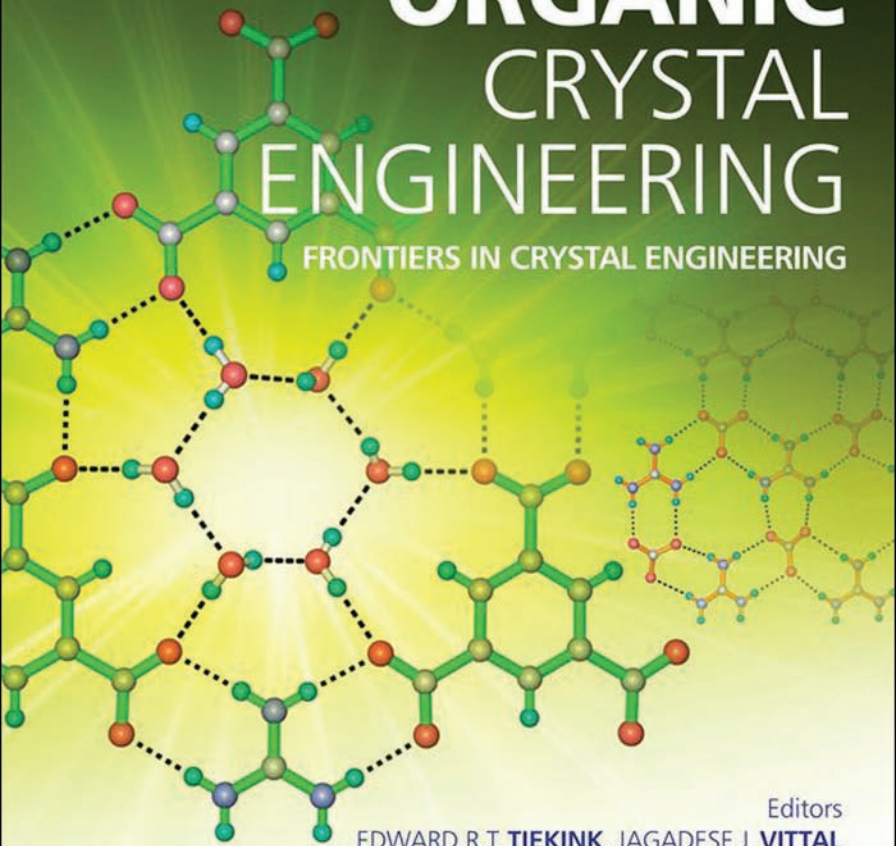


# ORGANIC CRYSTAL ENGINEERING

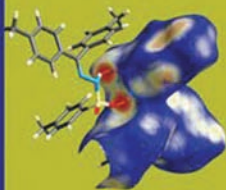
FRONTIERS IN CRYSTAL ENGINEERING



Editors

EDWARD R.T. **TIEKINK**, JAGADESE J. **VITTAL**  
and MICHAEL J. **ZAWOROTKO**

 **WILEY**





# **Organic Crystal Engineering**





# Organic Crystal Engineering

Frontiers in Crystal Engineering

Editors

EDWARD R.T. TIEKINK

*School of Materials Science and Engineering, Nanyang  
Technological University, Singapore*

JAGADESE VITTAL

*Department of Chemistry, National University of Singapore,  
Singapore*

MICHAEL ZAWOROTKO

*Department of Chemistry, University of South Florida, USA*



A John Wiley and Sons, Ltd., Publication

This edition first published 2010  
© 2010 John Wiley & Sons, Ltd.

*Registered office*

John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex PO19 8SQ, United Kingdom

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

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

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

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

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

The publisher and the author make no representations or warranties with respect to the accuracy or completeness of the contents of this work and specifically disclaim all warranties, including without limitation any implied warranties of fitness for a particular purpose. This work is sold with the understanding that the publisher is not engaged in rendering professional services. The advice and strategies contained herein may not be suitable for every situation. In view of ongoing research, equipment modifications, changes in governmental regulations, and the constant flow of information relating to the use of experimental reagents, equipment, and devices, the reader is urged to review and evaluate the information provided in the package insert or instructions for each chemical, piece of equipment, reagent, or device for, among other things, any changes in the instructions or indication of usage and for added warnings and precautions. The fact that an organization or Website is referred to in this work as a citation and/or a potential source of further information does not mean that the author or the publisher endorses the information the organization or Website may provide or recommendations it may make. Further, readers should be aware that Internet Websites listed in this work may have changed or disappeared between when this work was written and when it is read. No warranty may be created or extended by any promotional statements for this work. Neither the publisher nor the author shall be liable for any damages arising herefrom.

