \left(R_y^2 - R_{0y}^2\right) + \left(R_z^2 - R_{0z}^2\right)\right] \\ &= -\frac{3k_B}{2N\ell^2}\left[\left(\lambda_x^2 - 1\right)R_{0x}^2 + \left(\lambda_y^2 - 1\right)R_{0y}^2 + \left(\lambda_z^2 - 1\right)R_{0z}^2\right]. \quad (7.22)\end{aligned}$$

To find the total change in the entropy of the whole network, we have to sum contributions like (7.22) for all the subchains. In other words, we can average over $\mathbf{R}_0$, and multiply by the number of subchains, νV, in the network. (Here V is the volume of the sample, and ν is the concentration of subchains per unit volume.)

$$\Delta S = -\frac{3k_B\nu V}{2N\ell^2}\left[\left(\lambda_x^2 - 1\right)\left\langle R_{0x}^2\right\rangle + \left(\lambda_y^2 - 1\right)\left\langle R_{0y}^2\right\rangle + \left(\lambda_z^2 - 1\right)\left\langle R_{0z}^2\right\rangle\right] . \quad (7.23)$$

Now we can take into account that

$$\left\langle \mathbf{R}_0^2\right\rangle = \left\langle R_{0x}^2\right\rangle + \left\langle R_{0y}^2\right\rangle + \left\langle R_{0z}^2\right\rangle = N\ell^2 \quad (7.24)$$

(see (6.11)). We also know that all the three directions (x, y, and z) have equal rights, therefore $\left\langle R_{0x}^2\right\rangle = \left\langle R_{0y}^2\right\rangle = \left\langle R_{0z}^2\right\rangle = N\ell^2/3$. So we finally get:

$$\Delta S = -\frac{k_B\nu V}{2}\left(\lambda_x^2 + \lambda_y^2 + \lambda_z^2 - 3\right) . \quad (7.25)$$

It is interesting that the answer does not depend on the parameters N and ℓ which describe an individual subchain. This indicates that Equation (7.25) is universal. It works whatever the particular structure of the subchains (for instance, regardless of whether they are freely-jointed or wormlike), for whatever contour lengths and Kuhn lengths, and so on. If we glance again at our calculations, we can see that basically all we needed to draw the main conclusion (7.25) was just to regard the subchains as ideal.

We can use (7.25) to find the stress caused by the "entropic" elasticity, for all sorts of deformations. Obviously, one of the most important types of

deformation is the uni-axial elongation (or compression). Let's see what we can get out of (7.25) in this case. Suppose we have elongated the sample by the factor of λ along the x-axis, i.e. $\lambda_x = \lambda$. The size of the network along the y and z coordinates may change freely. Can we find the relative deformations λ_y and λ_z in this case?

Remember that we are talking about a polymer in a highly elastic state. It seems a sensible assumption that its volume has not changed under the strain. Then, both the y-size and the z-size of the sample ought to have shrunk by a factor of $\lambda^{-1/2}$, that is, $\lambda_y = \lambda_z = \lambda^{-1/2}$. Thus the total volume after the deformation would not change:

$$V = \lambda_x a_{0x} \lambda_y a_{0y} \lambda_z a_{0z} = \lambda_x \lambda_y \lambda_z V_0 = V_0 \ . \tag{7.26}$$

How can we justify, physically, that the volume has to be constant? A highly elastic polymer is usually a sort of fluid (a polymer melt). Its chains are linked with chemical bonds. So, if squashed in all directions, such a polymer is bound to behave as an ordinary liquid. In particular, a 1% change in volume can only be achieved with a pressure of roughly 100 atm $\sim 10^7$ Pa. At the same time, the elastic modulus of such polymer is fairly small. Therefore, the sample can be stretched a few times its length with a much smaller stress ($\sim$ 10^5 or 10^6 Pa). So it is only natural to assume that the volume does not change under such low stresses.

From this point of view, elastic polymers are different form ordinary solid crystals and glasses, which change their volume just because their length changes. At a molecular level, this difference is not surprising. When crystals are elongated, their atoms are pulled further apart. Meanwhile, polymers increase their length by merely disentangling, uncoiling, and stretching out their wiggly subchains; this way the distances between the atoms are kept unchanged.

If we substitute $\lambda_x = \lambda$ and $\lambda_y = \lambda_z = \lambda^{-1/2}$ into Equation (7.25), we obtain

$$\Delta S = -\frac{k_B \nu V}{2} \left(\lambda^2 + \frac{2}{\lambda} - 3 \right) \ . \tag{7.27}$$

There is no problem in finding the elongating force here, using a formula similar to (7.12):

$$f = -T \, \frac{\Delta S}{\Delta a_x} = -\frac{T}{a_{0x}} \frac{\Delta S}{\Delta \lambda} = -\frac{T}{a_{0x}} \frac{\partial S}{\partial \lambda} \ . \tag{7.28}$$

More often we are not interested in the force as such, but rather in the stress, i.e. the force per unit cross-sectional area. There is a little subtlety

of which we should inform the reader: in defining stress, one can divide force by the actual cross-section at the current state of deformation, or by the initial cross-section of the undeformed sample; the former quantity is called true stress, while the latter is usually dubbed as engineering stress. Engineering stress is much easier to find in practice, which is why it is most commonly used. We are also using engineering stress throughout this book (see, e.g., Figure 4.4 and its caption). For our present task, engineering stress is computed as follows:

$$\sigma = \frac{f}{a_{0y}a_{0z}} = -\frac{T}{a_{0x}a_{0y}a_{0z}}\frac{\partial S}{\partial \lambda} = -\frac{T}{V}\frac{\partial S}{\partial \lambda} \ . \tag{7.29}$$

Therefore, let's rewrite our answer in terms of σ:

$$\sigma = k_B T \nu \left(\lambda - \frac{1}{\lambda^2}\right) \ . \tag{7.30}$$

The result (7.30) is a major one in the classical theory for the high elasticity of polymer networks. If the elongation is small (i.e. λ is close to one), Equation (7.30) can be used to estimate Young's modulus of a polymer network (see (4.1)). Indeed, in the limit of $\lambda \approx 1$:

$$\lambda - \frac{1}{\lambda^2} = (\lambda - 1) + \frac{(\lambda+1)(\lambda-1)}{\lambda^2} \approx (\lambda - 1) + \frac{(2)(\lambda-1)}{1^2} = 3(\lambda - 1) \ . \tag{7.31}$$

Meanwhile, the value $\lambda - 1 \equiv (a_x - a_{0x})/a_{0x}$ is just the relative elongation. In other words, it plays the same role as the parameter $\Delta\ell/\ell$ in Equation (4.1). Comparing (4.1), (7.30), and (7.31), we end up with Young's modulus:

$$E = 3k_B T \nu \ . \tag{7.32}$$

Thus, E turns out to be the same as the pressure of an ideal gas whose molecular concentration is 3ν (i.e. three times the concentration of the cross-links). It means that the more cross-links there are in a highly elastic sample, the less elastic it is. Therefore, the value of E does not indicate a specific polymer. It varies dramatically depending on the density of the cross-links.

However, (7.30) can be used not only to find Young's modulus. It also describes the nonlinear elasticity, which takes up quite a lot of room on the stress versus strain curve. (In Figure 7.1, it spans from point A where the elasticity ceases being linear up to point B where the reversibility is lost.) What is more, Equation (7.30) is just as good for uni-axial compression. You only need to bear in mind that λ will be less than one in this case. Another warning is that when compressed along the x-axis, the sample will

automatically stretch in both the y and z directions. Even more complex deformations, such as two-dimensional elongation, torsion, shear, and so on are covered by the general relationship (7.25). Although we shall not do it here, you can derive equations similar to (7.30), revealing nonlinear behavior of the stress.

Thus, the result (7.30) $\sigma(\lambda)$ dependence is pretty general — but how accurate is it when compared with experiments? Figure 7.8 brings together both a typical experimental curve and the theory. You can see that up to about $\lambda = 5$ the agreement is far from perfect, but more or less tolerable. Then, for $\lambda > 5$, the discrepancy grows more and more. This is not surprising. Expression (6.16) for $P_N(\mathbf{R})$ ceases to work for long end-to-end distances R (or, equivalently, large elongations). Why? Because it does not take into account that there is a limit to how much the chains can actually be stretched. Namely, the distance R can never exceed the total contour length $N\ell$. This is why Equation (7.27) and all the consequent results will not hold for the case of strong elongation.

Let's look at the range of moderate elongations: $1.2 < \lambda < 5$. For most polymer networks, typical discrepancies between the theoretical and experimental $\sigma(\lambda)$ are not that high (about 20% or so), but they tend to be systematic (Figure 7.8). These are explained by the so-called topological constraints to the subchains' conformations (see Section 2.6).

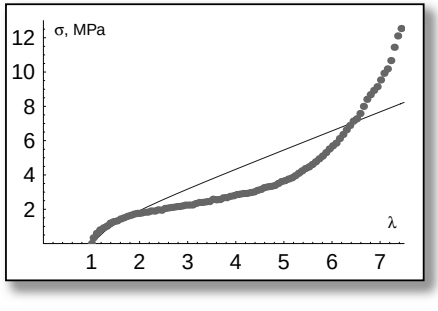


Fig. 7.8 The dependence of (engineering) stress σ on the strain λ for a highly elastic polymer network material. Solid line is the theory (7.30); dots show a typical experimental curve (see caption of the Figure 4.4 about the definition of engineering stress). Equilibrium module used to plot the theoretical curve is $3\nu k_B T \approx 3.3$ MPa, which corresponds to $\nu \approx 0.27$ nm^{-3} — roughly one cross-link per four cubic nanometers. Although the data presented are quite typical in the sense that data first go below theoretical curve and then shoot up above it, the value of ν and the corresponding values of stress can easily change by an order of magnitude either way in different materials. Experimental data courtesy of A.A. Askadskii were obtained at room temperature using the sample of polyurethane derived from methylenediphenyl diisocyanate and polyester based on hexanedioic acid and 1,2-ethanediol (the chain extender is 1,4-butanediol). The sample of initial length 25 mm was stretched at the rate 0.16 mm/s, which is very slow. The stress, which has the dimensionality of pressure, is given in the units of megaPascal: 1 MPa $= 1 \frac{\text{pN}}{\text{nm}^2}$.

Despite all the imperfect agreement with experiment, this approach to the description of high elasticity has proved ideologically quite successful, especially because it is universal. All the predictions are satisfactorily accurate, whether it is absolute values of Young's moduli, or their temperature dependencies, or the shape of the nonlinear stress versus strain curve. There are not many other examples in the physics of disordered solids and liquids where such simple arguments have helped to understand so much. The reason is rather obvious. The way the chains are entangled and rolled up into coils is not described by the short-scale chemical structure or interactions of individual atomic groups. It is determined by the very fact that the monomers are grouped into chains. So it is a long-scale feature. Soon we shall have a chance to discover that some other most interesting and peculiar properties of polymers have the same sort of origin.

7.11 The Guch–Joule Effect and Thermal Aspects of Rubber Deformation

Until now, all we have deduced from (7.30) was the dependence $\sigma(\lambda)$ at constant temperature. However, it also contains the dependence on T. So, let's analyze how the elasticity of polymer networks is affected by the temperature.

Suppose we hang a weight on a rubber string. The string will become elongated. Now let's increase the temperature. According to (7.30), as long as $\sigma = \text{const}$ (because hanging weight does not change), an increase in the temperature should lead to a decrease in λ. Thus, a stretched rubber string, in contrast to most non-polymer materials, contracts on heating! This strange behavior was discovered by Guch as early as 1805 when he experimented with strips of natural rubber. Half a century later, Joule (the same person who, being professional bear brewer, established equivalence of heat and mechanical energy and thus helped discovering the energy conservation) carried out a careful set of measurements and confirmed Guch's result. Thus this phenomenon is usually known as the Guch–Joule effect.

This effect demonstrates, perhaps in the most dramatic way, that the high elasticity of rubber and other polymers is related to entropy. Indeed, as the temperature goes up, all sorts of interactions start to lose their importance. This is because the characteristic energy of such interactions, ε, becomes much less than k_BT (i.e. $\varepsilon/k_BT < 1$). Meanwhile, the entropy contribution gains more and more significance. (According to (7.4), the

entropic elasticity is proportional to the temperature.) Therefore, the fact that the "bouncing" reaction is enhanced by heating suggests that it is the entropy to blame for the high elasticity of rubber.

What happens when the system cools down again? The sample may become partially crystallized, and in that case the Guch–Joule effect is replaced by the opposite: the partially crystallized rubber would expand with heating. This is yet another sign that only the highly elastic polymers exhibit the very peculiar elasticity of special entropic nature.

A common but very impressive way to demonstrate the Guch–Joule effect in a class setting is the following. (It can be reproduced easily in an ordinary school laboratory.) Take a bicycle wheel and replace the spokes by elastic strings made of soft rubber (Figure 7.9). Preferably, the strings should be stretched to about three times their original length. Now fix the axle in a horizontal position, so that the wheel can rotate in the vertical plane with little friction. Place an electric heater pointing at a certain

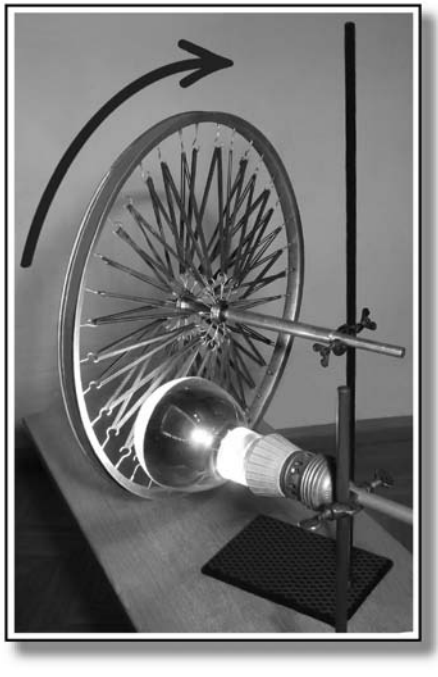


Fig. 7.9 An experiment to demonstrate the Guch–Joule effect. The device in this photo belongs to the lecture demonstration facility of Physics Department, Moscow State University; the image is courtesy of A.A. Aerov and S.B. Ryzhikov.

part of the wheel. (You can also use a powerful electric light bulb for this purpose, shining on some section of the wheel.) The heat makes the rubber strings shrink, and the center of mass shifts. As a result, the warmed sections of the wheel move up, and other sections take their place at the spot being heated. They are warmed up in their turn, and so on. As you might have guessed, the wheel starts rotating at a constant angular speed[2].

Here is an even simpler experiment. Hang a weight on a rubber band, and start warming up the band, either using a light bulb or some other sort of electric heater. Alternatively, you may place the whole thing in water (say, in a bucket), and heat it up. When heated, the rubber will pull the weight up, and after cooling down it will let it down again.

In the Guch–Joule phenomenon, the change in size of an elastic string is a consequence of the temperature variation. It turns out that it works the other way around, too: The temperature of a string will change if its size is changed, for example, if it is rapidly squashed or elongated. You can check it yourself. Stretch a strip of rubber quickly, then touch it with your lips to feel how it is warming up! Conversely, when stretched rubber is rapidly released, it will noticeably cool down. This is totally opposite to what happens with an ideal gas, which gets warmer when rapidly compressed and cools when allowed to expand rapidly.

In spite of the contrast, the two types of behavior — those of the gas and those of the rubber — have the same cause. Say a system is quickly extended (expanded) or compressed. Obviously, there will be no time for heat to be exchanged with the surroundings. To put it in other words, all such processes can be regarded as adiabatic. So why should an ideal gas warm up when it is adiabatically squashed? The answer is this. The squashing requires some work to be done by external forces. Where does this work go? As no heat is leaking out, all the work is transformed into the internal energy of the gas. Therefore, the temperature rises. The same happens when a piece of rubber is adiabatically (rapidly) stretched. Again, external forces do the work, which is all used to increase the rubber's internal energy (and the temperature). In contrast, when either rubber is compressed or an ideal gas is expanding, the work has to be done by the system itself. As no extra heat comes from the outside, some of the system's own internal energy has to be used, and so it cools down.

[2]At the first glance, this device may seem to work as a *perpetual motion machine*, violating the thermodynamics. In fact, of course, the wheel rotates on the expense of energy provided by the electric heater, so there is no contradiction.

7.12 Single Chain Stretching Revisited: Worm-Like Chain Model and dsDNA

When the first edition of this book was in preparation, the idea of single molecule stretching was a rather abstract theoretical concept. But this turned out to be the place of huge breakthrough of experimental techniques. At the present time the so-called force-spectroscopy experiments are routinely performed in many laboratories around the world. And this all started with the simple minded but technically sophisticated experiment by C. Bustamante and his colleagues, when they tried to pull dsDNA by its ends — pretty much in the way shown in our cartoon in Figure 7.3. Despite scores of textbooks describing a thought experiment on stretching a single polymer chain, the results of a real experiment proved largely unexpected. Mother Nature once again taught researchers that every theory has its limits of applicability...

We mentioned already more than once the nice property of universality in many cases characteristic of polymers. In this spirit, researchers got used to thinking that all polymer chains are alike, each characterized by just two parameters, total length and persistence length. Indeed, isn't that what we described above in this chapter? Formulas (7.20) or (7.21) relate pulling force f to end-to-end distance R in terms of the segment length ℓ and the full contour length of the chain $L = N\ell$; they represent, of course, quite a universal relation – but it is valid if and only if the chain is not too strongly stretched. If you go back and re-trace the derivation, you will see that formulas (7.20)–(7.21) represent an offspring of Gauss distribution (6.16), which is a very good approximation for modest end-to-end distances R, but certainly breaks down if R is too large (for instance, Gauss distribution predicts small but non-zero probability that a polymer chain stretches from here to the Moon, which is an obvious absurdity). Therefore, we should re-write formula (7.20):

$$f = \left(\frac{3k_BT}{L\ell}\right) R \quad \text{if} \quad R \ll L \ . \tag{7.33}$$

This universal result is useful to describe network elasticity, because almost none of the subchains in the network is close to full stretching. But as soon as people started doing single molecule experiments, the stretching of DNA to almost its full contour length became possible and brought interesting unexpected results.

Indeed, how should we address chain stretching when end-to-end distance R approaches full stretching length L? The easiest approach is to

solve this problem for the freely-jointed chain model (Figure 2.5 *b*). And people assumed — wrongly, as it turned out — that the result should be universal. But the result was in gross discrepancy with the dsDNA stretching experiment. Of course, it did not take long to realize that there is no (and should not be) universality in the strong stretching regime, and the analysis of worm-like chain model (Figure 2.5 *a*) showed excellent agreement with dsDNA data, as shown in the Figure 7.10. This was the proof that dsDNA does indeed possess the worm like mechanism of flexibility.

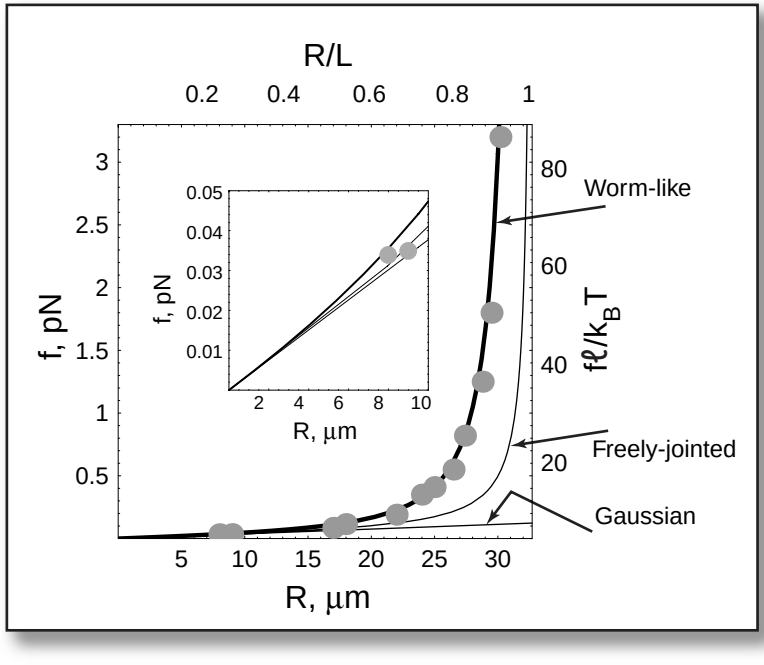


Fig. 7.10 Results of the experiment on stretching single DNA chain by its ends. The double λ-DNA was used, its length is about 97,000 base pairs, or $L \approx 32.7$ μm. On the lower axis, the end-to-end distance, R, is labeled in microns; on the upper axis it is labeled in terms of unitless ratio R/L. The force is seen to increase rapidly as R approaches L, or as molecule approaches full stretch. On the left axis, the force is marked in pico Newtons (10^{-12} N); on the right, the force is labeled in terms of unitless quantity $f\ell/k_BT$. The inset shows the region of small forces. The worm-like chain model is seen in excellent agreement with the data. The freely-jointed chain model significantly underestimates the force in the strong stretching regime. Gauss theory is only valid in the region of small forces, when, as expected, all theories yield basically identical results. The data are taken from the paper: S. Smith, L. Finzi, C. Bustamante, "Direct Mechanical Measurements of the Elasticity of Single DNA Molecules by using Magnetic Beads", Science, v. **258**, n. 5085, p. 1122, 1992.

Interestingly, there was a sort of over-reaction among researchers: now many people consider worm like chain model either more realistic or just "real" — while in fact many polymers (e.g., proteins) should be described by other models, with rotational isomers etc.

7.12.1 *Strong Stretching of a Chain is akin to its Confinement in a Narrow Tube*

Because of great importance of this subject, we will sketch below the derivation of $f(R)$ for both freely-jointed and worm like chain models and explain the physical source of difference between them. In both cases our approach will be based on the following idea.

If the chain is almost fully stretched, its excursions in perpendicular directions are very small, they are suppressed. This is illustrated in Figure 7.3 *b*: when the force is strong, the end-to-end distance R is large and approaching the full length of molecule L, while the characteristic distance of perpendicular excursions, D, gets small. Therefore, it will prove useful to consider an artificial model, depicted in Figure 7.11: the chain confined in a very narrow tube of some diameter D which is so small that it is smaller than the chain segment (or effective segment) length ℓ: $D \ll \ell$. Surely, the relevant tube diameter D should depend on the amount of stretching and we will have to determine it in some self-consistent way. But for now let us think about the tube and let us ask: how much work should we perform to squeeze chain in such a tube? That will help us determine how much force we have to apply in real situation to achieve the end-to-end separation R.

We should also mention that the problem of chain confinement in a narrow tube, introduced above as an artificial model, is of great interest in its own right: think about DNA translocation through a narrow channel (see Figure C5.7), or DNA pushed from virus head into the cell through a narrow tube. As a note in passing — it is probably true in general that a good model in physics will always find some application: in the words of Charles Dickens, "Nothing of what is nobly done is ever lost" (we will return to this in Chapter 11).

7.12.2 *Strong Stretching of a Freely-Jointed Chain*

If a freely-jointed chain is confined in a narrow tube, each of its straight segments makes a small angle with the tube axis. That is unfavorable in terms of entropy: as the tube gets more narrow, every segment has fewer

orientations available to it. We can apply Boltzmann principle to analyze this "orientational entropy". Since we talk about orientations only, we can imagine that one end of the segment is fixed in some point in space, then the other end chooses any point on the sphere of radius ℓ, its surface is $4\pi\ell^2$. But if the segment is confined in a tube, its end is restricted to the much smaller surface proportional to D^2, as it should be clear from the lower part of Figure 7.11 *a*. Therefore, dropping as usual the numerical coefficients[3], we can say the entropic price of confining one segment into a tube is about $k_B \ln(D^2/\ell^2)$, and for all N segments we get $\Delta S = N k_B \ln(D^2/\ell^2)$. The physical meaning of this quantity is revealed by noticing that $-T\Delta S$ is the minimal amount of work necessary to confine polymer into the tube, and $\Delta S < 0$ means that the chain resists being squeezed.

Interestingly, it does not matter how we are going to perform work: we can squeeze polymer from the sides or we can pull it by the ends, the amount of work should not depend on that (the beauty of entropy and Boltzmann principle!). But if we are talking about pulling the ends, it is more convenient to use end-to-end distance R instead of D. How are they related? For each segment, its projection along the tube axis is $\overline{\ell^2 - D^2}$ $\ell - D^2/2\ell$ (since $D \ll \ell$). Therefore, total end-to-end distance is R $N\left(\ell - D^2/2\ell\right) = L - LD^2/2\ell^2$. By doing now simple algebra we can re-express D in terms of R and then re-write entropy or the corresponding free energy for all L/ℓ segments as an explicit function of R:

$$\Delta F \sim -k_B T \frac{L}{\ell} \ln\left(1 - \frac{R}{L}\right) \ . \tag{7.34}$$

The necessary pulling force is found from here by differentiation: $f = -\partial\Delta F/\partial R$:

$$f \sim \frac{L}{\ell}\frac{k_B T}{L - R} \ \text{(freely jointed, strong stretching)}, \ L - R \ll L \ . \tag{7.35}$$

We see that force blows up as we approach full stretching. It is always difficult to devoid any system of its last pieces of freedom, which is why it is impossible to reach absolute zero of temperature, and which is why it is impossible to reach complete stretching of a polymer with finite force; in fact, the chain will break far before it reaches complete stretching.

This is all good, but formula (7.35) did not agree with experiment on dsDNA.

[3]We drop numerical coefficients inside the ln, such as 4π etc., because, for instance, $\ln(4\pi\ell^2/D^2) = \ln(\ell^2/D^2) + \ln(4\pi)$ and, since $\ell \gg D$, we neglect the second term.

7.12.3 *Strong Stretching of a Worm-Like Chain*

Let's try to analyze the worm-like chain model in the tube. We will do it using the idea of Figure 7.6, namely, finding the fluctuation with energy about k_BT. As it is seen in Figure 7.11 *b*, conformation of a worm-like chain in a tube represents a kind of succession of arcs. Let's concentrate on one such arc. If λ is the length of it, then simple geometry, shown in the lower part of Figure 7.11 *b* suggests that curvature radius ρ of the arc is about

$$\rho \sim \frac{\lambda^2}{D}, \tag{7.36}$$

(again dropping numerical coefficients). What is then its energy?

Well, we never said so far how to find bending energy of worm-like chain. As a matter of fact, we can borrow the prescription for this task from mechanical engineers, for they know everything about bending of elastic

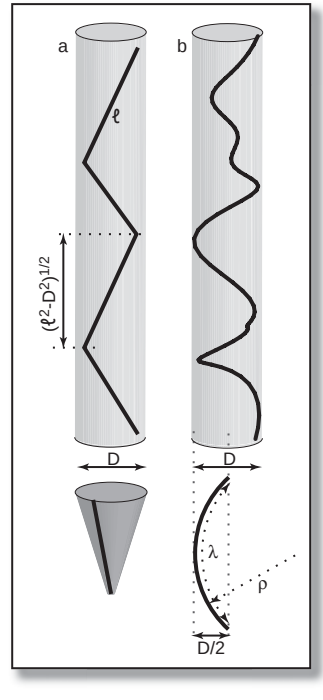


Fig. 7.11 Confined in a narrow tube freely-jointed chain (*a*) and worm-like chain (*b*). The lower figures represent auxiliary geometric constructions explained in the text. Notice that worm-like chain hits the walls in more points than freely-jointed.

beams. But we can also guess ourselves: bending energy should be proportional to the squared curvature $1/\rho^2$ (because it should vanish for the straight shape with no curvature, and it should not depend on the sign of curvature); it should be proportional to the length of the piece in question λ; and there should be some coefficient describing the rigidity of the material. If we take bending energy in the units of k_BT, the result should be then unitless and, therefore, the material coefficient in this case can be nothing else but persistence length or effective segment (up to a numerical coefficient). Thus, bending energy of the arc reads $E_{\rm bend} \sim k_BT\ell\lambda/\rho^2$. Since bending energy should be just about k_BT (see Figure 7.6), we get that

$$\frac{\ell\lambda}{\rho^2} \sim 1 \ . \tag{7.37}$$

Equations (7.36)–(7.37) together determine the important length scale λ — the so-called undulation length (also called Odijk length after the author of the concept — Theo Odijk of Leiden University in The Netherlands):

$$\lambda \sim \ell^{1/3} D^{2/3} \ . \tag{7.38}$$

Notice that $\lambda \ll \ell$; that means, worm-like chain hits the tube walls more frequently, in more places, than freely-jointed chain; one can say, worm-like chain requires more guidance to keep going straight.

And now we can estimate confinement entropy. We will argue that this entropy is of order unity per each arc of the length λ. Indeed, we can say each arc can bend to either right or left, which means the relevant number of conformations is estimated as $2^{L/\lambda}$, where L/λ is the number of arcs. According to the Boltzmann principle, we obtain then entropy proportional to L/λ. Transforming this to the function of end-to-end distance R (which is geometrically related to λ in the same way as R was related to ℓ for freely jointed chain, i.e., $R \quad L - LD^2/2\lambda^2$, or given formula (7.38), $R - L$ $-L\lambda/\ell$), we arrive at the free energy

$$\Delta F \sim -k_BT \frac{L^2}{(L-R)\ell} \tag{7.39}$$

and pulling force:

$$f \sim \frac{k_BTL^2}{(L-R)^2\ell} \ \ (\text{worm-like}, \text{strong stretching}),\ L - R \ll L \ . \tag{7.40}$$

As in case of freely-jointed chain, Equation (7.35), the force also blows up at the approach to complete stretching, but does so much stronger — as

$(L-R)^{-2}$ instead of $(L-R)^{-1}$. And this difference in power does the trick — formula (7.40) agrees with the data very well.

It is important to understand the physical meaning of the difference: why stretching worm-like chain becomes much more difficult at large elongations than stretching a seemingly similar freely-jointed polymer? This is actually yet another demonstration of entropic nature of polymer elasticity. Indeed, to keep the freely jointed chain in the confines of a thin tube it is necessary and sufficient to control the direction of every segment to within proper accuracy, but the segment has to be oriented at one point only. By contrast, the direction of worm-like chain has to be controlled everywhere: as we decrease the diameter of the confining tube, we have to correct the orientation at an increasing number of points; in other words, as we squeeze the chain in an increasingly narrow tube, or pull it by the ends with an increasing force, we must suppress increasingly short wave length motions of the chain, which requires more and more entropy.

A prepared student would be well advised to think through the analogy of this situation with the well-known Einstein and Debye models of heat capacity in solid state physics. The analogy is surely incomplete, and in some ways even reverted, but still very instructive. Let's remind that in Einstein model a solid body is represented as a set of oscillators, all having the same frequency; this is similar to freely-jointed chain, in which all possible ways to bend the chain have one and the same wavelength — the segment length ℓ. Einstein model played an important historical role in physics, for it showed how heat capacity can violate the classical mechanics prediction (which says that heat capacity does not depend on temperature); but Einstein model did not agree with experiments. Debye model proved immensely more successful, because it predicts a certain universal (for all bodies) behavior of heat capacity at low absolute temperature, proportional to T^3 (for regular three-dimensional crystals), which agrees very well with numerous experiments. The key feature of Debye model is the realization that oscillators in the crystal are nothing else but sound waves, or phonons in quantum language. As we lower temperature, we remain with increasingly long wave length waves (their frequencies and energies are lower), and these are universal because long wave envelops many atoms and hides the difference between them. In this sense strong stretching of worm-like chain is somewhat opposite to lowering temperature of a solid, because in the former case we suppress the long wave lengths, while in the latter case we suppress the short ones. (As a matter of precaution, we warn the reader that these "waves" in a worm-like polymer, unlike regular solid, are just

a mathematical way to think about bending fluctuations, these waves are overdamped by the friction in the viscous solvent, and they would not exist if not for the constant excitation by the surrounding molecules.)

7.12.4 *Force Spectroscopy*

The difference in stretching behavior between different polymers proved extremely informative and gave rise to the experimental method called "force spectroscopy". For instance, the data of C. Bustamante and his co-workers shown in the Figure 7.10 are considered the experimental proof of the fact that dsDNA is a worm-like chain. Strictly speaking, formula (7.40) is not enough to make this conclusion, for it only describes the situation at rather large force; one needs a more sophisticated formula that connects smoothly between the universal behavior at small forces (7.33) and the blow-up at large ones.

We will not derive general formulas here, but just write them down for both freely-jointed and worm-like chains — in case our reader wants to play with them. For the freely-jointed chain the formula expresses end-to-end distance as a function of force and reads

$$R = N\ell \left(\frac{e^{\xi} + e^{-\xi}}{e^{\xi} - e^{-\xi}} - \frac{1}{\xi} \right) , \quad \text{where} \quad \xi = \frac{f\ell}{k_B T} ; \tag{7.41}$$

this formula is called after P. Langevin (1872–1946) (he discovered the formula while studying the magnetization dependence on the applied magnetic field — yet another physical analogy). For the worm-like chain the corresponding formula is called after J. Marko and E. Siggia (who first used it to interpret Bustamante data), it expresses force in terms of the extension and reads

$$f = \frac{k_B T}{2\ell} \left[\frac{L^2}{(L-R)^2} - 1 + \frac{4R}{L} \right] . \tag{7.42}$$

The reader should check that these formulas have proper limiting behavior at small and large forces.

Worm-like chain model is good not only for dsDNA, but for a number of other polymers. The notable example is so-called F-actin. Strictly speaking, calling F-actin a polymer is a little bit of a stretch: it does not have a covalent backbone; it is actually a chain-like assembly of protein globules (called G-actin). The diameter of F-actin "chain" is about 5 nm, and the chain is rather stiff, its effective segment is close to 30 μm, about three

orders of magnitude larger than for dsDNA. Actin fibers are important components of cytoskeleton supporting shapes of eukaryotic cells.

The success of worm-like chain model for dsDNA made this model and formula (7.42) fashionable among the scientists (somewhat surprising fact of life is that there is such thing as fashion in science!); nowadays formula (7.42) is often used to fit the data for which it is not at all the most appropriate. We sincerely hope that Langevin formula (7.41) and its underlying freely-jointed model, as well as rotational isomers and other models will soon regain their rights and will be used where appropriate.

Meanwhile, every meaningful scientific theory, from very large (like relativity) to rather small (like worm-like chain model) has its limits of applicability. Finding these limits is not a penny less important than formulating the theory itself. Of course, worm-like chain model does not work for many polymers (having joints of some kind). Furthermore, if we pull on DNA stronger and stronger, we can eventually start deforming the double helix itself — which means we can start testing the energetic component of the DNA elasticity in addition to the entropic part. Indeed, under usual conditions of temperature and ionic strength, something serious happens to DNA at the force about 65 pN (see below Section 8.2 about the estimates of forces; you will see that 65 pN is actually a huge force). Definitely worm-like chain model fails above this force, but what exactly happens is not clear, and the debate continues whether double helix unwinds or undergoes some other transformation.

Unfortunately, this book is not the place to describe all the beautiful tools employed in force spectroscopy experiments, such as optical tweezers, magnetic tweezers, atomic force microscope, and several others. But the fact of the matter is that force spectroscopy methods which were born and gained popularity in the experiments on dsDNA are used with increasing success. People pull on proteins to learn how they fold and unfold (Section 5.7 and Chapter 10); on DNA to study its translocation (Figure C5.7), or examine helix-coil transition (called unzipping in this context), or to see how DNA tail is being packaged in the virus head (Figure C9.11); researchers pull on molecular motors (see Section 5.8) to find how strong they are; and the list of applications continues to grow rapidly.

Chapter 8

The Problem of Excluded Volume

> I hate to tell you, but there ain't any chance for but one of us. Bolivar, he's plenty tired, and he can't carry double.
>
> O. Henry,
> *The Roads We Take*

8.1 Linear Memory and Volume Interactions

What are the chances that one or another theoretical study will be a success? As history shows, it greatly depends on whether theorists can think of a nice, manageable model idealizing the real world. Of course, there are no ideally simple systems in nature. However, we can use our imagination and *invent* an ideal gas (whose molecules do not interact at all), an ideal crystal (with no defects at all to the regular atomic structure), and so on. As a matter of fact, you can say that all these models are ideal indeed, meaning that they are the best for physicists. This is because they are the simplest — but they are simultaneously the most basic ones. So one has to master them first, before moving any further in either statistical mechanics, or hydrodynamics, or solid state physics, or whatever chapter of physics.

How crude are the results we might get from such "ideal" models? Are there some cases where they work well, and some where they fail? There is a special trick that often helps us to decide. It involves finding some dimensionless parameters, either large or small, which describe the system. For example, a gas can be characterized by the fraction of volume which is taken up by the molecules. If this parameter is much less than one,

then the molecules are typically very far away from each other, and the gas can be treated as ideal. Similarly, a crystal is nearly ideal if the dimensionless fraction of incorrectly occupied sites of the lattice is small. Notice also that these small parameters emphasize the connection between "idealizations" and "approximations": under certain real circumstances, when proper parameter is small (or its inverse is large), a real system might be well approximated by an idealized model.

What sort of large or small dimensionless parameters can describe a polymer? One of them we have actually used already, and not just once. It is the large number of monomer units ($N \gg 1$) in a chain. We have shown that a huge N can account for many things. It explains, for example, the low concentration of monomers in a coil (see (6.14)), the existence of semi-dilute solutions (Figure 4.7 *c*), and the high elasticity of polymers.

Another special polymer parameter comes from the hierarchy of interactions. The energy E_1 of a covalent bond between two neighboring monomers in a chain is normally about 5 eV $\approx 0.8 \cdot 10^{-18}$ J. This is much higher than the typical energy E_2 of any other interactions (say, between the polymer and the solvent, or between monomers which are not nearest neighbors along the chain, etc.) Roughly $E_2 \sim 0.1$ eV $\approx 1.6 \cdot 10^{-20}$ J. Therefore, the ratio $E_2/E_1 \ll 1$ is just the type of small parameter we were seeking. It allows us to introduce an ideal polymer chain approximation.

Indeed, let's see what happens near room temperature ($k_BT \approx 2.6 \cdot 10^{-2}$ eV $\approx 0.41 \cdot 10^{-20}$ J). This region is the most interesting one as far as polymer properties are concerned. Covalent bonds cannot be broken due to thermal fluctuations, since $E_1/k_BT \approx 200 \gg 1$). This means that the sequence of units is "cemented" into the chain by the high energies of the backbone covalent bonds. Each unit "remembers" its own number which it acquired when the chain was formed. To put it briefly, a polymer chain has a fixed linear memory.

Having sorted out the covalent bonds between the neighbors, we can now concentrate on all the other interactions. These are frequently referred to as "volume interactions". As we have said, they have a typical energy E_2, and are much weaker than those in charge of the linear memory. In the crudest theory, we may completely neglect them. Then we shall end up with exactly what is called an ideal polymer chain. This is just how we handled all the calculations in the previous chapters. It worked fairly well, and we coped with quite a number of problems. We described how a chain rolls up into a loose coil, and we revealed the peculiar entropic nature of the high elasticity of polymers.

Nevertheless, the ideal polymer chain approximation — just as well as ideal gas or ideal liquid or any other idealization in science — proves not to be enough for many purposes. The properties of real polymers are much richer and more diverse than idealization predicts. If you are not convinced, think back to Chapter 4. There we talked about the various physical states of polymers. In order to understand fully how all those states are formed, and why, we need to allow for volume interactions. These include, in particular, interactions between different macromolecules. Monomers in the same macromolecule will also interact, even if they are not close neighbors along the chain, but somehow come close to each other in space due to the chain bending and wiggling. In this chapter, we shall look at interactions of both types — between monomers belonging to the same or to different macromolecules.

What can we say in general about interaction between two monomers? We will discuss it in some further details in the next Section 8.2, but it will be handy to make some preliminary arguments here. The interaction certainly depends on the type of chain, and on the solvent too. However, we can roughly sketch the potential energy of this interaction $u(r)$, as a function of the distance r between the monomers (Figure 8.1 *a*). (In general, the potential energy does not only depend on r, but also on the mutual orientation of the monomers, and bulky monomers can have some flexibility of their own to affect the u — see, e.g., Figure 4.2. We do not take this into account directly, since the main qualitative features are well enough represented by the simplified Figure 8.1 *a*.) Qualitatively, main features of $u(r)$ are quite common for all types of molecules and monomers: If r is small, $u(r)$ is positive and very large. This is because the monomers cannot penetrate into each other. In other words, the volume taken up by each monomer is automatically excluded from that available to any other one (hence the phrase "excluded volume"). As r becomes larger, monomers usually start to attract each other. This is the region on the right-hand side of the minimum in Figure 8.1 *a*. Usually, the crossover distance r_0 between the two regimes (corresponding to the minimum) should have the same magnitude as the size of a monomer unit, i.e. $r_0 \sim 1$ nm $= 10^{-9}$ m. What is the physical meaning of $u(r)$? To bring two monomers together, as close as r, some work has to be done. This work is stored in $u(r)$. It is done against the solvent molecules, as they need to be squeezed out of the way. Hence, the potential energy $u(r)$ represents the effective interaction of monomers through the solvent. It should depend, therefore, on the contents and state of the solvent, as well as on the temperature.

By the way, to describe interaction of monomers through the solvent in terms of potential $u(r)$ is also an idealization of sorts. It works well, but not always, and more ideas have to be brought to bear if it fails — in cases such as strongly elongated particles, when dependence on orientation is important; charged monomers, because Coulomb interaction is strong and long-ranged and involves many particles collectively; monomers forming covalent chemical bonds in addition to those along the chain backbone, etc. We will touch quite a few of those special cases later in the book, but the good practice in science (and, we believe, in life in general) is to examine simple things first, to gain insight and intuition which we can later bring to bear on complications.

8.2 Four Forces in Molecular World; Scales and Units

Before we move to our next subject, which is polymers with excluded volume, it is prudent to digress and review the forces operating between molecules and monomers in a more specific way. We realized already and will see more all the time that properties of polymeric substances are dictated by the interactions between monomers or between different chains. What are those interactions?

The reader may have heard about the four fundamental forces in nature (gravity, electromagnetic, weak, and strong are their names), but this is not what we are talking about. In molecular world all the relevant forces are fundamentally various forms of interplay between electromagnetism and quantum mechanics. But these can masquerade in different costumes, traditionally grouped into four categories, as illustrated in Figure 8.1: panels *b* through *e* represent, respectively, van der Waals interactions (see, e.g., Section 9.3), hydrophobic interactions (Section 5.1), Coulomb attraction or repulsion between charged groups (Section 2.5.3), and hydrogen bonds (Section 5.1). Van der Waals forces are the most generic, they are always present. Panel *a* represents a sketch of characteristic potential profile, including repulsion at short distances and attraction at longer ones. This arises typically of combination of several types of forces, including the interactions with the solvent.

To be quantitative, we should indicate the characteristic scale of these forces. This is a delicate matter, because we have to choose the units. Indeed, hundred dollar bill is inconvenient to operate a public telephone, while change or small coins are equally inconvenient to buy a pair of shoes;

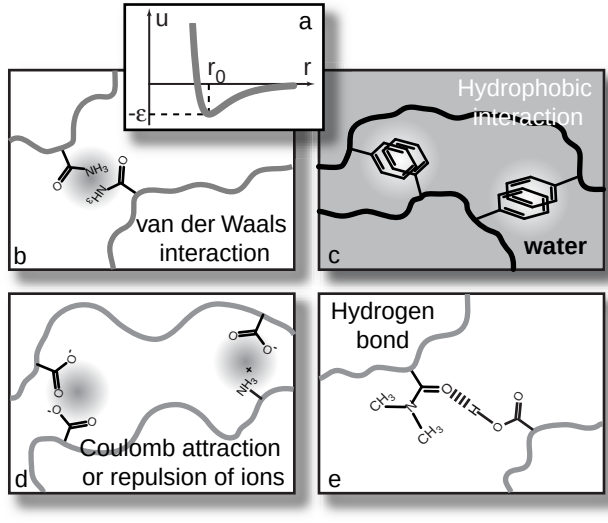


Fig. 8.1 This cartoon illustrates four types of forces operational in molecular world, such as (*b*) van der Waals interactions (which include both repulsion at short distances and attraction at longer ones), hydrophobic interactions (*c*) between nonpolar molecules and groups in water medium, Coulomb forces (*d*), which might be very strong and require special attention, but present only when some groups are charged, and hydrogen bonds (*e*), which are directional and saturate (unlike others). Panel (*a*) represents a typical potential of interaction between two monomers in the solvent which usually results from the interplay of several of the above mentioned interactions: repulsion at short distances is always present due to the van der Waals component, while other aspects might be influenced by the solvent and other circumstances.)

similarly, the common units (meters, kilograms, etc), chosen for their congruency to the human body scale, are inconvenient for the molecular world. The natural scale of energy in molecular world, as we already saw, is k_BT, because polymers and biopolymers usually exist under the conditions in which absolute temperature does not really change that much: for instance, the whole interval between 0°C and 100°C in terms of absolute temperature corresponds to room temperature plus or minus 10 or 20%. Therefore, for the purposes of rough estimates the T is always close to 300°K.

In terms of this natural energy unit, the typical strengths of the interactions are as follows. To begin with, covalent bond is as strong as about 100 or $200k_BT$; this is why covalent bonds do not break on their own. This is also why covalent bonds are not included among four types of forces in molecular world: unbreakable covalent bonds are always among the determining circumstances of all non-covalent interactions. Among the latter,

hydrogen bond (Figure 8.1 *e*) is usually an order of magnitude weaker than covalent, about $10k_BT$. Van der Waals interaction for atoms is even weaker, a fraction of k_BT — but one has to keep in mind that when two molecules or two monomers of a polymer approach each other, then several atoms come into contact; therefore, typical van der Waals energy for the atomic groups like monomers (which is ε in Figure 8.1 *a*) is typically several k_BT, close to hydrogen bond. Characteristic energies of hydrophobic and Coulomb interactions are more dependent on circumstances, and we will touch upon them below.

Last but not least, the energy k_BT allows us to estimate also the characteristic values of forces involved. Indeed,

$$k_BT \approx 4.1 \text{ nm} \cdot \text{pN} \quad \text{at room temperature.} \tag{8.1}$$

Notice that there is no powers of ten involved, in the chosen units k_BT is just about 4. So what are these convenient units? It is nanometer (10^{-9} m) times picoNewton (10^{-11} N), distance times force. Nanometer is a natural unit for distance, because, for instance, atom size is about 0.1 in these units, while the DNA diameter is 2. Therefore, given that k_BT is the natural scale of energy, and nanometer is natural for distances, we discover that picoNewtons is the natural scale of forces. The reader will be well advised to keep this in mind.

8.3 Excluded Volume — Formulating the Problem

Let's discuss how interactions of the type shown in Figure 8.1 *a* might influence the shape of an isolated polymer chain in a dilute solution (Figure 4.7 *a*). First of all, would volume interactions make the coil swell or shrink? This, it turns out, depends on the temperature of the solution.

Suppose the characteristic energy of attraction ε (Figure 8.1 *a*) is much greater than the thermal energy k_BT. Then attraction will dominate. As a result, the macromolecule will shrink to become more compact than an ideal coil. This is a special polymer state, called a polymer globule. We shall come back to it in the next chapter.

It is not the same story if ε is smaller than k_BT. In this case, attraction is not too important. Repulsion at shorter distances between monomers is the prevailing form of interaction. It makes the coil swell. Such swelling is called the excluded volume effect. (You presumably understand where the name comes from. As we have already said, the repulsion at short distances occurs because the volume of each monomer is excluded for all the others.)

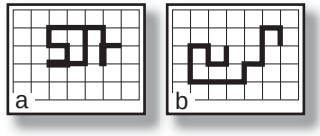


Fig. 8.2 A random walk (*a*) and a self-avoiding path (*b*) on the lattice in two dimensions (on the plane). Notice that random walk frequently retraces its own path, while self-avoiding walk never does so.

In this chapter, we are going to tackle the problem of excluded volume, that is, we shall try to picture how a polymer coil swells.

For an isolated polymer chain, the problem is purely geometrical. Indeed, the spatial shape of an ideal chain resembles the path of a randomly wandering Brownian particle (see Chapter 6). What new features will the shape of the chain acquire, if we allow for the excluded volume? Clearly, since the "private space" of each monomer is not available to the rest, the chain cannot possibly cross itself at any stage. This sort of behavior can be described as self-avoiding. For example, if there were an equivalent Brownian particle, it would not be allowed to cross its own track. A two-dimensional version of such a trajectory is sketched in Figure 8.2. Thus, we have made it a purely geometrical problem of self-avoiding random walks.

This problem can be quite successfully approached by computer simulation. The simplest way to set it up is to use a random number generator to try out various trajectories of a polymer chain (just as described in Section 2.4). Then whenever we obtain a trajectory with a self-crossing, we merely ignore it. Thus, we only keep the self-avoiding paths, and when we have enough of them, we can look at some average features. Although more sophisticated (and more efficient) algorithms are normally used these days, in principle they are not that different from what we have just described. Typical result is shown in Figure 8.3.

So what has been gathered from the computer simulations of self-avoiding walks? It appears that the conformational properties of a polymer coil are quite significantly affected by the excluded volume. The coils become looser, and the fluctuations in the segment concentration become more severe. The mean-square size of the coil increases. Moreover, the mean-square end-to-end distance $\langle R^2 \rangle$ now depends differently on the number of segments in the chain. Instead of the familiar $\langle R^2 \rangle \sim N$ (which we derived for an ideal chain in Chapter 6), we now get

$$\langle R^2 \rangle \sim N^{2\nu} \ , \quad 2\nu \approx 1.176 \approx 6/5 \text{ in } 3D \ , \tag{8.2}$$

for self-avoiding walks in three dimensions, and

$$\langle R^2 \rangle \sim N^{3/2} \ \text{ in } 2D \ , \tag{8.3}$$

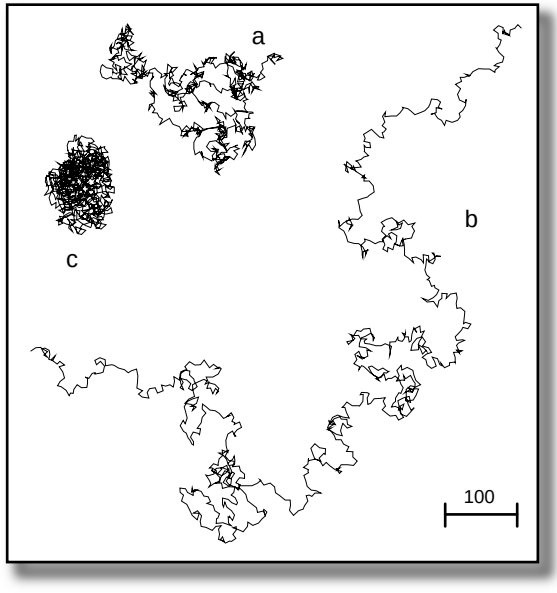


Fig. 8.3 Three types of conformations for the same polymer are shown for comparison. The polymer has 1,000 segments of length 1 each. A typical Gaussian conformation is shown in (*a*). A typical conformation of a polymer that is self-avoiding in real three-dimensional space is shown in (*b*). While the self-avoiding chain never crosses itself, its projection on the plane can (and frequently does) have crossings. This is why this figure is very different in spirit and in its meaning from the Figure 8.2. A typical collapsed globular conformation is shown in (*c*); we will discuss globules later in Chapter 9. The figure is courtesy of S. Buldyrev.

for self-avoiding walks on a plane (i.e. in two dimensions). Of course, the accurate values of indices 1.176 and 3/2 do not appear as such in the computer simulations, particularly on simple minded ones. But within a certain accuracy, the indices produced by a computer are close to these values. Relationships (8.2) and (8.3) confirm that a polymer coil with excluded volume is swollen compared to an ideal coil. You will find it handy to introduce a swelling coefficient α, such that

$$\alpha^2 = \frac{\langle R^2 \rangle}{\langle R^2 \rangle_0} , \tag{8.4}$$

where $\langle R^2 \rangle_0 = N\ell^2$ is the size of an ideal polymer chain. As N increases, α grows approximately as $N^{1/5}$ in three dimensions, and as $N^{1/2}$ in two dimensions.

8.4 The Density of a Coil and Collisions of Monomer Units

The problem of excluded volume did not succumb to the efforts of theorists for more than 20 years. The way to tackle it, or, to be more precise, the

way to reduce it to some other, better explored problems, was found by P.G. de Gennes in 1972. His solution goes far beyond what we can explain in this book. However, we do not have to jump straightaway to Equations (8.2) and (8.3). To convince you, we offer a simple explanation known as Flory's theory, although the way we are going to present it does not quite follow Flory's original version.

First of all, the mean spatial size of a coil R is obviously of the same order of magnitude as $\langle R^2 \rangle^{1/2}$. This is why we can write that $R \sim \ell N^{1/2}$ for an ideal polymer (see (6.11)). Meanwhile, for a coil with excluded volume we get $R \sim \alpha \ell N^{1/2} > \ell N^{1/2}$. The volume taken up by such a coil is $V \geq \ell^3 N^{3/2}$ (we omit the factor $4\pi/3$ as usual). However, a polymer chain never uses up all the space inside the coil. This is clearly seen in Figure 2.6. You can show it is true by the following argument. Let the volume of a single monomer segment be v. Then the total volume of the coil is Nv. Since $N \gg 1$, we have $V > \ell^3 N^{3/2} \gg Nv$. In other words, the fraction ϕ of the volume of the coil taken up by the monomer segments is really very small:

$$\phi \sim \frac{Nv}{V} < \frac{Nv}{\ell^3 N^{3/2}} \sim N^{-1/2} \left(\frac{v}{\ell^3} \right) \ll 1 \ . \tag{8.5}$$

(In Section 6.6, we used the same sort of argument for an ideal polymer.) The same thing can be said about the mean concentration of the segments in the coil, $n \sim N/V \sim \ell^{-3} N^{-1/2}$ (cf. (6.14)). At first glance, you may think it implies that a polymer with excluded volume is always ideal. Indeed, if the segment concentration is so low, their encounters are very rare, and one can be tempted to neglect them. On the other hand, we know that the coil is very pliable, and its elastic modulus is small.

This suggests that the question should be treated with more subtlety. Let's make a crude estimate of how many encounters (i.e., collisions) between the segments of the coil may occur at the same time. Assume that the coil is a cloud of totally independent particles (segments) spread over the volume V. It would be wonderful if we could take a three-dimensional photo of this cloud. We would then be able to count all the collisions between two, three or more bodies, caught at a moment in time.

Unfortunately, we cannot do this, so we have to use another approach. There are N particles all together. The probability that each particle has a close "partner" is ϕ. This is why the number of pair collisions is of order $N\phi$. In the same way, the number of three-body collisions is roughly $N\phi^2$, and so on. In general, the number Y_p of p-body collisions can be estimated

as $N\phi^{p-1}$. From (8.5),

$$Y_p \sim N\phi^{p-1} < N^{(3-p)/2} \left(\frac{v}{\ell^3}\right)^{p-1} \quad . \tag{8.6}$$

You can see that $Y_p \ll 1$ if $p > 3$. This indicates that many-body collisions are really rare. Even the number of three-body collisions in a swollen coil is of order 1. So they cannot seriously affect the conformation of the coil. In contrast, the number of simultaneous pair collisions is about $N^{1/2}$. This is much less than N (so each particular segment seldom has a collision), yet it is a large number compared to 1.

Besides, as we showed in Chapter 7, a long polymer chain is very pliable, and its elastic constant is small ($\sim 1/N$, see (7.21)). Therefore, we have the right to suspect the pair collisions of making the polymer swell in the way implied by (8.2) and (8.3).

What is the free energy of a polymer like, given the excluded volume interactions? (See formula (7.19) for the definition of free energy.) Of course, it has the usual entropy term $-TS$ (which would be the only term in the case of an ideal gas or an ideal polymer). In addition, it includes the internal energy U of the segment interactions. This latter term is responsible for the swelling. In other words, it accounts for the excluded volume effect. All we need to know now is the contribution of the binary collisions to the internal energy U of the coil.

Here is how you can find it. The segment density n is very low, as we have seen. So U can be expanded as a series of powers of n:

$$U = Vk_BT\left[n^2B + n^3C + \ldots\right] \ , \tag{8.7}$$

where V is the volume of the coil, and B and C are expansion coefficients, or virial coefficients (i.e., B is the second virial coefficient, C is the third, and so on). These coefficients are fully determined by the form of the interaction potential $u(r)$ and the temperature T. Obviously, the first term in expansion (8.7) stands for the binary interactions. This is because it is proportional to n^2, which is just the pair collision probability. Likewise, the second term is related to three-body interactions, and so on[1].

[1]Here is a useful leisure time exercise for a very attentive reader. The purpose is to understand the connection between virial expansion (8.7) and the well known van der Waals equation of state (i.e., the relationship between volume, pressure, and temperature) for an ordinary imperfect gas. You may have studied van der Waals equation in general physics and/or general chemistry class, it reads $(p+a/V^2)(V-b) = Nk_BT$. Say, the volume is V, and the number of molecules in the gas is N. Then $n = N/V$. You can work out the pressure by differentiation: $p = -(\partial F/\partial V)$, where free energy F is defined by formula (7.19), $F = U - TS = U + U_{\text{eff}}$, the internal energy U is given by (8.7), and

Notice that B has the units of volume, while C has the units of volume squared. Thus, the energy of all the binary interactions between the segments of a coil is:

$$U = V k_B T n^2 B \ , \tag{8.8}$$

where n is the average segment density in the coil (the number of segments per unit volume inside one coil).

8.5 Good and Bad Solvents, and θ Conditions

We have already discussed the potential $u(r)$ in Figure 8.1 *a*. Repulsion between the segments dominates at higher temperatures ($\varepsilon \ll kT$) (the excluded volume effect), whereas at lower temperatures ($\varepsilon \gg kT$) attraction takes over. Let's look at the higher temperature region first. The most important values of r are those where $u(r) > 0$. So the internal energy of the coil (as well as the second virial coefficient) is positive. In contrast, at lower temperatures it is the "attractive" part of $u(r)$, where $u(r) < 0$, that gives the biggest contribution. So the internal energy of the coil U and B are both negative. In the former case we say that we are dealing with a good solvent, and in the latter case with a bad one. We are not being biased! If in a solvent the segments of polymer chains tend to repel each other, the polymer will dissolve. Conversely, if the segments attract each other, the polymer chains will be rather "sticky"; in other words, they will stick together and precipitate out rather than dissolve.

The quality of a solvent (i.e., whether it is good or bad) may change with its contents or with temperature. Hence, there has to be a special

entropy S or $U_{\text{eff}} = -TS$ can be thought to be the same as for an ideal gas. Here, we should warn the reader: we never in this book wrote the relation for ideal gas entropy, we only wrote formula for the *change* of this quantity between two states of different volumes but the same numbers of particles, this was formula (7.13) — pay attention to Δ on both sides of that formula! In fact, the corresponding formula in general should read $U_{\text{eff}} = -k_B T \ln(eV/N)^N$. The extra term $N \ln N/e$ arises because particles in the gas are identical; in this book, we do not discuss it, because monomers in the polymer chain are *not* identical, each of them has its own place along the chain! Besides, this extra term is also not relevant for the determination of pressure, for it cancels away upon differentiation. Thus, using equations above you can derive an equation of state. Check that this leads to the ideal gas equation $pV = Nk_BT$ if one takes $B = 0$ and $C = 0$. For non-zero B and C, compare your answer with the van der Waals equation of state; they are not the same, but prove that your equation, just like van der Waals one, indicates a single value of volume (or density) for any pressure at high enough temperatures (e.g., $B > 0$ and $C > 0$), but predicts two possible stable densities (i.e., phase segregation) in a certain interval of pressures at lower temperature. Good luck!

point where the second virial coefficient goes through zero: $B = 0$. It is usually called the θ-point (or θ-temperature — obviously this is the temperature when $B = 0$). At the θ-point, attraction and repulsion between the segments completely cancel out, and the behavior of the polymer becomes ideal. When $T > \theta$, repulsion dominates. This is the excluded volume (and good solvent) region. In contrast, when $T < \theta$ attraction prevails, making the solvent bad. We can now rephrase our initial problem. The swelling of a polymer due to the excluded volume effect is the same as the swelling of a polymer in a good solvent, that is, at $T > \theta$.

You may wonder why such θ-conditions are possible in the first place. Is it a mere coincidence that at a certain point repulsion and attraction are so perfectly balanced? For instance, such balancing, or compensation, never quite happens in a real gas. Historically, Boyle found that his law ($pV =$ const for a gas at fixed temperature) is followed at some temperatures more accurately than at others, but never quite perfectly; in modern language, we can say that the gas should be close to ideal at the temperature (called Boyle's point) when $B = 0$, but it is not quite ideal because $C = 0$. By contrast, compensation between attraction and repulsion is indeed nearly perfect for a polymer coil. Why? The answer is that the cancelation only works because three-body interactions (and all the higher ones) are not important. Their contribution to U is always very small. As for the binary collision term (8.8), it is proportional to B, so it falls to zero at the θ-point. Hence, all that really remains of the free energy F at $T = \theta$ is the entropy term (see (7.19)). This is why the coil's behavior becomes ideal.

Thus, the existence of the θ point (where the segment interactions have no influence on the shape of the chain) is yet another peculiarity of polymers. It is all to do with the very low segment concentration $n \sim N^{-1/2}$.

8.6 The Swelling of a Polymer Coil in a Good Solvent

Let's consider an isolated polymer coil in a good solvent ($B > 0$), and try to find its swelling coefficient α. The first calculation of this sort was done by P. Flory in 1949. His approach was as follows. The main cause of the swelling is repulsion between the segments inside the coil (the binary collisions). However, there is also an effect that hinders swelling, arising from the elastic forces whose origin is due to entropy (we discussed them in Chapter 7). These forces emerge because there are fewer different shapes

that the chain can take when it is straightened out (or swollen). So Flory's idea was to obtain the swelling coefficient α from a balance condition between the repulsive and elastic forces.

Both factors contribute to the free energy of a swollen polymer coil (with swelling coefficient α), $F(\alpha) = U(\alpha) - TS(\alpha)$ (see Section 7.8, Equation 7.19). The potential energy term $U(\alpha)$ is determined by the repulsive interactions (see (8.8)):

$$U(\alpha) = Vk_BTn^2B \sim \frac{k_BTR^3BN^2}{R^6} \sim \frac{k_BTBN^{1/2}}{\ell^3\alpha^3} \ . \tag{8.9}$$

To write (8.9), we used the following straightforward relationships: $V \sim R^3$, $n \sim N/R^3$, $\alpha = R/R_0 = R/N^{1/2}\ell$, where $R_0 \sim N^{1/2}\ell$ is the size of an ideal polymer coil. As usual, we leave out numerical factors as we are only making estimates. Likewise, the entropy term $S(\alpha)$ in the free energy of the swollen coil is in one-to-one correspondence with elastic forces. We can work it out from (7.4):

$$S(\alpha) = \text{const} - k_B\frac{3R^2}{2N\ell^2} = \text{const} - k_B\frac{3N\ell^2}{2N\ell^2}\alpha^2 = \text{const} - \frac{3}{2}k_B\alpha^2 \ . \tag{8.10}$$

Thus, the free energy $F(\alpha)$ should be obtained by combining (8.9) and (8.10). In doing so, we should be careful about the fact that we left out (and do not know) the numerical coefficient, e.g., in formula (8.9) for internal energy. To make sure that this unknown coefficient does not corrupt all our theory, let's denote it by some letter, say γ, and introduce it to the formula (8.9), replacing then the sign $\sim$ with definitive $=$; then we can write

$$F(\alpha) = U(\alpha) - TS(\alpha) = \text{const} + \gamma\frac{k_BTBN^{1/2}}{\ell^3\alpha^3} + \frac{3}{2}kT\alpha^2 \ , \tag{8.11}$$

where "const", as before, is independent of α.

The function $F(\alpha)$ is sketched in Figure 8.4. You can see a minimum in the curve at a certain α. The minimum of the free energy always gives the equilibrium state. So the equilibrium swelling coefficient is just the value of α at the minimum. Notice that the unknown numerical value of the coefficient γ does not affect the qualitative shape of $F(\alpha)$ curve, but does somewhat affect the value of α in the minimum.

Can we find out where exactly the minimum is? The usual way is to differentiate $F(\alpha)$ with respect to α, and to set the derivative equal to zero. We obtain

$$\frac{\partial F}{\partial \alpha} = -3\gamma\frac{k_BTN^{1/2}B}{\ell^3\alpha^4} + 3k_BT\alpha = 0 \ . \tag{8.12}$$

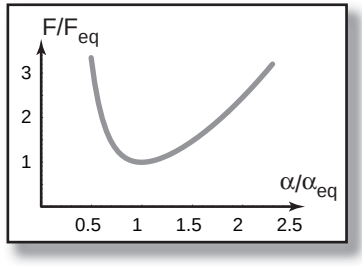


Fig. 8.4 The dependence $F(\alpha)$ given by Equation (8.11). In order to present the most universal plot possible, independent on the values of parameters, such as N, B, and ℓ, we re-write Equation (8.11) in the following way, using our result (8.13) for the equilibrium value of α: $F = \frac{5}{2}\left(\frac{\gamma B}{\ell^3}\frac{\overline{N}}{}\right)^{2/5}\left[\frac{2}{5}\left(\frac{\alpha_{eq}}{\alpha}\right)^3 + \frac{3}{5}\left(\frac{\alpha}{\alpha_{eq}}\right)^2\right]$ where we also dropped the irrelevant additive constant. In this form, the F/F_{eq} ratio becomes a universal function of the α/α_{eq} ratio; this function is plotted in the figure.

From here, the equilibrium value of α is

$$\alpha_{eq}^5 = \gamma\frac{BN^{1/2}}{\ell^3} \qquad \text{or} \qquad \alpha_{eq} \sim \left(\frac{N^{1/2}B}{\ell^3}\right)^{1/5} . \tag{8.13}$$

In the latter formula we returned to the use of the sign $\sim$ (i.e. "of order of"): we dropped again the numerical factor γ, because it does not affect the most important feature of the result, namely, the dependence of the equilibrium value of swelling parameter α on the number of monomers N and on the properties of monomers B/ℓ^3. We would not be able to find γ anyway within this simple theory. All that Flory's theory really gives are the power indices in equations like (8.2). Indeed, from (8.13), the size R of the coil is estimated as:

$$R \sim \alpha R_0 \sim \alpha N^{1/2}\ell \sim \ell N^{3/5}\left(\frac{B}{\ell^3}\right)^{1/5} . \tag{8.14}$$

The theoretical result $R \sim N^{3/5}$ agrees reasonably well with the outcome (8.2) of computer simulations.

In the future, we will ignore coefficients similar to γ when combining various contributions to the free energy.

Using a similar method, we can also explain Equation (8.3) which is the two-dimensional equivalent. Fortunately, the expression (8.10) for $S(\alpha)$ remains the same. As for $U(\alpha)$, we need to be careful. The two-dimensional "volume" is not $V \sim R^3$ as usual, but $V \sim R^2 \sim \alpha^2 N\ell^2$. Therefore, instead of (8.9) we arrive at:

$$U(\alpha) \sim \frac{k_B T R^2 B N^2}{R^4} \sim \frac{k_B T B N}{\ell^2\alpha^2} . \tag{8.15}$$

The free energy is calculated as $F(\alpha) = U(\alpha) - TS(\alpha)$, where $U(\alpha)$ and $S(\alpha)$ are given by formulas (8.15) and (8.10), respectively. This time around, we

don't have to worry about introducing the γ coefficient, because we know by experience that it will be dropped out at the end anyway. Therefore, we simply work out where $F(\alpha)$ reaches a minimum, using the same idea as before. The answer is:

$$\alpha \sim \left(\frac{BN}{\ell^2}\right)^{1/4} . \tag{8.16}$$

Finally,

$$R \sim \alpha R_0 \sim \alpha N^{1/2}\ell \sim \ell N^{3/4}\left(\frac{B}{\ell^2}\right)^{1/4} , \tag{8.17}$$

in total agreement with (8.3). In fact, it turns out that the result of Flory theory happens to be exact in two dimensions, while in three dimensions it is only approximate, even if reasonably accurate.

Thus, if we allow for the excluded volume effect, the average size of a polymer coil will no longer be proportional to $N^{1/2}$ (as for an ideal chain), but to $N^{3/5}$ in three dimensions, and to $N^{3/4}$ in a plane. So, just as we expected, the excluded volume effect is quite significant. This is despite the coil being extremely loose, and the collisions between the segments being very unlikely. Indeed, if N the swelling coefficient also grows without limit. The analogy with a Brownian particle no longer makes sense for a swollen coil. If a Brownian particle were not allowed to cross its own path, it would move much further away from its starting point in a given time.

8.7 The Excluded Volume Effect in a Semi-Dilute Solution

As we discussed in Section 4.6, isolated polymer coils are typical for dilute solutions, where the volumes taken up by the coils do not overlap (Figure 4.7 *a*). Things change when the polymer concentration exceeds the threshold value $c^\star$ (which is defined by Equation (6.14) for an ideal polymer). In this case we have a semi-dilute solution (Figure 4.7 *c*). Although the fraction of the volume taken up by the polymer is still rather small, the coils are already highly intermingled. Can we work out what the excluded volume effect does to the coils in this case (i.e., when the polymer concentration $c \gg c^\star$)?

First of all, we need to know the value of $c^\star$ for a semi-dilute solution. When $c = c^\star$, the average segment density in the whole solution becomes equal to the average segment concentration inside each coil (see Section 6.6).

Therefore, taking into account the excluded volume, we get:

$$c^\star \sim \frac{N}{R^3} \sim \ell^{-3} \left(\frac{B}{\ell^3}\right)^{-3/5} N^{-4/5} \ . \tag{8.18}$$

Since $N \gg 1$, the threshold density $c^\star$ is fairly small (just as it was for an ideal polymer solution; see (6.14)). Thus the semi-dilute regime is appropriate for a wide range of concentrations.

In the same way as for a dilute solution, we can describe the swelling of the coils by means of $\langle \mathbf{R}^2 \rangle$. (As always, $\mathbf{R}$ is the end-to-end vector of the coil.) Obviously, the average size of the coil R is estimated as $R \sim \langle \mathbf{R}^2 \rangle^{1/2}$. Now we need to calculate the value $\langle \mathbf{R}^2 \rangle$ in a semi-dilute solution.

Choose a certain polymer chain and fix in space one of its monomer units. Now look at a strand of the chain near the fixed unit. Suppose this strand contains g effective segments. If there were no other chains around, the excluded volume effect would swell the strand roughly to a size $\ell g^{3/5}(B/\ell^3)^{1/5}$ (see (8.14)). The volume taken up by such a g-strand would be of order $\left[\ell g^{3/5}(B/\ell^3)^{1/5}\right]^3 \sim \ell^3 g^{9/5}(B/\ell^3)^{3/5}$. The monomer concentration in this volume would be estimated as $g/\left[\ell^3 g^{9/5}(B/\ell^3)^{3/5}\right] \sim \ell^{-3} g^{-4/5}(B/\ell^3)^{-3/5}$. It decreases as g is increased. This is not surprising. Since the neighbors are tied to the fixed monomer with a piece of polymer chain, there is a sort of "correlational" thickening round this area. (We call it "correlational" because it results from interactions or correlations between the monomers in the chain.) We must not forget that this strand is not alone. It is in an ocean of intermingled, overlapping polymer chains, with monomer concentration c. Do the surrounding chains manage to penetrate into the densest region near the fixed monomer? The answer is no. There is just no room, since the monomers cannot go through each other (the excluded volume effect). In this region, the correlational density is higher than c. We can find the size $\xi^\star$ of this region (i.e., where there is correlational thickening), and the number of monomers in it $g^\star$ from the following conditions: $\ell^{-3}(g^\star)^{-4/5}\ (B/\ell^3)^{-3/5} \sim c$, and $\ell(g^\star)^{3/5}(B/\ell^3)^{1/5} \sim \xi^\star$. Hence,

$$\xi^\star \sim \ell(c\ell^3)^{-3/4} \left(\frac{B}{\ell^3}\right)^{-1/4} \qquad ; \qquad g^\star \sim \left(c\ell^3\right)^{-5/4} \left(\frac{B}{\ell^3}\right)^{-3/4} \ . \tag{8.19}$$

Note that (8.18) and (8.19) lead to $g^\star < N$ provided that $c > c^\star$.

Now that we have a clearer picture, let's draw some conclusions. In the case of a semi-dilute solution ($c > c^\star$), it helps if we divide each chain into a sequence of strands, or blobs, of a certain length $g^\star$. Each blob,

taken separately, looks like a normal isolated polymer chain, swollen by the excluded volume effect. The size $\xi^\star$ of the blob provides an important length scale. That is, there is no correlation between the monomers of the chain at distances longer than $\xi^\star$. You could say that, due to the excluded volume effect, the "inner lives" of neighboring blobs are screened from each other. So a blob in a semi-dilute solution does not really "care" whether the neighboring blobs belong to the same chain as itself, or not. Now, this is interesting. Let's zoom out, and look at a chain at a less detailed scale. We shall see a sequence of blobs. What if we regard the blobs as new, bigger monomer units? The chain of blobs will behave as an ideal one, so we can apply the usual theory of a Gaussian chain. The number of blobs in the chain is about $N/g^\star$, and the size of each of them is of order $\xi^\star$. Therefore, we have:

$$R \sim \left\langle \mathbf{R}^2 \right\rangle^{1/2} \sim \xi^\star \left(\frac{N}{g^\star}\right)^{1/2} \sim \ell N^{1/2} \left(c\ell^3\right)^{-1/8} \left(\frac{B}{\ell^3}\right)^{1/8} . \qquad (8.20)$$

Figure 8.5 gives an idea of how R depends on the concentration c of the solution. When $c < c^\star$, the size of the coil is not affected by c; it is simply described by Equation (8.14). If c is increased, we shall reach the regime $c > c^\star$, that is, a semi-dilute solution. Here the swelling coefficient starts dropping, just as (8.20) predicts. By the way, for $c \sim c^{\star\star} \sim B/\ell^6$, Equation (8.20) leads to $R \sim N^{1/2}\ell$. Thus, when the density becomes very high, the excluded volume effect no longer causes swelling. This conclusion refers, in particular, to polymer melts, where there is no solvent at all and c

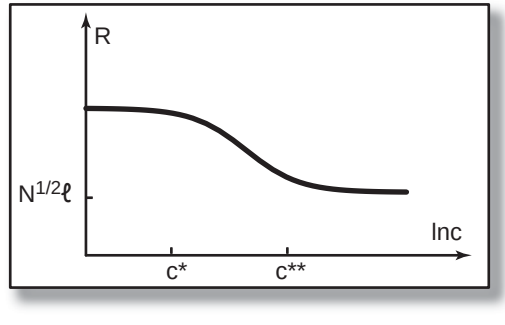


Fig. 8.5 A sketch of the dependence $R(c)$, where R is the size of the coil and c is concentration. Notice that concentration on the abscissa must be in the Log scale, otherwise the region of dilute solutions, $c < c^\star \sim c^{\star\star} N^{-4/5} \ll c^{\star\star}$ would look so narrow as not to be visible at all. The interval between $c^\star$ and $c^{\star\star}$ corresponds to semi-dilute solution.

reaches the highest possible value. If you still remember the Flory's theorem (Section 7.10), you will not be surprised by this.

8.8 The Near Immiscibility of Polymer Blends

We have a "debt" left over from Section 4.7: we have not yet proved that, if contact between monomers of types $\mathcal{A}$ and $\mathcal{B}$ in a polymer $\mathcal{A}$/polymer $\mathcal{B}$ blend is even slightly energetically unfavorable, phase separation into almost pure $\mathcal{A}$ and $\mathcal{B}$ phases will occur. Now, at last, we are equipped with all we need to prove this.

Let's take a mixture of $\mathcal{A}$ and $\mathcal{B}$ monomers that are not linked into chains. Can we compare the phase separation in this mixture (Figure 8.6 *a*) with that in a polymer blend (Figure 8.6 *b*)? In each, the number of energetically unfavorable $\mathcal{A}-\mathcal{B}$ contacts drops dramatically during the phase separation into almost pure $\mathcal{A}$ and $\mathcal{B}$ phases. These contacts may only take place along the surface separating the two phases. We can say that the gain in energy due to the phase separation is the same in both cases.

However, the phase separation not only gains energy, but also leads to a loss of entropy. This is because the number of possible conformations of the system is lessened. When mixed, $\mathcal{A}$ and $\mathcal{B}$ molecules had access to the whole volume. Once separated, they can only reach a part of it (compare Section 7.7). Are these losses in entropy the same for Figures 8.6 *a* and *b*? The answer is "no" — due to how many possible conformations can be realized in either case. Obviously, this number is many orders of magnitude greater in a low molecular weight mixture of $\mathcal{A}$ and $\mathcal{B}$ than in a polymer blend. Unattached monomers in the mixture can move independently of

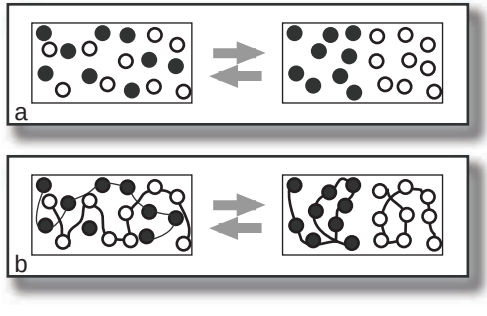


Fig. 8.6 A cartoon illustrating phase separation in (*a*) a low molecular weight mixture and (*b*) a polymer blend. $\mathcal{A}$ and $\mathcal{B}$ components are shown with empty and shaded circles, respectively.

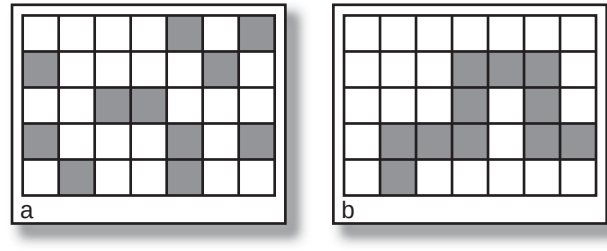


Fig. 8.7 Two examples of how to arrange N molecules on V cells of a square lattice: N independent molecules (*a*); N molecules linked together into a chain (*b*).

each other, whereas in the case of the polymer blend they are linked into chains.

To give an illustration, we will calculate the number of possible conformations of N molecules on a square lattice containing V cells (Figure 8.7). We shall consider two cases: N independent molecules, each taking up just one cell (Figure 8.7 *a*), and N molecules linked together into a chain (Figure 8.7 *b*).

In the first case, the answer is obvious. There are V possibilities for each molecule. (Let's assume that $N \ll V$, and neglect the excluded volume of the molecules.) Therefore, the total number of conformations is V^N.

Now, let's look at the second case. The first monomer unit can be positioned in V different ways, the second one in four ways (i.e., in the cells sharing edges with the first), the third and all the following ones in three ways (Figure 8.7 *b*). Thus, we get $4V \cdot 3^{N-2}$ ways altogether.

Obviously, if $N \gg 1$,

$$V^N \gg 4V \cdot 3^{N-2} . \tag{8.21}$$

This proves that the number of conformations is much less, and therefore the entropy is much lower for the polymer system.

This is why the entropy losses due to the phase separation are much smaller in a polymer blend (Figure 8.6 *b*) than in the corresponding low molecular weight system (Figure 8.6 *a*). However, the energy gain is the same (see above). Thus, it is the energy that dominates. Even a very weak repulsion between $\mathcal{A}$ and $\mathcal{B}$ monomers is enough to compensate for some minor entropy losses due to the phase separation. How weak can the repulsion be? Suppose you mix two polymers, $\mathcal{A}$ and $\mathcal{B}$, containing N units each. Theory shows that the contact energy ε between them, which is sufficient to induce the phase separation, can be estimated as:

$$\varepsilon \sim \frac{k_B T}{N} . \tag{8.22}$$

Clearly, for large N, this value is really very low.

Chapter 9

Coils and Globules

His kingdom was very small, but still
quite large enough to marry upon.

Hans Christian Andersen,
The Swineherd

He may be only little,
But he's a good boy.

V. Mayakovsky,
What Is Good And What Is Bad? (Russian children's poem)

9.1 What is a Coil-Globule Transition?

In the previous chapter, we focused on the excluded volume problem. We learned to appreciate that each monomer has a certain volume, and the monomers cannot penetrate each other. This leads to repulsion at short distances. In the case of a good solvent, repulsion is the prevailing tendency overall, so the polymer coils swell. But what if the quality of the solvent grows worse? For example, you could add some precipitant into the solvent, or change the temperature. As a result, the solvent may go through the θ point (as we discussed in Section 8.5) and the binary interactions between the monomers will become mainly attractive. Segments will tend to stick to each other from time to time, so there will be lots of temporary couples. What will this do to the coil as a whole?

W. Stockmayer (1914–2004) in Dartmouth College in New Hampshire was the first to predict, in 1959 that, if the attraction between monomers

becomes strong enough, the polymer undergoes a phase transition of the same sort as the transition from gas to liquid. Bits of the polymer "condense on to themselves", and instead of a loose coil you end up with a dense "drop" — a polymer globule. This is just what is meant by the name coil-globule transition.

A typical globule with strongly interacting monomers is portrayed in Figure 8.3 *c*. Compare it with Gaussian (*a*) and swollen (*b*) coils shown in the same figure. These images were obtained by computer simulation. How is attraction modeled on a computer? First, you can figure out the attractive potential energy of the binary interactions (of the same kind as shown in Figure 8.1 *a*). Then you will know the forces of interaction. Then you can use Newton's laws and trace, on a computer, how the chain would move under these forces. You will see some shapes similar to Figure 8.3 *c*. In contrast to the coil, the globule is very dense and compact, and there are no vast "holes" inside it. The only real difference between a globule and an ordinary liquid is that the globule's "molecules" (i.e., the monomers) are all linked together.

Another image of a globule, also produced by computer simulation, is shown in the Figure C9.1. One can notice that the structure is overall spherical, its surface layer consists of some loops, while the interior is quite dense in the sense, that most spatial neighbors of every monomer belong to the very distant parts of the chain. Fundamental difference between coiled and globular states of the polymer is also revealed by considering their fractal properties, as we will discuss in the Section 13.4.

Polymer globules and coil-globule transitions came in from the cold thanks to molecular biophysics. Some of the most important biological polymers — protein enzymes — usually appear in living cells in globular form (we mentioned them in Chapter 5). If something nasty happens to the solvent surrounding the proteins (say, it gets overheated, or the contents are disturbed), the proteins may be denatured. In other words, they lose all their biochemical activity. Denaturation of proteins usually implies a dramatic change in shape, and is accompanied by a strong absorption of heat. The first scientists who worked on the coil-globule transition were inspired by the thought that it might shed some light on the denaturation of proteins. It seemed quite plausible that when denaturation occurred, the dense globular structure is destroyed, and the protein takes on the shape of a coil.

Only later did it turn out that there is no straightforward analogy between the coil-globule transition and protein denaturation. However, the

coil-globule transition appeared to be quite extraordinary and exciting in its own right. This stimulated further interest in the problem. The studies have expanded, covering all kinds of polymers. A globular state has been discovered for many other systems. Not only proteins, but also DNA molecules and macroscopic polymer networks, for example, can have a globular structure. This explains quite a few unusual polymer effects, such as the existence of the compact form of DNA, and the so-called collapse of polymer networks. We shall talk about some of them a little later. This very broad view of the coil-globule transition was initiated in 1968 by the pioneering work of the Russian physicist I.M. Lifshitz.

9.2 The Free Energy of a Globule

Let's now take an isolated polymer molecule, and try to build a simple theory for the coil-globule transition. For our purposes, we do not need to go into the details of Lifshitz's consistent approach. Instead, we will stick to the same logic as when we used Flory's approximate arguments to tackle the excluded volume problem (see Section 8.6). As you remember, we introduced the swelling coefficient α; the free energy of a swollen coil was written as the sum of two terms (see (8.11)). One term was the free energy associated with stretching the coil by the factor α, $U_{\text{eff}} = -TS(\alpha)$. The other term was the energy of the monomer interactions in the coil, $U(\alpha)$. Then we found the equilibrium value of α; it was the α corresponding to the minimum of the free energy $F(\alpha)$.

Where did the term $U_{\text{eff}}(\alpha)$ come from? It has to do with the poorer choice of shapes that a straightened polymer can take. Fewer possibilities means lower entropy, and a lower probability of the elongated state (see Section 7.5). Meanwhile, we have already said that the other term, $U(\alpha)$, is the energy of the monomer interactions. Thus, when we write the free energy in the form (8.11), we automatically distinguish its entropy and energy parts. Notice that we never use the fact that $\alpha > 1$. It applies just as well if the molecule shrinks ($\alpha < 1$) instead of swells ($\alpha > 1$). We still have the free energy in the same form:

$$F(\alpha) = U_{\text{eff}}(\alpha) + U(\alpha) \ . \tag{9.1}$$

Here $U_{\text{eff}}(\alpha)$ is determined by the entropy of the final state of the coil, when it is either swollen or shrunk by the factor α. (Although in the case of shrinking, α is less than one, we would like to keep its previous name,

the "swelling coefficient". After all, its mathematical definition (8.4) is still the same.)

9.3 The Energy of Monomer Interactions

If we want to cater for both regimes — a swollen coil ($\alpha > 1$) and a shrunken globule ($\alpha < 1$), we need to express the two terms in (9.1) in a slightly different fashion than Section 8.6. Consider the energy $U(\alpha)$ first. We used to estimate it, as in (8.9), by taking into account only the binary interactions between monomers (described by the second virial coefficient B). We had the right to do this, since the monomer density is very low in both ideal and swollen polymers. However, when a polymer shrinks (i.e., $\alpha < 1$), the monomer density goes up. Many-body collisions may now become important. This is why we can no longer stop after just the first term in expansion (8.7). Let's see what we gain if we keep the second term too. We shall have:

$$\begin{aligned} U(\alpha) &\sim R^3 k_B T \left(Bn^2 + Cn^3\right) \\ &\sim R^3 k_B T \left[B\left(\frac{N}{R^3}\right)^2 + C\left(\frac{N}{R^3}\right)^3\right] \\ &\sim k_B T \left[\frac{BN^{1/2}}{\alpha^3 \ell^3} + \frac{C}{\alpha^6 \ell^6}\right] \end{aligned} \tag{9.2}$$

(cf. (8.9)), where $R \sim \alpha \ell N^{1/2}$ is the size of the molecule, and C is the third virial coefficient, which represents three-body interactions. (In estimate (9.2), as before, we have dropped all numerical coefficients of order of one.) Will it do just to include the three-body interactions and ignore the rest? The answer is yes. If we did a more detailed calculation, we would see that the first two terms in expansion (8.7) are enough to give a correct account of the coil-globule transition. In principle, higher terms would be needed to describe the actual state of a dense globule. However, as we shall show later, a globule tends to swell before turning into a coil. So during the actual transition a globule's density is not so high.

What are the signs of the two terms in (9.2) in the transition region? Since the transition can only happen in a bad solvent (i.e. below the θ-temperature), the second virial coefficient $B < 0$, and the binary interactions are mainly attractive. As for the third virial coefficient C, it turns out normally that $C > 0$ in the transition region. So repulsion is the predominant type of three-body collision. In general, the higher the

order of interaction, the wider the range where they are effectively repulsive. Roughly, we can explain this in the following way. Suppose a particle (or a monomer) interacts with a clump of m other particles. The excluded volume for this particle will be proportional to m. However, attraction emerges only in the surface layer. The volume of this layer is proportional to $m^{2/3}$. This is why, as long as m is large enough, repulsion will always prevail. It is actually quite fortunate — otherwise no substance would be stable, and all things would shrink without limit. To conclude, the energy $U(\alpha)$ can indeed be approximated by (9.2); in particular, it works in the case $B < 0$, $C > 0$, which we are considering in this chapter.

9.4 The Entropy Contribution

Now let's concentrate on the entropy contribution $U_{\text{eff}}(\alpha)$ to the free energy. In the case of a swollen coil, this contribution was described by Equation (8.10) resulting from (7.5). Would it be valid for $\alpha < 1$ as well? Let's think. Equation (7.5) gives the free energy of an ideal coil whose end-to-end distance is of order $R \sim \alpha \ell N^{1/2}$. This is the only condition on the coil's shape that we used when deriving (7.5). Now suppose we have a shrinking coil ($\alpha < 1$). In this case, the number of available conformations is reduced not only due the fixation of the end-to-end distance, but also because the whole coil has to be fitted into a volume of linear size R (see Figure 9.2 *a* and *b*). Hence, Equation (7.5) and its consequence (8.10) would be no good in the case $\alpha < 1$. It would seriously underestimate the entropy loss.

So how can we find a reasonable estimate for $U_{\text{eff}}(\alpha) = -TS(\alpha)$? Let's look at the Boltzmann equation (7.2). It suggests that the entropy (as well as the entropy loss) should not depend on the actual cause of the

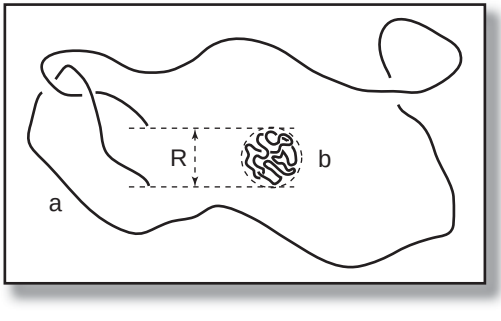


Fig. 9.2 (*a*) — An ideal polymer chain with a short end-to-end distance *R*; (*b*) — A polymer chain squashed to the size *R* in all directions.

shrinking. This is good news. We no longer have to worry about a real polymer with some attractive forces between the monomers. We might as well just consider an ideal chain with no interactions between the monomers whatsoever. All we need to do is to imagine that this ideal chain has been squeezed into a cavity the sides of which are of length $R = \alpha \ell N^{1/2}$.

Take a certain monomer unit of the chain. Suppose it is currently far away from the cavity walls. What can we say about a strand of the chain near the chosen monomer? It does not seem to "know" anything about the surroundings. It can "sense" neither the walls of the cavity (since they are far away), nor the presence of other bits of the chain (since the chain is ideal). Therefore, this strand merely acts as a Gaussian coil. If its length is g, then its size will be about $ag^{1/2}$. Of course, this would only be right if $ag^{1/2} < R$, that is, $g < g^\star \sim (R/a)^2$.

Now let's look at a $g^\star$-strand. The monomers deep inside it can arrange themselves in any sort of shape. Therefore, they do not contribute to the entropy (since they do not reduce the choice of possible conformations of the chain). On the other hand, the ends of the $g^\star$-strand must be near the walls, even though they cannot leave the cavity. This restricts the number of possible conformations, so that each end loses a bit of entropy, of order k_B. (To see this suppose the number of conformations Ω of a monomer segment in (7.2) drops by half. Then the entropy decreases by $k_B \ln\, 2 \approx\, 0.76\, k_B$). In a chain of N monomers, there are $N/g^\star$-strands altogether. So the loss of entropy is

$$U_{\text{eff}}(\alpha) = -TS(\alpha) \sim k_B T \frac{N}{g^\star} \sim k_B T \frac{N\ell^2}{R^2} \sim k_B T \frac{1}{\alpha^2} \ . \tag{9.3}$$

For comparison, let's remember what we had for a swelling polymer, (i.e., for $\alpha >\ 1$) (see (8.10)):

$$U_{\text{eff}}(\alpha) \sim k_B T \alpha^2 \qquad \text{when} \qquad \alpha > 1 \ . \tag{9.4}$$

In (9.4) we left out a constant term (independent of α), since it does not affect any physically measurable quantities.

Now we know how the function $U_{\text{eff}}(\alpha)$ looks when $\alpha < 1$ or $\alpha > 1$. Can we use this to figure out the form of $U_{\text{eff}}(\alpha)$ in the intermediate regime $\alpha \sim 1$? Since we only want a qualitative answer, we can just do a simple interpolation (as first suggested by T.M. Bistein and V. Pryamitsyn in St. Petersbourg):

$$U_{\text{eff}}(\alpha) \sim k_B T \left(\alpha^2 + \alpha^{-2}\right) \ . \tag{9.5}$$

Estimate (9.5) gives the right result for $\alpha \ll 1$ and $\alpha \gg 1$, and is approximately correct (to an order of magnitude) at $\alpha \sim 1$. By the way, the

function (9.5) has a minimum at $\alpha = 1$. This is not surprising. Indeed, if there are no excluded volume interactions, the chain has to behave as if ideal; in other words, when $U(\alpha) \equiv 0$, and so $F = U_{\text{eff}}(\alpha)$, which, therefore, must have a minimum at $\alpha = 1$.

9.5 The Swelling Coefficient α

Now we can summarize what we have learned. Based on (9.1), (9.2), and (9.5), we can write the free energy $F(\alpha)$ of a polymer molecule of size $R = \alpha N^{1/2}\ell$ taking into account both attractive and repulsive volume interactions:

$$F(\alpha) \sim k_B T \left[\alpha^2 + \alpha^{-2} + \frac{2}{3}\frac{x}{\alpha^3} + \frac{1}{3}\frac{y}{\alpha^6}\right] . \tag{9.6}$$

Here $x \equiv \gamma_1 B N^{1/2}/\ell^3$ and $y \equiv \gamma_2 C/\ell^6$, where γ_1 and γ_2 are numerical constants of order one. This estimate gives a qualitatively correct answer for the whole range from $\alpha > 1$ (the swelling of a coil), through $\alpha \approx 1$ (the coil-globule transition), and to $\alpha < 1$ (collapsed globule). What do we do next? Just as before, we need to minimize the free energy (9.6) as a function of α (see Section 8.6). The condition for a minimum, $\partial F(\alpha)/\partial\alpha = 0$, leads us to the equation for the equilibrium value of α:

$$\alpha^5 - \alpha = x + y\alpha^{-3} . \tag{9.7}$$

This equation determines the size of a coil, $R = \alpha\ell N^{1/2}$ as a function of two characteristic "governing" parameters, x and y.

A graphical interpretation of (9.7) is shown in Figure C9.3. Let us start with Figure C9.3 *a*, where we have plotted a set of curves $\alpha(x)$ for different values of y. How did we manage to do it? We could not possibly solve Eq. (9.7) to obtain an explicit expression for $\alpha(x)$. The trick is that you can easily find the reciprocal function $x(\alpha)$ from (9.7), for every given α and y. It is single-valued, and you can plot it on the graph of $\alpha(x)$.

To understand the meaning of $\alpha(x)$ curves, we have to get an idea about the parameters x and y? The sign of x merely matches the sign of the second virial coefficient B. In a good solvent, $B > 0$ and so $x > 0$. At the θ-point, $B = 0$ and so $x = 0$. And, finally, in a bad solvent (i.e., a precipitant), $B < 0$ and $x < 0$. What is the magnitude of x? In a very good solvent, the second virial coefficient B is neither especially large or small; in this case $x \gg 1$ (because $x \sim N^{1/2}$ and $N \gg 1$). In a similar way, in a very bad solvent x is negative, and $x \gg 1$. From all this we can

conclude that as long as $x \sim 1$ we can be pretty sure that we are close to the θ-point. It looks like the parameter x always mirrors the quality of the solvent. As x ranges from $-\infty$ to $+\infty$, the quality of the solvent varies from very bad ($x \gg 1$, $x < 0$) to moderately bad ($x \sim 1$, $x < 0$), then to the θ-solvent ($x = 0$), a moderately good solvent ($x \sim 1$, $x > 0$), and finally to a very good one ($x \gg 1$, $x > 0$). According to Section 8.5, the quality of the solvent is controlled, in particular, by the temperature. So you can relate the parameter x to the temperature. The range $x < 0$ corresponds to temperatures $T < \theta$, and $x > 0$ to $T > \theta$; x is a monotonically increasing function of temperature. This is why the curves $\alpha(x)$ plotted in Figure C9.3 reflect the dependence of a polymer's swelling coefficient on the temperature or on the quality of the solvent.

As you can see from Figure C9.3, the main changes in the function $\alpha(x)$ occur in the vicinity of the θ temperature. This is the region where the parameter $y = \gamma_2 C/\ell^6$ hardly varies, so we can set it to some constant value (as in fact we did when making Figure C9.3). By definition, y depends on the third virial coefficient C, that is, on the make-up of the polymer chain. We shall skip the details of the theory, and just tell you some results. The study of the coefficient C for different types of monomers has shown the following. If a polymer chain is flexible (which means that the Kuhn segment ℓ is of the same order as the characteristic thickness of the chain d), then $C \sim \ell^6$, so that $y \sim 1$. For rigid chains ($d \ll \ell$), the stiffer the chain, the smaller the parameter y. Thus, all the different curves in Figure C9.3 correspond to different amounts of rigidity of the chain. The parameter y describes the rigidity.