***Library of Congress Cataloging-in-Publication Data***

Organic crystal engineering : frontiers in crystal engineering / editors, Edward Tiekink, Jagadeś Vittal, Michael Zaworotko.  
p. cm.

Includes bibliographical references and index.

ISBN 978-0-470-31990-1 (cloth)

1. Molecular crystals—Research. 2. Organic solid state chemistry. 3. Crystal growth. I. Tiekink, Edward R. T. II. Vittal, Jagadeś J. III. Zaworotko, Michael.

QD921.O65 2010

548—dc22

2009052144

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

ISBN 978-0-470-31990-1

Typeset in 10/12pt Times-Roman by Laserwords Private Limited, Chennai, India  
Printed and bound in Great Britain by CPI Antony Rowe, Chippenham, Wiltshire

## **Dedication**

**Fumio Toda**, a pioneer in organic crystal engineering



# Contents

<i>List of Contributors</i>	xiii
<i>Preface</i>	xv
<b>1. The Role of the Cambridge Structural Database in Crystal Engineering</b>	<b>1</b>
<i>Andrew D. Bond</i>	
1.1 Introduction	1
1.2 Organisation and Management of Crystallographic Information	2
1.2.1 Validation of New Crystal Structures: Mogul	3
1.2.2 Validation of Structural Discoveries: What is New and What is Not?	5
1.3 Organisation of Crystallographic Information for Crystal Engineering	6
1.3.1 IsoStar	6
1.3.2 CSDDContact	8
1.3.3 CSDSymmetry	10
1.4 New Tools for Database Research	13
1.4.1 A General Tool for 3-D Searches of the CSD: 3DSEARCH	13
1.4.2 Ad hoc Software	17
1.5 Search for Functional Group Exchanges: <i>GRX</i>	18
1.6 Search for Solvated and Unsolvated Structures: <i>Solvates</i>	20
1.6.1 Tools to Examine Structural Similarity	22
1.7 Clustering and Classifying CSD Search Results: <i>dSNAP</i>	23
1.8 The PXRD Profile as a Structural Descriptor	25
1.8.1 IsoQuest	26
1.8.2 Self-organising maps	28
1.8.3 Discrimination of polymorphs and redeterminations	30
1.9 Identifying Supramolecular Constructs: <i>XPac</i>	32
1.10 Concluding Remarks: The Future Role of Crystallographic Databases	36
References	38
<b>2. Computational Crystal Structure Prediction: Towards <i>In Silico</i> Solid Form Screening</b>	<b>43</b>
<i>Graeme M. Day</i>	
2.1 Introduction	43
2.2 Methods used to Predict Crystal Structures	45
2.2.1 Search Methods	46