You may have noticed that all the curves $\alpha(x)$ in Figure C9.3 *a* fall into two significantly different groups, depending on the value of y. If $y > y_{\text{cr}} = 1/60$ (i.e., if the chain is fairly flexible), α grows monotonically with x, although the rate of this growth varies. It changes slowly in the range $x \gg -1$, and also when $x > 0$, but goes up rather rapidly for $x \sim -1$, (i.e., just below the θ temperature. On the other hand, if $y < y_{\text{cr}}$, there is a characteristic "loop" just below the θ temperature. It looks pretty much like the loops on isotherms of a van der Waals gas familiar from physics or chemistry. So the function $\alpha(x)$ becomes multi-valued: there are three values of α for each value of x within a certain range of x. This is because the free energy $F(\alpha)$ in (9.9) has as many as three local extrema (Figure C9.4), two local minima and one maximum.

The two local minima we are interested in correspond to the smallest and the largest of the three values of α for each given x. Now we need

to compare the values of F at the two minima, in order to discern which is smallest. It will give us, as usual, the equilibrium value of the swelling coefficient α. This is illustrated in Figure C9.3 *b*. Let's look at one of the curves. You can see that, as long as x is less than some critical value $x_{\rm cr}$, one of the two branches of the "loop" in Figure C9.3 *a* provides the equilibrium solution for α. However, as x increases, there is a sudden "swapping over" between the branches. Now the equilibrium solution is represented by the other branch. Thus, if $y < y_{\rm cr}$ (which means that the chain is fairly stiff), the polymer suddenly rearranges its own shape just below the θ point. When it does this, it changes in size in a very abrupt, jump-like manner. The smaller the parameter y the more dramatic this "jump" (see Figure C9.3 *b*).

9.6 The Coil-Globule Transition

Let's explore the dependence $\alpha(x)$, given by (9.7), in more detail in the range $y < y_{\rm cr}$. When $x > 0$ (so that the solvent is good), this dependence provides the correct qualitative description of how a polymer coil swells due to excluded volume interactions (see Section 8.6). In this case we can neglect the second terms on both sides of Equation (9.7); mathematically, this neglect is obviously justified at $\alpha \gg 1$, but one can check that it works OK everywhere at $x > 0$. Then we shall have:

$$\alpha \sim \left(\frac{BN^{1/2}}{\ell^3}\right)^{1/5}, \quad \text{i.e.,} \quad R \sim \alpha \ell N^{1/2} \sim \ell N^{3/5} \left(\frac{B}{\ell^3}\right)^{1/5}, \qquad (9.8)$$

in complete agreement with (8.14).

Now, what if $x < 0$? Let's look at the region $x > x_{\rm cr}$ first, which corresponds to higher temperatures than that where the jump in polymer size occurs. We can deduce from (9.7) that α is close to one in this case. This means that the molecule takes the shape of a nearly Gaussian coil, and is hardly at all disturbed by volume interactions. What happens if $x < x_{\rm cr}$ (on the bottom branch of the "loop")? In this case, normally, the equilibrium swelling coefficient is very low ($\alpha \ll 1$). So the molecule looks terribly "squashed" by attraction between the monomers, when compared with an ideal coil. The terms on the left-hand side of Equation (9.7) will be much smaller than those on the right. Where did the terms on the left come from? You can easily trace that they have to do with the entropy, Equation (9.5).

From here we conclude that the entropy contribution $U_{\text{eff}}(\alpha)$ to the free energy is not significant for $x < x_{\text{cr}}$. Thus, the equilibrium size of the molecule is only controlled by the free energy of the monomer interactions $U(\alpha)$. If we neglect the terms α^5 and α in (9.7), we shall have:

$$\alpha \sim \frac{C^{1/3}}{(-B)^{1/3} N^{1/6} \ell} \ . \tag{9.9}$$

(Since $x < x_{\text{cr}} < 0$, the second virial coefficient stands for attraction and so must be less than zero.) We can rewrite it for the equilibrium size R:

$$R \sim \alpha \ell N^{1/2} \sim \left(\frac{C}{-B} \right)^{1/3} N^{1/3} \ . \tag{9.10}$$

Then the concentration (or density) n of monomers inside the globule is estimated as:

$$n \sim \frac{N}{R^3} \sim \frac{-B}{C} \ . \tag{9.11}$$

According to Equation (9.11), the shape of the molecule for $x < x_{\text{cr}}$ is totally different from that of a typical polymer coil (e.g., Figure 2.6). First of all, the monomer density does not fall as N grows (compare (9.11) with (8.5)). Moreover, the size of the molecule is proportional to $N^{1/3}$ (not to $N^{1/2}$ as for an ideal polymer, or $N^{3/5}$ as for a polymer with excluded volume interactions). Such unusual properties are actually similar to those of an ordinary liquid drop of constant density. This suggests that if $x < x_{\text{cr}}$ the molecule might have a globular structure, just like the ones in Figures 8.3 (c) or C9.1. This turns out to be true.

The zoomed part of Figure C9.1 explains why density of the globule is independent of polymer length N: we see that there are many completely different pieces of the chain, how close they approach each other — which is another way to talk about density — is determined by the balance of attractive and repulsive forces, and this balance establishes itself locally, in every piece of globule.

Thus, it is the coil-globule transition that is reflected by the "jump" in the molecule's size in Figure C9.3 for $y < y_{\text{cr}}$. You can see that this transition occurs at $x_{\text{cr}} \sim -1$ which is only slightly below the θ temperature. What does this mean? Suppose you have a loose polymer coil at the θ point. Then it does not take much to make the coil "condense onto itself" and form a globule. Just a slight worsening of the solvent quality — that is, just a tiny bit of attraction between the monomers — would be enough.

So far, we have only considered the coil-globule transition for $y < y_{\text{cr}}$. This is when it is accompanied by a "jump" in the molecule's size. Can the

coil-globule transition also happen in the case of $y > y_{\rm cr}$, that is, when the polymer chains are fairly flexible? Certainly it can. If the temperature falls far below the θ point, attraction between the monomers becomes strong enough in this case too. As a result, a condensed globular state is formed. Mathematically, we can describe it in the same fashion as before, skipping the terms on the left-hand side of Equation (9.7). We end up with the same estimates (9.9)–(9.11). Still, there is an important difference from the previous case $y < y_{\rm cr}$. Now the formation of the globule is not step-like, but is a smooth and gradual process. However, although it happens smoothly, it only spans over a fairly narrow temperature interval, somewhat below the θ point. You can see this in Figure C9.3. The transition region for $y > y_{\rm cr}$ is just where the function $\alpha(x)$ changes most rapidly. All this region lies within $x \sim 1$, which corresponds to a temperature variation of the order of

$$\frac{\theta - T}{\theta} \sim N^{-1/2} \ll 1 \ . \tag{9.12}$$

If the relative temperature variations are much greater than $N^{-1/2}$, you can be absolutely sure that the molecule is in the globular state, even for $y > y_{\rm cr}$.

9.7 Pre-Transitional Swelling

There is an extremely important feature of the coil-globule transition that we can draw out of Equation (9.9). We have already said that the transition occurs in the vicinity of the θ temperature. However, at the θ point the second virial coefficient B goes to zero, and is very small around that temperature. What does this tell us? Suppose we approach the θ temperature from below (i.e., we move towards the transition region). We shall notice that the globule grows significantly in size. Meanwhile, the average concentration n of the monomers inside the globule decreases just as significantly. In other words, the globule gradually swells. (This is certainly because B decreases and tends to zero as $T \quad \theta$.)

This is good news. It means that we need to keep no more than the two first terms in the virial expansion (see (9.2)). That will do for globules near the θ temperature (i.e., near the coil-globule transition point). Indeed, since the monomer concentration in the globule is fairly low, we can ignore many-body interactions. Thus, all we need to describe a globule near the θ temperature are the second and the third virial coefficients, B and C.

Now you see that the analogy between the coil-globule transition and an ordinary gas-liquid transition has its limits. A liquid condensed from a gas always has a fairly high density. In contrast, a "newly born" globule is usually quite tenuous at the transition point. This explains why a theory for the coil-globule transition is much more straightforward than for a gas-liquid one. In the case of the globule, we are equipped with a nice small parameter — the monomer concentration inside the molecule. This really helps when constructing a strict mathematical description.

Thus, when nearing the θ point, the globule swells. This gives rise to fluctuations. An interesting question is this. At what stage will the growth in fluctuations cause the actual transition? In fact, when the globule is swelling it remains a globule, until it gets really very close to the θ temperature. In other words, the correlation radius of the concentration fluctuations remains much shorter than the size of the molecule. The actual transition to the coil (i.e., to much stronger fluctuations) only occurs at the temperature T estimated by (9.12). If the chain is fairly flexible (i.e. $y > y_{\rm cr}$), the transition is smooth. The globule goes on swelling more and more, until it reaches the size of a normal polymer coil at the θ point. In contrast, if the chain is stiff (i.e., $y > y_{\rm cr}$), the transition has the form of a jump, which happens at some critical value $x_{\rm cr}$ (i.e., at the critical temperature $T_{\rm cr}$).

9.8 Experimental Observation of the Coil-Globule Transition

There have been a fair number of experiments on the coil-globule transition with decreasing temperature. The most detailed studies were carried out by scientists from three laboratories. One was the group of E. Anufrieva at the Institute of High Polymers in St. Petersburg where they used a polarized luminescence technique. The other was the group of T. Tanaka (1946–2000) at the Massachusetts Institute of Technology (MIT) near Boston, by means of inelastic scattering of laser beams by polymer solutions, and lately the main contributions were made by the groups of B. Chu at State University of New York (SUNY) at Stony Brook and Chi Wu at The Chinese University of Hong Kong, also by the light scattering. We shall not go into the details of how these experiments were done. In general, they measured some quantities related to both translational and rotational diffusion of the molecules, from which the average size R of a molecule can be deduced.

As to the system examined, most of the initial experiments were done on polystyrene macromolecules diluted in cyclohexane. The θ temperature for this system is within the convenient range; it is close to 35°C. What was noticed in the experiments? Below the θ point, the molecule exhibits a very abrupt shrinkage in an interval of only a few degrees. Its volume changes by ten times or more. Obviously, this is where the molecule turns into a globule. However, at the very point of the coil-globule transition, the monomer density inside the globule is still much less than for a dry polymer. In other words, the globule is still rather loose, just as the theory predicts.

Some other methods have also been employed to observe the coil-globule transition for isolated polymer molecules (e.g., ordinary light scattering, viscosity and osmotic pressure measurements, and elastic neutron scattering off polymer solutions). However, the two techniques we mentioned before are the best for this purpose. They are very sensitive and allow measurements of solutions at extremely low concentrations.

The reason the low concentrations are so crucial is this. Suppose we reduce the temperature below the θ point. Attraction between the monomers starts to prevail. This certainly enhances the "condensation" within the molecule (i.e., the formation of a dense globule). This is not the only tendency, however. Another is for different molecules to stick together to form huge "lumps", or aggregates, and these molecular aggregates precipitate out. Obviously, we want to avoid such a process. Therefore, we need to make sure that — at least at the transition point — the condensation of monomers inside the molecule is much more likely than that between macromolecules in solution. The only way is to restrict ourselves to low concentrations. The less concentrated the solution we choose, the further below the θ temperature we can go without worrying about molecular aggregates. In reality, the experiments go to a concentration that is as low as $c = 10^{-2}$ g/l for the chains that are as long as about 10^7 monomer units each. Nevertheless, it remains unclear whether this concentration is low enough to study the entire region of the coil-globule transition.

Despite various tricks, the problem of chain aggregation has not quite been solved. The experts still argue. Research continues, and experimentalists have now reached remarkably low concentrations. In fact, at these concentrations averaged macromolecule has to diffuse for almost ten minutes before it meets another macromolecule. It seems then that the most promising direction for further studies is to try making all measurements within these ten minutes, before macromolecules can aggregate

fully. Unfortunately, even this strategy is not free of difficulties, because although most molecules keep wandering around still not seeing each other, there are relatively few atypical ones which happened to be close at the beginning of the experiment, which can aggregate — and contaminate the signal (because aggregates scatter light much stronger than non-aggregated molecules).

Very interesting experimental results on the coil-globule transition were obtained by the group of K. Yoshikawa in Kyoto by taking DNA instead of a synthetic polymer. We will discuss this a little further in the Section 9.12.

9.9 Dynamics of the Coil-Globule Transition

Whatever transition you explore, it is not just the initial and the final states that you are interested in, but the actual process of the transition. You are not only concerned with water, steam, and so on, but also with boiling, vaporization, and condensation; nor only with ice, but also with melting and solidification. We know that the kinetics scenarios of phase transitions are quite diverse and vary quite wildly depending on the circumstances. For instance, water is water and vapor is vapor, but the transformation of water into vapor may be a slow hardly noticeable process (as, e.g., a wet road after a rain) or it may be a violent explosion (as, e.g., in an overheated steam boiler); boiling may be by formation of bubbles throughout the volume of water, or it may proceed only from the surface — and so on.

What can we find out about "globulisation"? How does it proceed? How do polymer networks collapse? How does coil-globule transition develops in biopolymers, such as DNA and proteins? It turns out that the scenarios are about as diverse as in the case of water and vapor — or more diverse. We will discuss later some other cases, but here we would like to concentrate on one particular possibility for a single chain collapse which we like because we think it is beautiful.

Figure 9.5 sketches the first stages of one of the mechanisms of globule formation. At the very start, lots of little "droplets" emerge. They are the "embryos", or — in a more scientific language — nuclei of the globular phase. Then the "embryos" grow and merge with each other, until a larger spherical globule is formed.

But theorists think that this may not yet be the equilibrium globule, because its chain has not yet become entangled. To form the entanglements,

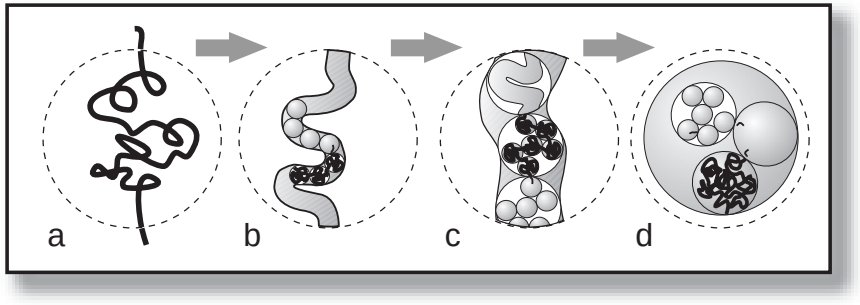


Fig. 9.5 A few initial stages of the coil-globule transition as it develops in time. This looks self-similar! Compare at what we write about the self-similarity in the Chapter 13. The figure is courtesy of S. Nechaev.

it needs to go through an extra stage, which is to do with the so-called reptation motion. We shall discuss reptation a little later, in Chapter 12. The interesting thing is that there is a very simple experiment which you can do yourself to check this theoretical idea. Take a piece of rope (a "macromolecule"), crumple it (the "coil-globule transition"), and try to disentangle it without much shaking ("thermal motion"). People like mountain climbers, who are familiar with ropes, know that this is quite straightforward, as long as you don't pull the ends! This suggests that mere crumpling (i.e., collapse) is not enough to create the knots.

It sounds quite a nice theoretical idea, that the chain has to be crumpled first and then entangled. Subtle experiments by B. Chu in SUNY Stony Brook offer some indirect support for this view. However, if you believe it, you will have to face another bunch of questions. Imagine a very entangled globule. What if we tried to transform it back into a swollen coil? Using the same analogy with a rope, we can anticipate that the knots will tighten up, and... Well, we all know how long it often takes to undo a tight knot! Speaking of the molecular chain, would thermal motion ever be able to undo the knots? Would it take only a microscopic instance of time, or a minute? Or a couple of hours? Many years? Eternity? We do not know yet and the uncertainty is tantalizing!

9.10 Some Generalizations

Let's summarize what we have found out about the coil-globule transition in an isolated homopolymer molecule. Its experimental and theoretical

investigation is certainly very important. It is the simplest of the intermolecular condensation phenomena. If you understand it, you will be able to move on to more complex situations. On the other hand, it can only be observed under very special conditions (like low concentrations of polymer in solution). Fortunately, it turns out that there are many other transitions of a very similar sort in the physics of polymers and biopolymers, and not all of them are so capricious; quite often the problem of precipitation is nonexistent, or not that crucial.

We compared the coil-globule transition with a gas-liquid one. For more complex polymer systems, you can also think of some analogies. The local microstructure of a globule may sometimes resemble a liquid or plastic crystal, an amorphous solid, a glass, an ordinary crystal, a solid or liquid solution, and so on. This is why such strange things as a globule-globule phase transition become possible. (What happens is that the globule's core just rearranges its structure.) Another interesting thing is liquid-crystalline ordering in a concentrated solution of rigid polymer chains. (This is when the chains have a predominant orientation, see Section 4.5.) It turns out that this ordering can also be regarded as the formation of a globule! You only have to imagine it in a special sort of space, the space of the segments' orientations, not in the usual three-dimensional space.

In the next few sections, we shall look at three different effects that have something in common. In a way, they are all similar to the coil-globule transition. These effects are the collapse of polymer networks, the formation of a compact DNA, and denaturation of proteins. Of course, there are many more such phenomena. If you are intrigued, you can find lots of details, for example, in our book (Ref. [3]).

9.11 The Collapse of Polymer Networks

Suppose we have a piece of polymer network, swollen because it is in a good solvent. Let's look at one of the subchains (that is, a part of a chain between two adjacent cross-links, see Section 7.5). Naturally, it tends to take the shape of a loose polymer coil typical of a good solvent. Now, say the solvent becomes worse. The subchains will shrink, which leads to the shrinking of the whole network. If the temperature drops below the θ point, each of the subchains will undergo a coil-globule transition. As a result, the entire network will rapidly collapse. Unlike single molecule case, the gel collapse is easy to observe by the naked eye, as illustrated by the cartoon

in Figure C9.6. When we are dealing with the collapse of networks, there is no such problem as the precipitation of molecular aggregates. We have only one macroscopic sample which shrinks as a whole when all the subchains collapse.

This is exactly the phenomenon called the collapse of polymer networks. It was discovered by T. Tanaka (1946–2000) and his colleagues at the Massachusetts Institute of Technology (MIT) in 1978. They used networks of polyacrylamide diluted in a mixture of acetone and water. In these experiments, the temperature was not varied. To make the solvent worse, they just poured some extra acetone into the solution. (This worked because acetone, in contrast to water, is a bad solvent for polyacrylamide.) Figure C9.7 gives an idea of what was found. It sketches how the size of the network depends on the acetone concentration. You can see that if you "dump" 42% of acetone, the network collapses suddenly. Its volume drops by a factor of nearly 20.

It seems that to develop a theory for the collapse of networks, we could follow the same logic as for the coil-globule transitions. After all, they both have the same cause. However, this does not work. Such a theory was indeed created, but experiments did not support it. You can even spot a contradiction in Figure C9.7. The data are for polyacrylamide which is a flexible polymer. So we are in the regime with $y > y_{\rm cr}$, and the coil-globule transition should happen smoothly. Yet, the curve in Figure C9.7 *b* drops down in a step, and exhibits hysteresis on return.

Furthermore, T. Tanaka carried out a detailed investigation and found the following. The height of the step depended strongly on the time interval between preparing the network, and starting the experiment. The longer the delay, the higher the step. If a network was kept for two months after it was made, changes in volume by factors like a few hundred were observed. In contrast, freshly prepared networks manifested a nice, smooth collapse.

Can we find an explanation for all these oddities? The clue is that polyacrylamide chains are not stable in water. They are prone to the chemical reaction called hydrolysis. As a result, monomers which were initially neutral dissociate. This means that small light ions split off from the monomers, leaving behind segments of the opposite charge. (Such small ions are usually called counterions; we have talked about them in Section 2.5). The released counterions float on their own inside the swelling network. The hydrolysis of polyacrylamide occurs extremely slowly. So, over a short period of time, only a very small proportion of the monomers will gain an electric charge. However, the "older" the network (i.e., the longer ago it was prepared),

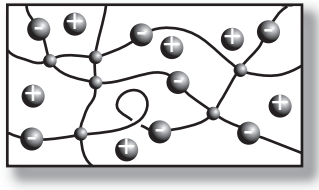


Fig. 9.8 The "gas" of counterions in a charged polymer network.

the higher will grow the proportion of charged monomers. Now we can explain T. Tanaka's experiments, if we make just one assumption. We need to assume that even a small proportion of charged monomers can affect the collapse rather strongly, and that the step becomes even higher as this proportion increases.

What gives us the right to make such an assumption? Let's see. Together with the charged monomers, there are floating counterions in the swollen network (Figure 9.8). Note that the counterions do not drift out of the network into the pure surrounding solvent. Why not? If they did, the system would lose its electrical neutrality on a large, macroscopic scale. Strong Coulomb interactions would arise between the charges on the network and the counterions outside. The energy of these interactions is extremely high. This is why such a state is energetically unfavorable and never occurs.

Thus, the counterions move freely inside the network, but are not allowed outside. You could say that the "shell" of the network (i.e., its outside surface) stops them. Evidently then, a crowd of counterions exert some pressure on the "shell". This pressure favors stretching the network in all directions. We are going to show now that this is exactly what makes the collapse so different from what you might expect.

The free energy of the whole network is the sum of the free energies of all the individual subchains. The free energy of each subchain, in its turn, consists of entropy and energy terms, $U_{\text{eff}}(\alpha)$ and $U(\alpha)$ (Equation (9.1)). (Here α stands as usual for the swelling coefficient — this time, swelling coefficient of either any particular subchain or the entire network.) The dependencies $U_{\text{eff}}(\alpha)$ and $U(\alpha)$, calculated from Equations (9.5) and (9.2), are plotted in Figure C9.9 *a*, together with their sum $F(\alpha)$, for an electrically neutral polymer network. The chosen value of the parameter $x = \gamma_1 B N^{1/2}/\ell^3$ lies in the coil-globule transition region. At the same time, the value of $y = \gamma_2 C/\ell^6$ is set to just over y_{cr}, so that the function $F(\alpha)$ has only one minimum (this corresponds to the "no loop" case in Figure C9.3).

What will change if we create just a small proportion of charged monomers in the network, and the same number of counterions? The energy $U(\alpha)$ will hardly alter. There will certainly be some extra contribution to it, due to Coulomb interactions, but it will not really matter, given that the number of charges is small. As for $U_{\text{eff}}(\alpha)$, we must not forget what we have just discussed. The counterions cause osmotic pressure, which inflates the network. What difference does this make? First of all, the minimum in $U_{\text{eff}}(\alpha)$ is shifted towards higher α (Figure C9.9 *b*). As a result, the function $F(\alpha)$ may now appear as in Figure C9.9 *c*. This corresponds to the "loop" in Figure C9.3 and implies a step-like collapse.

Thus, the charged monomers in the network extend the range of parameters for which a step-like collapse may occur. Even networks of flexible chains ($y > y_{\text{cr}}$) which are normally expected to have a smooth collapse (when they are neutral, Figure C9.9 *a*), may exhibit a step (Figure C9.9 *c*). Calculations show that if you have just a few charges per subchain, you can nearly guarantee that the transition will be step-like. Now you can see why partially charged networks, although made from flexible polyacrylamide chains, collapse so abruptly (Figure C9.7). Look at Figure C9.9 again. The more charged monomers you have (i.e., the higher the inflating osmotic pressure from the counterions), the higher the step. This fully explains the observations.

The collapse of polymer networks has recently attracted a lot of attention. This boom is partially due to some important applications, which all stem from the fact that you need only slightly change the quality of the solvent to make the network collapse rapidly. It is especially useful that the collapse is very sensitive to the presence of charged monomers and counterions in the solution. Thus collapsing networks can be adapt to detect small ion impurities in a solution, as well as to clear the impurities away. Besides all this, the collapse of networks can also serve as a good model for some other processes in biology (e.g., in the vitreous body in the eye).

T. Tanaka's group at MIT have assembled a sort of a large "collection" — cases of network collapse transitions in a variety of systems and circumstances. For example, the collapse can occur if the temperature changes — to complicate matters, some networks collapse on cooling, and some on heating! It can also be caused by ionic strength, by adding certain molecules, by light, etc. This is pictorially illustrated in the Figure C9.10 drawn by T. Tanaka himself. Moreover, a network can be made to collapse in patches, to give an irregular density. In this case we would obtain a special kind of diffraction grating, or a hologram of the object.

We should not forget also gel swelling. As you should be ready to appreciate, all the subchains of the gel will swell once placed in a good solvent, and this will make the macroscopic gel sample swell. What is interesting about it is that good solvent diffuses rather slowly into the volume of the gel, and thus surface portion of the gel swells faster than the balk. This leads to the formation of beautiful and sometimes strange looking patterns of buckles on the gel surface, making polymer scientists to joke about swelling gel looking not unlike a brain.

To conclude, we will mention just one more peculiarity of polyelectrolyte gels. It can be easily understood if we use some things that we discovered before. We saw that the counterions of polyelectrolyte gels cause an extra pressure that makes the network swell. This pressure is really rather high. So the gel can swell quite dramatically, to the extent that there will be no more than 0.1% of polymer inside the gel, and the rest will be taken up by water. In other words, 1 g of a "dry" polymer gel can absorb up to 1 kg of water! This is why they say sometimes that polyelectrolyte gels are super-absorbers of water. This property has found plenty of applications. Perhaps, the most impressive one is babies' nappies (disposable diapers). How do they manage to take in so much liquid? What happens is that the water is absorbed by granules of polyelectrolyte gel made of polyacrylic and polymethacrylic acids; the dry diaper is a collapsed network, water makes it swell abruptly through a very sharp step thus increasing its volume dramatically and thus opening up the room for a large amount of water. Such gels are also used in agriculture, to keep the upper layers of soil humid in dry areas.

9.12 The Globular State of the DNA Double Helix

Each cell of your body contains about ten centimeters to one meter of DNA, and this DNA must fit into a micron-sized cell nucleus. The DNA is far too big to fit into the cell not only as a straight stick, but even as a Gaussian coil (we mentioned this already in Chapter 5). Therefore, DNA in the cell must be somehow condensed and, from a polymer physics point of view, it must be in the state of globule, albeit quite peculiar one (see also Section 9.13 about the concept of globule in this case).

This conclusion is true not only for our cells and other eukaryotic cells (i.e., sophisticated cells which store their DNA in the special cell organelle called nucleus), but also for more primitive procaryotic cells (which do not

have nuclei), such as bacteria. You can see this in Figure C2.7: when the outer membrane of bacteria is almost completely destroyed, DNA — and the impressive amount of DNA! — spills out. Even this released DNA looks like it remains a globule, it is definitely much more compact than any of the typical coils shown, say, in the Figure 8.3 *a* or *b*; it must have been a much denser globule while still inside. Hence, DNA in bacteria has to be stored in a compact globular shape. In fact, it makes a very complex globule.

Furthermore, the same dimension argument applies even to viruses. As an example, consider a bacteriophage (meaning, virus which infects bacteria) called λ — one of the most studied ones. It represents a sort of a box made of protein subunits and called capsid, with double helical DNA genome stored inside. The capsid diameter is about $D = 55$ nm, while DNA length is 48,000 base pairs, or about $L = 16,300$ nm (about 0.34 nm per base pair). If you remember the diameter of the double helix (close to $d = 2$ nm), you can estimate the fraction of volume occupied by the DNA inside the capsid. It turns out huge:

$$\phi = \frac{L(\pi(d/2)^2)}{(4\pi/3)(D/2)^3} \approx 60\ \% \ . \tag{9.13}$$

This is astoundingly high degree of compactness. If you want to use units more traditional for biochemistry, this corresponds to DNA concentration about 500 $\frac{\text{mg}}{\text{ml}}$. By every measure, DNA inside the capsid is very dense indeed! One may want to know what is the structure of such a dense DNA globule formed inside the virus capsid. Of course, it has pretty little to do with the loose homopolymer globule close to the θ-point. In fact, as you see, the capsid size is not only tight for the given amount of DNA, it is pretty close to the DNA persistence length, which is about 50 nm. That means, it must be pretty difficult to bend DNA tightly enough to place it inside the capsid. Both theoretical considerations and experiments indicate that the DNA is organized inside the capsid as an inverse spool, with turns going parallel to each other in a very ordered manner. Nevertheless, the detailed structure of this spool, as well as the processes by which DNA unpacks from the virus head and infects the cell, remain the subject of current research and heated debate among the experts. In this sense, Figure C9.11 is only a cartoon, showing that the DNA is very densely packed inside the capsid.

The image of an elastic DNA tightly bent inside the capsid may remind to you "The Story of Keesh", by Jack London. There, the main character, named Keesh, invented a way to kill a polar bear by freezing a tightly coiled thin strip of the whalebone into a round ball of chunks of blubber. Once

the ball was swallowed, the blubber melted — with the dire consequences for the poor animal.

Organization of DNA in the cell is more complex, because more DNA has to be stored, and stored in a manageable form. Overall, it seems fair to say that the organization of DNA in either a prokaryotic or eukaryotic cell is not completely understood. It is believed, however, that DNA is organized in a sort of hierarchial fashion. To be specific, in eukaryotic cell there are special proteins, called histones; eight histone molecules of a certain types assemble together, and this histone octamer serves as a spool on which roughly two turns of DNA (containing 147 base pairs) are tightly wound — forming a "bead" on the DNA "string", called nucleosome. And now we can imagine that the whole DNA represents a string with beads; a physicist would call it a new "renormalized" polymer, while biologists call it a ten nanometer fiber (because 10 nm is about the diameter of nucleosome). The structure of nucleosome is known to atomic details, but the positioning of nucleosome beads along the DNA is hotly debated — the main question being whether (and/or to which extent) nucleosomes prefer certain features of the underlying DNA sequence. What happens to the ten nanometer fiber on a larger scale is not completely clear; the consensus among the researchers seems to be that it crumples somehow to form a thicker thirty nanometer fiber. The DNA forming a ten nanometer fiber and then ten nanometer fiber forming a thirty nanometer one is not entirely unlike the initial stages of our theoretical speculation Figure 9.5). And how the thirty nanometer fiber is organized on larger scales, up to the entire chromosome, is at present anybody's guess. One can find pictures of it in the literature or on the web, some of the pictures are truly beautifully executed — but they reflect more on the artist's imagination than on the solid scientific knowledge.

It is very hard to study natural organization of DNA in a cell or even in a virus. This is why it is interesting and makes a lot of sense to start with a model system — to place DNA in a poor solvent, and to see what we can find out about the globules and the coil-globule transition in this case. When experimenters thought to try this idea, the first rather obvious problem was this: It is not easy to force DNA to collapse, because DNA is strongly charged. Under normal conditions, almost every base pair in the DNA double helix in water solution carries two elementary negative charges (on the phosphate groups, that is, on the outside surface of the double helix) — which is why, a chemist would not miss to add, DNA has an A in its abbreviated name (indeed, A stands for acid). We discussed the role of

counterions and their pressure in the collapse of weakly charged networks (see Section 9.11). The counterions pressure is at play for DNA, too; but DNA is so strongly charged that Coulomb repulsion between segments is also of huge role, it has to be overcome for DNA to collapse. Of course, this problem exists not only for experimenters wishing to model DNA collapse, this problem exists also for nature which must pack and store DNA in tight spaces. Thus, scientists — as they usually do — can look at nature for inspiration.

Once you give it a thought, it becomes not surprising at all that the nucleosome core particles, the histone octamers, are strongly positively charged. Isn't it beautiful that negatively charged DNA wraps around a positively charged spool, with Coulomb forces stabilizing the whole thing? Similarly, internal surface of a virus capsid is also positively charged. Thus, a good idea to collapse DNA in a model system is to add multiply charged ions in solution. It turns out that even ions with charges +4 (spermine) or even +3 (spermidine) work pretty well and do help DNA to collapse. Another useful trick is to add some neutral polymers (usually polyethyleneoxide, PEO); the coils of this added polymer are comparatively short and flexible, so that they will be expelled when DNA shrinks and condenses, and so they will not get in the way of the condensed DNA inner structure: the "gas" of such coils causes a kind of external pressure on the "walls" of the DNA globule.

What was observed in these model DNA condensates is pretty interesting. First of all, DNA "globules" typically have the shape of a torus (like a doughnut) in which segments of DNA going around the torus hole are arranged rather closely and more or less parallel to each other — see Figure 9.12. Linguistically, this sounds admittedly bizarre (which is why we took "globule" in the quotation marks): the word "globule" comes from Latin *globulus*, diminutive of *globus*, which is a sphere, so when a physicist says that DNA globules are of doughnut shape — it can be translated into English as "DNA sphere is toroid". Nevertheless, however clumsy this terminology may be, as far as physics is concerned it is not surprising and quite natural that DNA condensates adopt a torus shape. Indeed, there are no places in the double helix where it can be easily kinked. (In other words, DNA behaves as a worm-like chain, see Section 2.3.) This is why it cannot possibly fill in the core of the spherical globule, and we end up with a hole in the middle. Of course, the finding of doughnut-shaped DNA condensates was very exciting, and it did help to establish the DNA organization in virus heads.

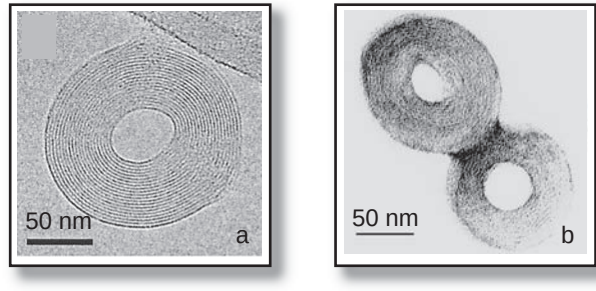


Fig. 9.12 Toroid-shaped "globules" of DNA. Panel (*a*): Cryoelectron micrograph of λ-DNA toroid-shaped globule (λ-DNA is the dsDNA from bacteriophage called λ, the length of this DNA is close to 50,000 base pairs). The turns of DNA are clearly visible. The image is courtesy of N. Hud, reproduced with permission from the paper: N. Hud and K. Downing, Proceedings of the National Academy of Sciences USA, v. **98**, n. 26, pp. 14925–14930, 2001. Copyright 2001, National Academy of Sciences, USA. Panel (*b*): Electron micrograph of two torus-shaped DNA globules attached together. Each torus is formed by λ-DNA. The image is courtesy of J.-L. Sikorav.

9.13 Why do We Call Them Globules?

Here we must digress and discuss an important point — so important in fact that we made it a separate section. We mentioned already that toroid globule is a linguistic nonsense. One could argue — and some people do argue — that even the DNA toroid condensates, not to mention much more complex and sophisticated in organization DNA in the cells, have rather little to do with the simple collapsed homopolymer globule which we discussed earlier in the this chapter and described in terms of second and third virial coefficients (see, e.g., Sections 9.5 and 9.6). In the next chapter we are going to discuss protein globules, also in no way the subject of a primitive theory.

Not entering any terminological disputes, we would insist that both DNA organized in cells, viruses, and model toroid systems, as well as globular proteins do share something fundamental with the homopolymer globules. One way to explain it is to discuss the balance of forces. Consider one monomer deep inside homopolymer globule, or one aminoacid residue deep inside protein globule, or one piece of DNA also deep inside. In all these cases the forces acting on a selected unit are balanced *locally*, they are forces of interactions with near neighbors in space. Accordingly, for instance, the density of a homopolymer globule is estimated by the formula (9.11), which does not involve the chain length N. This is to be contrasted

with the coiled polymer in which the density is a function of N (like $N^{-1/2}$ in Gaussian case, see (8.5)), bearing witness to the fact that forces are only balanced on the scale of the polymer molecule as a whole.

In a more formal language, the signature property of the globular state is the short correlation length — much shorter than, and independent of, the overall molecule size. Once again, this is to be contrasted with the coil in which correlation involves the coil as a whole. In this deep sophisticated sense both DNA in cells and viruses and proteins in their native form are globules indeed.

9.14 What is the Order of Coil-Globule Transition?

We begun our discussion of globules from the very simplest flexible homopolymer chain, and found that globule interior in equilibrium looks similar to the disordered solution of independent chains (see Figure C9.1). Subsequent analysis revealed the immanent features of globular state in a variety of polymer systems, for instance, in a toroidal DNA; we will meet some other examples of globules later on in this book. Similarly, coil-globule transition we considered in Section 9.6 for the same simplest homopolymer chain, assuming that monomers stick (attract) pairwise ($B < 0$), but repel in triple or any higher order collisions ($C > 0$). Using this model, we established, that the transition between states of coil and globule occurs rather smoothly if the chain is flexible, but becomes sharper for the stiffer polymers. In fact, even rather stiff macromolecules in this scenario undergo sharp, but not very sharp transition: globule swells significantly before the transition (Section 9.7), while the amount of absorbed or released heat, i.e., so-called latent heat of the transition[1], as it turns out, is rather small, proportional to $\overline{N}$ and not to N. But this is the case only for the simplest basal model of a macromolecule. In reality there are a number of polymer systems in which coil-globule transition is so sharp that it represents, in physics parlance, the first order phase transition. In this type of transition, the latent heat is large and it can be straightforwardly measured by a calorimeter, while pre-transitional swelling of the globule is practically absent; upon heating, globule does not feel any signs of the approaching catastrophe — and then suddenly breaks apart and completely unfolds. Not going here into any details, it might be useful to present a list of systems

[1]For the readers who do not remember the concept of latent heat we explain it in Chapter 10.

and situations when this scenario is realized:

- Polymer chains with peculiar interactions, in which third (or higher) virial coefficient becomes negative at a temperature (or solvent conditions) where second virial coefficient is still positive. An example is the system in which monomers can form complexes, similar to micells, with a defined number of participants. Following de Gennes, this is sometimes called p-cluster model.
- Macromolecules capable of internal local orientational ordering, usually of nematic type, in the globular state, like in toroid DNA case.
- Macromolecules with other types of symmetry or organization in the globule, up to a crystal.
- Polymers in which monomers can have two different states, such as helical or non-helical, when globule can be formed due to the jump in, e.g., the degree of helicity.
- Polymers in which monomers can absorb ligands from the solvent, such that globule can be formed due to a jump in the number of absorbed ligands.
- Polymer chains in multicomponent solvents, when globule formation can be achieved by the re-distribution of solvent components between the globule interior and the outer solution.
- Polymer chain or network with ionizable groups and in a poor solvent (e.g., with hydrophobic monomers), when osmotic pressure of counterions is responsible for the sharpness of the transition.
- Last but not least, the heteropolymers with properly selected sequences. The latter is the most interesting case, and it is the subject of the next chapter (see Section 10.6).

Chapter 10

Globular Proteins and Folding

I am practically perfect in every way.

P.L. Travers,
Mary Poppins (screenplay)

10.1 Anfinsen's Experiment: Renaturation

There are many globular proteins in a living cell, and they play a key role. We have already discussed this in Chapter 5. However, the theory of such systems is extremely hard; a protein globule is perhaps one of the most complex objects in modern physics. What is striking and unusual is that proteins have a strictly defined spatial tertiary structure (see Sections 5.6.4 and 5.7). In a protein globule, not only averaged density, but the entire spatial structure of the whole chain is fixed.

You may wonder why we are so worried about the tertiary structure in particular? There are other things, like the primary structure, for instance, which means that the whole sequence of monomers is also fixed! It all has to do with how different structures are formed. To produce protein chains with the right primary structure (as determined by the genetic program, DNA), there is a special complex "machine" in the cell, called a ribosome. Unfortunately, we do not yet know how to synthesize specific proteins without cells. Scientists certainly hope that they will eventually figure out how things work in a cell, but, at present we here cannot do much better than brush the question aside, saying: "Well, there is some mechanism of biosynthesis..."

By contrast, how is the fixed tertiary structure created? Perhaps there is some other mysterious "machine", of which we know nothing at all, but

which is actually in charge of packing protein chains into globules of the right shape? Note: this is where we reach the crucial point. The answer is that such a "machine" does not have to exist. We can pretty well manage without it. This was first shown in 1961 by the biophysicist Christian Anfinsen at the National Institute of Health near Washington DC (he later received a Nobel prize for this work).

We can explain Anfinsen's idea in the following way. We have already mentioned denaturation of proteins. Denaturation, like coil-globule transition, can be caused by heating (and sometimes by cooling); it can be caused by adding a special substance to the water (such as urea), to reduce the hydrophobic effect (see Sections 5.1 and 5.7); yet another way is to add some alkali or acid (in the latter case, certain amino acids gain a positive electric charge in an acidic medium, and others gain a negative charge in an alkaline medium, making globules unstable in either case due to repulsion between similarly charged segments). Denaturation is a sharp conformational transition. The stiff globular structure is destroyed, and the protein — say an enzyme — ceases to be chemically active. Anfinsen wondered if the protein could be returned to the native state. He tried — and succeeded: in a dilute solution, away from any cell machinery, when every protein molecule was left on its own in the medium of water and salt and nothing else, proteins were able to *renature* — provided that temperature change and other conditions were sufficiently slow and smooth. It is wonderful discovery indeed! It means that the single molecule of protein, completely unsupervised, is capable of correctly reproducing its own spatial structure. We do not have to employ a living cell, or to borrow some special living "machine". The correct tertiary structure can be restored or renatured *in vitro*. All you need is to be careful and make sure that the process is very gradual, and that the concentration is low, and so on.

As a matter of fact, as usual in biology, every rule seems to have at least some exceptions. Some complex proteins fold with the help of special molecules, called chaperones. Nevertheless, a firmly established fact is that many proteins do not need any assistance and are able to do this amazing job easily and reliably on their own. Moreover, from the physics point of view, does it really make a great difference whether it is just one protein molecule that organizes itself, or a pair of molecules, such as a protein and a chaperone?

Thus, in contrast to the primary structure which can only be produced in the living "factory", the tertiary structure is capable of organizing itself. In this sense, the formation of primary structure is in the subject of biology,

while self-organization of spatial ternary structure is physics. And it is extremely important: the capability of self-organization is exactly what enables all proteins to function, and all life to live. To try to understand it has been a challenge for physicists for over a quarter of a century.

10.2 Aperiodic Crystal or Equilibrated Glass?

At first glance, self-organization seems very straightforward. Suppose we have a certain chain with a fixed sequence of monomers. In the right sort of circumstances, left on its own, it will always roll up into a coil in exactly the same way. We take up the same thing every time, and obtain the same outcome; is this really so strange? Yet this phenomenon is unique, it is completely unlike anything else in physics. Let's discuss this problem in more detail.

Look at a cartoon in the Figure C10.1: a sequence of units of a protein chain is similar to a sequence of letters in a certain text or message, but we immediately discover that we cannot read it — simply because we do not understand the language! It turns out that the meaning of the message can be revealed when the fluctuating coil collapses and self-organizes. You could say that the tertiary structure reveals the meaning of the primary text. Thus, in addition to a simple collapse of a chain, which may not be so much unlike the familiar homopolymer condensation, a self-organized collapse of a protein chain can be described as reading *with understanding*, or as decoding of a message.

In one way, a protein globule has something in common with a solid crystal. They both have a very well defined three-dimensional structure, and rather small fluctuations of atoms around their "right" positions. However, this analogy does not extend much further than just the rigidity, the well defined character of the spatial arrangement. There is no similarity between the two structures themselves. Crystals are distinguished for their periodicity. In contrast, a protein globule consists of assorted amino acid residues, and so is completely non-uniform and irregular in shape.

Erwin Schrödinger (1887–1961), one of the founders of quantum mechanics, in his famous essay [37] on "What is Life?" based on lectures delivered in Dublin in 1943, at the height of the war (by the way, in our opinion — a must reading for anyone interested in biological physics), coined the special term: aperiodic crystal. This is intentionally an oxymoron. Aperiodic crystal is a crystal in the sense that every atom or

molecular group somehow "knows" its space position up to small fluctuations, and it is aperiodic in the sense that the defined positions are not obtained by simple translations. Contextually, Schrödinger suggested this concept in quite general way to describe biological cell constituents in terms of physics; he did not know anything about globular proteins at the time — they were not yet discovered as such, but his idea fits them aptly.

But, as we also mentioned, protein globule consists of so diverse molecular groups, that their arrangement in space is inevitably very irregular. In this sense, a protein globule resembles so-called disordered systems, also known to physics, such as an amorphous solid or a glass. Alas, we shall see shortly that this analogy is very limited too. It does not go much deeper than the very fact of the lack of spatial periodicity.

To make the point clear, we have to digress and summarize briefly what do we really mean by "a glass"? In Chapter 4 we talked briefly about both polymer and non-polymer glasses. We can describe them as substances that have been "frozen" in a somewhat non-equilibrium state. Glasses have enormous relaxation times (i.e., the time for the system to reach equilibrium) — far longer than any sensible physical experiment you can think of: for instance, ancient Greek vases were made of glass which did not come to equilibrium by now. We could say that the glass has "memorized" the structure (i.e., the positions of the atoms) that it happened to have when it was made (i.e., as it was cooled). If we melt the glass and then cool it down again, a completely new (although statistically similar) microstructure is created, and so the previous "memory" will be totally washed away and lost. Equilibrium is never reached. That is exactly why glasses are disordered.

Thus, the analogy with glasses does exist, because arrangement of atoms in a protein globule is irregular (or seems irregular to an untrained eye), but at the same time protein globule, as Anfinsen's experiment suggest, is in equilibrium[1]. Thus, we should talk about equilibrated glass — positively, another oxymoron.

In fact, there is a remarkable, even though also limited, resemblance between a glass and the primary structure of biopolymers. Indeed, if we take a protein molecule and place it in the solution conditions when it is stable, then more or less spontaneous rearrangements of the primary structure

[1]Strictly speaking, it might be that protein globule is the long living metastable state, like overheated water, but let's brush aside this detail for simplicity, it does not affect the overall logic of our discussion.

take unthinkably long times; in practice, it never happens. Thus, we can also talk of "memory": in glass, there is a memory of spatial positions of molecules; in protein, there is memory of positions of amino-acids along the sequence. The analogy goes further: as soon as the primary structure is destroyed (i.e., the chemical bonds of the chain are broken), the "memory" is completely lost. But then there is the difference: for proteins, we have cell machinery capable of producing a new exact copy of the same sequence, while for glasses we do not have any practical way to re-create exact same positions of all atoms.

Unlike primary sequence, tertiary structure of proteins does not have that much in common with a glass, except merely the lack of periodicity or other spatial regularity. The tertiary structure is never "forgotten", even after denaturation. If it were, it would never be able to reorganize itself, Anfinsen's experiment would not work. Of course, all the "memories" are kept in the primary structure; the reason why tertiary structure is not forgotten upon denaturation is because primary sequence is preserved. Therefore, all the information needed to reconstruct tertiary structure is contained in the primary sequence, engraved in some as yet unknown secret language. We see that something novel enters physics: we need to understand the behavior, specifically — collapse behavior, of a system with engraved information. Never before was there anything like that in physics.

Thus, the tertiary structure can rebuild itself just like a stable crystal, and is irregular in shape just like a amorphous glass — but the crystal is aperiodic, and the glass is equilibrated.

10.3 Levinthal's Paradox

Many medicines are based on proteins; pharmacological companies would love dearly to be able to predict, by a cheap computation, what is the equilibrium spatial shape of a polypeptide chain with a given sequence. This *prediction* of tertiary structure based entirely on the sequence is really a multi-billion dollar problem. What people try to do to address it in practice is to invoke an additional information, a hint apart from the sequence itself — to find other known proteins with elements of sequence similarity, and then to guess the new tertiary structure based on the elements of the known ones. This might be a nice practical solution, particularly when it works, but here in this book we are not interested in such things, for us it is almost like cheating. We want to discuss the heads-on approach; in the end, this

is now purely physics problem: having a molecule one needs to find its minimal energy conformation.

Can we approach this problem computationally? Since it is so important a problem, we could perhaps take a very large supercomputer and just compute energies of all conformations! Does this sound convincing? Not really. Let's make an estimate. Say the chain is as short as 100 units. For the sake of argument suppose further that each bond of the chain can take two different conformations only, for example, "right turn" and "left turn". (This is certainly an underestimate!) Even then the chain can have as many as $2^{100} = (2^{10})^{10} = (1,024)^{10} \approx (1,000)^{10} = 10^{30}$ conformations in total. Now, suppose the conformation is computationally updated at every tick of a computer clock (take it generously at about 10^{-9} s) and even neglect the time needed to compute energy. The process will then require $10^{30} \times 10^{-9}$ s $= 10^{21}$ s $\approx 10^{13}$ years — really not very practical! Needless adding, if you take a parallel computer with 100 or 1,000 processors (or as many as you can practically think of) — it won't change the verdict: enumeration of all conformations in order to find the lowest energy one is impossible.

But how about proteins themselves, how do they manage to find their lowest energy conformation? And they *do* manage, as Anfinsen's experiment suggests. Let's stay with the same estimate 10^{30} for the number of conformations, and suppose conformation changed by every single atomic collision, that is, every 10^{-11} s (definitely, a much too generous overestimate). How long would it take for a protein molecule to go through all the conformations in search of the stable state? Our calculation gives the incredibly long time of $10^{30} \times 10^{-11}$ s $= 10^{19}$ s $\approx 3 \cdot 10^{11}$ years. Just for comparison, the age of the Universe is about 10^{10} years!

This problem is known as Levinthal's paradox (as it was formulated by Cyrus Levinthal (1922–1990) at Columbia University in New York): protein molecule certainly cannot search through all of its conformations, yet it does find the particular one with lowest energy. It *is* a paradox, isn't it? How does the chain manage to find the equilibrium?

Alas, we seem to be getting nowhere. We had hoped to learn about self-organization through some analogies with other physical objects. However, we have found none. That, roughly, was the state of affairs in the field in 1970s and 1980s.

10.4 Denaturation and Renaturation are Sharp Cooperative Transitions, with Latent Heat

Remember, we are talking about physics of protein folding. Surely, Anfinsen-style experiments were repeated numerously, on many systems, and huge body of knowledge was accumulated on various proteins — but that is not what we are talking about. We want to know how protein utilizes the information encoded in its primary sequence and finds the lowest energy state overcoming the Levinthal paradox. In this regard, the extremely important hint came from the simple minded experiments first performed by P.L. Privalov in Pouschino near Moscow. He measured the calorimetric effect of protein denaturation and renaturation.

The name "calorimetry" goes back to the time when people did not yet realize that heat was just another form of energy. The technical sophistication in the field grew enormously, but the assortment of quantities measured remains pretty much the same as almost two centuries ago. Remember, if you heat a piece of ice in an open air, and start at pretty low temperature, then your sample gradually warms up, and the speed of temperature growth is proportional to the amount of heat power (energy per unit time) provided (the coefficient is heat capacity of the sample, that is, the specific heat of ice times the sample mass). This boring situation continues up until the sample reaches the melting temperature — 0°C at normal atmospheric pressure. At this point, what happens is you keep providing energy at the same rate as before, but the temperature does not grow and remains at 0°C; instead of warming the sample up, the energy provided is spent on melting, on destroying the ice. Only when all of the ice is melted, the temperature starts growing again. Thus, the point is that you have to provide a certain amount of energy, called *latent heat* to destroy (melt) the solid ice and transform it into liquid water. Conversely, the same amount of energy has to be taken away to freeze water and make it solid. By the way of example, the latent heat of ice melting is about 80 $\frac{\text{cal}}{\text{g}} \approx 330\ \frac{\text{J}}{\text{g}}$; latent heats of melting for most other simple substances are of the same order, give or take a factor of 3.

Privalov discovered, that protein denaturation, when it is caused by elevating temperature, requires some latent heat, just like melting of a solid, and in about the comparable amount per unit mass; knowing the latent heat per unit mass of ice melting and the molecular mass of each amino acid monomer (around 110 Dalton), you can estimate the latent heat of denaturation: for a typical protein of about 200 amino acids, it

is about a few hundred of k_BT per one protein macromolecule (remember that absolute temperature is always not too far from 300°K when we talk bio-molecules).

This is intriguing. The heat absorbtion at the transition tells us that in the transition point and in its vicinity both states (water and ice; native and denatured) exist as two distinct local minima of free energy (look again at the Figures C9.4 and C9.7). We should then be able to see both of them simultaneously; do we see them? As it turns out, yes, we do. How? Imagine, for instance, that you measure an optical absorbtion spectrum and it peaks at one wavelength for denatured protein and at another wavelength for native one; then, in the vicinity of the transition, we do see both peaks. Some fraction of the material is in the denatured state, some in the native state, and the proportion changes as we change temperature: not that the states move towards each other and become similar, no; only the relative population of each state changes. For instance, if the energy difference between the two states is ΔE, then the ratio of their populations should be proportional to $e^{-\Delta E/k_BT}$. Therefore, watching how the relative strength of the two absorbtion peaks changes with temperature, we should be able to measure ΔE. This was of course done, and the result is truly beautiful: ΔE measured this way is equal — up to the experimental errors, but for pretty much every protein tested — to the amount of latent heat per macromolecule. This means that protein globule denatures as a single cooperative unit.

10.5 Random Sequence Heteropolymers are Not Protein-Like, for They Have No Latent Heat

Discovery of latent heat and cooperative character of protein denaturation transition added much more food for thought for physicists working on protein folding.

First, cooperativity of denaturation is to be contrasted to globule-coil transition in a simple homopolymer — see Section 9.6. As a rule, that transition is accompanied by a rather small heat absorbtion, which changes with chain length N proportional to $N^{1/2}$. At the same time, the study of various proteins of different lengths did not reveal any systematic tendency of latent heat reduction for longer species. Second point which we will not develop in any details here is that there are some types of monomer

interactions in homopolymers which lead to large latent heat transitions (see section 9.14), but they, too, did not seem relevant to proteins.

Third and most important, theorists were able to look at the denaturation transition of a random sequence heteropolymer, and to show that it, too, does not have a significant latent heat. This was achieved first in 1987 by J. Bryngelson and P. Wolynes at the University of Illinois at Urbana-Champaign, who suggested a phenomenological description, and soon thereafter (and independently) by E. Shakhnovich and A. Gutin in Pouschino near Moscow, who developed a real microscopic model. Let's discuss briefly the main idea of their arguments.

The central point is to adopt the properly coarse-grained view of the protein conformations and their energies. Think for instance of the interaction energy between two aminoacid residues, of chemical names $\mathcal{A}$ and $\mathcal{A}'$. Of course, energy depends in general on the mutual orientations of the given residues, as well as positions of other residues around, and also water molecules nearby; furthermore, most of the aminoacids are themselves bulky molecules, with quite a few rotational isomeric states of their own. It is admittedly quite complex — but let's brush all these details aside, and say that two residues are either in contact, and then their interaction energy is some known quantity $\epsilon_{\mathcal{A}\mathcal{A}'}$, or they are not in contact, with vanishing interaction. One can think that we used interaction potential energy of Figure 8.1 (*a*) and further simplified it by making "rectangular" — exact zero at large distances, then vertical drop to the flat bottom of the potential well at the level $-\epsilon_{\mathcal{A}\mathcal{A}'}$, followed by the vertical "wall" signifying the excluded volume repulsion. Such simplification is very much open to criticism from a number of directions, but it is better to have some theory and criticize it than having nothing at all — so let's adopt the simplification and see what we can do with it.

As soon as we know contact energies for all pairs of amino acids, $\epsilon_{\mathcal{A}\mathcal{A}'}$, we can write down, at least formally, the expression for the total energy of a molecule having a certain sequence $\mathcal{A}(i)$ and folded into a particular conformation:

$$E\,(\mathrm{seq}, \mathrm{conf}) = \sum \epsilon_{\mathcal{A}(i)\mathcal{A}(j)} C_{ij} \ . \tag{10.1}$$

This is simple, but we have to explain the notations. First, i and j label monomers along the chain, from 1 to N. Second, sequence is described in terms of $\mathcal{A}(i)$ (or $\mathcal{A}(j)$): chemical name of aminoacid for each position along the chain. Third, and final, C_{ij} is the so-called contact matrix, it is simply 1 if i is in contact with j and it is zero otherwise. Thus, formula (10.1)

says the following: take all contacts presented in the given conformation and sum together their energies. If one imagines having composed a table of contact energies $\epsilon_{\mathcal{A}\mathcal{A}'}$ for all pairs of aminoacids (and such tables can be found in the literature), then we can compute energy (10.1).

Alas, this does not look promising on the first glance, because we remember Levinthal's lesson — we cannot compute energies of all conformations, and, therefore, cannot find the lowest energy conformation. But at least we can estimate the lowest energy itself, that is, presumably, the energy of the most stable conformation. For that, we first argue that the average energy over all conformations is proportional to N, let's say it is $\overline{E} = N\overline{\epsilon}$. This is because the total number of contacts in the globule is about N. Of course, in the coil conformations there are fewer contacts, but energies of those conformations are small and we do not have to worry about them. In fact, the number of contacts in the globule is several times larger than N (due to the fact that each monomer in the globule makes several contacts), but we can, for simplicity, absorb this extra factor in the definition of energy $\overline{\epsilon}$ — all we really need to know about this energy is that it does not depend on N. Admittedly, though, average energy over all conformations is not a very interesting quantity. The real question is how far below in energy is the lowest energy state. We now argue that this most important energy must be also proportional to N:

$$E_g - \overline{E} \quad -N\Delta \quad \overline{2s}\ . \tag{10.2}$$

where s is the factor of order unity which will be defined below, while Δ is the energy scale characterizing the diversity of different contacts, which we imagine distributed over the interval $\overline{\epsilon} \pm \Delta$. The formula (10.2) means that in the lowest energy state a significant ($\sim N$) number of contacts are the favorable ones, each having energy about $\overline{\epsilon} - \Delta$. In general, it might be difficult to find a conformation which realizes all favorable contacts and avoids all unfavorable ones. For instance, if monomers i and j are strongly attracting ones, that is, $\epsilon_{\mathcal{A}(i)\mathcal{A}(j)}$ is strongly negative, while $i \pm 1$ and $j \pm 1$ strongly attract some other completely different parts of the chain, then an attempt to realize all these contacts would create an impossible congestion, because all the monomers are connected by the chain. Factor s describes this effect, as we shall see.

Furthermore, formula (10.1) contains a hint of a different sort: this formula is a sum of many-many contributions, and such sums are subject to Central Limit Theorem — see Section 6.7, and so the value of the sum — the energy E — should be Gauss distributed. It is almost like the polymer

is again a Gaussian random walk — except this time it is "walking" along the single coordinate which is energy. Once again, this view is open for criticism, the main one being the questionable independence: it is conceivable that energies of contacts are random for random sequence polymer, but how independent are they? For instance, if one particular aminoacid $\mathcal{A}$ is in contact simultaneously with $\mathcal{B}$ and with $\mathcal{B}'$, then $\epsilon_{\mathcal{AB}}$ and $\epsilon_{\mathcal{AB}'}$ are unlikely completely independent. This is of course a serious question, but we mentioned already in Chapter 6 that Central Limit Theorem is very robust; besides, once again, it is better to have something to be improved than nothing at all — therefore, let's accept the idea of Gaussian distributed energies and see what happens[2].

As the reader may realize from Chapter 6, the Gauss distributed energy should obey the square-root law, it should typically deviate from the average by about $\sim \Delta \sqrt{N}$. Comparing this to our earlier estimate (10.2), we see that the lowest energy state is actually much lower than typical — the former is proportional to N, while the latter to $\sqrt{N}$. This means, if the energies are treated as random, then the lowest energy state must belong to the very far tail of the probability distribution. We can confirm this quantitatively. Indeed, given that energy is the sum of about N contributions which we view as independent, the probability density of the energy is proportional to $p(E) \sim \exp\left[-\left(E-\overline{E}\right)^2/2N\Delta^2\right]$. If we take one conformation and look at its energy, it is typically within the range of energies such that $\exp\left[-\left(E-\overline{E}\right)^2/2N\Delta^2\right] \sim 1$; this of course returns the familiar square root law. But if we take some $\mathcal{M}$ conformations, and look at the lowest of their energies, it should correspond to the condition $\mathcal{M}\exp\left[-\left(E-\overline{E}\right)^2/2N\Delta^2\right] \sim 1$. Remembering that the number of conformations is exponential in the chain length, $\mathcal{M} \quad e^{sN}$, and performing a few lines of simple algebra, one arrives exactly at the estimate (10.2). Thus, s in the formula (10.2) is the conformational entropy per monomer: previously, we made estimates assuming two possible states for each monomer leading to $2^N = e^{N\ln 2}$ conformations; here we are a bit more general, assuming e^{sN} conformations. Thus, s characterizes flexibility of the chain. It is then not surprising that s appears in the formula (10.2): the more flexible is the chain, the larger is

[2]We should mention here that independent Gauss distributed energies were first considered by Bernard Derrida (in the French Center for Nuclear Research in Saclay near Paris) in a completely different context, and this is known in the literature as Random Energy Model (REM).

s, the more favorable contacts can be realized, the lower is the lowest energy.

We made a good progress, but we should not forget the goal. The goal was to understand if the random sequence heteropolymer exhibits a large latent heat characteristic of real proteins. For this purpose, it is not enough to know the energy of the lowest conformation. When the globule "melts", the lowest energy conformation competes with other states. Therefore, we need to know energies of other low-lying states, for instance, the second lowest one. To this end, we just give here the result (we will hint on the origin of this result in the next Section 10.6): the gap between lowest and second lowest energies of conformations is independent of N, it is about $\Delta/\sqrt{2s}$. Naturally, it is governed by the energy scale of chain heterogeneity, Δ, and decreases with increasing chain flexibility, s (because more favorable contacts can be made also in the second lowest state).

Thus, heteropolymer is a better candidate to model proteins than a homopolymer — because heteropolymer globule does have a unique lowest energy state, at least in the coarse grained consideration. But random sequence heteropolymer is not good enough, for it does not have the latent heat as proteins do.

10.6 Selected Sequences

Our results of the previous section can be rephrased in the following useful way. Since the lowest energy conformation is about $\Delta/\sqrt{2s}$ below second lowest, random sequence heteropolymer has to freeze at some temperature $T_{\rm fr}$ such that $k_B T_{\rm fr} \sim \Delta/\sqrt{2s}$: below this temperature even the second lowest conformation is prohibitively expensive and the system can afford nothing else but staying in the lowest energy state. When first discovered, this freezing transition caused considerable excitement among the theorists working on proteins. Indeed, does not it sound like a protein — the system stays in one particular conformation, unique, albeit in a coarse grained sense?

Alas, this frozen state is not good enough — although unique, it does not stand any further tests. We mentioned already that it melts without latent heat. Furthermore, imagine that we subject our molecule to one point mutation, that is, replace one amino acid with a different species. It would mean changing several of the $\epsilon_{\mathcal{A}\mathcal{A}'}$ terms in energy (10.1), which can easily overcome the $\Delta/\sqrt{2s}$ gap, thus leading to a complete revolution: the

former ground state ceases to be the lowest energy and some completely new unrelated conformation assumes that role. Proteins are surely not like this, they have a significant mutation stability and typically keep essentially unchanged folded state after several mutations. Even worse, energies $\epsilon_{\mathcal{A}\mathcal{A}'}$ can be affected by the environment. Imagine, for instance, that you ate a pickle, the amount of salt in some places in your body changes a bit, leading to some changes in $\epsilon_{\mathcal{A}\mathcal{A}'}$ and ... to renewal of protein conformations, with subsequent loss of protein functions? No, luckily to all of us (even those who do not like pickles), this is not how real proteins work. Indeed, compare it with the opinion of Miguel de Cervantes: "It is a true saying that a man must eat a peck of salt with his friend before he knows him." (Don Quixote, Lockhart's Translation).

Thus, we need something better than the random sequence heteropolymer. But this is not really surprising. We compared folding of a protein with reading and understanding of a message. But not every message, not every sequence of letters, can be understood — meaningful ones have to be written by someone knowing the same language. Indeed, elementary school kids are taught reading and writing practically at the same time, concurrently. How to write, or design, heteropolymer sequences with protein-like properties is the subject of active current research. People try to do it computationally and experimentally, with the goal to reproduce one or several of protein properties, not only unique folding, but also mutation stability, stability against precipitation, ability to selectively absorb molecules or particles of certain shapes and sizes, and many others. Unfortunately, all of this goes far beyond the scope of this book. But we cannot resist mentioning that in 2003 D. Baker and his co-workers at the University of Washington in Seattle were able to "design" a completely artificial sequence such that when the molecules were synthesized with that sequence — they did fold into the prescribed three-dimensional structure, which was intentionally chosen to be unlike any other known in living nature[3].

In line with our desire to outline only the most basic physical principles, we will discuss just one question: how many foldable sequences are there? The full importance of this question will become apparent in Chapter 14 on the origin of life problem. For now let's just make an estimate. For that, we have to understand better the formula (10.2). We said E_g was the lowest energy (ground state) of a random sequence heteropolymer. But

[3]B. Kuhlman, G. Dantas, G. Ireton, G. Varani, B. Stoddard and D. Baker, "*Design of a Novel Globular Protein fold with Atomic-Level Accuracy*", Science, v. **302**, n. 5649, pp. 1364–1368, 2003.

what if we take two, or three, or several *different* random sequences — will their lowest energies be *exactly* the same? Of course, not. In fact, E_g given by the formula (10.2) is only the most probable value of the lowest energy. We need to work a little harder and find not only the most probable value, but the whole probability distribution of ground state energy $W(E)$; thus, E_g (10.2) is where $W(E)$ is maximal.

Determining this probability distribution requires slightly more mathematics than we use in this book. The necessary mathematical technique is called extreme value statistics. But the idea is simple. First, we will argue that ground state energy for any sequence is hardly ever higher than E_g (10.2). Indeed, for that to happen, *all* $\mathcal{M} = e^{sN}$ conformations should have energies above E_g — this is extremely improbable. This is as if you were to ask how tall is the tallest teacher in your school? Presumably teachers are hired irrespective of their heights, which creates a good case for independence. Therefore, teachers are like energy levels, except upside down — we are looking for tallest teacher, but lowest energy. Thus, what we are saying is this: it is unlikely that the tallest teacher at school is below, say, 170 cm, because for this to happen *all of the other teachers* — and there are many — must be smaller. On the other hand, the probability for the tallest teacher to be 220 cm is also small but for a completely different reason — simply because such people are rare. Similarly, the probability of the ground state energy to be below E_g (10.2) is given simply by $p(E)$ — the Gaussian distribution resulting from summation of contact energies. Since E_g is itself already in the tail of $p(E)$, we can simplify[4] the expression for $p(E)$ in this region and obtain

$$W(E) \quad \begin{cases} \frac{\sqrt{2s}}{\Delta} \exp\left[(E-E_g)\,\frac{\sqrt{2s}}{\Delta}\right] & \text{when } E < E_g \\ 0 & \text{when } E > E_g \end{cases} . \tag{10.3}$$

[4]Suppose ground state energy equals $E_g - \mathcal{E}$, that is, some interval $\mathcal{E}$ below E_g, such that $\mathcal{E} \ll E_g - \overline{E}$. Then the expression in the exponent in $p(E)$ reads

$$\frac{\left(E_g - \mathcal{E} - \overline{E}\right)^2}{2N\Delta^2} \simeq \frac{\left(E_g - \overline{E}\right)^2}{2N\Delta^2} - 2\frac{\mathcal{E}\left(E_g - \overline{E}\right)}{2N\Delta^2} \simeq \frac{\left(E_g - \overline{E}\right)^2}{2N\Delta^2} + \frac{\mathcal{E}\sqrt{2s}}{\Delta} ,$$

where we used formula (10.2) in the last transformation. Plug this into the exponential, find normalization — and arrive at formula (10.3). More accurate formula, called Gumbel distribution, reads

$$W(E) = \frac{\sqrt{2s}}{\Delta} \exp\left[(E-E_g)\frac{\sqrt{2s}}{\Delta} - \exp\left((E-E_g)\frac{\sqrt{2s}}{\Delta}\right)\right] .$$

We encourage the reader to plot it carefully alongside the approximate version to see how similar they are.

First of all, this formula allows us to return the debt to the reader and to explain our statement about the gap between lowest and second lowest energy states: characteristic scale at which $W(E)$ decays is $\Delta/\ \overline{2s}$ — exactly as we claimed before. Second, we are now in a position to estimate the number of sequences, as we planned.

For instance, if we wish to have sequences with some value of the latent heat Nq (that is, q per monomer), these sequences should deliver ground state as low as $E_g - Nq$, which means that the fraction of such sequences should be proportional to $W(E_g - Nq)$ or to $\exp\left[-Nq\ \ \overline{2s}/\Delta\right]$. These same sequences exhibit also mutation stability and environmental stability, all governed by the same energy scale Nq. Thus, only an exponentially small fraction of all sequences are protein-like in the sense of all these properties (latent heat, stability, etc). To find and select protein-like sequences is a hard work indeed! But the total number of sequences is also exponentially large, say, $Q^N = \exp[N \ln Q]$, where Q is the number of monomer species ($Q = 20$ for proteins, which gives $Q^N \approx e^{3N}$).

Therefore, we can draw a number of very interesting conclusions. First, there are no sequences with too large latent heat q (with $q > \left(\Delta/\ \ \overline{2s}\right) \ln Q$). Second, the larger is q, the fewer sequences there are, and the more difficult is finding them. Third, when q is modest, there is still an exponentially large number of sequences to choose from. Fourth, to make stable proteins the alphabet should be rich enough, i.e., Q should not be too small; for instance, the favorite toy of every protein theorist, heteropolymer with just two monomer species, is definitely not good enough. All these conclusions are of great importance to understanding the protein evolution, which we will discuss some more in Chapter 14.

10.7 Memorizing (and Confusing) More Than One Conformation

The above ideas are really far reaching. As an example of their application in a more subtle context, let's consider the possibility of a protein having not one but several distinct low energy folded states. Such phenomenon is known to biologists and medical doctors: some proteins have two or sometimes even three different folded states, and depending on the conditions (or even on chance) they can fold into either one of them. Sometimes this has bad medical consequences: in one of the folds protein performs some useful function, while in the other it causes some deadly disease, such as,

e.g., so-called "mad cow disease" (the discovery of these proteins, called prions, in 1982 was the Nobel Prize winning work by S. Prusiner).

A useful starting point to explain the physics of this phenomenon is to resort again to the idea that protein folding is the decoding of the message contained in the sequence. In this sense, protein is an information processing device, while folding process is akin to pattern recognition. But we know that pattern recognition is sometimes ambiguous. For instance, some of the sentences in our common language have more than one meaning [5]. Some visual images are also like that, as, for instance, some drawings by M. Escher, where you can see in one picture either white birds flying to the right or black birds flying to the left. Similar thing can happen also in a protein, and we are in a good position to make even quantitative estimates about it.

Note that protein sequence *can* serve as a code because the number of possible sequences, Q^N, is greater that the number of possible conformations, e^{sN}. Indeed, imagine that in a country of e^{sN} inhabitants you want everyone to have a distinct, unique last name; having an alphabet of Q characters, you must have the length of last names, N, such that $Q^N > e^{sN}$. In reality, for proteins $Q = 20$ and e^s is usually about 5; therefore, Q^N is indeed far greater than e^{sN}. This is why there is extra "information capacity" left in the proteins to memorize more than one conformation. We can imagine that in fact some part of the sequence of only $M < N$ residues is sufficient to memorize (to code for) any conformation of an N-monomer protein, that is $Q^M = e^{sN}$, or $M = sN/\ln Q$ (these M need not be in any way close to each other in the sequence). The other residues are then free to memorize another information, and if there are M or more of them — they can record a second, entirely independent conformation! Generally, the number of conformations possible to memorize is, therefore, $K \leq N/M = \ln Q/s$.

This truly exciting result was first obtained in 2001 by T. Fink and R. Ball at Cavendish Laboratory, Cambridge University in England[6]. We mention three interesting consequences of this result. First, the number of possible memorized conformations does not depend on the chain length, N. Second, given that $\ln 20 \approx 3$, we see that regular proteins can memorize

[5] An example from a scientific biography: "XYZ was a great scientist, but his addiction to smoking finally caused his death, and appropriately a meeting of ... Society was dedicated to his memory". It was appropriate to dedicate the meeting to his memory because he was a great scientist or because of his addiction to smoking?

[6] Physical Review Letters, v. **87**, p. 198103, 2001.

typically not more than two independent conformations (for the attentive reader, we emphasize the word "independent" here!). Third, we see that two letter alphabet ($Q = 2$) is not enough, two letter heteropolymers — these favorite toys of theorists — cannot be protein like, they cannot reliably memorize even a single conformation.

One can then speculate that the existing alphabet of $Q = 20$ amino acids is the result of evolutionary trade-off: having smaller Q would not allow sufficiently stable proteins (restricting the latent heat q); having a larger Q would make too many possibilities for prions.

10.8 Landscapes and Funnels

The coarse grained view of protein conformations, as we see, turned out extraordinarily fruitful. It gave us at least a glimpse of physics understanding for otherwise seemingly mysterious properties of proteins. And the potential of this approach is far from exhausted.

Nevertheless, people think of course about more detailed and more specific aspects of proteins and protein folding. A popular metaphor in the discussions of this subject is that of an energy landscape. The idea is to imagine a plot of potential energy of a protein — not the coarse grained model (10.1), but the real full potential energy — as it depends on all the internal degrees of freedom of the molecule, such as coordinates of all atoms, valence and dihedral angles, and so on. Since it is not a one-dimensional plot, one imagines a mountainous country, such as Switzerland, Caucasus, Colorado, or Chile, with peaks, ridges, valleys, passes, etc. And now one can think that folded state corresponds to one deep valley, while unfolded states perhaps form a large shallow flatland, etc.

The landscape is by no means simple. To realize that, let's mention that apart from folded (native) and completely denatured state there is also so-called molten globule state. O.B. Ptitsyn and his colleagues in Pouschino near Moscow discovered in 1982 that, in many cases, denaturation is not a globule-coil transition, but rather a transition from the native globule to the molten globule. During this transition, the blocks of the secondary structure remain stiff, but open up a little. This gives more room to the amino acids' side groups. They can now oscillate and rotate more freely. However, the globule as a whole remains stable. The openings are too narrow to let in water molecules, so the hydrophobic effect is not disturbed. There is some evidence that molten globule might be an intermediate state

on the way of renaturation. If that is the case, we can imagine that the first stage of folding is reasonably quick, it roughly repeats the scenario sketched in Figure 9.5 and ends up with a molten globule. Molten globule already has some vague features of the tertiary structure, correctly outlined but unfinished (e.g., particular positions of atomic blocks of the secondary structure, etc.) The second stage takes then much longer. This is when all the structural details are properly set (e.g., positions of individual atomic groups, etc.) This is a possible scenario, but likely not the only one.

One important conclusion one may draw from thinking about the landscapes is the danger of local minima, which can serve as traps slowing down the finding of the minimal energy native state. To emphasize the necessity of sliding down without significant traps and barriers, yet another metaphor is popular, that of a funnel: the landscape should be funnel shaped in order for the native state to be readily accessible.

Landscape and funnel are nice images to keep in mind when thinking about proteins. The trouble with them is that they exist in the space of many-many dimensions. We know that the image of a mountainous country (two-dimensional landscape) is much more rich than just a curve with maximums and minimums (one-dimensional landscape). It is reasonable to think that things become proportionally more rich every time that we add one more dimension. The complexity of landscapes arising when the dimension becomes as large as proportional to N is tremendous, our brain is not equipped to imagine it in any useful way. In this sense, landscapes and funnels are even dangerous, for they often create a deceiving illusion of understanding where real understanding is still elusive and very far. In practice, whenever people say they use landscapes and funnels, they in fact also use something else — they exercise physical intuition and guess the few most important degrees of freedom to construct the imaginable low dimensional landscapes. This is easy to say, but very difficult to realize, moreover, it is not always possible, and there are no systematic ways to do it...

10.9 Nucleation, and the Resolution of Levinthal's Paradox

Since we know from the coarse grained considerations that the protein-like sequences provide for a very low lying ground state energy, corresponding obviously to the very deep valley, we can use the landscape metaphor to hint on the resolution of the Levinthal paradox (see Section 10.3).

To explain the idea, let's point out that a similar paradox could be formulated for a more trivial case — formation of a crystal, say, crystallization of water: how do molecules find their unique positions while they obviously cannot manage to try all of the exponentially large number of possible arrangements? Indeed, there are exponentially many e^N configurations, so an attempt to test them all would require an exponentially long time.

How is this later paradox resolved? In general, one useful way to resolve a paradox is to reformulate it. In this case, we can view the exponentially long time as arising from crossing a very high free energy barrier. Indeed, let's remember boiling water for the morning coffee and think how liquid is transformed into gas. Obviously, liquid and gas are vastly different in density, and we can, therefore, try plotting a "free energy landscape" as a function of density. We should remember that in liquid phase molecules are close to each other and enjoy the attractive interactions; this is energetically favorable. By contrast, in the gas phase, molecules enjoy the freedom of independent relative motion, this is favorable entropy wise. Let's now start from liquid phase and decrease density a few percent; the system will loose virtually all of the energetic advantage, because molecules are not close enough any more, and will not gain entropy anywhere close to compensate the energy loss. It means, free energy will be quite large, larger than in either liquid or gas phase — which means there will be an energy barrier. Importantly, both energy loss and entropy disadvantage in this scenario happen to each and every molecule in the system, which means the height of free energy barrier will be proportional to N, the number of molecules. Given that the time to cross the barrier is exponential in barrier height, we arrive at the conclusion that liquid-gas transformation should take an exponentially long time... This is another formulation of Levinthal's paradox, and it is an obvious absurdity: we do know that water does boil in a not so long time!

The resolution of this later paradox is well known: crystallization, condensation, boiling and every other phase transformation of this type (so called first order) occurs via nucleation. Indeed, rain water falls in little droplets instead of the whole cloud suddenly condensing and falling on us all at once. This means, plotting free energy as a function of density was a bad idea; the landscape constructed this way turned out dramatically misleading — and in this example it is pretty clear why: the system can circumvent the high barrier by moving around it in a multi-dimensional space. The landscapes should not be taken lightly! But with nucleation in

mind things become clear: in crystallization, for instance, only a modest size nucleus of the crystal has to be formed, the corresponding barrier height is determined by the size of the nucleus instead of the whole system, and once the nucleus overcomes the barrier — the rest is the downhill process in the landscape language. In other words, the system does not perform random walk in search for the crystal state, but rather it is forcefully driven there, because every new arriving molecule is forced to the right place.

Similarly then one would expect the landscape of a foldable protein to feature not only a very deep valley with the ground state at the bottom, but also a funnel shaped region around this deep valley, such that molecule slides straight down, not meeting significant energy barriers on the way and not encountering large flatland to do the time consuming random walking (the later also corresponds to entropy barrier).

We should emphasize that completely denatured coil-like protein chain, when it is placed under the conditions of folding (e.g., temperature at which folded state is stable), has to perform some search through a flatland (or it has to overcome an entropic barrier), but it does not search for the particular ground state conformation (which would be prohibitive according to Levinthal). Instead, it searches for a very large region in the space of conformations, the region which is the basin of attraction of the ground state, or, for those who like the funnel metaphor, the upper entrance of the funnel. In other words, the height of the barrier is not determined by the size of the entire protein (which would correspond to the Levinthal's paradox), but rather by the much smaller nucleus size.

10.10 *In vivo*, *in vitro*, *in virtuo* ...

Not only in the field of proteins, but virtually in any part of biological physics scientists are usually not unanimous even about how to begin. Some say, "To learn about living things, we have to study them while they are still alive". This is called *in vivo*. Meanwhile, others argue: "Life is too complicated. We will understand nothing if we don't experiment with relatively simple pieces of living nature in the lab." This is known as *in vitro*. The discussions have been going on for decades, causing mutual irritation, as well as progress on both sides. However, nowadays a third way has appeared, which could be called *in virtuo*. We are talking about the

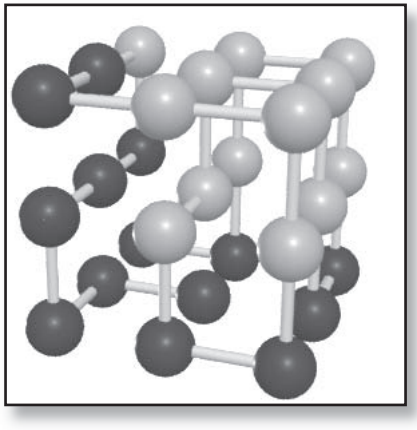


Fig. 10.2 Heteropolymer of 27 different monomers can fill a $3 \times 3 \times 3$ domain on the cubic lattice.

so-called "virtual reality" that does not exist anywhere but in the memory of a computer[7].

Let's give an example from a familiar field. Imagine a "polymer" that consists of 27 units (monomers). Suppose the monomers can be positioned at the vertices of a cubic lattice (Figure 10.2). In a close-packed state, such a polymer would occupy the volume of a $3 \times 3 \times 3$ cube (this is why we chose the number 27 in the first place). Now, it turns out that a "27-mer" can be arranged on a cubic lattice in surprisingly many ways. There are about a hundred thousand possibilities (103,346, to be precise)! So you could say that there are 103,346 different globular conformations. Although the number is large, it is still possible for a computer to "churn out" the energies of the polymer in all its conformations.

How can this be done? Say, there are monomers of Q different sorts (we can think of them as different "colors"). How do they interact? Let's assume that, if two monomers of the same sort happen to be "neighbors" on the lattice, they attract each other with a certain energy, $-J$. Meanwhile, two neighboring monomers of different sorts repel each other with an energy $+J$. We could also look at various other kinds of interactions.

[7]Some people call it *in silico*, because present day computers use silicon-based semiconductor materials. We believe that the idea of computational studies does not depend on the particular hardware; maybe, there will some day be other, non-silicon based computers? What could they be based on? For this book, we must mention that there are polymer-based organic semiconductors — maybe they will replace the silicon! Thus, we prefer to use *in virtuo*.

As a result, we would find that, for very many primary sequences, one of the conformations has a much lower energy than any of the others.

The reader should of course notice that this lattice model is very closely related to the coarse grained view of proteins that we discussed earlier in this Chapter. Physicists often call lattice models toys — but this is actually just a joke of restraint: humor is the best way to avoid pompous seriousness which is incompatible with science. Lattice proteins is a very serious business. They are used to test the theories, to get the hints on how to improve the theories. For instance, we mentioned that only properly selected sequences are foldable; this idea was tested with lattice models and beautifully proved right.

Computer models (i.e. experiments *in virtuo*) can take us even farther, but not really that far, unfortunately. Suppose you wish to explore a longer polymer. The "magic" numbers that you have at your disposal are the following: 36, 48, 64, 80, 100, 125, ... A "36-mer" can be close-packed on a $3 \times 3 \times 4$ lattice, a "48-mer" can be fitted on to a $3 \times 4 \times 4$ lattice, etc. No modern supercomputers have managed even to list all the possible states of a "64-mer". And you do need to list them all, before trying to discover the state with the lowest energy. Nature may have learned somehow to disregard Levinthal's paradox, but we and our computers have not yet! A "48-mer", for example, has as many as 134,131,827,475 close-packed states. This is too big a number to handle, in terms of calculating all the energies. Thus, a "36-mer", with its 84,731,192 close-packed states, and a "27-mer" are the only two models that are manageable so far. They (along with some models on a flat surface) have become the workhorses of *in virtuo* studies of lattice proteins.

Not only the difficulties, but also the possibilities of the model grow really fast with the chain length. For example, "36-mer" is the shortest of the "magic" once that can have a knot, as shown in Figure C11.7.

The models can be improved in a different way. Look at the picture Figure C10.3. Here is a lattice globule with a pocket, where we can put a "substrate". Figure C10.4 shows that this globule is able to "renature," i.e. to self-assemble its correct structure. In this sense, it is indeed similar to a real protein molecular machine. Moreover, one can easily imagine the whole machine shop, as in the Figure C10.5. In a way, this reminds us of a well-known joke about a theoretician who decided to do biology. This is how he started: "Suppose a horse has the shape of a cube, with a 1 m edge, and weighs 1 kg..." Indeed, just compare Figures C5.14 and C10.3! Well, whatever you say, sometimes even a cubic horse might be useful.

Sure enough, computational studies of proteins are not restricted to cubic (or any other) lattices. Off-lattice, the most ambitious approach is to do the all-atom simulation, integrating Newton's equations of motion for all atoms in a protein along with the thousands of surrounding water molecules (this is called molecular dynamics). Unfortunately, in the straightforward approach, even the most powerful of modern supercomputers can only follow the averaged size protein for at most several nanoseconds — about six orders of magnitude short of a typical folding event, and this is not to mention the host of other difficulties, like the choice of potentials, the role of electrostatic forces, etc. People use a large variety of inventive strategies to overcome these problems. One nice idea is to use thousands and thousands of idling computers around the world — the so-called Folding@Home program; we encourage the reader to visit the web site of the program and to see for him- or herself.

Thus, together with the familiar *in vivo* and *in vitro* experiments, studies of proteins *in virtuo* are going on as well at full speed.

10.11 Do We Understand Protein Folding?

Let's summarize: do we understand folding?

In the opinion of a maximalist, the folding problem is the one of *prediction*: if we knew the primary structure, could we in principle predict the tertiary one? As we mentioned before, the latter question is a practical one. If the answer were "yes", we would no longer need complicated and expensive X-ray and NMR analysis of proteins. Unfortunately, the ideal full prediction is far from being true despite impressive achievements in some cases. At present, we can guess the secondary structure with decent accuracy (i.e. α- and β-segments), but predicting the tertiary structure remains elusive. In this sense, protein folding is definitely not resolved.

An analogy might be useful here. At the dawn of computer era people started working on computer chess — trying to instruct computer to play chess. In fact, the overarching goal was to learn how human intellect is working. Eventually computers became so powerful, they became able to memorize so many positions and parties played by great chess masters, that a specially build machine won a match against greatest of human chess players, Kasparov. In our opinion, this is a highly unfortunate outcome: the interest in computer chess is lost perhaps forever, while the goal to understand human intellect remains as elusive as before — and the pursuit

of this goal is, therefore, left for other approaches. Similarly, protein structure prediction may some day be solved by brut-force computing. If this happens, it might become a useful achievement, but, unfortunately, it may contribute nothing to the understanding of fundamentals.

But in terms of these fundamentals, we would dare to say that an important progress had been achieved over the last several years. Previously, at the time of Anfinsen, protein folding seemed a fundamental physics mystery. People could not imagine how it could be happening even in principle. Now, there is at least an overall understanding of the basic physics behind folding, as we tried to outline in the present chapter. There are lattice models, which are very much unlike proteins in many respects (too many and too obvious to list) — but which are like proteins in two most important aspects: they have the same fundamental difficulties, such as Levinthal paradox, and they do fold. And we understand this model pretty well! But, of course, the study continues in many directions...

10.12 Wooden Toy

Some people like brain teasers, and there are special stores (including some on the internet) selling puzzles of various kinds. We want to describe here one of such puzzles, called "snake cube", as it presents a surprisingly deep analogy with protein folding problem. Sure enough, the analogy is limited, and the reader should certainly exercise his or her sense of humor — but the story is interesting. Here is how it goes.

The toy consists of 27 equal size wooden cubic blocks, usually about 1 centimeter each, connected as a snake (you can Google it as "snake puzzle"). Each pair of neighboring cubic blocks in the snake is connected by an axis around which both of them can rotate freely; axis connects centers of cubic blocks, and its length ensures that neighboring faces touch each other. The rotation around these axes is similar to the rotation around σ-bonds in real polymers (see Chapter 2). Due to this rotation the snake as a whole is "flexible", it can be easily shaped into zillions of different shapes, or conformations; in that it is also similar to a polymer chain with rotational isomer mechanism of flexibility. In particular, among all other conformations, the snake can adopt a fully folded conformation in the shape of a $3 \times 3 \times 3$ cube. This is illustrated in the Figure C10.6, where photographs are presented of both unfolded (left) and folded (right) conformations.

The toy is subject to Levinthal paradox: it is virtually impossible to test all of its conformations, there are too many of them. Some people find the way to fold the toy quite easily; others, some of them very clever, cannot do it easily or fail altogether. We do not know what kind of mental ability controls the toy folding success, and, like in proteins, we do not know how to formulate the algorithm leading to successful folding and beating the Levinthal estimate.

What makes this toy exciting for us is the fact that it has a sequence. Apart from the two ends, where cubic blocks have only one connection axis each, the two types of "monomers" assembled along the chain are shown in the lower panel in Figure C10.6: they are blocks with two axes for two neighbors going either in opposite directions on two opposite faces (one type), or forming the 90° angle on two adjacent faces (another type). The chain in the corresponding places is straight (first type) or makes a 90° turn (second type), accordingly, we call these two types of monomers $\mathcal{S}$ and $\mathcal{C}$, respectively. The specific sequence of $\mathcal{C}$ and $\mathcal{S}$ for our particular toy can be read off the left panel of the Figure C10.6. Interestingly, we examined several copies of the toy, manufactured in a number of different countries, and all of them, although different in sizes, color, and material, have nevertheless the same sequence. This seems highly unlikely to have happened by chance, for the number of possible sequences is too large, almost 17 million.[8] It is more plausible that someone designed the first toy, and then people copied it (similar events can also happen in biological evolution of protein sequences — see Chapter 14).

This leads us to the question: are all of the possible almost 17 million sequences foldable? The answer is, of course, no. For instance, if there are even two $\mathcal{S}$ in a row anywhere in the sequence — such snake cannot fit into the $3 \times 3 \times 3$ cube. It is not completely trivial, but possible to calculate the number of sequences having no double-$\mathcal{S}$ places; the result is substantially smaller than 17 million, namely, it is 98, 514. Of course, this is only an upper bound for the number of foldable sequences, because there might be some other less obvious restrictions. This situation is similar to the

[8] At the first glance one may think that the number of sequences is 2^{25}, because for the two blocks at the ends there is no choice, while for all other 25 positions there are two choices for each. However, this way we double count, because we include for every sequence also its reverse, so the closer estimate would be 2^{24}. In fact, we double counted all of the sequences except the palindromic ones. The number of palindromes is 2^{13}, because we have to choose arbitrarily only first 13 monomers, while the other 12 are determined uniquely to make a palindrome. Therefore, total number of distinct sequences is $\frac{1}{2}\left(2^{25} - 2^{13}\right) + 2^{13} = 2^{24} + 2^{12} = 16,781,312$.

real proteins: not all of the sequences are foldable, and in many cases we can formulate some "grammar rules" (similar to "no $\mathcal{S}-\mathcal{S}$ pairs" above) substantially reducing the estimate of the number of foldable sequences down from the enormous 20^N. In real proteins, it is difficult to identify *all* such rules.

Another similarity, as well as the difference with real proteins become apparent if we compare the number of foldable sequences with the total number of possible conformations. The former number, as we have just seen, is not greater than $98,514$; the latter number (conformations of the lattice 27-mer filling the $3\times3\times3$), as we mentioned already (Section 10.10), is $103,346$. The conclusion is that there are some sequences which can fold into two (or maybe even more) different conformations! This is indeed similar to the story we told in Section 10.7. Of course, the analogy with proteins is limited, for instance, in proteins the number of sequences is larger than the number of possible conformations. Still, the snake cube is a great fun for someone thinking about proteins!

Color Figures for Chapters 9–10

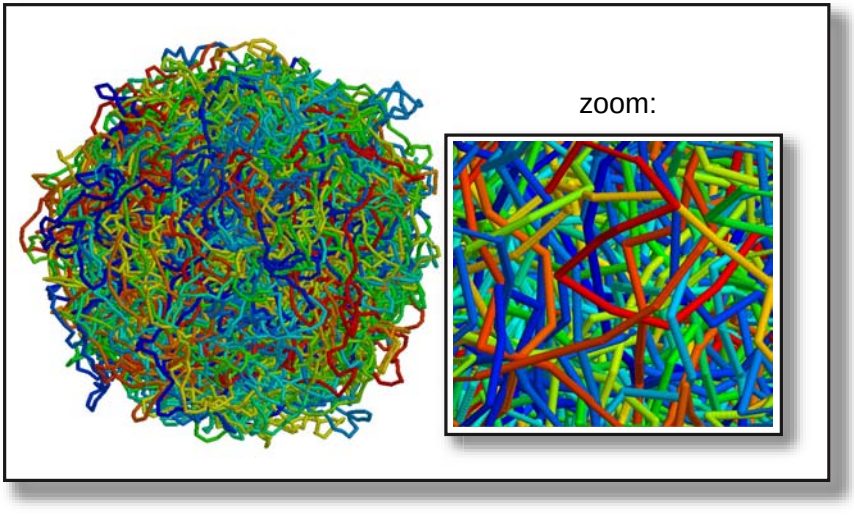


Fig. C9.1 A computer simulated globule of a long chain. Notice that the globule is pretty accurately spherical, its surface consists of loops, while its interior, particularly well seen in the zoomed part, reminds a concentrated solution of different chains (compare Figure C12.3) — even though in reality they are all distant parts of the same chain. The chain is a homopolymer in terms of chain flexibility and monomer–monomer interactions being the same for all monomers. However, to help the eye, the chain is colored, smoothly going through the rainbow colors from one end to the other (e.g., one end is red and the other is violet, with all intermediate colors in between). What one should notice is that any particular color is not located in a particular region of the globule; just the opposite, every color is reasonably uniformly distributed throughout the globule, and the local surrounding of any monomer is full of all sorts of different colors, confirming that very distant parts of the chain form contacts in the globule. The figure is courtesy of L. Mirny.

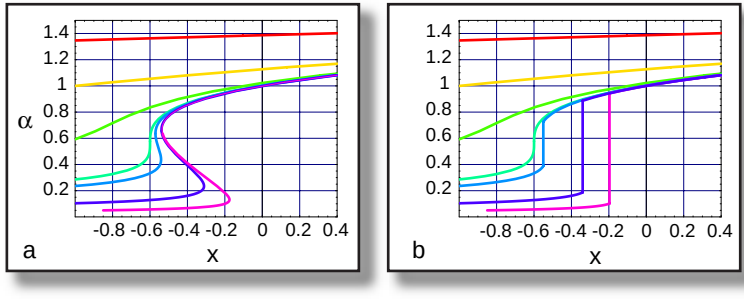


Fig. C9.3 The dependence $\alpha(x)$ given by Equation (9.7) for different values of y; from top to bottom, the curves correspond to the following values of y: 10, 1, 0.1, 1/60, 0.01, 0.001, 0.0001. Here α is the swelling parameter, that is, the ratio of the actual polymer coil size to ideal coil size; $\alpha < 1$ corresponds to chain collapse, or formation of a globule. Parameters x and y are defined such that x is controlled by the solvent quality and chain length ($x \sim BN^{1/2}/\ell^3$), while y is determined by the chain stiffness ($y \sim C/\ell^6$); small values of y correspond to rather stiff chains. In the panel (a), the dependencies $\alpha(x)$ for some values of y are multivalued for certain interval of x (i.e., they have van der Waals like loops). In the panel (b), one solution is selected for each x, such that the values of $\alpha(x)$ correspond to the absolute minimum of free energy for every x.

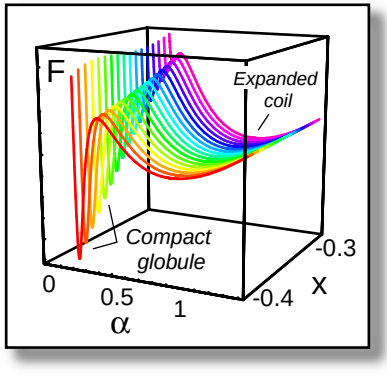


Fig. C9.4 The dependence of free energy $F(\alpha)$ on the swelling parameter α in the case where $\alpha(x)$ is multi-valued function of x, characterizing the solvent quality. As x changes (which can be controlled by, say, temperature change), the shape of $F(\alpha)$ dependence changes such that one minimum is getting deeper on the expense of the other. Deeper minimum corresponds to the more stable state. For this figure, we choose the value $y = 0.001$.

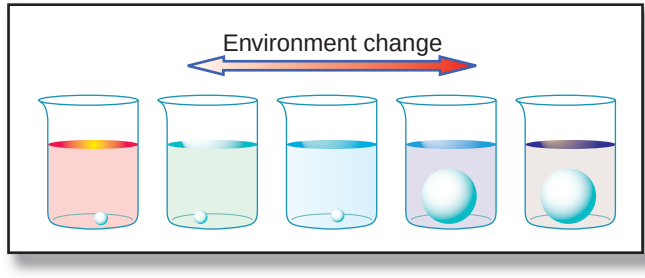


Fig. C9.6 A cartoon showing gel collapse and swelling upon change of environment conditions, such as, e.g., solvent composition, temperature and so on. The figure is courtesy of T. Tanaka.