2.2.2	Evaluating the Computer-generated Crystal Structures	47
2.3	Current Capabilities of Crystal Structure Prediction	48
2.3.1	The Blind Tests	48
2.3.2	Further Assessments of Crystal Structure Prediction	52
2.3.3	Flexible Molecules	57
2.4	Exploration of Crystal Forms. A Case Study: Carbamazepine	58
2.4.1	Polymorphism and the Influence of Small Molecular Changes on Packing	59
2.4.2	Solvate and Co-crystal Formation	61
2.4.3	Computational Solid Form Screening?	63
2.5	Summary	64
	Acknowledgements	64
	References	64
<b>3.</b>	<b>Multi-component Pharmaceutical Crystalline Phases: Engineering for Performance</b>	<b>67</b>
	<i>Matthew L. Peterson, Edwin A. Collier, Magali B. Hickey, Hector Guzman and Örn Almarsson</i>	
3.1	Introduction	67
3.2	Exploring Crystal Form Diversity	68
3.3	High-throughput Experimentation	72
3.4	Examples of 'Form and Formulation'	75
3.4.1	Indinavir sulfate ethanolate – a case of a salt hydrate/solvate	75
3.4.2	Norfloxacin – polymorphs, solvates, salts, complexes and co-crystals	77
3.5	AMG517 and Celecoxib – 'Spring and Parachute' Approach	78
3.5.1	$\beta$ -Lactam antibiotics and hydrates – the importance of crystallinity	82
3.6	Carbamazepine – Stabilization Against a Hydrate	83
3.7	Theophylline:Phenobarbital – Two is Better Than One	87
3.8	Delaviridine Mesylate – Material Misbehaviour	90
3.9	Summary and Outlook	91
	References	93
<b>4.</b>	<b>Complex Formation of Surfactants with Aromatic Compounds and their Pharmaceutical Applications</b>	<b>101</b>
	<i>Yuji Ohashi, Keiju Sawada and Nahoko Imura</i>	
4.1	Introduction	101
4.2	Structures of the Complexes Formed Between Surfactants and Aromatic Compounds	103
4.2.1	Crystal structure of I (CTAB- <i>p</i> -hydroxybenzoic acid)	103
4.2.2	Crystal structure of II (CTAB- <i>m</i> -cyanophenol)	105
4.2.3	Crystal structure of III (CTAB- <i>p</i> -cresol)	105
4.2.4	Crystal structure of IV (CTAB-hydroquinone)	106
4.2.5	Crystal structure of V (CTAB- <i>o</i> -iodophenol)	107

4.2.6	Crystal structure of VI (MTAB- <i>o</i> -iodophenol)	107
4.2.7	Crystal structures of VII (LTAB- <i>o</i> -iodophenol) and VIII (DTAB- <i>o</i> -iodophenol)	107
4.2.8	Crystal structure of IX (DTAB- <i>p</i> -hydroxybenzoic acid)	108
4.3	Complex Formation of Aromatic Compounds Containing an Hetero Ring	109
4.3.1	Crystal structures of X (CTAB-acridine), XI (CTAB-phenoxazine) and XII (CTAB-phenothiazine)	109
4.3.2	Crystal structures of XIII (CTAB-dibenzofuran) and XIV (CTAB-carbazole)	109
4.4	Complex Formation of Biphenyl with Cationic Surfactants	110
4.5	Complex Formation of Odd-Number Surfactants with Biphenyl	112
4.6	Common Packing Mode in the Complexes	114
4.7	Complex Formation by Grinding in a Mortar	117
4.8	C-H... $\pi$ interactions	122
4.9	Complex Formation of Anionic Surfactants with Aromatic Compounds	125
4.10	Increased Solubility of Insoluble Drugs	127
4.11	Enhanced Thermal Stability of Perfumes	130
4.12	Complex Formation with Surfactants other than Quaternary Alkylammonium Salts	134
4.13	Hydroquinone Complexes	136
4.14	Application to Whitening Agents	142
	Acknowledgements	147
	References	147
<b>5.</b>	<b>Hydrogen Bonding and Molecular Packing in Multi-functional Crystal Structures</b>	<b>151</b>
	<i>Ashwini Nangia</i>	
5.1	Introduction	151
5.2	Hydrogen Bonding in Ureas and Amides	154
5.3	Pyridyl Ureas and Amides	156
5.4	Nitrophenyl Ureas and Amides	160
5.5	Molecular Conformation and Hydrogen Bonding	166
5.6	Supramolecular HSAB Interactions	182
5.7	<i>gem</i> -Alkynols	184
5.8	Conclusions	186
	Acknowledgements	187
	References	187
<b>6.</b>	<b>Persistence of N-H...S Hydrogen Bonding in Thiocarbamide Structures</b>	<b>191</b>
	<i>Edward R. T. Tiekink</i>	
6.1	Introduction	191
6.2	Supramolecular Aggregation Patterns in the Thiocarbamides	194
6.2.1	Cyclic thiocarbamides – five-membered rings	194
6.2.2	Cyclic thiocarbamides – six-membered rings	196