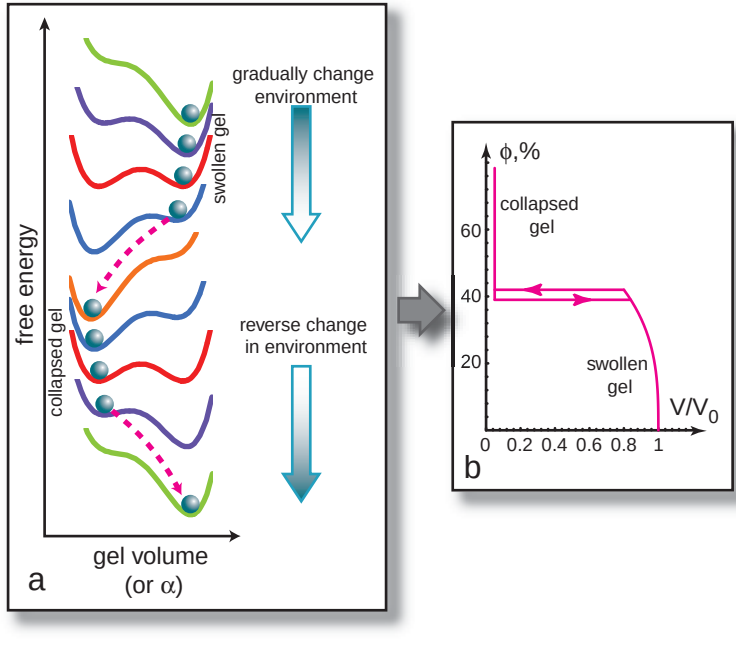


Fig. C9.7 (*a*): A cartoon showing how free energy profile changes with the change of environment, and how phase transition occurs when the barrier between states becomes very low. (*b*): The volume of a polyacrylamide network in a mixture of acetone and water, as a function of the percentage of acetone. (V_0 is the volume that the network had when just prepared.) The figure is courtesy of T. Tanaka.

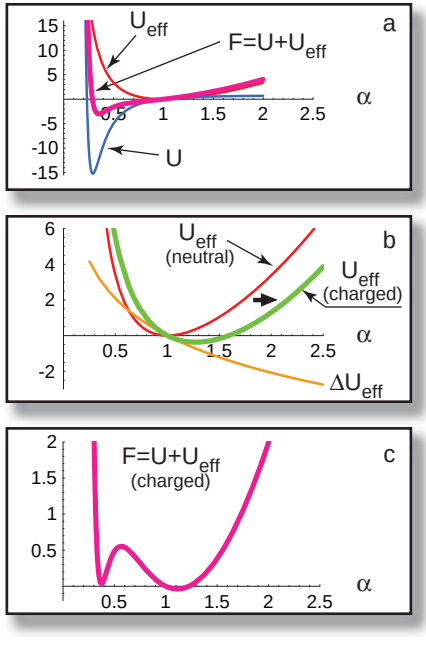


Fig. C9.9 (*a*): The dependencies $F(\alpha)$, $U_{\text{eff}}(\alpha)$ and $U(\alpha)$ for a neutral network. (*b*): The change in $U_{\text{eff}}(\alpha)$ when the network acquires an electrical charge. (*c*): $F(\alpha)$ for a charged network.

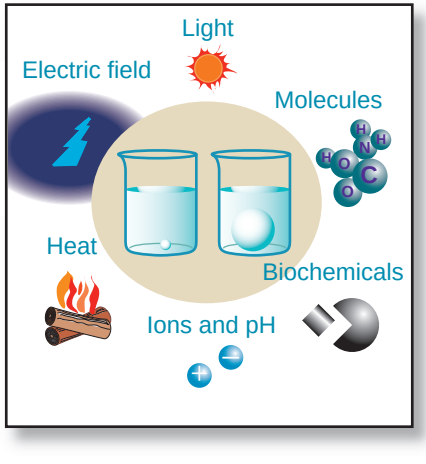


Fig. C9.10 A variety of factors that can cause gel to collapse. The figure is courtesy of T. Tanaka.

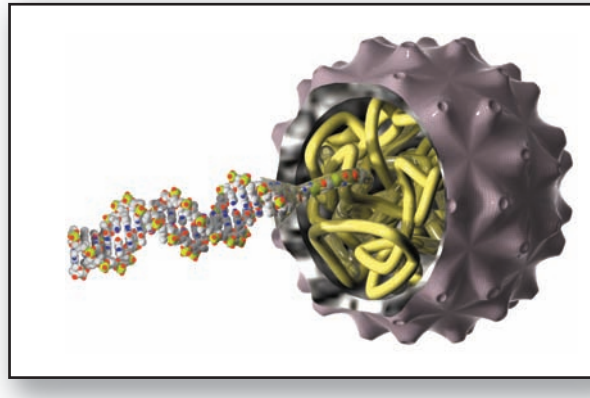


Fig. C9.11 A cartoon of dsDNA packing into a virus head. The double helix, shown as a worm-like polymer inside, is packed very densely, forming a globule. Figure is courtesy of P.G. Khalatur.

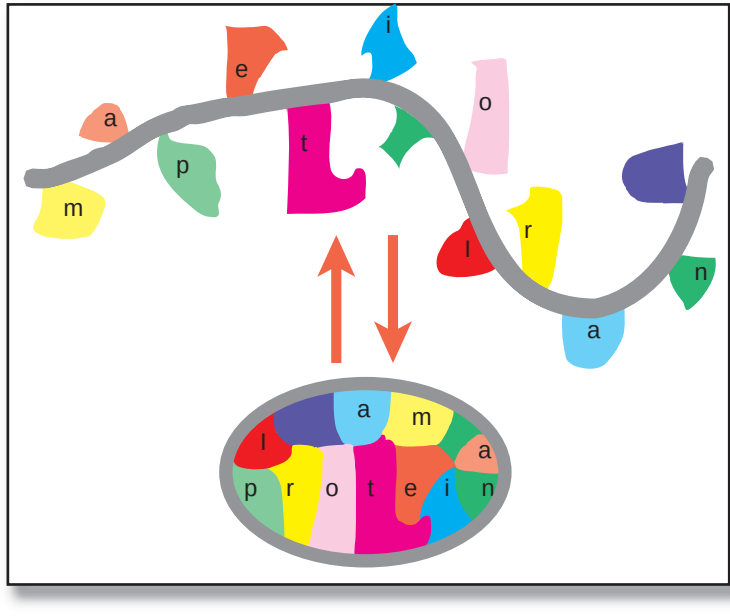


Fig. C10.1 This cartoon does not pretend to be very serious, but it explains why we compare the protein folding transition to "reading with understanding", or to the decoding of a message. Indeed, the extended chain is a meaningless string of letters, but when it has been folded correctly, it clearly states: "I am a protein". Notice that it may also fold incorrectly to ask: "Am I a piton?" — particularly if *r* and *e* masquerade themselves as a question mark.

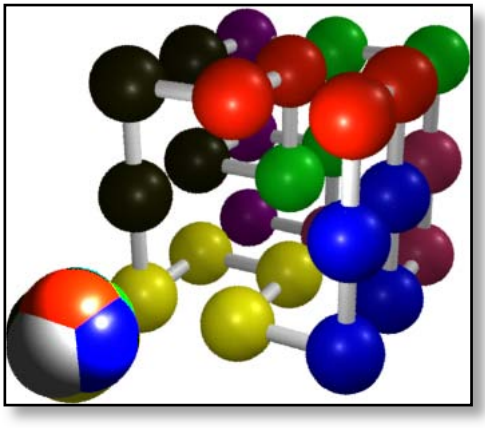
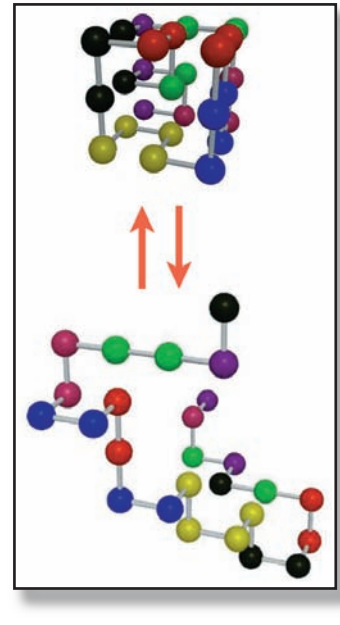


Fig. C10.3 Lattice globule with an "active site" capable to specifically recognize a "target molecule."

Fig. C10.4 If, for some reason, the globule shown in the Figure C10.3 is denatured, it is able to renature back, with restoring the correct "active site". It is important that it does not need any assistance to do that. In this sense, it is indeed like a molecular machine. The figure is courtesy of V.S. Pande.



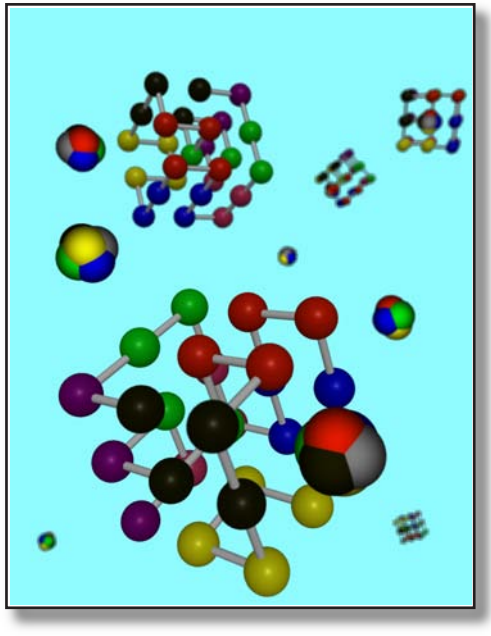


Fig. C10.5 Somewhat fictitious picture presenting many molecular machines working in the water medium. The figure is courtesy of V.S. Pande.

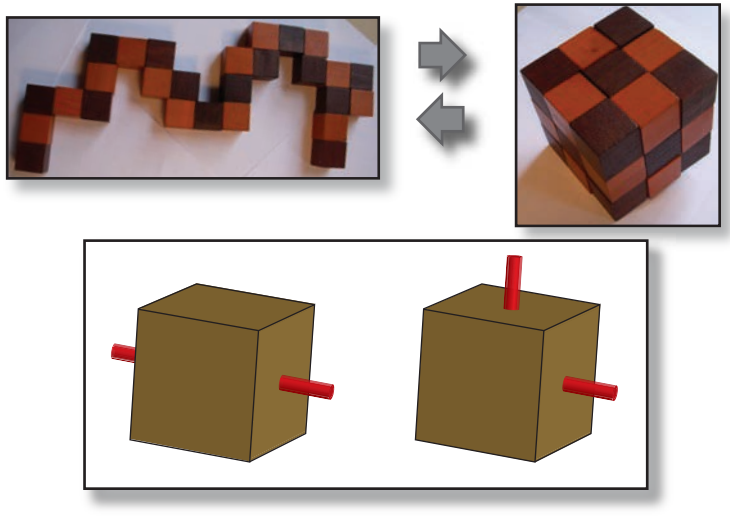


Fig. C10.6 Wooden Toy. The upper panels are the photographs of the toy in unfolded (left) and folded (right) conformations. The lower panel presents two types of cubic blocks in the toy, called $\mathcal{S}$ and $\mathcal{C}$ in the text, because they give rise to straight pieces and corners, respectively.

Chapter 11

To Knot or Not to Knot

My soul is an enchanted boat ...

Percy Bysshe Shelley (1792–1827),
Asia: From Prometheus Unbound.

We bet every reader of ours have been annoyed — and perhaps more than once — by a rope, or a thread, or a fishing line tangling and knotting out of control. Does not this also happen to molecular "ropes" — polymer chains? Do they spontaneously knot? This question was first asked by Max Delbrück (1906–1981) in the DNA context in 1962 and, independently, by Harry Frisch (1928–2007) and E.Wasserman for regular polymers in 1961. In this chapter, we will discuss what is known about knots in polymers, but first we have to digress into the exciting history of the subject.

11.1 Knots in Physics: What are Atoms?

Knots entertained people's imagination since the time immemorial. People used them for all sorts of purposes, as, e.g., in the legend about Gordian Knot and Alexander the Great, and people did observe knots in nature, as, e.g., some medieval sources mention finding knots in the umbilical cords of some babies[1]. But the start of the scientific study of knots can be dated pretty accurately to the year 1867, when William Thomson (1824–1907, later to become knighted and named Lord Kelvin for his role in the construction of Transatlantic telegraph cable) was thinking about the

[1]Modern statistics indicates that as many as about 1% of all newborn babies have their umbilical cords knotted; some sources even suggest that these babies statistically tend to have somewhat higher IQ later on in their lives.

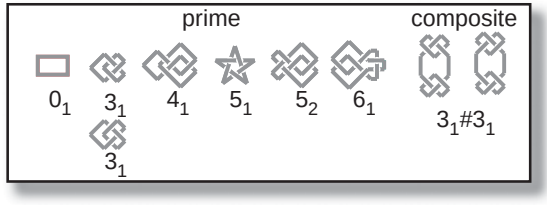


Fig. 11.1 Several simple knots. For the trefoil knot 3_1, two isomers are shown, they are mirror impages of one another. Also, two distinct composite knots are shown (called *granny* and *square*), combining different trefoil isomers.

nature of atoms and came up with the following idea. Remember that at the time perhaps the biggest puzzle in the whole of physics and chemistry was why there are discrete species of atoms, that is, why there is no intermediate forms between hydrogen and helium, between helium and lithium, and so on. At the time people imagined space filled with ether — fictitious fluid whose waves and ripples carry electric and magnetic forces. Thus, Thomson imagined several "if's": *if* there are vertex lines, like little tornadoes, in ether; and *if* ether is an ideal liquid, then the two vertex lines cannot cross (i.e., would we say now in this book, vertex lines behave like polymer chains); and *if* vertex lines could be closed — then there could be several sorts of species, sketched in Figure 11.1. And the nice thing is that these objects are definitely discrete — there are no intermediate forms. Perhaps we can speculate that the knots presented in our Figure 11.1 could be the atoms of hydrogen, helium, lithium ... and so on, all the way down the periodic table of chemical elements.

Isn't it a nice idea? In our opinion, it is extremely beautiful; it is a pity it did not prove true, but maybe it will play out at some other level some time, maybe in the theory of strings in subatomic particles — we can only hope that such beauty would not be wasted. "Nothing of what is nobly done is ever lost." — Charles Dickens.

Returning to the story, W. Thomson got understandably excited and asked his friend and collaborator P.G. Tait (1831–1901) to make a table of possible knots and also try to compute the frequencies at which the knotted strings could oscillate — maybe, they hoped, it could explain the atomic spectra? Tait worked hard, made first large table of knots and formulated several conjectures about classification of knots. The excitement among physicists continued for some years, but eventually nobody lesser than James Clerk Maxwell (1831–1879) grew sceptical, for there was no experimental support for the idea, and in 1878 he wrote in his letter to Tait:

My soul is an entangled knot,
Upon a liquid vortex wrought
By Intellect in the Unseen residing.
And thine doth like a convict sit,
With marlin spike untwisting it
Only to find its knottiness abiding;
Since all the tools for its untying
In four-dimensional space are lying.

It is rather obvious that Maxwell was paraphrasing Shelley — see our epigraph to this chapter. Apparently, these lines by Shelley were a sort of cultural cliche at the time in Victorian England; a popular song by M.V. White and a famous painting by Walter Crane were both called "My soul is an enchanted boat", so Maxwell was not alone to adore it. But what is even more interesting is that Maxwell thought about spaces of the dimension other than three — even though he still did not know about non-integer dimensions (see Chapter 13).

The work by Tait jump-started the mathematics of knots. But as far as physics is concerned, knots went largely off the physics horizon for almost a century, until Max Delbrück and Harry Frisch, in the beginning of 1960s, revived the interest in knots in a completely new context of polymers and biopolymers.

11.2 Table of Knots

Although we do not plan to delve into mathematics, a few words are necessary about classification of knots, just to introduce terms. The problem is that there is an incredible variety of different types of knots. Traditionally, starting from Tait, they are presented in the form of tables, similar to the one shown in the Figure C11.2.

Knots are presented by their two dimensional projection indicating the over- and underpasses. Surely, every knot can be drawn in many different ways, with the different numbers of crossings, but the table uses for each knot the projection with minimal possible number of crossings. For instance, an unknot and a trefoil have minimal crossing numbers 0 and 3, respectively.

Each knot in the table is denoted by its minimal crossing number. In most cases, there is more than one type of knot with any given minimal crossing number; for instance, two knots with five crossings. These different

types of knots with the same crossing number are then labeled by different values of the index. The choice of index has no significance and is determined by the tradition only. Thus, there are knots 5_1 and 5_2 with five crossings, while an unknot and a trefoil are denoted 0_1 and 3_1, respectively.

Some knots are chiral, i.e., knot is different from its mirror image. For instance, there can be two distinct types of the trefoil, left and right, they cannot be transformed into one another continuously. Other knots, such as 4_1, are not chiral, they are mirror symmetric.

Table of knots includes only prime knots. Obviously, several knots can be tied on a single rope, in which case one talks of a composite knot (this is very similar to prime and composite numbers). Figure 11.1 shows two ways to combine two trefoils. Importantly, one can easily convince oneself that there is not such thing as an anti-knot or knot annihilation: given any knot on the rope, one cannot tie another knot on the same rope such as to make the composite knot an unknot (in this sense composition of knots is like multiplication of integers, there is no analog of division, therefore, no inverse).

Equipped with this terminology, we can return to physics.

11.3 Are Knots Common?

Delbrück and Frisch hypothesized knots in various polymers based on the common sense considerations: all other known strings tangle, why not polymers? Indeed, we should remember that polymer chains are not phantom (see Section 2.6) and cannot cross. While accepting this logic, people thought that knots is perhaps something exotic for polymers. Although general expectation was that the probability of knotting, $P_{\rm knot}$, must be small, for a while nobody could suggest any practical method to find out for sure, neither experimentally nor theoretically. The difficulty is, given a chain, to find out if there is a knot or not, and, if yes, then which kind. The problem is aptly illustrated by the Figure C2.7: the reader is invited to guess whether the DNA spilled out of a bacteria is knotted or not... Indeed, polymer coil typically is a complex random shaped object, a mess if we look at it — how can we find out whether it is knotted or not?

The breakthrough came around 1970 mostly due to the work of M.D. Frank–Kamenetskii and his young co-workers A.V. Lukashin (1945–2004) and A.V. Vologodskii in the Molecular Genetics Institute in Moscow. These researchers realized that a piece of very abstract mathematics, called

algebraic topology (or the theory of polynomial topological invariants) and thought totally useless for any applications, is actually perfectly suited to recognize knots by a computer. Therefore, if one could generate many-many closed random walks on a computer, one could measure what fraction of them is knotted. And generation of these "polymer conformations" is not really that difficult: we discussed this procedure already (see Section 2.4) for the linear chains; it is not completely trivial, but possible to modify the algorithm such that it produces closed trajectories, coming exactly to the beginning after the given number of steps, N.

We should mention that polynomial topological invariants were invented in 1920s by Princeton University mathematician James Waddell Alexander II (1888–1971). What are these invariants? Theoretically, given any continuous and not-self-intersecting closed curve in three-dimensional space, one can compute a polynomial such that this polynomial would not change upon any continuous deformation of the curve, but would change upon crossing — that is, this Alexander polynomial is a topological invariant. For instance, for any shape of the trefoil knot the polynomial is $\Delta(t) = t^2 - t + 1$ while for the unknot it is $\Delta(t) = 1$: whatever the shape of the unknot, however bent and crumpled it might be, its Alexander polynomial is guaranteed to remain $\Delta(t) = 1$; but the moment two pieces cross and the knot becomes a trefoil, its polynomial changes to become $\Delta(t) = t^2 - t + 1$. And what is t here? Nothing! It has no physical relevance whatsoever. Alexander $\Delta(t)$ is just an abstract algebraic object... This is the type of mathematics to which most physicists are usually deaf. But science teaches us time after time that prejudice of any sort is a bad advisor, that a scientist, to deserve the name, should keep his or her eyes open... In our story, the researchers in Frank–Kamenetskii group realized that their computer could work out Alexander polynomial for several values of t for every loop generated and, for instance, if the result was $\Delta(-1) = 3$, then the loop is most likely the trefoil. (Most likely and not for sure because some other rather complex knots also have the same Alexander polynomial as the trefoil; also, Alexander polynomial does not distinguish left from right; this was not a significant problem, so let's skip it here.)

That was truly an exciting idea. For the first time instead of vague qualitative arguments people started measuring knot probabilities quantitatively, as concrete numbers. And sure enough, knots were found in random closed loops — in small but perfectly noticeable and measurable quantities. When plotted against the chain length N, the knot probability P_{knot} showed clear, more or less linear, tendency to increase. This was an

encouraging sign, for it promised a higher harvest of knots in longer loops, but it was also a very disturbing sign: if it keeps increasing, it will soon hit 100%!

The first simulations tested loops up to about $N = 70$. As computers were rapidly getting more powerful, longer and longer loops were studied, and soon it became clear that $P_{\rm knot}$ does indeed approach 100% with increasing loop length. Knots are not exotic, but rather typical for long enough polymers. Tacitly the mood and interest of researchers flipped the sign, and the question became now about the unknots — how common are they in long loops, or what is the probability $P_{\rm unknot} = 1 - P_{\rm knot}$. Year after year, loops of up to about $N = 3000$ have been studied, and a simple looking result emerged:

$$P_{\rm unknot} = e^{-N/N_0} \ . \tag{11.1}$$

Thus, unknots are exponentially rare in long loops, almost all long loops are knotted.

But what is N_0 — a parameter appearing in formula (11.1)? Well, formally it is a characteristic chain length at which the probability of unknot decays by a factor of e. It depends, as it turns out, on the chain mechanism of flexibility, on its excluded volume, and perhaps on other properties. But in any case it is surprisingly large. It is about $N_0 \approx 250$ for very thin freely-jointed chain of straight segments, and it becomes even much larger for the "thick" chains (with excluded volume), reaching $N_0 \approx 200,000$ for the chains on a lattice. In this sense, knots become dominant only for very very long loops, particularly when excluded volume is significant.

This was about knots in coils. As excluded volume reduces knotting, the "negative excluded volume" — the poor solvent effect or globular conformation — dramatically increases the knot probability, which is not inconsistent with our mundane experience with ropes and the like.

This prompts us to make another remark, to avoid disorienting the reader. Among present day professionals the greatest familiarity with knots, perhaps, can be claimed by surgeons (with fishermen and mountaineers close behind). For obvious reasons, surgeons are taught to tie reliable knots which will not loosen under any circumstances, even if the stitches become wet and slippery. This hints on the important role of friction in such macroscopic knots. We are not talking about this friction, dealing with molecular knots only; for instance, DNA segments are negatively charged, avoid approaching each other too closely, and so dry friction is out of question.

The result (11.1) was proven as a mathematical theorem for several different models of loops (on cubic lattice, freely jointed, etc) — one of the

very few analytical theoretical results in the field (we recommend interested reader the review article [34] where rigorous facts about knots statistics are explained for physicists and where complete list of references is provided). Nevertheless, a simple hand waving argument explaining exponential dependence (11.1) is still missing, and there is no analytical understanding whatsoever of the parameter N_0. Lion's share of our knowledge about knots in polymers comes from computer simulations.

11.4 Knots in DNA

And how about experiments on polymer knots? The most important and, luckily, also the easiest subject of such experiments is double helical DNA. One nice experiment can de done using DNA with "sticky ends" – a long double helix with each chain extending at one end by 15 or so unpaired nucleotides beyond the counterpart chain. If the sequences of these extending pieces are complementary to each other, they will stick upon first collision due to the random fluctuations of the double helical coil. Can we then determine the topology of the product?

One can also extract the ring DNA plasmid from bacteria and ask what are their topological states. Again, the question is how to determine the topology. As in the theoretical studies, this is the most difficult part.

A possible approach is to cover DNA with some suitable proteins (such as recA) to increase its thickness, then to adsorb it on a suitable surface, and then it becomes possible to make either an electron microphotograph, or an atomic force microscope image of it. One fruit of this very labor consuming procedure, coming from the lab of Nicholas Cozzarelli (1939–2006) at the University of California at Berkeley, is shown in the Figure 11.3. The result is interesting, for it shows, that native DNA is frequently knotted. But surely this is not the way to address any statistical questions.

Another method is based on the fact that DNA is negatively charged and, therefore, moves when the electric field is applied (see below Section 12.10). It is easy to believe that DNA with a more complex knot is, on average, more compact and, therefore, moves faster through the gel. It is this electrophoresis method that is behind most of the experimental discoveries in the field. In particular, all knots with up to six crossings have been positively identified both in native plasmid and in experiments on DNA with sticky ends.

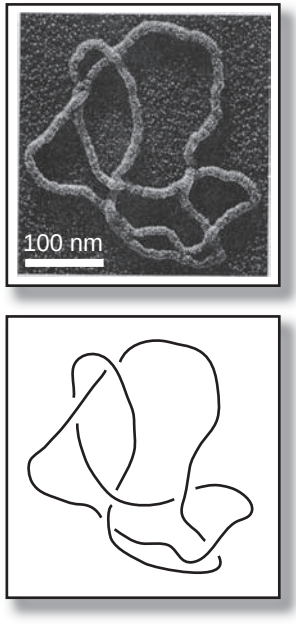


Fig. 11.3 Electron micrograph of circular DNA. This particular molecule is seen to be knotted; the lower image shows DNA in the style of a knot diagram and allows one to identify this knot as 6_2 (see Table of Knots, Figure C11.2). This particular knot was produced as a result of an enzyme-catalyzed reaction, the so-called site specific genetic recombination. Determining the topological properties of products of such reactions proved a very efficient tool in the study of their mechanisms. The figure is reproduced with permission from the paper: S. Wasserman, J. Dungan and N. Cozzarelli, "Discovery of a Predicted DNA Knot Substantiates a Model for Site-Specific Recombination", Science, v. **229**, p. 171, 1985.

Furthermore, the probabilities of knots computed in simulations for chains of various thicknesses and measured in experiments for DNA under different salt conditions agree quantitatively almost perfectly well. (Salt ions screen the Coulomb repulsion between DNA segments and thus control the effective diameter of DNA.)

We want to note in passing that this creates a situation somewhat unprecedented in the whole of history of science: researchers claim a rather complete understanding of DNA knotting, based on the agreement between simulations and experiments, but we have no theory. Will it stay that way, or somebody will eventually be able to crack a theory — remains to be seen.

11.5 Plectonemic DNA and Topological Enzymes

Double helical DNA is of course just a polymer, but a very peculiar one – in many respects. One peculiarity is that double helix has twisting rigidity. Usual chemical polymer chains, such as the ones shown in Figure 2.1, or protein chains in Figure 5.4, if we twist one end with respect to the other, can relax the deformation by turning around the single covalent bonds of

their backbones. DNA is decidedly not like that. A useful way to think about it is to imagine preparing closed circular DNA from a piece of linear double helix: we should bend the double helix to align its two ends against each other, and then connect the ends. Chemically, $5'$ end of one chain can be connected to the $3'$ end of the same chain (see Figures 5.5 and 5.6 about these ends). This is why closed circular DNA always has two strands topologically linked to each other, like it is sketched in Figure 2.10 (*e*). Still, this leaves one degree of freedom: the loop can be prepared with different values of the linking number between two strands. We recommend the reader to practice this process with a long narrow strip of paper (say, 25 cm by 0.5 cm). Label the two long sides of your strip with markers of two different colors (representing two strands of the DNA) and now try making a loop by gluing the ends together matching the colored sides. The later "color-matching" condition rules out Möbius stripe as well as other similar one-sided constructs with half-integer number of turns, and leaves only the possibilities to make an integer number of turns. In reality, the double helical DNA already has turns, about one turn per ten base pairs, so what we can do is we can force some integer number of super-turns, positive (if we tighten the double helix) or negative (if we loosen it).

By making a certain number of super-turns we fix the linking number between DNA strands. Notice that linking number is a topological invariant and cannot be changed without breaking the chemical bonds. A beautiful mathematical theorem establishes that the amount of twist and the linking number are related to each other depending on the shape of the double helix in space. The discovery of this theorem and its subsequent use to decipher many biologically relevant properties of DNA is one of the most beautiful success stories of the interdisciplinary development involving biologists, physicists, mathematicians, etc. It is a pity that this story is too complex to be told here. We only mention that, since the linking number is a topological invariant and cannot be changed without breaking the bonds, the molecule minimizes its free energy by balancing the torsional deformation energy, the bending deformation energy, and conformational entropy. This leads to interesting forms which are called plectonemic and which can be realized not only in DNA, as exemplified in the Figures C11.4 (*a*) and (*b*), but also in a telephone cord C11.4 (*b*) (the latter happens because we take the receiver and put it back by different moves, introducing each time 360° of torsional deformation).

To conclude our brief visit to the subject of DNA topology, we should mention that the cells have developed the special machinery to deal with

topological constraints in DNA. This of course only confirms the importance of topological properties for molecular biology. The sheer number of different types of topologically-relevant enzymes and the list of cellular processes (including, e.g., DNA replication) in which these enzymes are involved goes far beyond the framework of this book. We will only mention that there are two broad categories of topological enzymes — the ones which do not, and the ones which do, consume energy. The latter ones are in many ways very similar to molecular motors (see Section 5.8). Why some topological processes require energy while others do not is an intriguing question; some answers to it are known, some are not, and the reader who continues to study the subject beyond the present book has a good chance to learn lots of exciting stuff.

11.6 Knots in Proteins

The logic of a biologist is unlikely to make parallels between DNA and proteins, for they are of completely different roles in the cell. By contrast, the logic of polymer physics makes such comparison quite natural. Thus, are there knots in proteins? Proteins are, of course, much shorter, but they are globular, so the question is delicate.

This question is delicate also in a more fundamental sense: proteins are never loops, they are linear polymers with free ends. Because of that, the question of knots in proteins is not legitimate in the strict mathematical sense. But we should not be too rigid applying mathematical rigor — anyone with shoe laces will confirm that there is difference between a knot and an unknot for polymers with available ends. More to the point, the more compact is the linear polymer with open ends — the less ambiguous is the question of its knots. Indeed, consider a closed loop consisting of a polymer and a straight segment connecting its terminals. Its topological state is perfectly defined, and it is believable that the role of straight connector is small when the ratio of its length to the polymer length is small, which is particularly well obeyed for globules. In fact, there are better and more sophisticated ways to close the loop, using a not-necessarily-straight connection of terminals — but we shall not touch upon this here. The essence is that the question of knots in proteins can be formulated and should be addressed.

The answer is positive: yes, there are proteins with knots. Sometimes, these are simple trefoil knots, but a few cases are known with rather complex

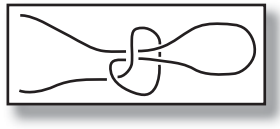


Fig. 11.6 A slip knot. It is not really a knot, because you can untie it if you just pull the ends. Nevertheless, it is an interesting long lived feature of some proteins. Some data on slip knots in real proteins can be found in the paper: J. Sulkowska, P. Sulkowski and J. Onuchic, Proceedings of the National Academy of Sciences, USA, v. **106**, p. 3119, 2009.

knots. For instance, Figure C11.5 shows the protein called human ubiquitin hydrolase featuring the knot as complex as 5_1. Other proteins feature so-called slip knots, as illustrated in Figure 11.6. Not much is known beyond the fact that the knots exist. For instance, the fraction of proteins with knots in the protein data base is much smaller than one would expect if conformations were random; the reason for this fact is not known, although there is no shortage of speculations. Are knots of some special importance for protein function? There are speculations that this is the case for the human ubiquitin hydrolase mentioned above. But in general the question of the role of knots in proteins is wide open.

Knots can of course be addressed also for toy lattice "proteins" discussed in section 10.10. The shortest cubic lattice knot has 24 monomers (Figure C11.7 *a*), but it is not space filling. The shortest space filling (open ended) lattice polymer to have knot is 36-mer. Its conformation with a trefoil knot is shown in the Figure C11.7. By the way, if the sequence of monomer species in it is properly selected (see Section 10.6), it folds (*in virtuo*, of course) quite successfully, and not much slower that the corresponding chain without a knot — which opens even wider the question as to why real proteins are statistically less likely to have knots than random.

Chapter 12

Dynamics of Polymeric Fluids

> So he started to climb out of the hole. He pulled with his front paws, and pushed with his back paws, and in a little while his nose was out in the open again... and then his ears... and then his front paws... and then his shoulders.. and then...
>
> A. Milne,
> *Winnie The Pooh*

12.1 Viscosity

What do we mean by a polymeric fluid? It is a viscous liquid, made of heavily entangled polymer chains. In particular, it could be a polymer melt, a concentrated or a semi-dilute polymer solution. You can easily get a feel for what these are like. All you need to do is melt a piece of ordinary plastic, so that it starts flowing. Obviously the most significant reason why polymeric liquids are important is because they are encountered in all technological processes of plastic production. Polymeric fluids are quite peculiar. In many ways, they are nothing like water or any other ordinary fluid that we are used to.

The first thing that strikes you is the high viscosity. It is normally much higher than for water. The physical cause of viscosity is internal friction. This acts between adjacent layers of the flowing fluid. Thus, we can say that internal friction in fluid polymers is greater than in water.

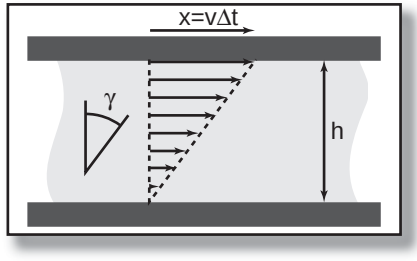


Fig. 12.1 A liquid layer between two parallel plates. The upper plate is moving at a speed v causing a simple shear flow.

Let's bring some maths into it. Figure 12.1 shows a very simple experiment. Some liquid is confined between two flat horizontal plates, a distance h apart. The lower plate is at rest, but the upper plate is moving at a constant speed v. How will the liquid at different heights move? A thin layer of liquid at the very bottom will stay at rest, due to internal friction. This layer is effectively "glued" to the lower plate. Similarly, the uppermost layer will be dragged by the upper plate with speed v. The distribution of the speed across the liquid, shown in Figure 12.1, is called a simple shear flow. Of course, the upper plate does not move of its own accord, but is pulled by an external force. The question is: can we calculate the force f which must be exerted on the upper plate to maintain its motion with speed v and, therefore, to maintain a simple shear flow of the liquid? The first point is that the force necessary to exert must be equal (in absolute value) to the friction force — because we are interested in the motion with a constant speed (sometimes called terminal velocity in some physics textbooks). The second point is that the internal friction in liquid is usually proportional to the speed and, therefore, the necessary force must be proportional to the velocity v; further, it is quite natural that the force is proportional to the surface area of the upper or lower plate, A, because friction acts everywhere along the plate. Finally, force has to be inversely proportional to the spacing between plates, h, because the smaller is h the sharper is velocity profile, and greater is the friction between neighboring layers of liquid. This problem is one of the most classical in physics, it was studied in great detail by the likes of Newton and Stokes, they came up with equation

$$f = \eta \frac{Av}{h} \ , \tag{12.1}$$

where η is the property of liquid known as the coefficient of viscosity, or just the viscosity. It gives a measure of how viscous the liquid is. For example, water has $\eta \approx 10^{-3} \ \frac{\text{kg}}{\text{m}\cdot\text{s}}$ (at room temperature and normal atmospheric

pressure), ketchup or honey has $\eta \approx 10 \frac{\text{kg}}{\text{m}\cdot\text{s}}$, whereas polymeric liquids in practice often have η as much as $10^3 \frac{\text{kg}}{\text{m}\cdot\text{s}}$ — depending on how long the chains are, and whether there is any solvent[1].

12.2 Viscoelasticity

The unusually high viscosity is not the only surprise that polymeric fluids offer. Another interesting and maybe even more important property is the viscoelasticity. Depending on how rapidly the external force changes, polymeric fluids can behave either like normal, low molecular weight liquids, albeit very viscous, or like elastic solids.

A whole series of nice experiments to demonstrate viscoelasticity can be easily performed at home, as demonstrated in Figure C12.2. Take a toy widely known as "silly putty" (or sometimes "jumping putty"). It represents nothing more than a piece of the polymeric material called silicone, but it has a potential to entertain children and grown ups alike, for quite a while. To begin with, you can easily give the toy virtually any shape you like by, e.g., rolling it between your palms (Figure C12.2 *a*). If you make it into a ball (Figure C12.2 *b*) and drop it on the floor — it will bounce and jump just like rubber (Figure C12.2 *c* and *d*; remember the Native American balls made from natural, unvulcanized rubber?). Now shape the same piece of silicone into a sausage (Figure C12.2 *e*) and leave it for some time; in a few hours there remains no doubt that it is a flowing liquid (Figure C12.2 *f*).

How can we explain such a "double life"? The polymer behaves as a liquid when steadily affected by gravity over a long period of time (Figure C12.2 *f*). On the other hand, when the action of the force is very short (when hitting the floor, Figure C12.2 *c*), the reaction is elastic. This is viscoelasticity. In general, viscoelastic bodies tend to show a viscous response to a slowly changing force, and an elastic response to one which varies quickly.

Peculiar combination of viscous and elastic properties can be also see in the following siphon effect (Figure C12.2 *g*). Place a sample of viscoelastic polymer (such as properly shaped "silly putty") in a tilted glass container (A) such that a sufficient part of the material extends beyond the container

[1] A purist will say that glasses (see Section 4.3) are also liquids, and their viscosities are many orders of magnitude higher. Purist is right, as usual — but we are also right: polymeric liquids are very viscous even far from glass transition.

and continues below its bottom (B). It looks amazing, but the polymer will not stop oozing from the A to B, until A is completely empty! Thus, A and B behave as if they were connected by a tube (a siphon), except there isn't one — the role of the tube is played by the flow itself. Clearly, this would not work with water or another ordinary liquid whose behavior is mainly viscous. There must be some elasticity as well to achieve this effect.

There is yet another interesting experiment on viscoelasticity. Get hold of a cylindrical jar full of a concentrated polymer solution. Install another, smaller cylinder inside the jar, so that it can turn around the common axis of both cylinders. Make the inner cylinder rotate at a constant angular speed for a while, and then suddenly let it go. Guess what will happen. Before stopping, the inner cylinder will turn back a little in the opposite direction! The angle of this backwards turn can reach a few degrees. Such unusual behavior is certainly a sign of viscoelasticity.

All polymeric liquids are viscoelastic. This suggests that viscoelasticity is not caused by something special in the chemical structure; it is a universal property. This is why theoretical physicists have flocked to study viscoelasticity and, in general, the dynamics of fluid polymers.

Is the viscoelasticity of polymeric fluids really so unexpected? Imagine a bunch of very long, very mixed up, and entangled chains; a plausible image of this mess is shown in the Figure C12.3. How can it flow? Obviously, a certain chain, if it wants to move, has to slither along a little wiggly corridor inside the bunch, undoing the knots on its way. This sort of picture inspired the theory of reptations (named after the snaky motion of reptiles). As the first successful molecular theory of fluid polymer dynamics, it was developed back in the 1970s by the physicists P.G. de Gennes (1932–2007) at College de France in Paris, M. Doi at the university of Tokyo, and S.F. Edwards from Cambridge University in England.

Before we tell you more, we'll just make one comment. Some readers may have heard of L.D. Landau's scepticism about molecular theories of liquids. He thought it was impossible to create such a theory. All liquids are so different, they just don't seem to have enough in common for a common theory. Compare, for example, liquid helium and ordinary water. Another awkward thing about liquids is that there are no obvious small or large parameters. Most of the important dimensionless parameters are of order 1. This is a nuisance. As we have seen in Section 8.1, if you want to simplify a system's behavior, and build an ideal model for it, you need some small or large parameters.

However, it is not so bad with polymeric liquids. There is a natural large parameter, the number N of monomer units in a chain. This is why it would be helpful to know how parameters such as the viscosity and molecular diffusion coefficients depend on N, in the limit of $N \gg 1$. The form of this dependence will determine how the polymer behaves. Such a situation allows us to use a very common approach of theoretical physics.

12.3 The Reptation Model

Let's pick a test chain in a polymeric liquid. Imagine for a minute that all the other chains are "frozen" and cannot move. What can the test chain do in such a "frozen" jungle? It cannot go through other molecules. So it will be confined in a sort of tube formed by the neighboring chains (Figure 12.4). This is a fundamental concept. The chain cannot move through the walls of the tube, so all it can do is to crawl along. This is very clearly seen in a two-dimensional version in Figure 12.5. (In this figure, the "frozen" surroundings are modeled by fixed obstacles on the plane, which cannot be crossed by the chain whilst it is moving.) If there are no external forces, the motion along the tube is obviously purely diffusive. It is like Brownian motion (see Chapter 6): the chain makes random steps in one or other direction, with equal probability.

Now, let's "defrost" the surrounding chains. Then more opportunities for the test chain will arise. Some of the neighboring chains will start moving away. Therefore, some constraints and entanglements that had formed the tube (Figures 12.4 and 12.5) will gradually disappear (or "decay"). However, as P.G. de Gennes showed, this effect is not important. The chain will snake out of the "frozen" tube much sooner than it takes the

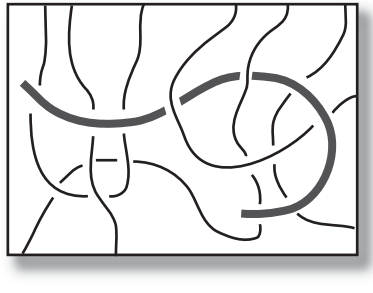


Fig. 12.4 A polymer chain among other chains, in a concentrated system.

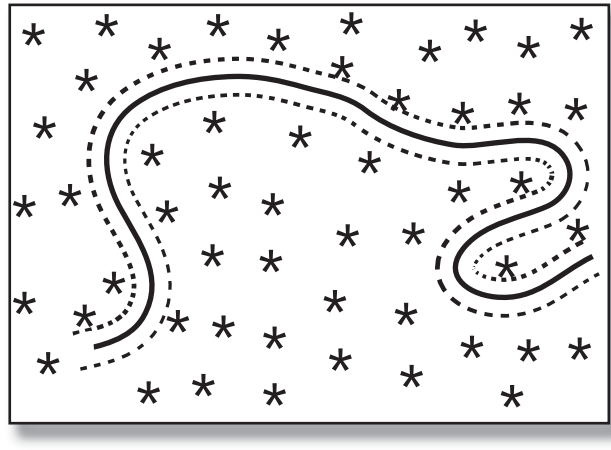


Fig. 12.5 A sketch of a polymer chain in a network of "frozen" obstacles (two dimensional case).

constraints to decay. This is why the motion in a fixed tube of obstacles is the main mechanism for the dynamics of a highly entangled chain.

The snakelike motion along the tube is called reptation, from the Latin *reptare*, "to crawl". The corresponding model of polymeric liquids is known as the reptation model.

12.4 The Longest Relaxation Time

It is interesting to see what we can learn from the reptation model. The best way to see it is to look at a simple experiment. Take a polymer melt or a concentrated solution and place it between two plates, similar to the geometry shown in Figure 12.1 (in practice it can also be a gap between two co-axial cylinders). At time $t = 0$, apply a constant shear stress σ, and measure the relative deformation, or strain γ, as it develops in time after the stress is switched on at $t = 0$. If σ is small, the deformation will be proportional to the stress:

$$\gamma(t) = \sigma J(t) \ . \tag{12.2}$$

The function $J(t)$ is called the compliance of the material. On a logarithmic scale, it looks like the curve in Figure 12.6 *a*. After a sharp rise at the start, it reaches a plateau, $J(t) = J_0 = \text{const}$. If we set this constant $J_0 = 1/G$,

then, in the plateau region, we have

$$\sigma = G\gamma \ , \tag{12.3}$$

which is essentially the Hooke's law, and G is shear modulus. Strictly speaking, the most common formulation of Hook's law Equation (4.1) is different from formula (12.3) in the sense that the former describes elongation while the latter describes shear. But since the volume of polymer samples hardly changes, we can brush aside the difference between these types of deformations; in particular, shear modulus G and Young modulus E are not significantly different for these systems. Thus, the system is elastic in the plateau region, with Young's modulus $E \sim 1/J_0$.

Only after a long enough time, $t > \tau^\star$ (Figure 12.6 *a*), does the deformation become irreversible, and the polymer starts to flow. In this case, the compliance is a linear function of time, $J(t) = J_1 t + J_2$, where J_1 and J_2 do not change with t. Let's compare this dependence with the definition (12.2) of the function $J(t)$. We can conclude that in the range $t > \tau^\star$ the stress is no longer proportional to the strain, but rather to the rate at which the strain increases

$$\sigma \sim J_1^{-1} \frac{d\gamma}{d\tau} \ . \tag{12.4}$$

This is the typical behavior of fluids. If you are not convinced, just compare Equation (12.4) with the Newton–Stokes law (12.1). Figure 12.1 indicates that $\tan\gamma = x/h$, where x is the displacement of the top plate from where it was at $t = 0$. Hence, if γ is small, $\gamma \quad x/h$. Remembering that $v \ = dx/dt$, we can re-write formula (12.1) as

$$\frac{f}{A} = \eta \frac{v}{h} = \eta \frac{d\,(x/h)}{dt} = \eta \frac{d\gamma}{dt} \ . \tag{12.5}$$

Further, we can replace the ratio f/A in the left hand side with the tangential shear stress σ. The most interesting result we can draw is that the coefficient J_1^{-1} in (12.4) is nothing else but the viscosity of the polymer liquid η. Thus,

$$\sigma = \eta \frac{d\gamma}{dt} \ . \tag{12.6}$$

Let's summarize. At $t < \tau^\star$ the polymer melt behaves as an elastic body (see (12.3)), whereas at $t > \tau^\star$ it is rather like an ordinary fluid (see (12.6)). Just for comparison, we have also sketched the compliance function $J(t)$ for a typical non-polymeric liquid (e.g., water) in Figure 12.6 *b*. You can see that this graph has no intermediate plateau region corresponding to elastic behavior (12.3).

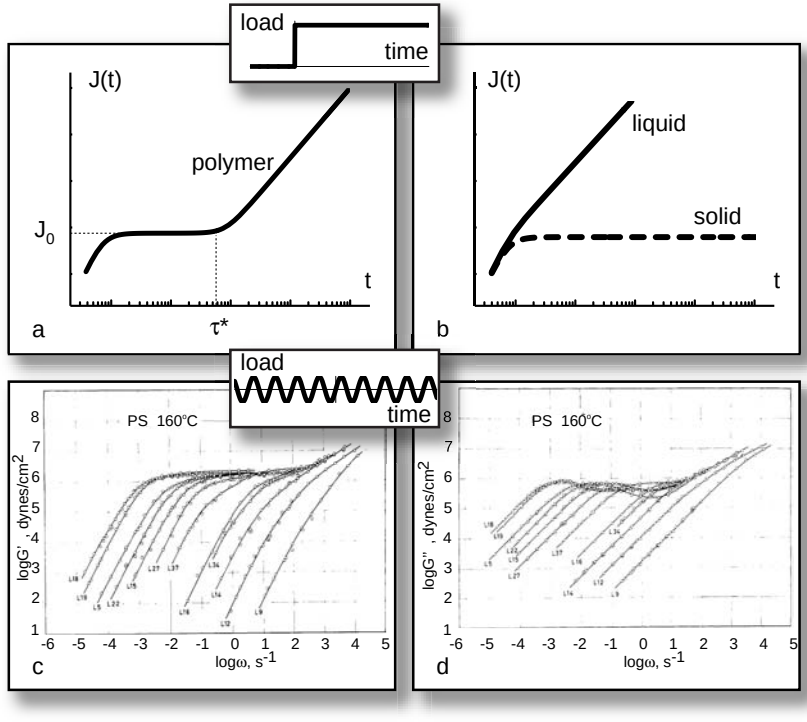


Fig. 12.6 Mechanical response of any material can be tested in many ways, of which the ideologically simplest is the "switch-on" experiment (panels *a* and *b*), while technically most robust is the "oscillating load" experiment (panels *c* and *d*). The corresponding load versus time protocols are shown in the insets. In a "switch-on" experiment, the time evolution of the strain after step-wise application of stress is described by the compliance function $J(t)$ plotted schematically in double logarithmic scale in panel (*a*) for a polymeric liquid and, for comparison, in panel (*b*) for a regular low molecular mass fluid (solid line) and for a solid (dashed line). The major feature of a polymer system is the presence of plateau region at $t < \tau^*$. Comparing with the dashed line in panel (*b*), the plateau of the compliance for polymeric system means elastic type of behavior at times up to about τ^* (compare bouncing ball in Figure C12.2 *c* and *d*). By contrast, the linear increase of polymeric compliance $J(t)$ at $t > \tau^*$ is characteristic of viscous behavior, similar to solid line in (*b*) (compare flowing liquid behavior in Figure C12.2 *e* and *f*). Panels (*c*) and (*d*) convey the same physical idea, they show the results of real oscillating load experiment. The results of such experiment are reported in terms of the two parts of the response, elastic part, which stays in phase with the driving force, and viscous part, which lags behind by $\pi/2$ in phase; they are called traditionally G' and G'', respectively, and they are plotted in the double logarithmic scale against frequency ω. The data are obtained for polystrene at 160°C, and exhibit clearly the characteristic plateau behavior at the range of frequencies whose periods correspond to t between microscopic time and τ^*. Different curves correspond to samples with different chain lengths. Panels (*c*) and (*d*) are reproduced with permission from the paper S. Onogi, T. Masuda and K. Kitagawa, "Rheological Properties of Anionic Polysterens", Macromolecules, v. **3**, n. 2, pp. 109–116, 1970.