6.2.3	Acyclic thiocarbamides – E configuration	196
6.2.4	Acyclic thiocarbamides – Z configuration	201
6.2.5	Co-crystals containing thiocarbamides	206
6.3	Conclusions	206
	References	211
<b>7.</b>	<b>Crystal Engineering with Molecules Containing Amide and Pyridine Functionalities</b>	<b>215</b>
	<i>Kumar Biradha and Lalit Rajput</i>	
7.1	Introduction	215
7.2	Primary Amides Containing a Pyridine Moiety	217
7.2.1	Amides directly linked to a pyridine moiety	218
7.2.2	Primary amide and pyridine linked by spacer units	220
7.3	Co-crystals with Primary Amidopyridines	220
7.3.1	Combination of amides for co-crystallization	223
7.4	Secondary Amides Containing a Pyridine Moiety	224
7.4.1	Monoamide and monopyridine derivatives	224
7.5	Bis-Amidopyridine Derivatives	225
7.5.1	C-Terminal bis-amidopyridines	227
7.5.2	N-Terminal bis-amidopyridines (reverse amides)	229
7.6	Two-component Structures Containing Secondary Amides and Pyridine Derivatives	231
7.7	Triamidopyridine Derivatives	232
7.7.1	2-Pyridyl based systems	232
7.7.2	3-Pyridyl and 4-pyridyl based systems	233
7.8	Conclusions	234
	Acknowledgements	236
	References	236
<b>8.</b>	<b>Urea/Thiourea-Anion Host Lattices, Stabilization of Labile Species, and Designed Construction of Rosette Ribbon and Layers</b>	<b>239</b>
	<i>Thomas C. W. Mak, Chi-Keung Lam, Jie Han, Qi Li and Feng Xue</i>	
8.1	Introduction	239
8.2	Inclusion Compounds with Urea/Thiourea-Anion Host Lattices	241
8.2.1	Isostructurality Based on Interchangeability of Supramolecular Synthons	241
8.2.2	Carbonate and Oxalate as Prolific Hydrogen-Bond Acceptors	243
8.2.3	Hydrogen-Bonded Networks Constructed with 5-Nitrobarbiturate	247
8.2.4	Generation and Trapping of Labile Anions	257
8.3	Stabilization of Cyclic Oxocarbon Dianions by Hydrogen Bonding with Urea/Thiourea	265
8.3.1	Cyclic Oxocarbon Acids	265
8.3.2	Trapping Cyclic Oxocarbon Dianions in the Crystalline State	265



8.3.3	Charge-localized and -delocalized Valence Tautomeric Forms of Cyclic Oxocarbon Dianions	278
8.4	Supramolecular Assembly Based on the Rosette Motif	289
8.4.1	Anionic Rosette Ribbon Built of Guanidinium and Hydrogen Carbonate Dimer	289
8.4.2	Anionic Rosette Networks Assembled with Guanidinium and $C_3$ -symmetric Oxo-anion Building Blocks	293
8.4.3	Anionic Rosette Layers Formed by Guanidinium and Non- $C_3$ Molecular Building Blocks	301
8.5	Conclusion and Outlook	306
	Acknowledgements	307
	References	307

<b><i>Index</i></b>	<b>313</b>
---------------------	------------



# List of Contributors

**Örn Almarsson**, Alkermes Inc., 88 Sidney Street, Cambridge, MA 02139, USA

**Kumar Biradha**, Department of Chemistry, Indian Institute of Technology, Kharagpur-721302, India

**Andrew D. Bond**, University of Southern Denmark, Department of Physics and Chemistry, Campusvej 55, 5230 Odense M, Denmark

**Edwin A. Collier**, TransForm Pharmaceutical Inc., a Johnson and Johnson Company, 29 Hartwell Avenue, Lexington, MA 02421, USA

**Graeme M. Day**, Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, UK

**Hector Guzman**, Alkermes Inc., 88 Sidney Street, Cambridge, MA 02139, USA

**Jie Han**, Department of Chemistry and Center of Novel Functional Molecules, The Chinese University of Hong Kong, Shatin, New Territories, Hong Kong SAR, PR China

**Magali B. Hickey**, TransForm Pharmaceutical Inc., a Johnson and Johnson Company, 29 Hartwell Avenue, Lexington, MA 02421, USA

**Nahoko Imura**, Department of Pharmaceutics, Niigata University of Pharmacy and Applied Life Sciences, 265-1, Higashijima, Niigata 956-8603, Japan

**Chi-Keung Lam**, Department of Chemistry and Center of Novel Functional Molecules, The Chinese University of Hong Kong, Shatin, New Territories, Hong Kong SAR, PR China