The time $\tau^\star$, when the polymer's type of response to the stress changes, is called the longest relaxation time.

What came before the reptation model? Experimental data like the one in Figure 12.6 *a* used to be explained in the following way. A polymeric liquid was thought to contain some kind of effective cross-links. In contrast to the usual chemical cross-links (formed by chemical bonds), the effective ones do not live for long. They can only last for a period of the order of $\tau^\star$. Then they break (or "decay"), and new cross-links are created at other places, and so on. Thus, when $t \ll \tau^\star$, the cross-links do not have enough time to vanish. They hold the sample together, so it behaves like an elastic body. In contrast, when $t \gg \tau^\star$, the cross-links start decaying, and the sample flows.

The reptation model makes this picture clearer. It tells us what these cross-links really are, at a molecular level. For example, pick two chains in a polymer melt. They are both confined in their own fixed tubes. Suppose these tubes pass near to one other (Figure 12.7). Then the two molecules will have to enjoy each other's company, until one of them abandons the part of its tube that comes close to the other tube. You could say that, while the two tubes are next to each other, there is an effective cross-link in that area. However, as soon as one of the chains leaves the neighborhood area, the cross-link will disappear.

Now you can see the microscopic meaning of $\tau^\star$, which we introduced as a typical relaxation time of effective cross-links. The cross-links decay

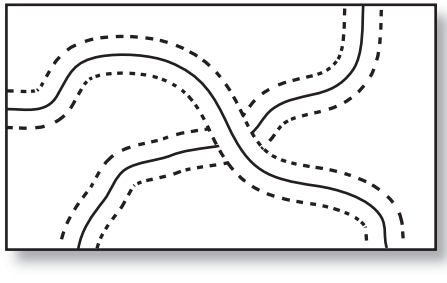


Fig. 12.7 Two polymer chains forming an effective cross-link.

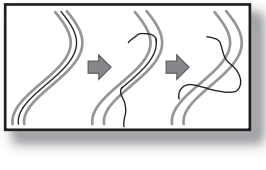


Fig. 12.8 Successive stages of a polymer chain leaving its original tube.

because of the chains' reptation, that is, because the chains slither out of their tubes. Therefore, $\tau^\star$ gives an idea of the time it takes for the chain to abandon its original tube (where it was at $t = 0$). After this period of time, the chain finds itself in a brand new tube into which the random motion of its ends has led it (Figure 12.8). You can say that the tube has been fully "renewed". All the original cross-links (i.e., the neighboring parts of different tubes) have totally vanished.

Let's go back to the simplest experiment shown in Figure 12.6, where we applied a small constant stress σ at $t = 0$. What estimates can we make for the compliance in the ranges $t \gg \tau^\star$ (viscous flow) and $t \ll \tau^\star$ (elasticity)? From (12.6) and (12.2), we deduce that $J(t) \sim J_1 t \sim t/\eta$ for $t \gg \tau^\star$. On the other hand, from (12.3) we have $J(t) \sim 1/E$ for $t \ll \tau^\star$. Can we tell what happens for $t \sim \tau^\star$? Obviously, the two estimates should merge smoothly into one another. This idea leads us to a very important relationship between the viscosity η, the longest relaxation time $\tau^\star$, and Young's modulus E for a network of effective cross-links:

$$\eta \sim E\tau^\star \ . \tag{12.7}$$

This relationship can help us to learn about the viscosity of a polymer solution. In particular, we can use it to figure out how the viscosity depends on the number of monomer units N in a chain (in the limit $N \gg 1$). Presumably, we need to know first how E and $\tau^\star$ depend on N. So we should venture a little investigation. Let's consider E and $\tau^\star$ separately, and concentrate on the case of a polymer melt, to make it easier. In principle, the same sort of logic should apply to concentrated and semi-dilute solutions.

12.5 Young's Modulus of a Network of Effective Cross-Links

A network of effective cross-links behaves as a normal elastic network for $t \ll \tau^\star$. We discussed the classical theory of high elasticity in Chapter 7. The Young's modulus of a network, as you remember, is of the order of $k_B T$ multiplied by the density of cross-links. (As usual, k_B is Boltzmann's constant, and T is the temperature.)

Thus, we have to work out roughly how many effective cross-links there are in a polymer melt. The tricky bit is to decide what exactly is an effective cross-link, and what is not. All the chains are highly entangled. An extreme view would be to regard any contact between a pair of chains as an effective

cross-link. This is not completely illogical. Whenever a pair of chains come close to each other, their further motion is constrained (since they cannot go through each other). This is why the number of conformations allowed for each chain is much less than it would be in free space. You could model such topological constraints by effective cross-links.

What sort of picture would we really get if we replaced each contact between the chains by a cross-link? As you can imagine, it would be a very densely woven structure. All the cross-links would make it extremely stiff, and so it would be nothing like an ordinary elastic body (i.e., nothing like our melt for $t \quad \tau^\star$). Suppose the chains in the melt are flexible, with Kuhn segment ℓ. The number of contacts per unit volume is approximately $1/\ell^3$. Let's accept for a moment the extreme view we have suggested. Then Young's modulus of the melt (i.e., of the network of effective cross-links) would be $E \sim k_B T/\ell^3$. How good is this estimate? We can check it if we use it to calculate the plateau value of the compliance, $J_0 = 1/E$, for various melts. It turns out that the answers it gives are far too high compared with experiments.

This is not surprising. In fact, there is a big difference between an effective cross-link and a mere contact between the chains. Look at Figure 12.9 *a*, for example. The two chains pass near each other, so we can say that they are in contact. However, this does not seriously restrict their freedom, that is, the choice of possible conformations. In contrast, the contact shown in Figure 12.9 *b* is much more constraining. It is really just the same as a cross-link. In this case, the number of allowed conformations is obviously much reduced.

Thus, there are contacts and contacts. Not all of them play the role of effective cross-links. Taking this into account, let's modify the estimate for

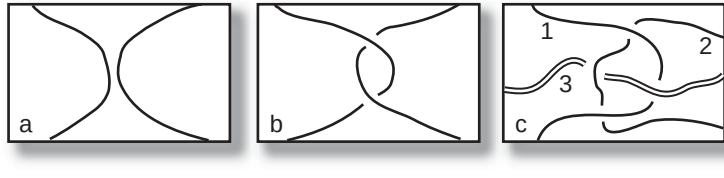


Fig. 12.9 Contacts between polymer chains: (*a*) without an effective crosslink; (*b*) with an effective crosslink. Panel (*c*) shows that in general presence or absence of an effective crosslink between two chains might depend on the other chains around: in this example, chains 1 and 2 are linked if chain 3 is there, but would not be linked without chain 3.

E:

$$E \sim \frac{k_B T}{N_e \ell^3} \ . \tag{12.8}$$

Here N_e is the average number of monomer units along the chain between two nearest effective cross-links. The parameter N_e is the only phenomenological one in the modern theory of polymeric liquids (i.e., it has to be found separately, from some other arguments or observations). Nobody has yet worked out how to calculate it from a knowledge of the microscopic structure. All we can say is that it must be somehow related to the ability of the chains to form knots with each other. Therefore, it must depend on the chain stiffness and geometry (e.g., whether it has any side branches, and so on). You can find N_e experimentally, from the value of Young's modulus corresponding to the plateau in Figure 12.6 *a*. Typically, N_e ranges from 50 to 500. In any case, $N_e \gg 1$. This confirms that only a small number of contacts work as effective cross-links.

Given that that the "entanglement length" N_e is large, one should ask — how long should the chains be to make for a highly entangled polymer melt? The reptation model talks about a chain in a tube. It only makes sense if there is a great number of effective cross-links per chain, that is, $N/N_e \gg 1$, or $N \gg N_e$. We shall bear this in mind when deriving how the viscosity η and the maximum relaxation time $\tau^\star$ depend on N.

12.6 The Tube

In order to find η and $\tau^\star$, we will pick a chain and explore its tube in more detail. The tube is created by other chains. If they come into contact with the test chain, they act as obstacles to the chain's motion. However, we have seen that only a small proportion of such contacts can really limit the chain's choice of conformations. This proportion is of the order of $1/N_e$. These are the contacts which can be regarded as effective cross-links.

Therefore, we come up with the following picture of a chain in a tube (Figure 12.10). First of all, there is a characteristic size $d \sim \ell N_e^{1/2}$, which roughly gives the distance between two nearest cross-links along the chain. On the scale $r < d$, the chain does not "feel" the effective cross-links. So it has the full choice of allowed conformations. On the other hand, for distances $r > d$, the effective cross-links create the tube. This is why d must be the same as the diameter of the tube. Now we can regard the chain as a sequence of "blobs" of size d. Each blob contains N_e monomers,

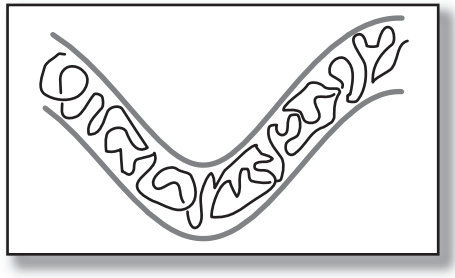


Fig. 12.10 A chain in a tube.

and behaves as an ideal polymer coil. (The blobs are ideal because the excluded volume interactions are completely screened in a polymer melt (see Chapter 8).) They fill the tube, lining up along its axis. Hence, the total contour length of the tube axis is $\Lambda \sim (N/N_e)d$, since N/N_e is the number of blobs per chain. Remembering that $d \sim \ell N_e^{1/2}$, we get:

$$\Lambda \sim \ell N N_e^{-1/2} \ . \tag{12.9}$$

Notice that this result for the length of the tube is much less than the full length of the chain $N\ell$. This is because $N_e \gg 1$.

12.7 The Dependence of the Longest Relaxation Time on the Chain Length

Now let's calculate the longest relaxation time $\tau^\star$ for a polymer melt. As we have said, it is the time that a reptating chain takes to leave its original tube. To do this, the chain obviously has to diffuse along the tube axis by a distance of order Λ.

When a chain moves in a dense system (like a polymer melt), the frictional forces acting on each monomer are totally independent. Hence, the total frictional force experienced by the moving chain is simply the sum of the frictional forces on each individual monomer. How can we find these frictional forces on the monomers? Let's focus on one monomer; suppose it has a velocity $\mathbf{v}$. This is the velocity of diffusion, so it is not too high. (To be more precise, it is of the same order as the thermal velocity of the monomer.) This gives us the right to take the force $\mathbf{f}$ of viscous friction to be proportional to the velocity: $\mathbf{f} = -\mu\mathbf{v}$. Here μ is the coefficient of friction for a single monomer. Since the total friction is the sum over all the monomers, the same must be true for the coefficients of friction. The

friction coefficients for the single monomers add up to give the friction coefficient for the whole chain. Say we have an N-unit chain crawling through a tube. Then its coefficient of friction μ_t will be just N times the monomer coefficient of friction μ: $\mu_t = N\mu$.

How do we normally describe diffusive (or Brownian) motion? An important quantity is the diffusion coefficient D. It determines the mean-square displacement $\langle x^2 \rangle$ of a Brownian particle over a period of time t (along one of the axes):

$$\langle x^2 \rangle = 2Dt \ . \tag{12.10}$$

(It is just because the motion is Brownian that $\langle x^2 \rangle$ is proportional to t; cf. (6.2).) How does friction come into this picture? Evidently, the greater the coefficient of friction μ for some particle, the lower will be the diffusion coefficient D, and vice versa. The exact relationship between the two was found in 1905 by Albert Einstein, and is called the Einstein relation. It states that

$$D = \frac{k_B T}{\mu} \ . \tag{12.11}$$

Incidentally, this happens to be the most cited paper by Einstein - more than relativity and more than everything else. The physics of why the temperature T appears in Equation (12.11) is quite clear. For given values of μ and t, the mean-square displacement (12.10) must increase with growing temperature, that is when the thermal motion becomes more intense.

Now we can come back to the problem of the reptation of a long polymer chain in a melt. We will estimate the diffusion coefficient D_t which describes lengthwise diffusion of the chain along the tube. According to (12.11):

$$D_t = \frac{k_B T}{\mu_t} = \frac{k_B T}{N\mu} \ . \tag{12.12}$$

As we know, the longest relaxation time $\tau^\star$ is roughly the time it takes the chain to diffuse along the tube by a distance equal to the length of the tube axis, Λ (Equation (12.9); see also Section 12.4). Therefore, using (12.9), (12.10), and (12.12), we obtain:

$$\tau^\star \sim \frac{\Lambda^2}{D_t} \sim N^3 \ell^2 \frac{\mu}{N_e k T} \ . \tag{12.13}$$

We can tell from this that the longest relaxation time increases dramatically with N, the number of monomer units in the chain: $\tau^\star \sim N^3$. This explains why relaxation is so slow in polymeric liquids (compared to ordinary low molecular weight liquids). As a result, polymeric liquids have a long-lasting

memory of the previous history of the flow. (If there were no such memory, for instance, the experiment shown in Fig. C12.2 would be impossible.)

How much can the factor N^3 really slow things down? Let's make some estimates, to compare polymeric and low molecular weight liquids. We can rearrange Equation (12.13) in the following way:

$$\tau^\star \sim \tau_m \frac{N^3}{N_e} \ , \tag{12.14}$$

where $\tau_m \sim \ell^2\mu/kT$ is the microscopic relaxation time, typical of a low molecular weight liquid. Using (12.10), we can write: $\tau_m \sim \ell^2/D$, where D is the diffusion coefficient of a single molecule in such a liquid. Now we can see the meaning of τ_m. It is the time taken for a molecule to move a distance equal to its own size ℓ. Let's take some typical values $\ell = 0.5$ nm $= 5 \cdot 10^{-10}$ m, and $D \approx 2 \cdot 10^{11}\ \frac{\text{nm}^2}{s} = 2 \cdot 10^{-7}\ \frac{\text{m}^2}{s}$. Then $\tau_m \sim 10^{-12}$s. Thus, we have found the typical microscopic relaxation time for a low molecular weight liquid.[2]

According to (12.14), the longest relaxation time $\tau^\star$ of a polymer melt is a factor of N^3/N_e greater than τ_m. Suppose the polymer chains are rather long, $N \sim 10^4$. Then, using the crude estimate $N_e \sim 10^2$ (see Section 12.5), we obtain $N^3/N_e \sim 10^{10}$. This leads to a longest relaxation time of $\tau^\star \sim 10^{-2}$ s, which is a completely macroscopic value. It can be even bigger. Strong interactions between the molecules may sometimes increase the coefficient of friction μ. This will, in turn, increase $\tau_m \sim \ell^2\mu/(k_BT)$, and hence $\tau^\star$. The longest relaxation time may become as high as a few seconds or even more. This is just what you are likely to observe in experiments measuring macroscopic relaxation times of viscous polymeric liquids.

High values of $\tau^\star$ are responsible for viscoelasticity in polymers, which we can witness even in the simplest macroscopic experiments, like the ones described at the beginning of this chapter. If an external force is quite

[2]The estimate $\tau_m \sim 10^{-12}$ s gives a natural scale of time for liquids at room temperature. In fact, we could have obtained it in a different way. In dense systems, the size ℓ marks the border between two important length scales. On shorter scales, the motion of each molecule can be described accurately as ballistic, very much like the free path of a particle in a low pressure gas. In contrast, at larger scales the molecules are engaged in diffusion. What we are interested in is the particle's displacement ℓ on the microscopic scale, which is the cross-over between diffusion, or random walk, type motion on large scales and the ballistic motion on the small scales between collisions. To work it out, we can use either the diffusive relationship $\tau_m \sim \ell^2/D$ or the formula $\tau_m \sim \ell/v$, where v is the average thermal velocity of the molecules. For light organic molecules at room temperature, $v \sim 500\ \frac{\text{m}}{\text{s}}$. Thus, $\tau_m \sim \ell/v \sim 0.5 \cdot 10^{-9}/500$ s $\sim 10^{-12}$ s.

abrupt — that is, it acts for a period shorter than $\tau^\star$ (e.g., when a silicone ball hits the floor), there is no time for relaxation to occur. So the polymer behaves as an elastic body. On the other hand, if the force lasts for longer than $\tau^\star$ (like gravity making silicone flow out of a jar), the viscous friction comes into play.

12.8 The Viscosity of a Polymer Melt and the Self-Diffusion Coefficient

Let's now use the reptation model to find the viscosity η of a polymer melt. We are going to use Equation (12.7) along with estimates (12.8) and (12.13) of Young's modulus E and the longest relaxation time $\tau^\star$. This gives:

$$\eta \sim E\tau^\star \sim \left(\frac{\mu}{\ell}\right)\frac{N^3}{N_e^2}\ . \tag{12.15}$$

If the chains are long enough ($N \gg N_e$), the viscosity of the melt goes up quite rapidly as N increases: $\eta \sim N^3$ (just like the relaxation time).

We will also calculate the coefficient of translational diffusion D_s of the chain as a whole moving in the melt. While the chain completely leaves its original tube in the time $\tau^\star$, its center of mass must move by the distance $R \sim \ell N^{1/2}$, that is, about the size of a coil. The displacements of the chain during each interval of length $\tau^\star$ are statistically independent. This is why on a large time scale we can talk about the diffusion of the center of mass. It is just the same as the Brownian motion of a particle that has a mean free time $\tau^\star$. Between collisions it moves a distance of order R in a random direction (see Chapter 6). Therefore, according to (12.10) we can write:

$$D_s \sim \frac{R^2}{\tau^\star} \sim \frac{k_B T}{\mu}\frac{N_e}{N^2}\ . \tag{12.16}$$

Thus, the reptation model predicts that D_s decreases as N^{-2} as the number N of monomers in the chain grows. When N is quite large, the diffusion coefficient is very low. As a result, if you bring two polymer melts together, they will tend to intermingle very slowly, even if the thermodynamics suggests that the mixed state is the most favorable one (i.e., if the two polymers are miscible).

12.9 Experimental Tests of the Theory of Reptation

Do the main results of the theory of reptation (12.13), (12.15), and (12.16) agree with experiments? As for the estimate $D_s \sim N^{-2}$, the agreement is usually very good. However, it is not quite as pleasing with the power laws for the longest relaxation time $\tau^\star \sim N^3$ and the viscosity $\eta \sim N^3$. Most experiments indicate slightly sharper dependencies: $\tau^\star \sim N^{3.4}$ and $\eta \sim N^{3.4}$. These are fairly close to the theory, yet not exactly the same. There have been many attempts to account for such a discrepancy. At present, the most widely accepted explanation is this. If the chains were infinitely long, experiments would give just what the theory predicts, $\tau^\star \sim N^3$ and $\eta \sim N^3$. But to the finite length of the chains that we observe the index 3.4; the number of monomers N per chain is not big enough compared with N_e.

The reptation model is more powerful than you might think. You can get much more out of it than just the simplest basic laws for the viscosity, the longest relaxation time, and the diffusion coefficient of a chain in a polymer melt. This model allows you to describe, for instance, the relaxation of a polymer after a stress has been released, or the response to a periodic force. As a result, you gain a fairly complete picture of the dynamics of polymer liquids, and of their viscoelasticity in particular.

The reptation model was the first to bring the large parameter N into play. As a result, a molecular theory of fluid polymer dynamics has been developed. All the previous theories of the dynamics of polymeric liquids were basically phenomenological.

12.10 Reptation Theory and the Gel-Electrophoresis of DNA

Now let's dwell at some length on a rather unexpected, and probably one of the most important, application of reptation theory. Genetic engineering, DNA sequencing, and other branches of biotechnology are all dependent on the high-precision techniques for analyzing DNA. In particular, it is important to learn how to distinguish DNA strands that differed slightly in length, or in the amount of twisting or knotting (the latter two only make sense for ring-shaped DNA), and so on. The method of gel-electrophoresis has turned out to be amazingly suitable for that purpose.

The idea is the following. Suppose you have a solution containing the bunch of DNA molecules that you wish to separate. Spread this solution

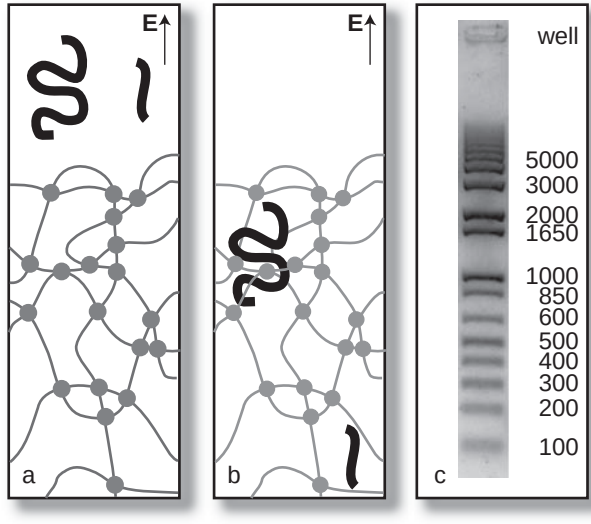


Fig. 12.11 Explanation of how DNA gel-electrophoresis works. DNA strands of different sizes are separated because they differ in mobility: while subject to the same field, they move with different speeds, and, therefore, in a given amount of time, cover different distances. Starting from the situation (*a*), the system after some time arrives at the situation (*b*), where short DNA moved much farther than the long one. Panel (*c*) represents a result of a real experiment separating a mixture of DNAs of many different lengths on a strip of polymer gel: the numbers to the right indicate lengths of the corresponding DNAs in base pairs, and it is clearly seen that each length produces well resolved separate band. Figure (*c*) is courtesy of A. Vologodskii.

on the edge of a polymer network (i.e., a gel). In water, each of the DNA monomers will dissociate and acquire a negative electric charge. Therefore, if you place the sample in an electric field of the right polarity (i.e., between the plates of a capacitor), you can make the DNA chains move through the gel. Such motion is called electrophoretic. You only have to hope that the speed of the chains depends on their length and structure! If it does, the problem is solved. Different chains will travel different distances and will become separated (Figure 12.11).

Let's consider, for example, separating linear DNA chains of different lengths. Can we work out how the speed of electrophoretic motion depends on the chain length, N? To get an idea, we will explore two limiting cases.

First of all, suppose the electric field is very strong. We follow the front end of a DNA molecule which moves forward and creates new bits of the tube. This end will spend more time traveling along the field direction

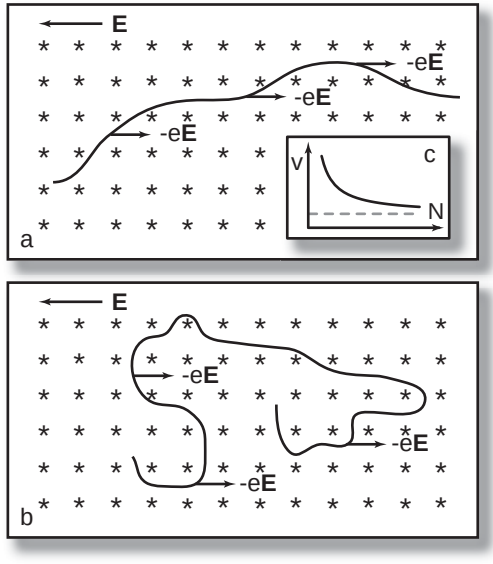


Fig. 12.12 DNA molecule in a gel in electric field is stretched if the field is strong (*a*), but remains an insignificantly deformed coil if the field is weak (*b*). The inset *c* shows the dependence of the electrophoretic speed v on the chain length N; the dependence flattens off for long chains, which means separation in a constant field only works for chains of moderate length.

rather than across it, or, even less likely, against it. As a result, the chain will tend to be stretched in the direction of the field (Figure 12.12 *a*). The stretching force f is proportional to N (because the total electric charge on the chain is proportional to N). Now, the coefficient of friction for the whole chain μ_t is also proportional to N, as we discussed when we talked about reptation. This is unfortunate. The speed of motion in a strong electric field is independent of N after all, $v = f/\mu_t$.

Maybe we shall be more successful with weaker fields? DNA molecules will not stretch in this case, but rather remain in the shape of Gaussian coils. The field pulls different parts of the chain in the same direction in space, but this may (and frequently does) correspond to the different directions along the tube (Figure 12.12 *b*). We end up with a sort of tug-of-war game. Who will win? Obviously, it will be the end of the chain which happens to be further forward in the field direction (and is therefore longer). The extra force it exerts is proportional to the displacement of this end, that is, to $N^{1/2}$. However, the coefficient of friction μ_t is still proportional to N. What a relief! The two dependencies do not cancel out this time, and we find that the speed of the motion along the tube, v_t, is:

$$v_t = \frac{f}{\mu_t} \sim \frac{N^{1/2}}{N} \sim N^{-1/2} \ . \tag{12.17}$$

(This is the speed of reptation. Don't confuse it with the speed of the chain as a whole which we are still trying to find!)

How fast will the molecule's center of mass move? Suppose the chain has crawled along the tube by a short distance Δ. The result of this motion is convenient to represent as if you chop a piece of length Δ at one end of the chain and stick it to the other end. When you chop the piece, you transport it by a distance $\sim N^{1/2}$. The center of mass will move by a distance $\sim N^{1/2}\Delta/N \sim \Delta/N^{1/2}$. Hence, the speed of the center of mass is a factor of $N^{1/2}$ slower than the speed of reptation. Thus we obtain that the speed of the center of mass is $v \sim v_t/N^{1/2} \sim 1/N$ in a weak field.

A more accurate calculation confirms our answer. It gives the following formula:

$$\mathbf{v} = \frac{q}{3\eta}\left\{\frac{1}{N} + \text{const}\left(\frac{Eq\ell}{kT}\right)^2\right\}\mathbf{E}\ . \tag{12.18}$$

Here $\mathbf{E}$ is the electric field vector, q is the charge per unit length of the DNA chain, N is the length of the chain (measured in Kuhn segments), ℓ is the length of the Kuhn segment, η is the viscosity of the medium, and "const" is a number of order one. The graph $v(N)$ is sketched in inset c in Figure 12.12.

When N is small, the dependence of v on N is rather strong. However, it flattens off as N increases, and becomes negligible. This means that only fairly short chains can be easily separated. Of course, you can increase the threshold length if you reduce the electric field. However, this is not terribly helpful. In a very weak field, the whole process will be far too slow, which is inconvenient, and may cause extra problems.

Many interesting little tricks have been devised in order to overcome some of these difficulties. The external field is periodically switched off (or rotated through 90°). The time taken for the field to go through one cycle should be roughly the same as the typical time of tube renewal, that is $\tau^\star \sim N^3$ (see (12.13)). In this case electrophoretic motion will only occur for chains that have about the right N. There is also two-dimensional electrophoresis and many other variants. Clever improvements proved to work splendidly and to give extremely precise results.

12.11 The Theory of Reptation and the Gel Effect During Polymerization

The theory of reptation helps us understand the gel effect during radical polymerization. We described how polymerization occurs in Chapter 3.

Suppose we have added some initiator to the solution of not-yet-polymerized monomers, and the reaction has begun. At first, the growing chains appear in a kind of dilute solution, in which the monomer molecules play the part of a solvent. With time, more and more molecules of the monomer become involved in the reaction. The concentration of the chains grows, and they begin to overlap. This is when the solution becomes semi-dilute. From this moment on, the character of the chains' motion changes — they start moving by reptation. As we have already shown, this means that diffusion of polymer chains slows down substantially.

On the other hand, a polymer chain stops growing when, as a result of diffusion, two free radicals at the ends of two chains happen to come together, react, and form a covalent bond (see Section 3). Obviously, if diffusion slows down, such encounters of the chain ends become less frequent.

Thus, you may expect that as soon as the chains start overlapping, polymerization should proceed much faster. The chains themselves should be able to grow longer, because, their growth is not stopped as often as before.

Indeed, all this can actually be observed, and is known as the gel effect during radical polymerization. The changes that occur are very dramatic. The rate of reaction jumps by a few orders of magnitude while the fraction of polymer increases only by a minute amount. This effect was noticed fairly long ago, well before the theory of reptation was proposed. However, it was the theory of reptation that enabled a proper mathematical description of the phenomenon.

Chapter 13

The Mathematics of Complicated Polymer Structures: Fractals

> And so he got the answer:
> Two and two thirds workers...
>
> S. Marshak,
> *(Russian children's poet)*

> "Very well," said Stuart, "what's the first subject you usually take up in the morning?"
> "Arithmetic," shouted the children.
> "Bother arithmetic!" snapped Stuart. "Let's skip it."
>
> E.B. White, *Stuart Little*

13.1 A Bit More About Maths in Physics: How Does a Physicist Determine the Dimensionality of a Space?

A good starting point for another very interesting yet unfinished story is Brownian motion. As you remember, the displacement of a Brownian particle (or the end-to-end distance of a polymer chain) is proportional to the square root of the time traveled (or the contour length of the chain). Surprising as it may seem in a book on polymers, the story is about the dimensionality of a space. Mathematicians have already been studying this topic for nearly one hundred years, and know quite a bit about it. However, it seemed of no particular relevance for physics until very recently, after two books by B. Mandelbrot appeared in 1977 and 1982 [47]. We shall avoid too much maths here, and basically talk about the physics side.

The space we live in is, of course, three-dimensional. We know this because three coordinates, e. g. x, y, and z, are needed to describe any position. You might have also heard that time is often regarded as the fourth coordinate. Thus, space-time is four-dimensional. A two-dimensional space is merely a plane, a one-dimensional space is a straight line. However, it turns out that there are also objects with fractional dimensionality!

Let's think of an object. As proper physicists, we can imagine that it consists of certain particles; let's agree to call them simply "atoms". For example, we can picture a volume lattice (like a crystal one) consisting of atoms, or a flat film, or a straight-line chain. The dimensions of these objects will be 3, 2, and 1, respectively. We can make sure this is true in the following way. Take a sphere of the radius R, and count how many atoms of the object there are inside the sphere. Say, this number is $N(R)$. For a volume lattice, $N(R)$ will be proportional to the volume of the sphere, $(4/3)\pi R^3$. Meanwhile, for a flat film, it will be proportional to the area of the cross-section through the center πR^2, and for a chain, to the length of the diameter $2R$. In all these examples, as you see, the dimension is given by some power of R, and the power in every case is equal to the dimensionality. In the general case, we can write:

$$N(R) = KR^{d_f} \ , \tag{13.1}$$

where K is a number independent of R. To get rid of this uninteresting constant, let's take the logarithmic derivative of each side:

$$\frac{d \ln N(R)}{d \ln R} = d_f \ . \tag{13.2}$$

The quantity d_f defined by this formula is known as the dimensionality of the object. More precisely, it is the so-called fractal, scaling, or Hausdorff dimensionality. (In maths, you may hear of lots of others, e.g. metric, topological, etc., but we shall not talk about them.)

13.2 Deterministic Fractals, or How to Draw Entertaining Patterns

"So what?" you may ask. "What's the use of Equation (13.2)? Instead of the simple idea that there are length, width, and height in three-dimensional space, we now have a complicated formula, with a derivative and logarithms. What's the point?"

Look at Figure 13.1. These patterns are called Sierpinski gaskets (after the Polish mathematician Waclaw Sierpinski (1882–1969) who invented

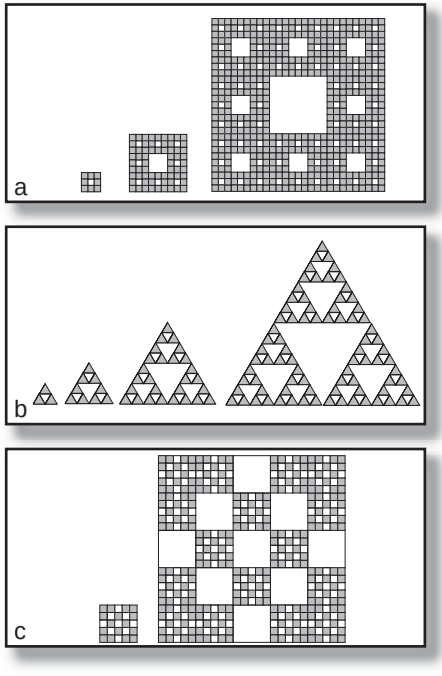


Fig. 13.1 Sierpinski gaskets — simple geometrical models of self-similar fractal patterns.

them in the beginning of 20th century). One can easily deduce the rule from the picture, and use it to create all kinds of similar patterns. In all the examples, there are two kinds of bricks, gray ones and white ones. (If you have colored pencils or a computer with a color monitor, colored bricks can be used.) The elementary bricks can be squares (e.g., see Figure 13.1 *a* and *c*), triangles (e.g., Figure 13.1 *b*), or any other sort of shape. Let's look at gasket figure 13.1 *a*, for instance. It is quite easy to make the shape on the left from the white and gray squares. We can now think of this shape as a new large gray brick. Now let's make the same kind of shape using these large gray bricks, with white bricks of the same size. Obviously, we can carry on like this *ad infinitum*. As we make bigger shapes, not only are the white "holes" larger, but so are the gray areas.

As a matter of fact, people knew about this kind of patterns many years ago. Look at Figure C13.2. This figure shows floor mosaics of the church in the village of Anagni, Italy, which was built in the year 1104; isn't it similar to a Sierpinski gasket?

Suppose the gray bricks are "atoms", whereas the white ones are just cavities. Can we work out how many "atoms" there are in the system?

Let's stick to gasket of Figure 13.1 *a*: Having made ℓ steps, we shall have a square of side 3^ℓ, therefore there are $\left(3^\ell\right)^2 = 3^{2\ell} = 9^\ell$ original elementary blocks. There are 8 "atoms" (i.e., original gray blocks) in the first figure. With every step, it gets multiplied by 8, so after ℓ steps it becomes 8^ℓ. Hence, if the size of a square is $R = 3^\ell$, then there are $N = 8^\ell$ "atoms" inside it. Simple algebra gives us $\ell = \log_3 R$, and $N(R) = 8^{\log_3 R} = R^{\log_3 8}$. Thus, the formulae (13.1) or (13.2) tell us that, in the case of Figure 13.1 *a*, the dimensionality of the Sierpinski gasket is $d_f = \log_3 8 = 3\log_3 2 \approx 1.89$.

Similar calculations lead to the dimensionality $\log_2 3 \approx 1.58$ for gasket in Figure 13.1 *b*, and $4\log_5 2 \approx 1.72$ for gasket Figure 13.1 *c*.

Thus Sierpinski gaskets are a simple model of objects with a fractional dimensionality. Of course, the naive ideas of length, width, and height cannot possibly help us when we try to determine the dimensionality of these gaskets. From the point of view of these naive concepts, the non-integer dimensionality is about as absurd as the answers with a fractional number of people that some careless primary school pupils are known to come up with occasionally. But the gaskets do have the fractional non-integer dimensionality!

So what is the physical meaning of fractal dimensionality? Since $N(R) \sim R^{d_f}$, then the greater the value of d_f, the more "atoms" can be fitted into a fixed volume of the system, hence the fewer cavities there are. In this sense, you can say that the fractal dimensionality shows how "holey" the system is. More accurately, since the maximal number of gray blocks in a Sierpinski gasket on a plane is proportional to $N_{\max} \sim R^2$, what really characterizes the "holeyness" of the system drawn on a two-dimensional plane is the value $2 - d_f$. Indeed, you can see by eye that the most "holey" gasket in Figure 13.1 is gasket *b*. And it really has the lowest dimensionality of the lot.

From what we have said, by the way, it follows that, if the system is situated on a plane, its dimensionality is $d_f \leq 2$. Similarly, in a three-dimensional space, $d_f \leq 3$, and so on. (If one fractal is placed on another one, $d_1 \leq d_2$!)

The question one may ask is this. Can the "holeyness" be described in a simpler way, for example, by means of density? Unfortunately, it can't. Take the gasket in Figure 13.1 *a*. At step ℓ, the gray blocks are spread over a fraction $8^\ell/9^\ell = (8/9)^\ell$ of the area. This is what can be most naturally thought of as the density. As you can see, since $8/9 < 1$, it tends to zero as ℓ grows. This is not surprising; it is a general law. The density is proportional to $N(R)/N_{\max}(R) \sim R^{d_f - 2}$, that is, it depends on R, and

tends to zero as R increases, as long as $d_f < 2$ (or, more generally, as long as $d_1 < d_2$).

It is very easy to write a computer program to draw Sierpinski gaskets. All you really need to do is design a subroutine that composes the very first shape out of the elementary bricks. Then you can just keep calling this routine at each subsequent stage. In other words, you use the idea of *matryoshkas*, little traditionally Russian dolls that you put into one another. This principle works not only for Sierpinski gaskets, but for many other patterns as well. Some of them are very beautiful, and they all are self-similar.

13.3 Self-Similarity

Imagine an ideal geometrical straight line. Take a piece of it, say, 1 cm long. Now "zoom in" and look at this piece on a larger scale. What do you see? Again a piece of the same straight line. You can do the same thing with a geometrical plane. To see this another way, examine a real physical straight line drawn in pencil or made of string or wire. You can look at it at higher and higher magnifications, through a magnifying glass, then through a microscope, etc. What you see still remains a straight line, until you can start to make out its width or the "atoms" of which it is made.

Sierpinski gaskets have the same property. There are two different gaskets in Figure 13.3. One of them was obtained from the other in two steps. First, the larger blocks were put together using the same "recursive" procedure as for the smaller ones, like in Figure 13.1. The resulting shape was then scaled down by a factor of 2. However, looking at Figure 13.3, you can hardly tell which figure is which! This is just what we call the property of self-similarity.

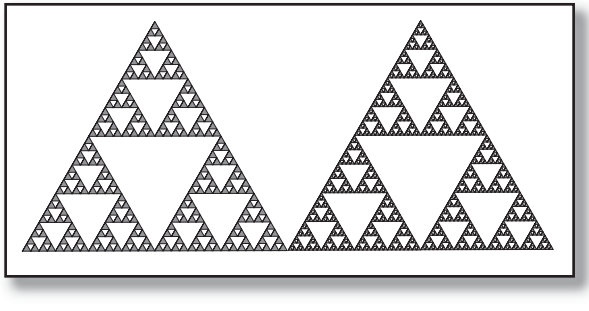


Fig. 13.3 The idea of self similarity. One of the two figures shown is the geometrically rescaled part of the other. It is really hard to say, which is which!

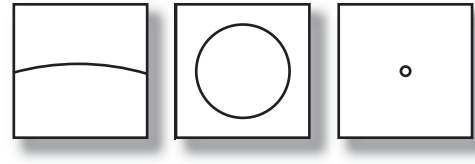


Fig. 13.4 A circle is *not* self-similar: this is how it looks like at different magnifications.

So we can see that the word *fractal* merely means a self-similar object.

Are there any fractals in nature? What is their significance in physics, if any? And what do polymers have to do with all this? Before attempting an answer, let's think about something slightly different — are there any geometrical shapes which are *not* self-similar? Of course, there are. Take, for instance, a circle. On a large scale it looks almost like a piece of a straight line, whereas on a small scale it is more like a single dot (Figure 13.4). Obviously, there is no similarity between the two whatsoever! This is precisely why the question about self-similarity is so important.

13.4 Natural Fractals

Figure C13.5 shows two photographs of what appears to be the head of a cauliflower. Actually, we took a picture of the whole head first, then cut a little floret out of the cauliflower and took a picture of it from much closer up. The pen in the photos shows the scale; if it were not there, it would be rather hard to tell one picture from the other. Moving the camera closer to the object is just the same as making a similarity transformation. Thus, a cauliflower is a self-similar object — a natural fractal.

This experiment had a nice side-effect — we have the tasty cauliflower left over! In our next experiment we shall end up with nothing but rubbish. Take a sheet of aluminium foil and cut it into little squares of different sizes. Then crumple them into little balls, and measure the balls' diameters. Now throw the balls away... (or recycle them). Figure 13.6 shows how a ball's diameter D depends on the size a of the square from which it was made. The graph is plotted on a logarithmic scale; that is, we have plotted $\ln D$ and $\ln a$ along the axes, rather than D and a, although the numbers indicated on the axes correspond to D and a themselves. You can see that the experimental points fit nicely on a straight line[1]. So, with a high degree of accuracy,

[1]The use of logarithmic coordinates is very useful whenever you search for a power law dependence. When we wrote in formula (13.1) that $N(R) = KR^{d_f}$, we definitely

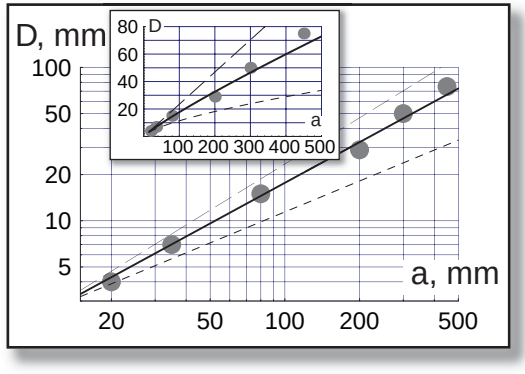


Fig. 13.6 Log–log plot of the size D of a crumpled piece of aluminium foil versus the piece's size a. Points represent the data of our measurements. The solid line gives a best fit of the form $\ln D \quad -1.18 + 0.88 \ln a$, which indicates a power law dependence $D \sim a^{0.88}$, thus manifesting the fractal structure of crumpled foil with fractal dimensionality $d_f \approx 2/0.88 \approx 2.27 < 3$. Two dashed lines are shown for comparison, they correspond to hypothetical systems with fractal dimensions 2 (unfolded foil) and 3 (dense piece of material), respectively. The inset shows the same data in linear scale.

$$\ln D \approx \alpha \ln a + b \text{ , hence} \qquad D \approx \text{const} \cdot a^{\alpha} \ . \tag{13.3}$$

We are not interested in the value of intercept b, while the slope of the line is such that $\alpha \approx 0.88$, and this reports on the value of fractal dimension.

What does it all mean? Let's go back to Equation (13.1). The amount of aluminium (the mass or the total number of atoms) in a square of side a goes as a^2, that is $N \sim a^2$. Now, this aluminium is packed into a sphere with diameter D, so $R \sim D$. Therefore, since $D \sim a^{\alpha}$, we get $N \sim D^{2/\alpha} \sim R^{2/\alpha}$. This means that the fractal dimensionality of the crumpled foil is $d_f = 2/\alpha \approx 2.27 < 3$. The foil is not completely squashed; there are many little cavities left inside the aluminium balls, and that is reflected in the fact that dimensionality is less than three. If we had a very sharp knife that could cut without squashing, we could chop the foil ball carefully and discover that the pattern at the cross-section is very much like the pattern of holes in a Sierpinski gasket. Thus, crumpled foil is also a fractal.

The inset in Figure 13.7 shows a satellite photograph of the Norwegian coastline. What can we say about the length of the coastline? We decided

view d_f as an important quantity, while K is frequently less important. Accordingly, if we take the logarithms of both sides, we get $\ln N(R) = \ln K + d_f \ln R$, and so expect the straight line the slope of which, d_f, is of interest to us, while the intercept contains information about K. Thus, whenever we need to fit to a power law, we use the log–log plot. Also, on the same note, we usually use the sign $\sim$ to indicate that we drop all the pre-factors like K and carry only the power law; in other words, we carry only the part of the formula which is responsible for the slope in the log–log plot. For instance, we will most frequently write $N(R) \sim R^{d_f}$ instead of the full version (13.1).

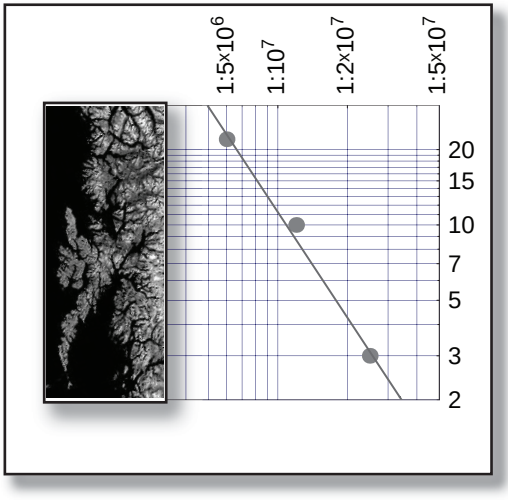


Fig. 13.7 The length of the coastline, measured from a map, depending on the scale of the map (on logarithmic axes). The inset shows the real coastline and gives the idea how wiggly it is. In the main plot, horizontal axes presents the scale of the map, while vertical baxis corresponds to the measurement of distance on the map in centimeters.

to make measurements and took three different maps of Norway. One map was fairly detailed, with the scale $1 : 6,000,000$. Meanwhile, two other maps were less detailed, $1 : 12,000,000$ and $1 : 25,000,000$, respectively. How do the three maps compare? The coarser maps miss out the finer details of the structure, such as little fjords and promontories. So the coastline drawn on a coarser map is not precisely the right shape; you could say the line is too thick to show all the little "wiggles" of the real structure (look at Figure 13.7 again).

What can we tell about the length of the coastline from the maps? Of course, the answer depends strongly on how detailed the map is. In other words, the distance traveled between two points on the coast will be different for a ship navigating in the open sea and for a little boat or canoe which has to keep close to shore and follow all the shoreline's ins and outs. Measurements of the lengths of the coastline between the towns of Bodö and Tromsö using the three maps mentioned above give 22, 10, and 3 cm respectively. These results are shown as dots in Figure 13.7. The graph itself plots the logarithm of the line's length s as a function of the logarithm of the map's scale m. As you can see, the dependence is nearly linear:

$$\ln s \approx -1.4 \cdot \ln m + \text{const} \ , \quad \text{hence} \qquad s \sim m^{-1.4} \ . \tag{13.4}$$

From here we infer the fractal dimensionality d_f of the coastline to be approximately 1.4. What does it mean that $d_f > 1$? It indicates that the actual line is "thick" (see above). Why is $d_f < 2$? That is because the

coastline is not really a surface (like a piece of fabric), but rather a border dividing the surface into two parts, sea and land. We should mention that there is nothing particularly special about Norway; you can take any other place of your liking, except Norway shoreline is so wiggly that the result could be seen with a very modest set of maps, other place may require more work.

The reader surely noticed that we mentioned the notoriously simple examples and intentionally described "experiments" (with cauliflower, crumpled foil, and geographic maps) which everyone can reproduce at home. In the books and on the internet the reader will easily find many more serious examples, in which the logarithmic plots not unlike our Figures 13.6 and 13.7, span many orders of magnitude. But we hope our examples are enough to prepare you to find it not so hard to believe that there are also "rough" surfaces with dimensionality somewhat greater than two. And some cosmologists say (and they are not joking!) that the Universe is a sort of a "foam" with dimensionality greater than four.

Now let's glance again at all the examples we have discussed. You may notice that they fall into two different groups. Some, just like Sierpinski gaskets, fit perfectly on top of themselves when you subject them to a similarity transformation. This is just the way they have been constructed. Such fractals are known as deterministic and are studied mainly by mathematicians. In contrast, fractals from the other group are self-similar only in some average statistical sense, judged by the general character of the pattern. The majority of "physical" fractals are like this.

Perhaps, for our purposes, the most important example of a fractal is the path of a Brownian particle which we discussed in detail in Chapter 6. How does an experimentalist detect a Brownian trajectory? He or she focuses a microscope, linked to a camera, on to a fluid of suspended particles which is lit by a flash, say, some k times a second. The result on the film is actually not a trajectory, but merely a sequence of points. The subsequent points are usually connected by straight lines — not because the particle traveled straight, but simply because experimenter cannot do any better; this gives a somewhat "straightened", or "coarse grained" representation of the trajectory. You can think this way about the image shown in Figure 13.8. Imagine now that we take a better equipment, for instance, a faster camera, and make flashes and photographs twice as frequently, $2k$ times a second. What we shall see is pretty much what is shown in the inset in Figure 13.8. We would see a trajectory reduced in size, but generally of the same kind. Hence, a Brownian particle's path is also a fractal.