**Qi Li**, Department of Chemistry and Center of Novel Functional Molecules, The Chinese University of Hong Kong, Shatin, New Territories, Hong Kong SAR, PR China

**Thomas C. W. Mak**, Department of Chemistry and Center of Novel Functional Molecules, The Chinese University of Hong Kong, Shatin, New Territories, Hong Kong SAR, PR China

**Ashwini Nangia**, School of Chemistry, University of Hyderabad, Hyderabad 500046, India

**Yuji Ohashi**, Ibaraki Quantum Beam Research Center, 162-1, Shirakata, Tokai, Ibaraki 319-1106, Japan

**Matthew L. Peterson**, Amgen Inc., One Kendal Square, Building 1000, Cambridge, MA 02139, USA

**Lalit Rajput**, Department of Chemistry, Indian Institute of Technology, Kharagpur-721302, India

**Keiju Sawada**, Toba National College of Maritime Technology, 1-1, Ikegami, Toba, Mie 517-8501, Japan

**Edward R. T. Tiekink**, School of Materials Science and Engineering, Nanyang Technological University, Nanyang Avenue, Singapore 639798

**Feng Xue**, Department of Chemistry and Center of Novel Functional Molecules, The Chinese University of Hong Kong, Shatin, New Territories, Hong Kong SAR, PR China

# Preface

The early promises of crystal engineering discussed in the first volume of this series, *Frontiers in Crystal Engineering*, are starting to be realised in the diverse areas ranging from co-crystal formation of Active Pharmaceutical Agents (APIs) to low pressure gas storage in Molecular Organic Framework structures (MOFs). In this volume, a snapshot of on-going research in ‘organic’ crystal engineering endeavours is presented. The utility of the Cambridge Structural Database (CSD), with now approaching half a million structures, as a primary resource for data mining studies is established early in the volume. Specialised methods, with worked examples, are outlined for the systematic analysis of the crystallographic literature, allowing comprehensive analysis and providing insight for rational crystal structure design. Literature knowledge is also pivotal for the emerging area of crystal structure prediction, *in silico* crystal engineering. A comprehensive review of the state of the art of computational crystal engineering is given, including the merits of various algorithms, pitfalls and successes. Reflecting the importance of polymorphism to the pharmaceutical industry, key examples are drawn from that genre. The formation of pharmaceutical co-crystals is thoroughly explored in a detailed expose of the rationale, principles and future prospects of co-crystal formation between APIs and pharmaceutically acceptable co-crystal conformers. This theme is elaborated upon in another contribution where surfactants are utilised as co-crystal conformers with drugs, as before, but also with perfumes, cosmetics, etc. The resulting co-crystal serves as a reservoir for the controlled/slow release of a therapeutic agent or perfume or cosmetic. The tailoring of organic crystal structure to meet the demands of specific applications is predicated upon the understanding of the principles of supramolecular aggregation patterns. Again, consistent with the demands of the pharmaceutical industry, where many molecules contain hydrogen bonding functionality, three chapters focus on providing an in depth profile of hydrogen bonding proclivities of various functional groups. Of particular interest are competitive studies between commonly occurring functional groups such as amide versus pyridine, etc. Non-traditional hydrogen bonding interactions, e.g. N-H...S, are shown to be important in structures related to pharmaceuticals. Intriguingly, iodo...nitro and iodo...ethynyl synthon interactions can compensate for more classical modes of association between molecules. The final chapter of this volume also concerns hydrogen bonding interactions but from a notably different and illuminative perspective. Using the principles of hierarchies of

hydrogen bonding, otherwise unstable organic anions are captured (stabilised) within a crystalline host environment allowing full structural characterisation that has previously remained elusive.

The full range of organic crystal engineering endeavours are encompassed herein, from theoretical, data mining, practical and applied perspectives – these are sure to provide a foundation for future studies in this ever emerging area of materials science.

# 1

## The Role of the Cambridge Structural Database in Crystal Engineering

*Andrew D. Bond*

### 1.1 Introduction

Although crystal engineering encompasses many aspects beyond structure determination, it is intimately linked to the acquisition and dissemination of crystallographic data. This Chapter considers the importance of effective organisation and management of crystallographic information and the extent to which a large number of crystal structures can be utilised as a tool for crystal engineering. At the present time, the collated body of crystallographic information, when referred specifically to the crystal structures of organic/metallo-organic molecular compounds, is contained within the Cambridge Structural Database (CSD) [1]. There are other crystallographic databases [2–5], but the CSD is of principal importance for crystal engineering, and it is for this reason that the CSD is the specific focus of this Chapter. At the time of writing, there are approximately 475 000 entries in the CSD [6], comprising effectively the complete body of published molecular crystal structures, gathered from almost a century of the single-crystal X-ray diffraction technique. There have been numerous reviews in the recent literature that have highlighted research applications of the CSD [7–9], and the web pages of the Cambridge Crystallographic Data Centre (CCDC) provide a comprehensive list of publications in which the CSD has been used [10]. This Chapter is not intended to serve as a review of that list since it is not possible to discuss specifically the many interesting examples where the CSD has been explored to address issues of crystal engineering. Rather, the Chapter attempts to pinpoint the specific value of the CSD for crystal engineering, to

identify contemporary trends in its usage within the field, and to extrapolate these trends to provide a glimpse of potential future developments.

As a conceptual framework for the Chapter, a distinction is made at the outset between effective organisation and management of crystallographic information and utilisation of that information to generate knowledge within the field of crystal engineering. The former amounts largely to a description of the CSD system, which is available in detail from other sources [1]. This Chapter will touch briefly upon some more recent aspects of organisation and management, mentioning specifically the role of the CSD for validating new crystal structures and new structural discoveries – that is, the role of the CSD in ensuring the integrity of new information that will be used to derive knowledge in the crystal engineering literature. The focus will then move to the manner in which suitably organised crystallographic information can find application within crystal engineering strategies: how can the information content of the CSD be harnessed to answer questions in crystal engineering? If there is a single obvious contemporary trend in this area, it is development of new tools for exploring crystallographic information through the development of derivative databases and new computer software. It is relevant to note that in a 2002 review of database research within crystal engineering, Nangia concluded that ‘certain questions that are somewhat difficult to address at the present time will be tackled with greater confidence as the database grows and the methods to mine the information become more refined’ [9]. This sentiment has been reiterated more recently with explicit reference to software development [11]: ‘as the scientific problems posed to the CSD become more sophisticated, there is a need for more extensive software facilities for database searching and for analysing search results’. A large part of this Chapter therefore focuses on developments in this area. It seems likely that the contents of this Chapter represent a snapshot of a research field that is on the brink of explosion – we might reasonably refer to it as ‘*in silico* crystal engineering’.

## 1.2 Organisation and Management of Crystallographic Information

In the ‘electronic era’, it is hardly necessary to highlight the importance of communicating knowledge in an electronic form. It is often said anecdotally that crystallographers have been amongst the primary innovators in electronic communication within science, since crystallographic data are so well suited to electronic representation. At the present time, the standard electronic means for communicating crystallographic data is the CIF [12]. Although the CCDC (and other organisations such as the International Union of Crystallography) [13] maintain an archive of CIFs, this is not the primary means by which crystallographic knowledge is disseminated – none of us store or consult an archive of 475 000 CIFs on a day-to-day basis. Instead, the CSD is distributed in the CCDC’s own binary ASER format [1]. Within the distributed archive, each individual CSD entry contains (almost)\* all of the information present in a CIF, plus a significant number of ‘value-added data items’, including bibliographic information and other

---

\* A CSD entry does not contain information regarding atomic displacement parameters. The reason for this is historical: in the early days of the CSD, all data were entered by hand from printed material that had itself been generated by hand, providing several stages at which errors could be introduced. Errors in atomic coordinates are usually detectable by anomalous structural features. However, errors in displacement parameters are



comments, derived information such as reduced cell parameters and  $Z'$  values, and chemical connectivity information, usually accompanied by a standard chemical diagram. The connectivity information is of prime importance for the CSD system, since it provides the means for structure and substructure searching. The distributed CSD therefore represents a collection of crystallographic information that is augmented in such a way that it becomes more valuable than a collection of 'raw' CIFs. The added value is not simply the individual items that are not present in the original CIF. Rather, the true value of the CSD comes from its completeness and uniformity, together with the tools that are provided to examine it.