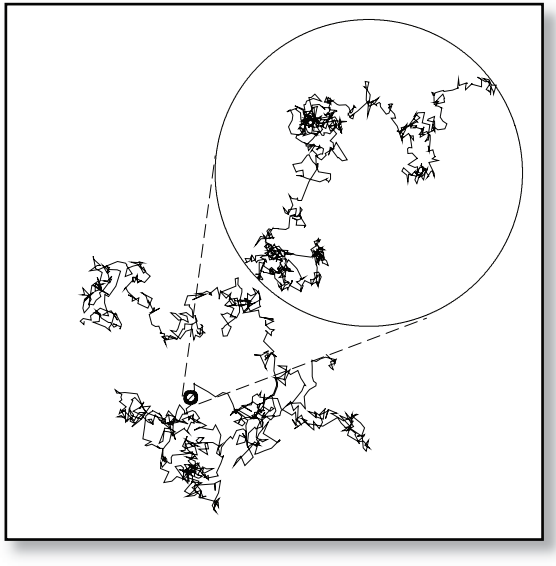


Fig. 13.8 A Brownian trajectory is self-similar on average. The random walk (or freely-jointed polymer) of 10^6 steps was generated computationally. In the main figure, every 10^3 steps are shown together as a single segment; there are $10^6/10^3 = 10^6$ of segments. In the inset, the "internal structure" of one segment is shown. It has 10^3 steps inside, and it looks very similar to the entire figure. The figure is courtesy of S. Buldyrev.

It is useful at this time to look again at Figure C9.1, particularly at the zoomed inset of that figure. Comparing it with Figure 13.8 reveals the fundamental difference between coils and globules: When we zoom in on a coil (Figure 13.8), we see a coil, but when we zoom in on a globule (Figure C9.1) we do not see a globule, we see a concentrated polymer solution instead. Some people think that fractal dimension of globule is 3, because its size scales as $N^{1/3}$; this is a mistake, globule is not a self-similar object.

Brownian motion and geography, Universe and cauliflowers — it is really astonishing how far mathematics can generalize... However, let's get back to polymers. How can all this stuff possibly be related to them?

13.5 Simple Polymer Fractals

The first "polymer" example is obvious. Indeed, we have already discussed how a polymer chain is bent and entangled, just like a Brownian particle's path. So it is bound to be a fractal!

So what is the fractal dimensionality of an ideal isolated polymer chain? The number of particles, that is, monomer units, is obviously proportional

to the contour length of the chain, $N \sim L$. At the same time, according to (6.3), the size of an N-monomer coil is $R(N) \sim N^{1/2}$. In other words, $N(R) \sim R^2$. Hence we obtain $d_f = 2$.

Thus, the fractal dimensionality of a free polymer chain turns out to be two. Although the chain is a sort of a line, its dimensionality suggests it must be more like a surface. To comprehend such a surprising result, imagine you flatten a polymer coil out on to a plane. This happens, for instance, when a polymer is adsorbed on to the surface of a solid. Alternatively, you could imagine random walks in two dimensions, for example, a rambler lost in a forest. If you have a long enough chain (or path), it will spread all over the surface more or less uniformly. (It is for precisely this reason that the rambler keeps coming back to places where he or she has already been!) You can think of this Brownian trajectory as being like a thread. It goes round and round, and gradually makes a piece of fabric, which is, of course, a two-dimensional object. In contrast, a molecule in the shape of an ordinary smooth line (such as a straight line) cannot possibly weave itself into any kind of fabric during adsorption. By the way, this is exactly why it has dimension of only one.

We shall give some more examples below. However, even now we can conclude that the self-similarity and the fractal structure are not an exception but rather a rule in polymers and other complex systems.

We were talking in Chapter 8 about the swelling of a real (not ideal) polymer coil — due to the fact that every monomer is not an infinitesimal point, but a body of maybe small, yet still finite, size. We have seen that the size of a swelling coil is $R \sim N^{3/5}$. A swollen coil is therefore also a fractal, with a fractional dimensionality $d_f \approx 5/3$.

We have talked about linear polymers so far, but what can we say about branched ones? Of course, a lot depends on what sort of branching exist. For instance, if you have a long chain with tiny side bits, you can really treat it as a linear polymer, except with somewhat peculiar monomer units. A more interesting case, though, is a randomly branched polymer. You can imagine it in the following way. Suppose a polymer molecule is gradually growing. At each point it either stops or splits into two branches. Then you get a "tree", a bit like the one in Figure 13.9[2]. As early as 1949, Bruno Zimm (1920–2005), at the university of California in San Diego, and

[2]There is quite a funny muddle of terms here: a *tree* is the generally accepted word for what we have described, yet its lattice model is known as a *lattice animal*. Does this indicate how well-informed physicists are about biology?

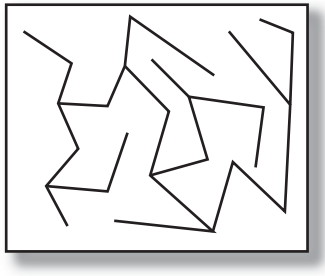


Fig. 13.9 Small piece of randomly branched tree. A much larger piece cannot be drawn on the paper, and cannot be fitted into space either.

Walter Stockmayer (1914–2004), at Dartmouth College in New Hampshire, showed that the size of such a tree containing N monomers is proportional to $R \sim N^{1/4}$. Hence, a tree of this kind is a fractal, and its dimensionality is $d_f = 4$.

"This is rather extraordinary!" an observant reader might remark. "We have never come across a dimensionality greater than 3 before." Clearly, a straight line or a Sierpinski gasket can be laid on a plane surface, since the dimensionality is less than 2. In contrast, a three-dimensional object cannot be fitted on to a plane. In the same way, it is only natural to suppose that a four-dimensional tree would not really fit into a three-dimensional space. This conclusion is correct. In the process of branching polymerization, a tree becomes thick, and either stops growing, or stops branching; in other words, it acquires longer and longer parts without any branches. (If you look around in a forest or in a park, you can convince yourself that this is actually true for real trees.) For molecular trees, this thickening of the structure appears very important in some cases. For example, it causes blood to become denser and clot in the presence of air. Anyone who has ever cut their finger will appreciate this effect!

Of course, we did not need to talk about dimensionality to draw the right conclusion about the thickening of a randomly growing chain in a three dimensional space. Suppose a molecule consists of N monomers, and its size is about $\sim \ell N^{1/4}$. Each monomer occupies a small, yet finite volume v. Then the fraction of the volume taken up by all the monomers is proportional to $Nv/\left(\ell N^{1/4}\right)^3 \sim \left(v/\ell^3\right) N^{1/4}$. So it increases with N indefinitely. On the other hand, in practice, a tree can only reach a volume fraction of order unity, and certainly no greater than 100%. So it is always true that $N < \left(\ell^3/v\right)^4$, which is interesting: linear polymer can, at least in principle, grow unrestrictedly long, while randomly branched polymer cannot, there is a solid limit for its growth; the limit might be somewhat

larger or somewhat smaller numerically depending on the branch length ℓ and volume v, but there is a limit.

Which of the two types of argument is better, the dimensionality one or the density one? We won't try to conceal that there are currently different views about this. Even the two authors of this book have had a number of discussions on the question. You are welcome to eavesdrop on one of our conversations (the names of the authors are abbreviated to A. and S.):

13.6 Why Worry About Fractals? (What the Two Authors Said to Each Other One Day)

A.: I wonder about this stuff on fractals. It feels like it's out in the cold. What really new ideas will readers have learned from this chapter?

S.: Why, they'll learn that such things as a Gaussian coil, a swollen coil, and a randomly branched polymer are all fractals. This is interesting in its own right. Mind you, there are more things in life than polymers. It might be interesting to hear about other fractals, the scale invariance of different objects, and the mathematical idea of fractional dimensionality.

A.: Perhaps you are right. But it doesn't follow from our text what you can do with fractals, what new problems they can help to solve.

S.: Yes, I see your point. But I don't think it's our fault! Suppose we weren't confined to simple examples, would we then be able to come up with such problems?

A.: The trouble is we wouldn't. As far as I know, no one has ever found anything new about polymers using fractal geometry. It was more about translating from an old language to a new one, rather than about deriving new things. Of course, it's a very beautiful language.

S.: Exactly! Remember Goethe's comment "Mathematicians are like the French..." that we put at the start of Chapter 6? But, to be serious, what I really want to emphasize is this. First, it is not just interesting — it is often useful to master different ways of describing the same thing. (Never mind that they are mathematically identical!) This is exactly what Richard Feynmann illustrated for the law of gravity, in his wonderful book, *Character of Physical Law* [38]. Second, there are loads of examples in Physics where new achievements (and sometimes rather exciting ones!) were not expressed in the language of fractals, yet were very closely connected with them, due to the use of power laws and fractional dimensions.

A.: Yes, of course, it would be hard to disagree with you on that one. One cannot help referring to the work by the Nobel Prize winner K. Wilson on the properties of strongly fluctuating systems. It turned out that the main problem was that the fluctuations do not "fit" into a three-dimensional space, but with increasing dimensionality, the situation simplifies. Even in just four dimensions it becomes trivial in some sense. So what did Wilson do? He looked at how the fluctuations behave for dimensionality $(4 - \varepsilon)$. If ε is small, than we are close to the simple case. He then found that the main features of what happens in reality ($\varepsilon = 1$) can be spotted even if you look at dimensionality 3.99.

S.: What a lovely example! I've just thought of yet another Nobel laureate, P.G. de Gennes. His ideas, all this stuff about scaling and blobs, which have proved so fruitful for polymers, are also connected with self-similarity and power laws, aren't they?

A.: Of course they are! But how much they have to do with all these light-hearted conversations about Sierpinski gaskets and cauliflowers is a matter of opinion. Mind you, we haven't even explained how power laws come into the question.

S.: Good point. Well, let's give the readers an opportunity to judge for themselves what has to do with what, and to what extent. And let's talk now about power laws.

13.7 Why Is Self-Similarity Described by Power Laws, and What Use Can be Made of This in Polymer Physics?

Starting with the very first Equation (13.1), we have come across quite a few power laws in this chapter. Just look at Figures 13.6 and 13.7. The graphs are linear in logarithmic coordinates, which means that they represent power-law dependencies. This led us to the conclusion that the objects in question are fractals, that is, that they are self-similar.

As far as polymers go, the problem is to decide what monomer units make up the chain. "What a stupid question", a chemist would say. "If the chemical formula of a compound is $\mathcal{A}_N$, and it is made up of lots of $\mathcal{A}$ molecules, then surely $\mathcal{A}$ is the monomer unit." However, why can you not connect the $\mathcal{A}$ molecules in twos, and then link all the $N/2$ dimers $\mathcal{A}_2$ together? Or, alternatively, why not link together groups of three monomer units, etc, etc? So:

$$\mathcal{A}_N = (\mathcal{A}_2)_{N/2} = \ldots = (\mathcal{A}_k)_{N/k} \ . \tag{13.5}$$

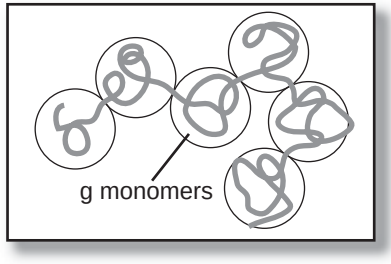


Fig. 13.10 An arbitrary number of links in polymer coil can be considered as a "new effective monomer", such new monomers are usually termed blobs.

There is even no need to talk about synthesis in this case. Let's just take a chain, choose a particular piece of it, say consisting of g monomers initially, and regard it as the new "monomer unit" (see Figure 13.10). However, the properties should not depend on the particular way we describe the structure, that is, on g. How does this come about?

Let's calculate the distance between the two ends of a polymer chain. We already know that $R \sim \ell N^{1/2}$ for a Gaussian coil — see formulae (6.3) or (6.11). Now, if we are talking in terms of the "new" monomer units of size g, we should still get the same size R, which would be expressed by the formula of the same structure, but via the size of the new "monomer", ℓ_g, and the number of new units, N_g:

$$R = \ell_g N_g^{1/2} \ . \tag{13.6}$$

So what are ℓ_g and N_g? First of all, $N_g = N/g$. As far as ℓ_g is concerned, we have to think in the following manner. Each new monomer is a tiny Gaussian coil in its own right. Therefore, ℓ_g plays the same role for this tiny coil as R does for the normal one. Thus $\ell_g = \ell g^{1/2}$. Putting all these arguments together, we obtain:

$$R = \ell_g N_g^{1/2} = \left(\ell g^{1/2}\right) \times \left(\frac{N}{g}\right)^{1/2} = \ell N^{1/2} \ . \tag{13.7}$$

Indeed, it does not depend on g!

We presume you might be interested in going through the same kind of proof for a swollen coil (which we discussed in Chapter 8); we proved then that $R = bN^{3/5}$, where b is proper N-independent, that is, associated with the monomer, length scale — see formula (8.14)), as well as for a random tree ($R = bN^{1/4}$). You would then be able to see for yourself that the power laws do indeed correspond to self-similar objects, that is, to those which have, say, a g-unit organized in the same way as the whole thing (obeying the same power law as the whole chain).

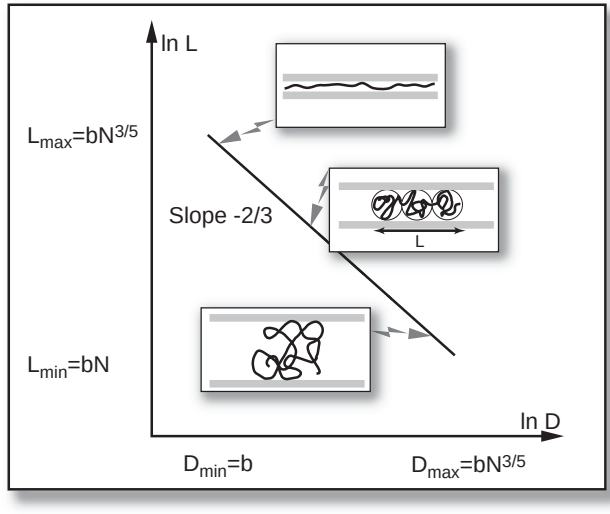


Fig. 13.11 A real (self-avoiding) polymer chain in a capillary tube. A log–log plot of the occupied length of the tube, L, vs. tube diameter, D, is linear, thus indicating the self-similar character of the structure. Typical chain conformations are shown schematically in the insets. At the smallest possible tube diameter, which is of the order of one monomer ($D_{\min} \quad b$), the polymer chain is almost completely elongated ($L_{\max} \quad Nb$). On the other hand, the polymer is not affected by the tube at large D ($D \geq D_{\max} \quad bN^{3/5}$). In between, the polymer can be viewed as a sequences of blobs, such that within each blob the chain is unaffected, while the chain of blobs is completely elongated.

Can we make any use of all this? Yes, of course, a lot! Look, for example, at the following problem. Take a real polymer coil. All we really need to know about it is just its size, $R = bN^{3/5}$. Now try to "squeeze" it into a capillary of diameter D, as Figure 13.11 demonstrates. You may just look at Chapter 8, to see that even the simple question about a real polymer coil with excluded volume is very complicated. Moreover, if the coil is placed in a capillary, it is hard even to think how to start. However, imagine that it is de Gennes himself who is tackling it, using the idea of self-similarity. He is free to choose the monomer units in any way he likes. Suppose g is such that the size of the unit is equal to the size of the capillary, that is, $bg^{3/5} = D$, and therefore $g = (D/b)^{5/3}$. The technical term for such monomer units is "blobs". Such blobs, obviously, go one after another along the capillary, just like the cars of a train[3]. Since the size of each blob is D, and their

[3]A question that might crop up here is this. What if we look at the same capillary problem for an ideal Gaussian coil? In that case, just as usual, we would choose a blob of the size of the tube, that is, $\ell g^{1/2} = D$, so $g = (D/\ell)^2$. However, the blobs will now

number is N/g, then the polymer has a length along the tube

$$L \quad D\left(\frac{N}{g}\right) \quad ND^{-2/3}b^{5/3} \ . \tag{13.8}$$

Is this all?? Yes, it is! Incredibly simple, and yet correct!

There is a useful habit which is certainly worth developing. As soon as you come across a formula, no matter how easy it is, it is a good idea to try and "get a feel for it" for some simpler limiting cases. What does this imply for Equation (13.8)? There are two limiting cases here. Imagine a spacious tube, whose width is comparable to the size of the coil. Obviously, it should make no difference to the coil. This is exactly what we get from (13.8): if $D = bN^{3/5}$ then $R = bN^{3/5}$. (For even larger values of D, this Equation (13.8) would no longer hold, of course.) Now assume the opposite: the tube is as narrow as just a single monomer. Then the chain cannot possibly meander, and has to stretch out into a straight line. This is also easily derived from (13.8). Indeed, if $D = b$ then $L = Nb$. Perhaps, our story about fractals is getting a little too long, but there are two things left which we must not miss out!

13.8 Other Fractals in Polymers, and Polymers in Fractals

Imagine little particles of soot flying out of a chimney, along with the smoke. These particles are sticky (if you don't believe it, touch them and then try to wash your hands!) When two bits of soot bang into each other, they get stuck together and make a larger particle. This flies on, picking up more and more sticky "mates" and growing in size. Thus, we end up with rather big flakes of soot. Some of them accumulate in the chimney, and others fly away. Yes, you are right if you are thinking that the structure of soot particles is self-similar, i.e. it is a fractal. Another example is snowflakes. (Talking about snowflakes, we are tempted to recommend you some reading, *Letters to A Certain German Princess* by Kepler, who is indeed *the* Kepler, the discoverer of the laws for the orbits of planets.) However, let's leave snowflakes and soot for good. What really matters for us is that many polymeric substances and materials are formed in a very similar way, when pieces of the substance stick together step by step.

pass freely through each other. This is why Equation (6.1) will not do for the random "walks" of the chain of the blobs in the tube, and we have to use (6.2) instead. It leads us to $L = D(N/g)^{1/2} = \ell N^{1/2}$. Thus, an ideal polymer squeezed into a tube does not become more stretched in the tube's direction.

Take, for instance, blotting or filter paper. It consists of a complex fractal structure of pores and channels. In a way, it looks like a piece of crumpled foil, although the whole thing is smaller (and the fractal dimensionality is slightly different).

Perhaps we should say no more about such materials, but rather discuss how they are used in practice, and that will lead us to the other side of the problem, "fractals and polymers".

A medium with complex branched pores can be used to purify and separate polymers.

This is because chains of different lengths and different chemical structure move in different ways in such a medium (cf. gel electrophoresis, Section 12.10). Thus the question arises, how does an ordinary linear polymer behave in a fractal medium? Let's take a simple model, random walks across the little gray shapes on a Sierpinski gasket. What will happen to equations like (6.2)? What will be the mean square displacement, i.e. the end-to-end distance, of the polymer chain? We hope that you are expecting to get a kind of power law for the dependence of the coil's size on the length of the chain, i.e. $R \sim N^{\nu}$. The only question is, what is ν equal to? We suggest you play this game on a computer. It is quite fun as well as instructive. For our part, we shall only say that $\nu = d/2d_s$. Here d_s is determined by what sort of sound waves would propagate along the fractal if it were made of little springs. It also depends on the heat capacity, C, of the "springy" fractal at low temperatures T ($C \sim T^{d_s}$. It would be interesting to talk about all this. Unfortunately, our book is not elastic, even though it is polymeric!)

13.9 Geometry and Classification

Perhaps every teenager goes through the period of excitement reading the adventurous novels by Jules Verne. One of the most charming characters of Jules Verne's is an absent minded scientist Jacques Paganel in "Les Enfants du capitaine Grant"; as a matter of fact, this same character appears in quite a few other Jules Verne's novels, even if under a different name. What Paganel does as a scientist, in Jules Verne's description, is he collects samples of everything — rocks, bugs, etc — and *classifies* them.

Indeed, classification of whatever subject of study is an important task of science, and finding *natural* classification is usually a big scientific leap. A few classical examples from very different fields will help to make the

point. Natural classification of chemical elements was found in 1869 by Russian chemist Dmitri Mendeleev and it has the form of periodic table, i.e., it can be drawn on the plane. Natural classification of geographical information is achieved by the map and, if we speak about Earth as a whole or its large parts, the natural map is a globe, as we know since the completion of Magellan expedition in 1522, i.e., it is drawn on a sphere. Natural classification of the living creatures was found by Carl Linnaeus in mid-1700s and it has the form of a tree; we do not want to scare the reader, but the natural place where the tree can be drawn is the non-Euclidean Lobachevsky plane.

Thus, we see that natural classification is always related to finding the proper underlying geometry. In this sense, the discovery of fractal geometry enormously widened our ability to classify various objects.

This page is intentionally left blank

Color Figures for Chapters 11–13

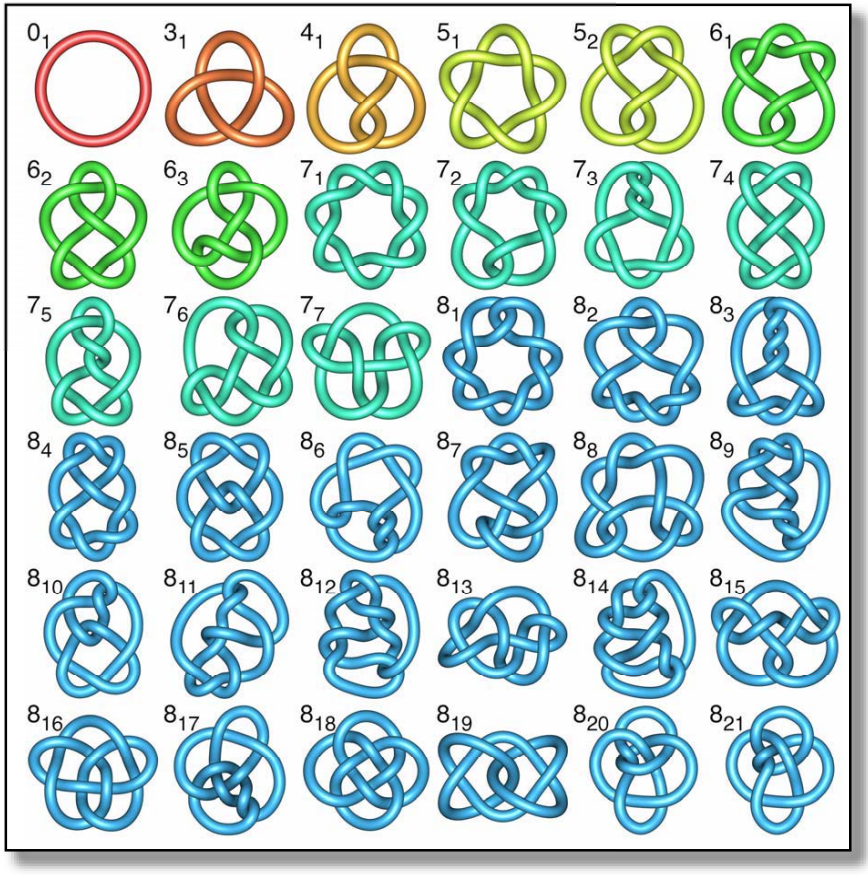


Fig. C11.2 This table shows all possible prime knots with up to 8 crossings on the projection. For chiral knots, only one of the mirror images is shown. There are many ways to identify certain separate classes of knots; for instance, knots 3_1, 5_1, 7_1 are called torus knots, because they can be nicely placed on the surface of a doughnut (and there is obviously a torus knot with any odd number of crossings). But the classification of knots relevant for their physics is yet to be developed. The figure is courtesy of R. Scharein; the knot images were produced by his software KnotPlot (see http://www.knotplot.com).

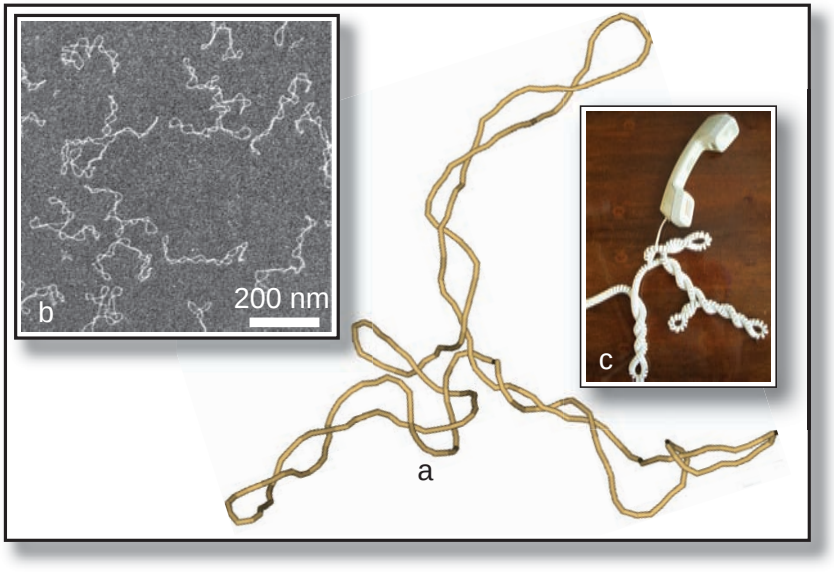


Fig. C11.4 Computer simulated plectonemic DNA in the main figure (*a*), electron microphotograph of real plectonemic DNA in the inset (*b*), and "plectonemic" telephone cord in the inset (*c*). The image (*a*) was produced by computer simulation of the 4000 base pairs dsDNA with the density of super-turns equal to -0.05; the latter means that the double helix was slightly untwisted before closing up the ring, by removing one turn of every $20 = 1/0.05$ turns of the original spiral; given 4000 base pairs, the original dsDNA would be expected to have about 400 helical turns, while the superhelical one was prepared to have about 380. The image is courtesy of A. Vologodskii. The inset (*b*) is the electron microscopy image of DNA rings about 3000 base pairs long with the same density of super-turns close to -0.05, which corresponds approximately to the linking number of about 285 instead of about 300 turns of relaxed double helix. The image is courtesy of D. Cherny.

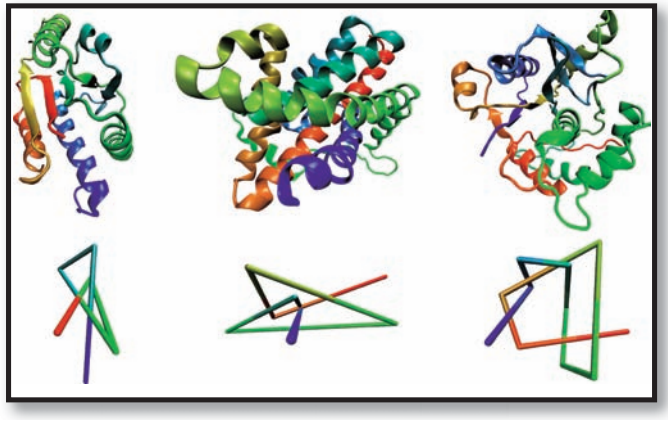


Fig. C11.5 Three examples of knots in proteins, including the very complex knot 5_1 in a protein called human ubiquitin hydrolase (the rightmost image). In each case, a simple model of sticks shows the same knot as in the protein. To help the eye, each chain is colored in rainbow colors from one end to the other. The figure is reproduced, with kind permission by the authors, from the paper: P. Virnau, L. Mirny and M. Kardar, "Intricate Knots in Proteins: Function and Evolution," PLoS Computational Biology, v. **2**, pp. 1074–1079, 2006.

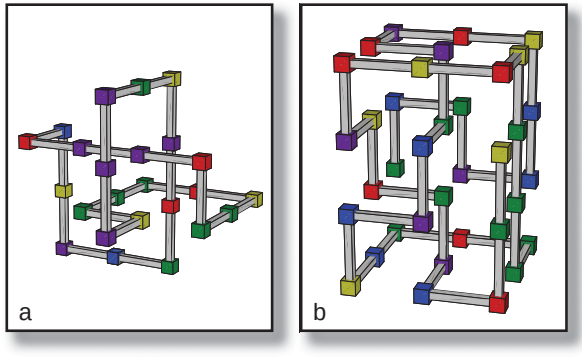


Fig. C11.7 (*a*): 24 is the minimal number of monomers on the cubic lattice to have a trefoil knot. (*b*): 36 is the minimal number of monomers for which compact polymer, filling a $3 \times 3 \times 4$ domain on the cubic lattice, can be knotted. The conformation shown does have a knot. You may wonder why monomers are shown as little cubes here and as spheres in Figures 10.2 and 10.3. This is just to emphasize that these shapes are a matter of aesthetic taste and of no real significance.

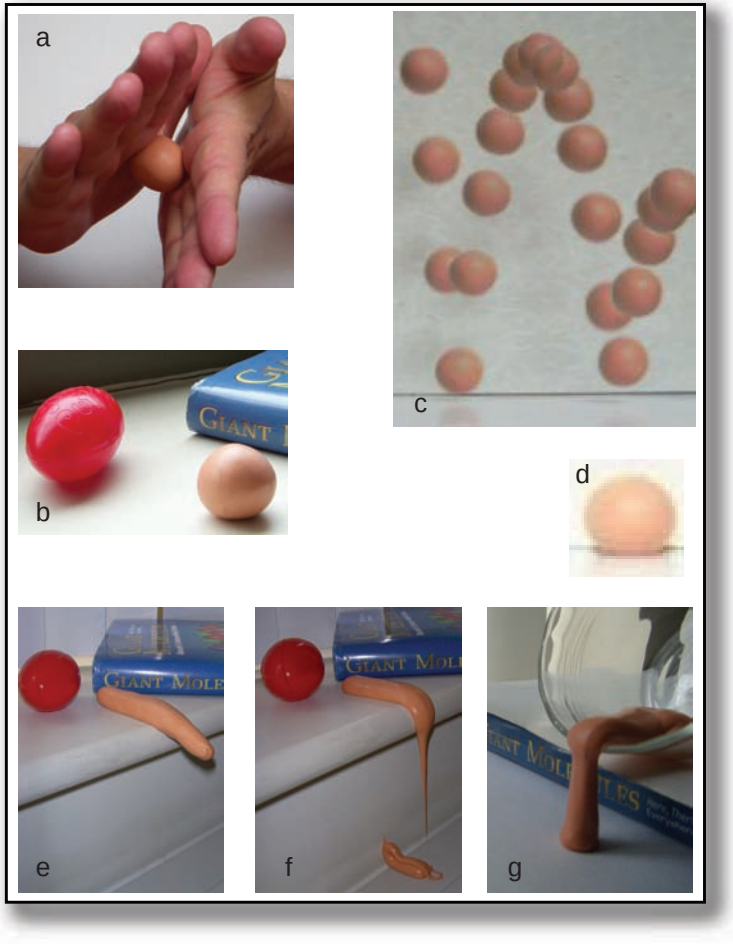


Fig. C12.2 Experiments with "silly putty" toy (a silicone) performed at home by one of the authors. By rolling between hands (*a*), the material can be easily shaped into a ball (*b*). The red plastic case in which the toy is sold as well as the first edition of this "Giant Molecules" book are shown for the scale. The ball-shaped "silly putty" bounces on the floor (*c*), it makes several jumps before stopping. The set of images (*c*) of a bouncing ball was made using regular digital camera in the movie mode with 30 frames per second. One of the frames, (*d*), by chance, captured the moment of ball impact against the floor, where we clearly see the temporarily deformed shape of the lower segment of the ball. Lower array of images illustrates liquid-like behavior of the same silicone. In the morning, the toy was formed into a sausage (*e*) and was left extending over the edge of a shelf; in the evening, it was abundantly clear that the material was a liquid flowing down to the lower shelf (*f*). Furthermore, the same sample was able to flow very slowly over the edge of the container, demonstrating the siphon effect (*g*).

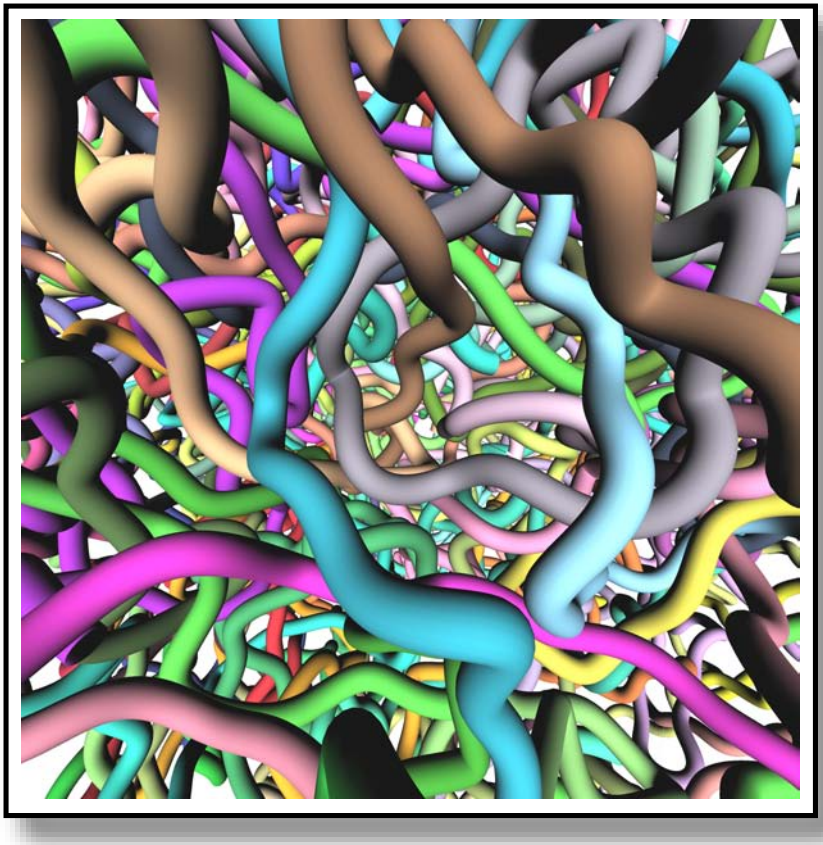


Fig. C12.3 A computer generated image of many tangled chains in a concentrated polymer system. Compare this image also with the zoomed interior of a single chain globule – see Figure 9.1. The figure is courtesy of A. Likhtman and T. McLeish. Reprinted with permission from T. McLeish, Physics Today, v. 61, issue 8, p. 40, 2008. Copyright 2008, American Institute of Physics.

Fig. C13.2 Fractal pattern that may be seen on the floor of the church in the village of Anagni, Italy (1104). The figure is courtesy of H.E. Stanley, reproduced with kind permission of Springer Science and Business media, from the book: Dietrich Stauffer and H. Eugene Stanley, "From Newton to Mandelbrot: A Primer in Theoretical Physics", Springer, 1995.

Fig. C13.5 A cauliflower and a part of it look similar to each other, except a pen shows the scale: the left and the right photos show the full cauliflower and one floret of it, respectively.

Chapter 14

Polymers, Evolution, and the Origin of Life

> What do we know about the fox?
> — Nothing. And not all of us know even that much.
>
> B. Zakhoder,
> *(Russian children's poem)*

> I cannot tell how the truth may be; I say the tale as't was said to me.
>
> Sir Walter Scott

There is no doubt that thinking about evolution and studying it truly elevates our spirits and broadens our horizons: "Living matter is the most interesting object for research for the living matter capable of researching!" (L.A. Blumenfeld). Also, this is a very big subject, and we have no choice but depict it in a much larger brush strokes than anything else in this book. Nevertheless, there is a very specific reason why we have to touch upon evolution.

14.1 Why Evolution in a Book on Polymers?

Charles Darwin published first edition of his "Origin of Species" in 1859. Ever since evolution remains the major organizing principle for all of life sciences; the cliche quotation from T. Dobzhansky (1900–1975) says it all: "Nothing in biology makes sense except in the light of evolution".

To see evolution happening one does not have to go to a paleontological museum: flu epidemic, for instance, happens every year, precisely because

viruses *evolve*; indeed, for them, our bodies is the place where they live, and their evolution selects the fittest of them, which are exactly the ones causing greatest trouble for us. Similarly, bacteria populations *evolve*, driven by us using the drugs, which is why, for instance, tuberculosis is on the rise after the decades of hope to defeat it by antibiotics.

Possibly the most impressive example of evolution is our body's reaction to a flu or a similar infection. The initial symptoms are familiar to everyone: head aches, body aches, eyes sore, nose runs; all we want is to stay in bed. This continues for a few days, and then ... then we feel better. How does this miracle happen? In fact, what happens inside our body is ... yes, evolution. We will not go into details, but in one sentence the cells of our immune system *evolve* to acquire a completely new ability to recognize and destroy the virulent intruder.

Thus, evolution is a fact, and many aspects of it do not require eons. But Darwin claimed much more in his book than the mere fact of evolution. He said that evolution is driven (mostly) by natural selection, famous "survival of the fittest" (the full title of Darwin's book reads "On the Origin of Species by Means of Natural Selection"). This aspect was decisively clarified between 1920s and 50s in what is called modern evolutionary synthesis, or neo-Darwinism, or synthetic view of evolution — the logical union of Darwinian evolution with Mendelian genetics. It was started by Ronald Fisher (1890–1962) in Great Britain, with the main idea being the role of genetic diversity carried by the population. In this picture, natural selection acts by providing relative competitive advantages or disadvantages to certain versions of the genotypes presented in the population, with the result that the population as a whole is statistically driven towards greater fitness. This idea is well illustrated by the concept of fitness landscape — something very similar to the free energy landscape which we discussed in Section 10.8, except physical systems tend to the lowest free energy, while biologists prefer the high fitness; minus free energy is thus an analog of fitness — the idea introduced in 1930s by Sewall Wright (1889–1988). The tricky part is that selective pressure is exerted on phenotypes, while genotypes are inherited, and the relation between genotypes and phenotypes is not one-to-one. Nevertheless, in the end evolution can be thought of as diffusion of a cloud of genotypes, presenting population, over the fitness landscape [22]. The development of such statistical theory of evolution was a real challenge, as evidenced by the fact that one of the world leading mathematicians of the time, Andrey N. Kolmogorov (1903–1987) of Moscow University, took active part.

When this evolutionary synthesis was formulated, people did not know what are genes. In this sense, the development of synthetic evolution theory can be compared to the initial steps of thermodynamics: Fourier, for instance, formulated correct heat conduction equation (and developed powerful methods to solve it — Fourier series) without any knowledge of what was heat. But once molecular nature of heat was understood, the science of thermodynamics received its natural foundation in statistical mechanics. Similarly, once the nature of genes, as the sequences of DNA coding for particular proteins, was understood — it opened up the doors for molecular understanding of evolution. And since the molecules involved are, of course, biopolymers of DNA, RNA, and proteins — we should touch upon this topic in this book. That is why we invite you to the discussion of physics underpinnings of evolution.

14.2 Molecular Phenomenology of Evolution

14.2.1 *Genealogic Tree and its Root: LUCA*

Thus, at the molecular level, biological evolution is — at least in part — about changes in the primary structures of biopolymers.

An important thing for physics is that a DNA double helix's shape does not really depend very much on the sequence of monomers. (This is because the pairs are mutually complementary and are hidden inside the double helix.) In this sense, DNA is like a piece of paper or a computer memory — a media suitable for recording *any* message. It is for precisely this reason that the DNA "texts" can be altered. Otherwise, the result of evolution would be not the best suited organisms, but merely DNA molecules with lower energy. In contrast, the tertiary structures of proteins strongly depend on their primary structures; this allows different proteins to carry out so many different functions, and does not allow proteins to serve as inheritable information storage.

Typically sequences of DNA and the set of sequences of proteins do not change during the life time of a particular cell, but changes do occur infrequently, from time to time, due to mutations, replication errors, and other mechanisms. What does it lead to?

Large scale sequencing of biopolymers and the growth of corresponding rich data bases produced an unprecedented wealth of data to address the genetic "closeness" of various species or even different individuals.

Take, for example, the sequence of amino acids from the hemoglobin (spelled also haemoglobin) — a globular protein present in red blood cells; it carries oxygen from the lungs or gills to other organs and tissues. In human population, about ten versions of hemoglobin sequence are known; some of them are functionally equivalent, but some present a moderate under-performance in terms of oxygen delivery function (partly compensated for the individuals by, e.g., the better tolerance to malaria — which is why this genetic version is more common in places where malaria is wide spread). This confirms the idea of genetic diversity in the population. But all these versions of hemoglobin sequences are different in only one or two positions. Compare now hemoglobin of a human, a horse and a shark. Out of 141 "letters" (aminoacids), the human and the horse have 123 in common and only 18 different ones. Meanwhile, the human and the shark agree about 62 and disagree about 79 of the amino acids. Presumably, this suggests that we are much more closely related to horses than to sharks!

Similar comparisons were performed for many thousands of proteins which have been sequenced for hundreds of species. Primary structures are being gathered and catalogued with much vigor. Without doubt, it is much easier to compare "texts" that are written out in a 20-letter alphabet, than to deal with real creatures, either living or extinct. Most importantly, comparing sequences can be formalized and entrusted to computers. The results of such analysis are used to reconstruct the "genealogical" trees of species. It is now established that there are three main domains of life — archaea, bacteria and eukaryotes, and there is very convincing evidence that the whole tree grows from one single root. That means, all of life descended from common ancestry, and, therefore, some time ago there must have been Last Universal Common Ancestor, an organism called LUCA.

The idea of LUCA started to crystallize in 1960s, when genetic code was cracked and turned out to be universal in all of the biosphere (with some marginal variations) — that was naturally interpreted as a sign of common ancestry. Now the idea of LUCA seems well established, although genetic and other features of it are not yet clear.

In fact, things are somewhat more complex than we just presented them. Indeed, there is evidence that some genetic material could be transferred in ways different from direct ancestry, this is called horizontal gene transfer. It may be realized by viruses or other means, but whatever the mechanism, in the end it means the genealogic tree is not really a tree, it is rather a

network of some sort. But a tree is a decent first approximation — good enough for us here.

14.2.2 *Further Observations*

What else can we learn from studying the sequences? There are many observations, but it seems rather hard to bring them into a system. We just choose several examples by our subjective taste to list here.

Some types of proteins have more common features, and some have fewer, when compared between different species. For instance, the so-called histones differ far less than, say, fibrinopeptides. This tells us something about the history of evolution. Indeed, this is perfectly logical that the job that histones do (i.e. packing DNA chains into chromosomes) must have emerged much earlier than the responsibilities of fibrinopeptides (blood-clotting); DNA packing is an old evolutionary invention, because it exists in all eukaryotic calls, including rather primitive ones, while blood and blood clotting is of course relatively much more recent evolutionary achievement. We can also conclude that DNA packing machinery did not evolve very much in the most recent stages of evolution.

Here is another example of an interesting observation. Imagine that you have a globular protein, and replace — not in an active center — one of the hydrophobic amino acid by another hydrophobic one, or hydrophilic by hydrophilic (see Section 5.7). Then, most likely, such a substitution will not cause much trouble: the three dimensional structure of the globule will not be violated, and the protein will still be able to do its job. Here, "most likely" does not mean "always", there are exceptions, some of them are unfortunate, for they have to do with inheritable diseases, but overall proteins have remarkable stability against mutations, as we already discussed in Section 10.6.

Genetic code also seems to have some error correcting capabilities. One particular property was noticed by M.V. Volkenstein, and can hardly be accidental. To explain the point, let's imagine that we have a gene — DNA sequence coding for a certain protein. If one DNA nucleotide is accidentally replaced by another, there is a more than even chance (in fact, about 2/3) that this will lead to the safest, that is, hydrophobic–hydrophobic or hydrophilic–hydrophilic substitutions in the coded protein.

14.2.3 *Power Laws*

Very recently, in 2003, E. van Nimwegen compared some hundreds of species and noticed something that we could have included in the Chapter 13 on

fractals, speaking about power laws[1]. It seems reasonable to think that more sophisticated organisms have generally longer genomes, and it does not sound like a great surprise that the number of protein types in the cell grows linearly with the genome length. But if we look specifically at proteins involved in regulation of DNA activities, such as transcription factors, their number in the cell scales as the square of genome length! Everyone understands that Ax^2 grows faster than Bx, irrespective of coefficients A and B, and that means the fraction of proteins used for anything but genome regulation *decreases*. Furthermore, if this scaling continues, at some genome length a catastrophe must happen when all proteins are regulators and nothing else. It is like in a business: if the fraction of managers among all employees increases with the company size, then the catastrophe gets inevitable when at a certain size company ceases to do anything, because all its workers are managing each other. From the study of fractals we know that the system must cross-over to a completely different scaling regime (see, e.g., Figure 13.11). How this plays out in evolution — we do not know yet.

Another group of evolution-relevant power laws is found in the relation between sequences and structures in proteins — molecular counterpart of classical biological relation between genotype and phenotype (see Section 5.9). Indeed, many — sometimes hundreds or even thousands — sequences can code for essentially the same conformation. To describe this mathematically, let us pretend that we have counted all sequences that code for any given conformation, S, and the number of such sequences is $k(S)$. Further, let us call p_n the number of conformations S such that $k(S) = n$; in other words, p_n is the number of structures realized by n sequences each. Then it turns out that $p_n \sim n^{-\gamma}$, where γ is usually around 2.

Furthermore, power laws and underlying fractal properties are also seen if one looks at the evolution from the point of view of protein conformations, not sequences. Here, we should mention that protein conformation appear to have been somehow selected. One aspect of it we already discussed in Section 11.6: conformations with knots seem to be under-represented compared at random. But quite apart from that, as C. Chotia of Cambridge University in England pointed out, only relatively few conformations, not more than several thousand, are featured in the proteins (his paper[2] had an interesting title: "*Proteins. One thousand families for the molecular*

[1]E. van Nimwegen, "Scaling Laws in the functional content of genomes", Trends in Genetics, v. **19:9**, pp. 479–484, 2003.

[2]Nature, v. **357(6379)**, pp. 543–544, 1992.

biologist"). Furthermore, it turns out that these selected protein conformations form a sort of self-similar network[3].

Here, we refer to a branch of fractal science which we did not touch upon before. To introduce it, let us imagine drawing a node on a (big) piece of paper for every person living on Earth, and then connecting every two nodes by a line if and only if the corresponding two persons ever had a handshake. We obtain a network. Another example would be a network of all physicists, where dots corresponding to two physicists are connected if they ever published a paper in co-authorship. Network of actors is obtained by making a bond for any two actors who participated in the same movie. World wide web is already a network, its bonds are links from one site to another. And there is also electric grid, a network of biochemical processes in the cell, etc. The remarkable discovery of around 1999, due largely to A.-L. Barabasi and colleagues at the University of Notre Dame in Indiana, is that all these networks are scale-free, which means, for instance, that in every network the number of nodes having some k bonds scales with k as $k^{-\gamma}$. Although γ might be different, in most cases between 2 and 3, the likely conclusion is that typical natural processes lead to formation of (very non-random) scale-free, self-similar networks — the ones with appreciable number of very strongly connected nodes. This of course goes along our everyday observations: some people around seem to know everyone, right?

It appears that protein conformations are also like that, they can be viewed as nodes of a scale-free network. What plays the role of bonds in this network, or "handshakes" between conformations? The answer would appear natural if the reader remembers the protein conformation contact matrix — we discussed it around formula (10.1) in Section 10.5. Indeed, the "distance" between two conformations can be characterized by the number of permutations by which the contact matrix of one conformation is transformed into the other. Then, two conformations are declared "neighbors", and are connected by a bond ("handshake"), if the distance between them is smaller than a certain threshold (and the results appear rather insensitive to the specific value of this threshold, within reasonable limits). With such definition, the network of protein conformations turns out to be scale-free. Obviously, this fact must be somehow the result of evolution. But how did it happen in evolution? And what does it lead to? Those are all topics of active current research.

[3]N.V. Dokholyan, B. Shakhnovich and E.I. Shakhnovich, "Expanding protein universe and its origin from the biological Big Bang", v. **99**, p. 13132, 2002.

14.2.4 *Statistics of Sequences*

Can statistics be of any use to analyze the "letter" sequences in protein "texts"? A similar problem exists for conventional human languages (for instance, if you want to decode a message, etc.) The Russian language, for instance, was studied from this point of view by the famous mathematician A.A. Markov around the beginning of the 20th century. There are quite a few useful things that mere statistics can reveal, with no knowledge of the language whatsoever. For example, you can distinguish poetry from prose. So what does "protein linguistics" tell us? It turns out that the protein "texts" look nearly random for statistical tests. Of course, we know that protein sequences are not really random (see Section 10.6), but their non-randomness must be of a more subtle character, difficult to reveal by mere statistics of aminoacids. O.B. Ptitsyn summed it up rather nicely: A protein is a slightly edited statistical copolymer.