The large amount of crystallographic information gathered in the CSD provides an overview of the manner in which chemical compounds arrange themselves within crystals – the very core of crystal engineering. If a new crystal structure is to be entered into this collection, it is helpful to make some early assessment of its likely validity by comparing it to all of the structures that have been established previously. Ideally, such an assessment should be applied at both molecular and supramolecular levels. Tools for validation at the molecular level are now well developed, and we will consider some recent developments in this area that will enhance the way in which the CSD ensures the integrity of the information that it contains. Tools for validation at the supramolecular level are rather more difficult to apply, principally because supramolecular geometry within crystals is vastly more variable than molecular geometry. We will consider some consequences for the crystal engineering literature that can be ascribed directly to this.

### **1.2.1 Validation of New Crystal Structures: Mogul**

One part of the process of validating a molecular crystal structure amounts to asking the question 'are there any features of the derived molecular geometry that appear to be unusual compared with all other molecular geometries that have been established so far?' For this purpose, the molecular geometry can be broken down into bond lengths, bond angles and bond torsion angles, and each one can be compared with all other known instances of that particular parameter in a similar chemical environment. To obtain a suitable set of structures for meaningful comparison will probably require a series of CSD searches, each one examining the parameter of interest (for example, a C–C bond length) in a particular chemical environment that must be specified as the search query. The length of time spent on this exercise will depend on how much effort is made to establish a similar chemical environment (if the molecule is new, we will not find one exactly the same), how many searches are required for that purpose, and the duration of each search (which will increase as the number of entries in the CSD increases). This procedure rapidly becomes tedious if it must be repeated several times for different bonds within a molecule. One approach that has been taken recently to address this

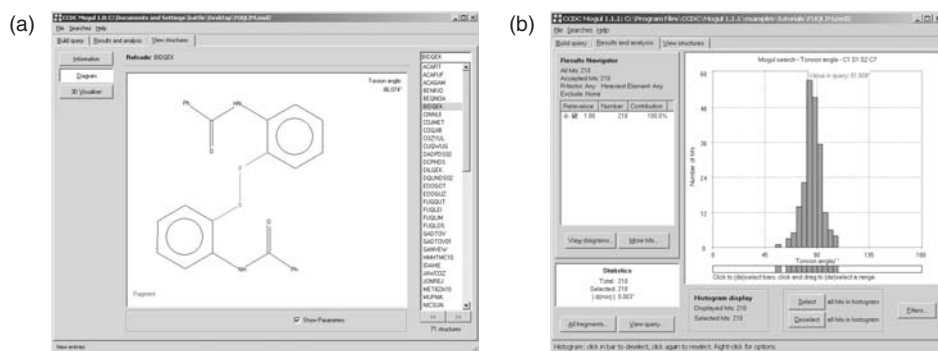
---

difficult to detect. For this reason, displacement parameters were not originally included in the distributed CSD, and have not been added since. In addition, for disordered structures the positions of all disordered atoms are usually retained, although their site occupancies and any information regarding which pairs of atoms are related by the disorder is lost. For more recent CSD entries with a corresponding entry in the CIF archive, it is possible to retrieve this information from the CIF. For older entries, it is necessary to refer to the original literature report.

type of problem is the preparation of ‘knowledge-based libraries’, which distil particular types of information only once from the entire CSD then permit subsequent searches to be made against the derived knowledge.

The CCDC’s *Mogul* package [14] provides one such example of a knowledge-based library, comprising details of intramolecular geometry classified according to chemical environment. The *Mogul* library contains over 20 million bond lengths, angles and (acyclic) torsion angles, derived from all entries in the CSD. Complete molecules or substructures can be drawn or imported. Bond distances, angles or torsion angles within those structures can then be specified, and distributions can be obtained for all comparable parameters in the CSD. Rapid searching is achieved by use of a tree structure that is indexed on the basis of chemical keys. For example, keys that refer to a single atom might be atomic number, number of connected non-hydrogen atoms, number of connected hydrogen atoms, and so on. Keys that refer to fragments might specify bond types between all atoms or denote whether the fragment is cyclic or acyclic. The complete key scheme, comprising the definition of each key and the information about how it should be applied, provides the means for classifying and retrieving fragments. The process of generating the *Mogul* library amounts to taking every example of a particular 2-atom, 3-atom and 4-atom combination in the CSD, applying the key scheme to classify each, then grouping those fragments with identical key values into distinct populations from which the required statistics (principally mean value and standard deviation for the parameter of interest) are derived. Where there are a vast number of comparable entries in the CSD, the *Mogul* library is restricted to 10 000 values selected at random from the complete population. A *Mogul* search amounts to evaluating the keys for a particular query fragment then matching the result to the *Mogul* distribution with the same set of keys. A typical query (input via the graphical interface) and resulting *Mogul* output are illustrated in Figure 1.1.