Here, we would like to outline the mathematical way to look at correlations in a long sequence of any nature — aminoacids in a protein, nucleotides in DNA, letters in this book, etc. The method is to map the sequence onto a random walk. It can be done in the following way. Let's classify all the symbols in the sequence into two groups (vowels and consonants; purines (A, G) and pyrimidines (C, T); etc), and assign to every symbol in the sequence the number $\xi_i = 1$ for the first group and $\xi_i = -1$ for the second. We can imagine that i labels clock ticks, and $\xi_i = \pm 1$ represent steps of the walker to the right or to the left. This allows us to use Equations (6.6–6.11), even in a simplified version, so random walks are confined to a straight line, and go only in two opposite directions, so we shall not even need any vectors. Let R_t be the displacement after t steps. Then $R_{t+1} = R_t + \xi_t$. When we did it before, we could square this equation straightaway. This is because the average of $\mathbf{R}$ was zero. However, now we may have a drift. (By the way, in the case of polymers, the drift exists when we pull the chain by its ends in opposite directions — see Figure 7.3.) So, bearing in mind the drift, and taking the average, we obtain: $\langle R_{t+1} \rangle = \langle R_t \rangle + \langle \xi_t \rangle$. From here $\langle R_t \rangle = t \langle \xi \rangle$ (obviously, the average value of $\langle \xi \rangle$ does not depend on t). Thus, the equation for the average displacement will be of the same sort as (6.1). It describes a simple uniform motion at a speed $\langle \xi \rangle$. You may want to suggest: choose a different frame of reference. Indeed, define the displacement as $S_t \equiv R_t - t \langle \xi \rangle$. Then $S_{t+1} = S_t + (\xi_t - \langle \xi \rangle)$. Obviously, S has no drift in this case. Its average value is zero. So we can turn back to the formulae from Chapter 6.

If the text is random, then no letter is determined by the preceding one, which means, S_t and $\eta_t = \xi_t - \langle\xi\rangle$ are totally independent, yielding the "square root law", i. e. $S \sim t^{1/2}$, where $S = \sqrt{\langle S_t^2\rangle}$. However, if the text is *not* random, then one might expect $S \sim t^{\alpha}$, where $\alpha \neq 1/2$. The value of fractal index α is one of the best measures of randomness ($\alpha = 1/2$) or non-randomness $\alpha \neq 1/2$) of a sequence.

In 1992, H.E. Stanley and his co-workers in Boston University claimed that while $\alpha = 1/2$ for coding regions of DNA, it appears that $\alpha \approx 0.7$ for non-coding DNA. This means the non-coding DNA sequence is not random. Moreover, it is a fractal!

14.2.5 *Meaningful and Meaningless, Random and Fractal*

"Fractal linguistics" of non-coding, i.e. the "meaningless" part of DNA texts remains a subject of controversy.

But how about texts that make sense? Does not it sound disturbing that the coding DNA has $\alpha = 1/2$, i.e. looks statistically random along with the corresponding proteins? It is not too surprising though. Meaning is too subtle a thing to be revealed at such a primitive level when we are only looking at the fractal power α, describing average over long chunks of the sequence. Any mechanism for producing a meaningful piece of text, when choosing the next letter of the text, relies entirely on the actual context, and totally disregards any physical side of the process, i.e. what the letters and symbols are made from (be it ink on paper, nucleotides, or anything else). Just imagine a self-similar text. The whole book looks statistically like each of its chapters (which means that all the chapters are statistically similar!), and every chapter is similar to any of its sections, etc. This would not be very meaningful, would it?

Trying to give an example of something most definitely meaningless, Arthur Eddington (1882–1944), famous British astronomer, suggested once to imagine what a monkey would produce on a typewriter — given the opportunity! In fact, as far as poor monkey is concerned, we have a second thought: the piece of literature it creates would, almost certainly, lack meaning; but we think, it would be far from random, likely a fractal with $\alpha > 1/2$. Typing monkey is an excellent metaphor for a mechanism producing meaningless text, but not a random uncorrelated sequence. In fact, generation of purely random sequences is an important task in many areas of computer science, and it turns out a very difficult problem; there are

many sophisticated computer programs called random number generators, none of them is perfect.

Thus, non-trivial fractal correlations in the sequence may indicate something interesting, but most likely not a meaningful sequence. By contrast, meaningful sequence will most likely look random for someone who does not know the language. For instance, sequences of digits in decimal form of π or e look statistically completely uncorrelated, but they are definitely very meaningful. Protein sequences, too, are not random as far as their folding is concerned (see Section 10.6), but appear random for blind statistical tests.

14.3 Entropy and Evolution

14.3.1 *Life in Evolving Universe*

Lying in wait for us is the next question, how did it all start? We have talked a little about the molecular picture of the evolution of life, in which less complex forms develop gradually into more complex ones. But where did the simplest forms of life spring from? In the spirit of Darwinism, the answer is obvious: they evolved, step by step, from the inanimate world. How did this happen? There are no surviving witnesses — but there are scraps of evidence! And, of course, we can turn to theory and experiments. So what do we know about evolution before life came on the scene?

Some scientists have suggested that life did not occur spontaneously on Earth, but was carried here from outer space. However, even if there were some firm evidence for this idea (which there probably isn't), it would not help. It just moves the goal posts. We would still have to answer the question of how life appeared out there. But since we have started talking about the cosmos, it is a good place to recall that the age of the Earth ($4.5 \cdot 10^9$ years) is comparable to the age of the whole Universe ($14 \cdot 10^9$ years). This is why it is natural to regard the appearance and evolution of life on the Earth in the context of the Universe's evolution.

Our Universe started about 14 billion years ago in a "Big Bang". At the very beginning, the Universe was unimaginably dense and hot, and all was light (photons). The pressure of light caused the Universe to expand. Whilst expanding, the Universe became more rarefied and cooled down (and this is still continuing). As it cooled, other particles started to materialize — electrons and positrons, then protons, neutrons and their corresponding antiparticles, etc. We cannot do justice to this topic here, except recommending the book by S. Weinberg entitled "The First Three Minutes" [42],

but the basic principle is very simple: each particle species "condenses out" at the maximum temperature it can stand without "falling apart"[4]. Note especially that at each temperature *all* the energetically possible particle species are formed.

Initially, matter was nearly homogeneously distributed in the Universe, but the Universe evolved and developed gradients — non-uniformities of everything — temperature, density, composition, etc. Just like strong imbalance of temperature and humidity in the atmosphere creates winds and hurricanes, the sufficiently steep gradients of all scales lead to non-equilibrium dissipative processes and structures, such as galaxies, black holes, and planetary systems. We will argue that life, from this very general view point, is just one of such non-equilibrium dissipative structures.

Some theorists argue that our Universe is just a part of the multiverse — one of a very large (or infinite) number of quite different universes. In any case, the conditions and physical laws in our part of our Universe are conducive for life evolution, because if it were not — we would not be here to discuss it. This is called *anthropic* principle, and we leave it for the reader to view it as either a deep philosophy or a truism.

14.3.2 *Life and the Second Law of Thermodynamics*

Theory of evolution and thermodynamics were developed at about the same time, in mid-nineteenth century. One of the tantalizing questions, much confused by poorly educated writers, was the seeming controversy between Darwinian evolution, with gradual development of increasingly organized life forms, and the Second Law of Thermodynamics, suggesting the increase of entropy in any isolated system. In fact, there is no contradiction whatsoever, because neither any particular organism, nor biosphere as a whole is an isolated system: they consume *free* energy, which is possible precisely because we live in the place of sufficiently steep gradients. In a sense, you can say that the Earth with a surface temperature around 300°K reminds us of a water wheel. What makes the wheel turn? Obviously, it turns because it is in the way of a stream of falling water. In the same way, the Earth is in the "stream" of light that rushes from the hot Sun (with the surface temperature close to 6,000°K) to cold outer space (where the

[4]Compare this with some facts from more everyday physics. Take water. At 287°K (0°C) ice crystals "fall apart" (melt); at 387°K (100°C) water droplets "fall apart" (evaporate); at 10^4°K molecules "fall apart" into separate atoms; at 10^5°K atoms lose their electrons, turning into a plasma, etc. The higher the temperature, the smaller the units which can exist.

background radiation has a temperature of about 3°K). We all know that this "stream" is still turning the "wheels" of life.

Sometimes people say that an organism or biosphere need to consume energy. This is of course not a very meaningful wording, because energy is conserved and so cannot be "consumed", "spent", or "wasted": your body, for instance, receives, on average, the same amount of energy as it releases in all forms, and the same is true for any bacteria, any organism, or a biosphere as a whole. But we mentioned already in Section 7.8 that some forms of energy are "more equal than others": life consumes energy at low entropy and gets rid of the same amount of energy at higher entropy. Indeed, we know that not all of the energy is available to extract a useful work of any kind, but only part of energy, called free energy $F = E - TS$. Unlike energy, free energy is not conserved, it can be consumed, because entropy can be produced, and, indeed, entropy is generated in all sorts of dissipative processes, ranging from ocean currents (such as Gulf Stream) to life.

Let us make this a little more quantitative. The amount of energy delivered by light from the Sun (mostly in visible and near infrared range) is known rather accurately: the flux of solar energy at the distance of earth orbit is called solar constant, it is about $s \approx 1400\ \mathrm{W\,m^{-2}}$. Given the radius $R \approx 6370$ km, Earth receives $Q = s\pi R^2 \approx 1.8\times10^{17}$ W of power (make sure you understand why it is πR^2 instead of the full surface $4\pi R^2$; what counts here is the cross-section which blocks the way of sunlight). Due to Earth rotation, this power is distributed over the entire Earth surface; besides, some fraction of the incoming light, called albedo α, gets reflected back into space (mostly by snow and ice, by white clouds, and by deserts); $\alpha \approx 0.3$. Overall, Earth receives about $s(1-\alpha)\pi R^2/4\pi R^2 = s(1-\alpha)/4 \approx 245\ \mathrm{W\,m^{-2}}$ of power per unit area. Very nearly the same amount of energy is emitted by Earth back into space (mostly in far infrared range), because there is not much accumulation of energy anywhere on our planet.

First of all, this balance determines the temperature on Earth, because a body of temperature T emits power σT^4 per unit area, where $\sigma = \frac{2\pi^5 k_B^4}{15h^3c^2} \approx 5.7\times10^{-8}\ \mathrm{W\,m^{-2}\,K^{-4}}$ is called Stefan–Boltzmann constant (we will not derive or explain this result, except mentioning that it was one of the major ingredients in Max Planck's introduction of quant hypothesis; h and c are Planck constant and the speed of light, respectively). Balancing incoming and outgoing power amounts to a simple equation for Earth temperature: $\sigma T_{\mathrm{Earth}}^4 4\pi R^2 = s(1-\alpha)\pi R^2$, yielding $T_{\mathrm{Earth}} \approx 256$°K. The result of our

estimate is some 30 to 40°K lower than real average surface temperature, but instead of trying to improve it[5], we say that our result is accurate to less than 20%, and proceed further on this level.

Second of all, analyzing the flow of energy from Sun to Earth and then to the outer space, we can estimate the amount of *free energy* consumed by all active processes on our planet. For this purpose, a rather unexpected analogy turns out useful, namely, between Earth and . . . heat engine, such as a steam turbine. The idea is as follows. Idealized heat engine receives some amount of heat energy Q_1 per unit time from a hot body with temperature T_1, for instance, from a hot steam; a smaller amount of heat energy Q_2 in the same time is dumped to a colder body with temperature $T_2 < T_1$, for instance, to the surrounding air; the difference, $W = Q_1 - Q_2$, is the amount from which the useful work can (does not have to, but can) be extracted. Quite analogously, Earth receives energy from the sunlight, from the hot gas of photons. They arrive to us directly from Sun, therefore, temperature T_1 is nothing but $T_1 = T_{\text{Sun}} \approx 5700°\text{K}$ — the temperature of Sun's surface layer. The amount of energy received from Sun is $Q_1 = (1 - \alpha)Q$ (because the part αQ is re-emitted into space "unprocessed"). The dumping of heat occurs into the Earth's surface layer, its temperature we just estimated: $T_2 = T_{\text{Earth}} \approx 300°\text{K}$. The question is what is Q_2 and, most importantly, W.

It may seem at the first glance that $Q_2 = Q_1$: indeed, we have said that all of the energy arriving from Sun gets eventually re-emitted into space. This is true that everything ends up re-emitted — but this happens at T_{Earth}, a much lower temperature than T_{Sun}, therefore, on the way from absorption at T_{Sun} to emission at T_{Earth} some part of energy can be used to drive all dissipative processes on Earth. Indeed, this situation is exactly like engine. Consider, for instance, a car on a horizontal road: *all* of the energy extracted from burning fuel is eventually dissipated as heat in the environment, but part of this energy on the way performs the work against friction forces — the work which car driver presumably considers *useful*.

Second Law of thermodynamics sets very rigid cap on the possible amount of useful work W. To see this one has to realize that entropy of a body receiving some heat δQ at temperature T increases by the

[5]For a better estimate, one has to include such details as the difference between temperatures on surface and in the clouds, the greenhouse effect, geothermal energy (arising from the fact that Earth interior in 4.5 billion years did not (yet!) come to equilibrium and keeps dissipating), tidal waves in ocean (which dissipate kinetic energy of the Earth rotation), etc. None of the neglected effects changes the qualitative conclusions.

amount $\delta Q/T$. We will not derive this result here; we hope that the reader already realized the necessity to study thermodynamics seriously[6]. But if the result is accepted, then it is easy to find the amount of useful work for a heat engine or the amount of free energy available on Earth. Specifically, the amount of directly dissipated heat must be sufficient to increase overall entropy. To achieve this, it would be sufficient to dump a fraction of the incoming heat such that $Q_2/T_2 \geq Q_1/T_1$, or $Q_2 \geq Q_1(T_2/T_1)$. For the regular heat engine this restricts the possible amount of useful work: $W = Q_1 - Q_2 \leq Q_1\,(1 - T_2/T_1)$ — the famous Carnot theorem. The same statement applied to Earth informs us that the amount of free energy available to drive the active processes is $(1-\alpha)Q\,(1 - T_{\rm Earth}/T_{\rm Sun}) \approx 0.66Q \approx 1.1 \times 10^{17}$ W, or 230 Wm^{-2}. It is this amount of free energy that drives all the dissipative processes on Earth, ranging from huge hurricanes, winds, tornadoes, thunderstorms and ocean currents, to the hydrological cycle (evaporating ocean water to produce clouds, then rains, and then rivers), and to all forms of life[7].

Let's emphasize that radiating all of the incoming energy, in the "processed" form at temperature $T_{\rm Earth}$, our planet releases also all of the entropy produced. Therefore, both energy of Earth and its entropy remain approximately unchanged.

Life is in fact only a marginal dissipator of free energy on Earth: all photosynthetic organisms together absorb about 1% of the incoming Sun light (and all other organisms, including us, receive their free energy from the photosynthetic ones), while ocean currents, hurricanes, etc. dissipate the remaining 99%. That means evolution has plenty of free energy available. This can be also confirmed by a simple estimate of entropy associated with the high organization of biological world.

Indeed, we can use Boltzmann formula (7.2), for instance, to estimate the entropic price of building a human body from its parts. There are about 10^{13} cells in a human body, and if we assume that all of them are different and each must occupy a uniquely defined position, the entropy loss due to their arrangement will be $k_B \ln\left(10^{13}!\right) \approx 10^{14} k_B$. Similarly, each

[6]As an exercise, the reader can use the formula $\delta S = \delta Q/T$ to establish that the heat goes from a hotter body to a colder one: imagine that two bodies at different temperatures are in contact, and exchange a small amount of heat; from the requirement that entropy of the whole system increases one can establish in which direction heat flows spontaneously.

[7]Note that "free energy driving dissipative processes" and "useful work" are the same in fundamental physics sense, except we could not bring ourselves to call hurricanes "useful".

cell contains about 10^8 molecules of biopolymers; again assuming that all of them are distinct and irreplaceable, we obtain $k_B \ln\left((10^8)!\right) \approx 10^9 k_B$ for each cell, or about $10^{21} k_B$ for all cells in the body. By far the largest amount of entropy is associated with the fact that numerous proteins in every cell have their sequences strictly fixed. All the cells together have about 10^{23} base pairs in their DNAs and about 10^{25} residues in their proteins. Even if all of them are uniquely positioned, that corresponds to entropy price, respectively, $k_B \ln 4^{10^{23}} \approx 10^{23} k_B$ and $k_B \ln 20^{10^{25}} \approx 10^{26} k_B$. Thus, all the entropy associated with high organization of a human body does not exceed $10^{26} k_B \approx 10^3 \text{ J K}^{-1}$. This is a very modest amount of entropy, the corresponding free energy at 300 K is provided by the Sun light in a few hours to every square meter of the Earth surface.

To summarize, life and evolution are realized on the expense of free energy delivered by the Sun light, this free energy is — and always has been in the history of Earth — plentiful. Of course, this only means that entropy is not sufficiently good measure to characterize biological organization. But now we have to ask: how could this plentiful free energy have been used by the evolution?

14.3.3 *Chemical Evolution on the Early Earth*

What was there on the early pre-life Earth? There was the atmosphere, water, and land. And there was certainly light coming from the Sun. Violent processes were occurring: winds blew, waves battered, rivers rushed, thunder and lightning rent the air, and volcanoes exploded.... The atmosphere consisted chiefly of the simplest gases, nitrogen, carbon monoxide, steam and hydrogen. (The latter rapidly escaped from the outer layers of the atmosphere.) Importantly, there was no oxygen, the atmosphere was reducing, and oxygen appeared later, due to life.

What could have happened? Nitrogen N_2 and carbon monoxide CO, together with hydrogen H_2, gave rise to ammonia NH_3 and methane CH_4 (with the release of water). Other gases... well, do we need more examples? After all, it is not that unlikely that any simple compound could have eventually been created, even given the scanty choice of original constituents. There were just so many possibilities how it could have happened. In one or other place, in deep or shallow water, in the air, or in the sand. At any time, during millions of millennia ($\sim 10^9$ years). In larger or smaller quantities. Through one or other chain of chemical reactions. If a particular reaction needed free energy, there was no problem with that. It could have

been supplied by the ultraviolet irradiation from the Sun, electrical lightning discharges, hot volcanic products, or shock waves, etc. Clays could have played the role of enzymes. There were 24-hour cycles of different conditions of light, temperature, and humidity...

No doubt, this offers a vast field for fantasies, but there are just as many opportunities for scientific research. One type of experiment can be done as follows. A hermetically sealed vessel is filled with the right mixture of solid, liquid and gaseous ingredients. Appropriate light and electrical discharges are provided. All the conditions (such as the brightness of the light, the average frequency of "lightning" discharges, etc.) are chosen in such a way that, say, a week of the experiment would be equivalent to some 50,000 years of history. At the end, the resulting mixture is analyzed carefully. What do these studies show? The answer is: huge variety of things, up to and including aminoacids in respectable quantities. You can learn all your chemistry if you just go through the list of the final products.

To summarize, the chemical evolution of the early Earth obeyed the same principle as the cosmological evolution that we mentioned before: at each stage, you get the particles (or molecules, as long as we are talking about chemical evolution) of all the possible sorts allowed by the energy conditions... All? Really? How about polymers? Well, some protein-like polymers — protenoids were found in experiments modeling prebiological evolution[8]. But as soon as we start talking polymers — we are in a big-big trouble. Polymerization could have really changed the whole character of evolution: this is the stage which fundamentally delineates chemical evolution from pre-biological and biological ones.

People with very keen insight never really doubted that biological evolution was far more complicated than chemical evolution. As an example, let's just quote the great German scientist and philosopher Immanuel Kant (1724–1804). (By the way, he was the first to suggest a scientific theory of the evolution of the Universe, which was identified with the Solar System at that time. In particular, he proposed that the planets were formed by the condensation of hot nebulous matter.) This is what Kant wrote about evolution, more than 200 years ago: "It is easier to understand the creation of all the celestial bodies and the cause of their motion, in other words the origin of the whole present-day organization of the universe, than to find out by means of mechanics how a little blade of grass or a caterpillar appeared."

[8]S.W. Fox and K. Dose, "Molecular Evolution and the Origin of Life", Marcel Dekker, 1977.

14.3.4 *Primary Polymerization*

Out of m monomer species you can make m^N different polymers each of length N. Although they will only differ in the sequence of monomer units, i.e. in the primary structure, this difference might be very important. Thus, the number of possible polymers grows exponentially with their length, $m^N = \exp(N \ln m)$.

Physicists know that if there is a large parameter in the argument of an exponential, it needs to be treated with care. We have already come across such a situation in this book. The expression m^N is not as simple as it looks. It tells us that the Earth is not old enough, nor is there enough material on it, for all the possible sequences of monomers to have been tried. Therefore the principle to which we referred when we talked about cosmological and chemical evolution, no longer holds. (As you remember, it was that all possible sorts of particles are formed.) This takes us to a new, utterly different stage of evolution, and we will argue that it was precisely at the stage of polymerization that the first "life-like" features started to show up.

Let's make some estimates to confirm what we have just said. For example, how many protein chains of a given length (say, 200 monomers) can exist? For proteins, $m = 20$, and if we take $N = 200$, we shall have $m^N = 20^{200} = 10^{200 \log 20} \approx 10^{260}$. This number is ludicrously huge. Even if the whole surface of the Earth (roughly $5 \cdot 10^8$ km^2 $= 5 \cdot 10^{14}$ m^2) were covered with a 10-kilometer thick layer of protein-like polymers, you would "only" manage to fit in about 10^{44} chains. (This is because the volume taken up by each 200-monomer chain is roughly 200×0.1 nm^3 $= 2 \cdot 10^{-26}$ m^3.) Now imagine that every molecular collision (lasting for about 10^{-11} s) throughout the history of the Earth ($4.5 \cdot 10^9$ years) led to renewal of the primary structures of all the chains (which sounds even more unlikely than anything else we have said so far!) Even then "only" about 10^{28} attempts would have been tried out by now. Hence, our super generous overestimates give an answer of order $10^{44} \cdot 10^{28} = 10^{72}$ of primary structures that could have existed, which is a huge number, but still far too far from the desired 10^{260}. As you see, exponentials are not to be trifled with![9]

Thus, there were too many sequences to try out. Far from it. By the way, what sort of chains are we talking about? Over the last years, more

[9]This fact — the danger of exponentials — became painfully known to the king in the legend about the invention of the chess game, but his problem was $2^{64} \approx 1.84 \cdot 10^{19}$, while ours is incomparably bigger.

and more evidence has been gathered that RNA played the main role in prebiological evolution, that it was a sort of "RNA world". Indeed, RNA can carry out DNA's "instructive" function, and could work as a catalyst (although not quite as good as proteins).

Interestingly, the role of RNA puzzled scientists since the very dawn of molecular biology. Already in 1954, Gamow[10] founded the elite "RNA Tie Club". The aim was "to solve the riddle of the RNA structure and to understand how it built proteins." The club consisted of 20 regular members (one for each amino acid), and four honorary members (one for each nucleotide). Each member received a woolen necktie, with a helix embroided in green and yellow (idea and design by Gamow). The club was quite influential, although as it happened the genetic code was cracked not by the members. In the very recent years, the interest in RNA grew up once again, this time it is mostly about the non-coding short microRNAs that play crucial role in regulation — one more interesting subject beyond the framework of this book.

Here, we are not that worried about the particular chemistry of the polymers. What is more important is the very fact of the chain structure. It is just because all the monomers line up in a chain that we end up with a horrendous number of possibilities, m^N, which cannot all be tried out.

The mixture of early evolutionary products on the Earth is usually called the primordial soup. In this soup, there were monomers which could join up with each other, given favorable conditions, and they formed some polymer chains. This is a well established fact, proved by laboratory experiments. Moreover, some of the polymers created had a slight ability to act as catalysts; this has also been confirmed by experiments. However, what could have happened next is much less clear.

One scenario discussed in the literature is the formation, by chance, of mutually catalyzing polymers. Manfred Eigen of Max Planck Institute in Göttingen (Germany) and his co-workers worked out a model, called hypercycle, which represents a sort of evolution on the level of polymers [25]. The model is dressed up in a beautiful mathematical clothes of differential equations, but its essence is simple.

[10]George A. Gamow (1904–1968) was a very non-usual scientific star, as he produced three first rate ideas in three unrelated fields of science: Gamow was the first to explain radioactive decay as quantum mechanical tunneling; he was the first to predict the cosmic background radiation — the relict of the Big Bang; and Gamow was the first to formulate the idea of genetic code. He was born in Odessa and educated in Leningrad (now St. Petersburg) in the early years of Soviet Union, where he shared friendship with Lev D. Landau and Matvei P. Bronshtein. In 1934, he managed to escape from the USSR and emigrated to the USA.

Suppose a polymer chain $\mathcal{A}_1$ was created accidentally in the primordial soup. Entirely by chance, it is a weak catalyst, and can speed up the production of another polymer $\mathcal{A}_2$; in other words, it stimulates the monomers from the soup to join each other, forming chains of $\mathcal{A}_2$. Similarly, we assume that $\mathcal{A}_2$ helps making $\mathcal{A}_3$, etc, up to some other polymer $\mathcal{A}_k$ that helps making $\mathcal{A}_1$. It should be clear why this is called hypercycle! As soon as such a structure appears in the soup, it will cause a kind of explosion! The system will start making its own copies, one after another. It will be a snowballing process, following a geometric progression, until all the monomers in the soup are used up. Due to mutations and mistakes, hypercycles will not be identical to each other. If one of the hypercycles has more efficient catalysis, it will reproduce itself more rapidly. Hence, there will be more chains of the corresponding type. Meanwhile, the stock of monomers in the primordial soup is limited, and is shared between all. Moreover, the chains tend to break spontaneously from time to time. The monomers which are released when some chains are broken are then reused to build new chains — most likely they will be used by the most efficient hypercycle, etc.

This story is remarkably similar to how Darwin's "survival of the fittest" is often described, at a primitive level. The monomers play the role of food, and self-copying chains or hypercycles play the role of living beings who reproduce themselves given enough food. As a result of the competition for fodder, only the most gluttonous and prolific species will survive. Recently, such mutual catalysis system was experimentally realized with RNA[11]. Nevertheless, how it could have started in a primordial soup conditions remains an open question.

The question is essentially that of the chicken and the egg, or, in molecular biology formulation, which function came first, that of a DNA carrying a blueprint of how to make a protein, or that of a protein synthesizing DNA? Since we know that not all of the sequences could be tried, we have to ask: what is the fraction of sequences that are capable of, for instance, mutual catalysis, hypercycle-style, strong and selective enough to start the feedback loop? Can we calculate the probability that a sequence of monomers, picked at random, will be a functioning polymer? This is a similar task, in a sense, to the prediction of the tertiary structure of a globular protein (see Chapter 10). Neither problem has been solved, but both are intensively studied. Who knows, maybe some of the readers of this book will manage to clarify the matter?

[11]T.A. Lincoln and G.F. Joyce, "Self-Sustained Replication of an RNA Enzyme", Science, v. **323**, n. 5918, pp. 1229–1232, 2009.

14.3.5 *Memorizing of a Random Choice*

We concluded previous section with a question. Just in case if anybody wants to give it a try, we ought to explain how is this problem related to what we said in Section 14.2.5 about the difference between meaningful and meaningless texts? The Eddington monkey experiment suggests that it is enormously unlikely to produce a meaningful text of any sensible length entirely by chance. In the primordial soup we seem to face that very problem — except ...

When Eddington monkey sits at the typewriter, someone knowing language watches over the monkey's shoulder and waits for a piece that makes sense in that language. But imagine that we ask the monkey to make a password for our new computer account. That would be a totally different story! Very likely, with probability close to 100%, the monkey will produce an excellent password very quickly. And as soon as the password is entered and memorized by the computer — it becomes highly meaningful, because it opens the account up! In this case, the monkey is successful, because it creates new information from scratch; you can say it creates the word and the language in which this word makes sense.

For present day biopolymers, "meaningful" means compatible with all other existing biopolymers, and its creation by chance is as unlikely as writing by chance a literature masterpiece. But the situation in primordial soup might have been different, when you only needed some polymer with a decent catalytic function: it isn't over until the fat lady sings.

By the way, if such spontaneous polymerization under any circumstances is possible — why does not it happen right now? Well, the conditions might have changed, present atmosphere, for instance, is oxidizing — but let's also recall a really wonderful piece from Darwin's letter to his friend J. Hooker: "They often say that at present there are precisely the same conditions for primitive living creatures to appear as there used to be some time ago. However, if now (oh, what a really big "if"!), in some warm little pond containing all the necessary salts of ammonia and phosphor, and accessible for the action of light, heat, electricity, etc., a protein resulted from chemical reactions that was capable of further, more and more complicated transformations, then this protein would be immediately destroyed or absorbed, which could not have happened in the period before the appearance of living creatures." In other words, if nowadays a chain of a biopolymer is accidentally synthesized, it is bound to be eaten before anything interesting can happen to it.

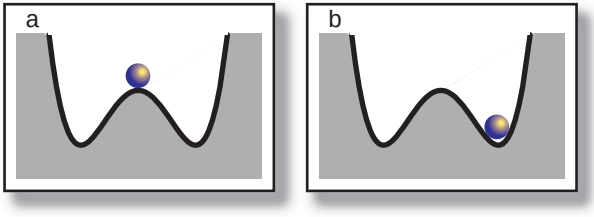


Fig. 14.1 A mechanical illustration of how a random choice is memorized.

This situation is a bit similar to a simple mechanical model shown in Figure 14.1. The first diagram, Figure 14.1 *a*, depicts a perfectly symmetric yet unstable system. Then the symmetry is broken at random (or spontaneously, as they often say). As a result, the system comes to a stable position (Figure 14.1 *b*). The stability means that the random choice that has been made is now "engraved" into the system's memory. Indeed, before the polymerization starts great many polymer sequences are equally likely — in this sense they were symmetric; similarly, all combinations of symbols are equally likely candidates for password. But as soon as the choice is made — because you have chosen the password or because some sequence have actually appeared and catalyzed their products — this random choice is memorized, and it gives things the meaning.

The memorizing of the random choice turns out to be a very interesting thing. Let's give a few examples.

14.3.6 *Right and Left-Handed Symmetry in Nature*

Most people have their hearts not in the middle of their bodies, but on the left-hand side. In contrast, a DNA double helix, a triple helix of collagen, and α spirals of globular proteins all have a right-handed structure. In engineering, except in some very special cases like the left-hand pedal of a bike, only right-handed screws are used. Why?

Let's start with engineering. A left-handed thread is no worse than a right-handed one. However, imagine a child's do-it-yourself kit where right-handed screws and nuts are mixed up with left-handed ones. The symmetry is, therefore, preserved. But it is quite awkward to play with, to put it mildly! This awkwardness comes from the instability of the symmetrical state. The direction in which the symmetry was broken (right-handed screws) was chosen more or less by accident in the past. However, now it is

well established and is retained by standards and tradition. In other words, this choice has been memorized, and the system is now fairly stable. It is quite unthinkable that left-handed screws will suddenly come into fashion!

The situation with molecular "screws" in nature is a little more complex. In atomic nuclei, there are mirror-asymmetric interactions which are called "weak". They are weak indeed. At least, they hardly affect the properties of the atoms. Some time ago, this influence was even disputed all together. However, in the 1970s scientists managed to detect it, in very refined optical experiments with bismuth vapor. Weak interactions make right-handed and left-handed spiral molecules differ slightly in energy. This difference is very small, and it is estimated to be about 10^{-17} of the characteristic energy. We personally subscribe to the view that such a minute discrepancy could not have played any role, and the modern biosphere is asymmetric in the way it is entirely by memorizing the random choice.

There is a curious fictional story on this subject. It is about a shipwreck. The victims die of starvation on an unknown island, although beautiful fruits grow there in abundance. The clue is that this island is really a land behind the looking glass. It is a country of left-handed DNA, collagen, and α-helices. The fruits grown there are no good for eating.

Yet another example is the driving. People drive on the right in most countries, though there are exceptions, such as England, Ireland, Japan, etc. You can find on the Internet a map showing in different colors countries with left- and right- hand roads. You will see that most of the left-driving countries are either on the islands or geographically quite isolated from the neighbors. We can speculate that in this sense our planet is like an island, where the initial random choice[12] of chirality is memorized, fixed by tradition, and remains unchanged without strong influence from outside.

Many more fascinating details about the left-right asymmetry can be found in the book [46].

14.3.7 *QWERTY*[12]

Do you recognize the "word" QWERTY? Look at the keyboard of your computer, upper left — see it? Why do (almost) all computers, at least

[12]For the driving, the initial choice in many cases was not independent, but rather imposed by the British colonial administration — which is not related to our subject here.

[12]This subsection was written by A. Aerov.

in the English speaking world, have the same order of characters on the keyboard?

Present day most common layout of the keyboard was first suggested in 1870s by Christopher Latham Sholes (1819–1890), a journalist and newspaper editor in Milwaukee, Wisconsin. He was concerned with the problem of arms jamming in the mechanical typewriters in case of pressing two neighboring buttons in a rapid succession. A common misconception is that he was trying to reduce the typing speed; a more likely version is that he was trying to allow for a high speed reliably, without jams. In any case, his idea was to place the most common (in English) pairs of characters, such as th or st, as far on the keyboard as possible.

Of course, jamming of the mechanical arms is not an issue for modern computers, but QWERTY layout is still by far the most common. Furthermore, August Dvorak, a professor of education in Seattle, showed already in 1940s that even a randomly reshuffled keyboard will most likely be more convenient for the typist's hands than QWERTY... But QWERTY still survives. Why?

Of course, the explanation is the memorization of random choice. So many people are so much used to the current keyboard layout, that changing it does not appear practical. And this is a very common situation in economics, politics, etc. Arrangement of major controls in a car; the width of the rail gauge; the set of accepted computer programs; the design of appliances; the association of a firm with a certain groups of consultants, customers, etc; voltage in the electric grid; obsolete laws and law implementation legal practices — all that and many other similar examples illustrate again and again the role of random choice of one alternative, which is very difficult if not impossible to change later. QWERTY is a quintessential example of such situation, and the very "word" QWERTY is an example of a meaningless string of letters acquiring "meaning" because of the memorization of random choice.

14.3.8 *Emergence of Novel Information*

The appearance and progress of life is not the only case of evolution based on impossibility to test out all variants. There are other examples, such as the emergence and development of languages, literature, art and science, and even the game of chess. All these systems, in a way, do the same things as living beings. You can say that scientific ideas and pieces of literature are also born. Some of them die (but not all!). Many leave offspring. (For

example, the first derivation of the heat conduction equation was based on the belief that heat is carried by some sort of medium or ether. Thus, the heat conduction equation is an offspring of a now dead concept. Similarly, our old friend Don Quixote may not have appeared without a whole bunch of medieval novels about knights, now happily forgotten.) What is most important for us is that all such systems develop by memorizing random choices. For instance, it is more or less by accident that English speaking people use the word "number" for the concept of number. We could have had another word instead (as, say, the Russians do; they call it "chislo".) However, once made, the random choice is cemented by books and people's memories. The attachment of the words to their meanings is no less stable than the position of the ball in the little hole in Figure 14.1 *b* or the technical standard of right handed screws.

It is interesting that in language there are some general laws which do not depend on random choices, i.e. on the meanings that particular words happen to have. Such are the laws of poetry, for example. This becomes clear from the following absurd yet "proper" poem:

> Hunkle, chinkle, mrony phar,
> Brough I junder whow mee dar?
> Up above the fye bo clar,
> Hunkle, chinkle, lubby phar.

Is there anything similar in the physics of biopolymers, any general laws that are not affected by the random choices? There certainly are! They control the formation of knots in DNA (see Section 2.6), the hydrophobic-hydrophilic separation of a globular protein (Section 5.7), and many other properties; most of these laws may still be unknown.

When you (or your a computer) solve equation $2x = 4$ and obtain $x = 2$, no new information is created. But when you start from $x^2 = 4$ and obtain $x = 2$ — you do create a bit of information, because you have chosen $+2$ and threw away -2. You brake the symmetry. Information is created by erasing one of the two *a'priori* equivalent possibilities.

As Henry Quastler (who started as a medical doctor in Albania, and ended up as biophysicist in Brookhaven National Lab, 1908–1963) showed in 1962, in his essay [23] which appears to be under-appreciated in the English speaking world, when a system memorizes a random choice — and only in this case! — it creates a novel bit of information. This information is about facts, which were never known or questioned, nor even existed before. For example, in which hole is the ball in Figure 14.1 *b*? What sort

of primary structure does the oxygen-carrying protein have, in human red blood cells? Which words are used for one or another concept? How exactly did Tom Sawyer manage to get the fence painted? And so on. As a matter of fact, the memorizing of random choices is also relevant to creativity, both in art and science — but we shall not discuss it any further.

14.4 Conclusion

From all we have said we can draw the following conclusion. There are systems where you cannot physically try out all the possibilities, there are just far too many of them. The only way such systems can evolve is by memorizing random choices. This is how new bits of information are created. The first in history as well as the simplest event of this sort was probably the synthesis of polymer chains in the primordial soup. This is why the entire field of evolution is intimately related to polymer physics.

There is another fascinating aspect of the origin of life. Is there life elsewhere in the Universe? Who can tell. From physics prospective, the main inhabitability condition is the presence of plentiful free energy, and there are such places. It is worth searching for the signs of primordial polymerization even on the Mars. What we think must be the most promising places in the Solar System are the methane seas on Titan, a satellite of Saturn, or an ice-covered ocean on Europa, a satellite of Jupiter. Unfortunately, we have not the foggiest idea of how to get there to check. This is why, as it is often said in such cases in scientific books, we leave this as an exercise for the reader...

To conclude, we would like to warn you that perhaps not all the experts would agree with what we have said in this chapter. Some may argue that whenever physicists try to discuss such matters it is merely naive and useless. We would like to respond to this with the following words from the famous *Feynman Lectures on Physics*: "The most important hypothesis in all of biology ... is that *everything that animals do, atoms do.* In other words, *there is nothing that living things do that cannot be understood from the point of view that they are made of atoms acting according to the laws of physics.* This was not known from the beginning: it took some experimenting and theorizing to suggest this hypothesis, but now it is accepted, and it is the most useful theory for producing new ideas in the field of biology." And more in the other place: "Certainly no subject or

field is making more progress on so many fronts at the present moment[13], than biology, and if we were to name the most powerful assumption of all, which leads one on and on in an attempt to understand life, it is that *all things are made of atoms*, and that everything that living cells do can be understood in terms of the jigglings and wigglings of atoms."

We can only encourage our readers to continue exploring this wonderful hypothesis. Good luck!

[13]It was written in 1963, and it is equally fair now, in 2009.

List of Suggested Further Reading

Monographs and Textbooks on Polymer Physics:

[1] Pierre-Gilles de Gennes, "*Scaling Concepts in Polymer Physics*", Cornell University Press, Ithaka, NY, 1979.

[2] Masao Doi and Sir Sam F. Edwards, "*The Theory of Polymer Dynamics*", Oxford University Press, Oxford, 1986.

[3] Alexander Y. Grosberg and Alexei R. Khokhlov, "*Statistical Physics of Macromolecules*", AIP Press, NY 1994.

[4] Michael Rubinstein and Ralph H. Colby, "*Polymer Physics*", Oxford University Press, 2003.

Soft Matter Physics:

[5] Jacob Israelachvili, "*Intermolecular & Surface Forces*", Academic Press, 1992.

[6] Paul M. Chaikin and Tom C. Lubensky, "*Principles of Condensed Matter Physics*", Cambridge University Press, 2000.

[7] Thomas A. Witten and Philip A. Pincus, "*Structured Fluids. Polymers, Colloids, Surfactants*", Oxford University Press, 2004.

[8] W.C.K. Poon and David Andelman, "*Soft Condensed Matter Physics in Molecular and Cell Biology*", Taylor & Francis, 2006.

Molecular Biology and Biological Physics:

[9] Bruce Alberts, Alexander Johnson, Julian Lewis, Martin Raff, Keith Roberts and Peter Walter, "*Molecular Biology Of The Cell*", Garland Publishing, 2007.

[10] Bruce Alberts, Dennis Bray, Karen Hopkin, Alexander Johnson, Julian Lewis, Martin Raff, Keith Roberts and Peter Walter, "*Essential Cell Biology*", Garland Publishing, 2009.

[11] Philip Nelson, "*Biological Physics. Energy, Information, Life*", W.H. Freeman and Company, NY, 2004.

[12] Rob Phillips, Jane Kondev and Julie Theriot, "*Physical Biology Of The Cell*", Garland Publishing, 2008.

[13] Kim Snappen and Giavanni Zocchi, "*Physics in Molecular Biology*", Cambridge University Press, 2005.

[14] Charles Cantor and Paul R. Schimmel, "*Biophysical Chemistry*", W.H. Freeman & Company, NY, 1980.
[15] Mikhail V. Volkenstein, "*Molecular Biophysics*", NY Academic Press, 1977;
"*Physics and Biology*", NY Academic Press, 1982;
"*General Biophysics*", NY Academic Press, 1983;
"*Physical Approaches to Biological Evolution*", Berlin & NY, Springer Verlag, 1994.

Introductory Physics with Biophysical Emphasis:

[16] Robijn Bruinsma, "*Physics for Life Scietists*", (Physics 6A, 6B, & 6C), UCLA, 1998.
[17] Ken A. Dill and Sarina Bromberg, "*Molecular Driving Forces: Statstical Thermodynamics in Chemistry and Biology*", Garland Science, 2003.

Protein Physics:

[18] Carl Branden and John Tooze, "*Introduction to Protein Structure*", Garland Science, 1999.
[19] Alexei V. Finkelstein and Oleg B. Ptitsyn, "*Protein Physics*" (a course of lectures), Academic Press, 2002.
[20] Gregory A. Petsko and Dagmar Ringe, "*Protein Stucture and Function*", New Science Press, 2003.
[21] Eugene I. Shakhnovich, "*Protein Folding Thermodynamics and Dynamics: where Physics, Chemistry, and Biology Meet*", Chemical Reviews, v. **106**, pp. 1559–1588, 2006.

Polymers and Evolution:

[22] Ronald A. Fisher, "*The Genetical Theory of Natural Selection*", Oxford University Press, 1930.
[23] Henry Quastler, "*The Emergence of Biological Organization*", Yale University Press (New Haven), 1964.
[24] Manfred Eigen, "*Selforganization of Matter and the Evolution of Biological Macromolecules*", Die Naturwissenschaften, Springer, 1971.
[25] Manfred Eigen and Peter Schuster, "*The Hypercycle, a Principle of Natural Self-Organization*", Berlin-NY, Springer Verlag, 1979.
[26] Motoo Kimura, "*The Neutral Theory of Molecular Evolution*", Cambridge University Press, 1983.
[27] Andrew H. Knoll, "*Life on a Young Planet: The First Three Billion Years of Evolution on Earth*", Princeton University Press, 2003.
[28] Konstantin B. Zeldovich and Eugene I. Shakhnovich, "*Understanding Protein Evolution: From Protein Physics to Darwinian Selection*", Annual Review of Physical Chemistry, v. **59**, pp. 105–127, 2008.
[29] Daniel F. Styer, "*Entropy and Evolution*", American Journal of Physics, v. **76**, n. 11, pp. 1031–1033, 2008.
[30] Charles H. Lineweavera and Chas A. Egan, "*Life, Gravity and the Second Law of Thermodynamics*", Physics of Life Reviews, v. **5**, n. 4, pp. 225–242, 2008.

[31] Eugene V. Koonin, Yuri I. Wolf and Georgy P. Karev, "*Power Laws, Scale-Free Networks and Genome Biology*", Springer, 2006.

Topology in DNA and Polymers:

[32] Alexander V. Vologodskii, "*Topology and Physics of Circular DNA*", CRC Press, 1992.
[33] Andrew D. Bates and Anthony Maxwell, "*DNA Topology*", Oxford University Press, 2005.
[34] Enzo Orlandini and Stuart G. Whittington, "*Statistical topology of closed curves: Some applications to polymers*", Reviews of Modern Physics, v. **79**, pp. 611–642, 2007.

Biophysical Classics:

[35] Edward M. Purcell, "*Life at Low Reynolds Number*", American Journal of Physics, v. **45**, pp. 3–11, 1976.
[36] Howard C. Berg, "*Random Walks in Biology*", Princeton University Press, 1993.

Popular Books:

[37] Erwin Schrödinger, "*What is life?: the Physical Aspect of the Living Cell*", Cambridge - NY, University Press, McMillan, 1945; Enlarged edition, 1992.
[38] Richard P. Feynman, "*Character of Physical Law*", Cambridge, M.I.T. Press, 1967.
[39] Jacques Monod, "*Chance and Necessity: an Essay on the Natural Philosophy of Modern Biology*", NY Knopf, 1971.
[40] James Watson, "*Double Helix*", A Norton Critical Edition, edited by Gunther S. Stent, Norton, 1980. [This edition is particularly interesting, because it contains, along with the original book, also several brilliant reviews of it, along with interesting discussion.]
[41] Freeman Dyson, "*Origins of Life*", Cambridge University Press, 1985.
[42] Steven Weinberg, "*The First Three Minutes: a Modern View of the Origin of the Universe*", NY, Basic Books, 1988.
[43] Francis Crick, "*What Mad Pursuit: A Personal View of Scientific Discovery*", Basic Books, NY, 1990.
[44] Manfred Eigen and Ruthild Winkler, "*Laws of the Game: How the Principles of Nature Govern Chance*", NY, Harper & Row, 1983;
Manfred Eigen and Ruthild Winkler-Oswatitsch, "*Steps Towards Life: a Perspective on Evolution*", Oxford-NY, Oxford University Press, 1992.
[45] Maxim D. Frank-Kamenetskii, "*Unraveling DNA*", NY, VCH, 1993.
[46] Chris McManus, "*Right Hand, Left Hand: The Origins of Asymmetry in Brains, Bodies, Atoms and Cultures*", Harvard University Press, 2002.

Fractals:

[47] Benoit Mandelbrot, "*Fractals: Form, Chance and Dimension*", SF Freeman, 1977; "*The Fractal Geometry of Nature*", SF Freeman, 1982.
[48] H. Eugene Stanley, "*Fractals in Science*", Springer, 1994.

[49] Dietrich Stauffer and H. Eugene Stanley, "*From Newton to Mandelbrot: A Primer in Theoretical Physics*," Springer, 1995.

[50] R. Albert and A.-L. Barabasi, "*Statistical Mechanics of Complex Networks*", Review of Modern Physics v. **74**, pp. 47–97, 2002.

Nanopore Sequencing:

[51] D. Branton, D.W. Deamer, A. Marziali, H. Bayley, S.A. Benner, T. Butler, M. Di Ventra, S. Garaj, A. Hibbs, X. Huang, S.B. Jovanovich, P.S. Krstic, S. Lindsay, X.S. Ling, C.H. Mastrangelo, A. Meller, J.S. Oliver, Y.V. Pershin, J.M. Ramsey, R. Riehn, G.V. Soni, V. Tabard-Cossa, M. Wanunu, M. Wiggin and J.A. Schloss, "*The Potential and Challenges of Nanopore Sequencing*", Nature Biotechnology, v. **26**, pp. 1146–1153, 2008.

Optical Tweezers:

[52] S.P. Smith, S.R. Bhalotra, A.L. Brody, B.L. Brown, E.K. Boyda and M. Prentiss, "*Inexpensive Optical Tweezers for Undergraduate Laboratories*", American Journal of Physics v. **67**, pp. 26–35, 1998.

Historical:

[53] Mitchell Wilson, "*American Science and Invention: a Pictorial History*", Simon and Schuster, New York, 1954.

[54] William Irvine, "*Apes, Angels, and Victorians: Darwin, Huxley, and Evolution*", Meridian Books, Cleveland and New York, 1967.

[55] George L. Trigg, "*Crucial Experiments in Modern Physics*", NY Van Nostrand Reinhold Co, 1971;
"*Landmark Experiments in Twentieth Centrury*", NY, Crane Russak, 1975.

[56] Herbert Morawetz, "*Polymers: The Origins and Growth of a Science*", John Wiley & Sons, 1995.

Index