One especially important feature of *Mogul* is the possibility to generate distributions for *all* bond lengths, angles and torsions within a query molecule almost instantly in a single run. This permits convenient assessment of the validity of all molecular parameters



**Figure 1.1** Input and output for the knowledge-based library Mogul: (a) a diaryldithiole fragment is input via the graphical interface; and (b) the C–S–S–C torsion angle in the query fragment is compared instantly with all similar S–S bonds in the CSD. Reproduced by permission of The Cambridge Crystallographic Data Centre

during refinement of a new crystal structure. This feature has been integrated, for example, into the *CRYSTALS* [15] and *DASH* [16] programs for determination and refinement of crystal structures against single-crystal or powder diffraction data. Thus, *Mogul* provides an immediate indication of any molecular features that are anomalous by comparison with all existing crystallographic knowledge, encouraging more careful examination of these features before they are committed to the literature. In this way, *Mogul* can make a significant contribution to ensure the integrity of new crystallographic information, at least at the molecular level. This is of great importance, since the conclusions of database studies, and any knowledge that might be derived from them, is crucially dependent on the integrity of the crystallographic information from which it is obtained.

### 1.2.2 Validation of Structural Discoveries: What is New and What is Not?

As the number of known crystal structures continues to grow at a dramatic rate, one function of the primary crystal engineering literature must be to distil and unify this vast amount of information. To maintain the integrity of the research field, it is important that literature reports should be based on accurate and complete assessments of the state of crystallographic knowledge. The sub-heading of this section is taken (almost) literally from a provocative article that appeared in the literature in 2006 [17], in which Mascal, Infantes and Chisholm noted that a substantial proportion of literature papers that claim to describe extraordinary or even unprecedented assemblies of water molecules in crystals actually report motifs that are already present in the CSD. One key question posed by Mascal *et al.* is: ‘how can the novelty of any experimentally observed water oligomer be substantiated without consulting the CSD?’ To generalise the point, ‘water oligomer’ might be replaced by any structural feature. It is clear that there can be only one answer to this question: a structural feature cannot be described as novel unless it is first established that it is not already present amongst the existing body of crystallographic information. In many ways, this single statement embodies the value of the CSD for ensuring the integrity of the crystal engineering literature. It is not possible to operate as a crystal engineer without being aware of existing crystallographic information, and it is clear that it is not possible to be aware of all crystallographic information unless it is collated and organised in an efficient way. The article of Mascal *et al.* considered specifically 17 recent literature reports chosen at random from a selection that highlighted interesting assemblies of water molecules, and showed that at least 14 of these articles reported motifs that were already present in the CSD. In 11 cases, the articles included specific claims of novelty that were shown to be false. It was noted that ‘the problem is not so much one of intentional publication of misinformation as it is the relative difficulty in establishing what is already known’. Thus, these misleading reports in the primary crystal engineering literature may be viewed as a direct manifestation of a lack of suitable tools to validate and explore crystallographic information at a supramolecular level.

In that context, we might return for a moment to the role of the *Mogul* library, which provides an efficient and automated means to identify anomalous structural features at a molecular level. Following one of the basic themes of crystal engineering, namely the extension of molecular concepts to a supramolecular level, we might ask if the *Mogul* methodology could be extended to explore *supramolecular* novelty and validity. Since

the variety of supramolecular environments is very much greater than molecular chemical environments, it is very much more difficult to imagine a robust set of keys that could be defined at the supramolecular level. However, this seems to be a challenge that must be taken up in one way or another if we are to ensure the accuracy and integrity of the primary crystal engineering literature.

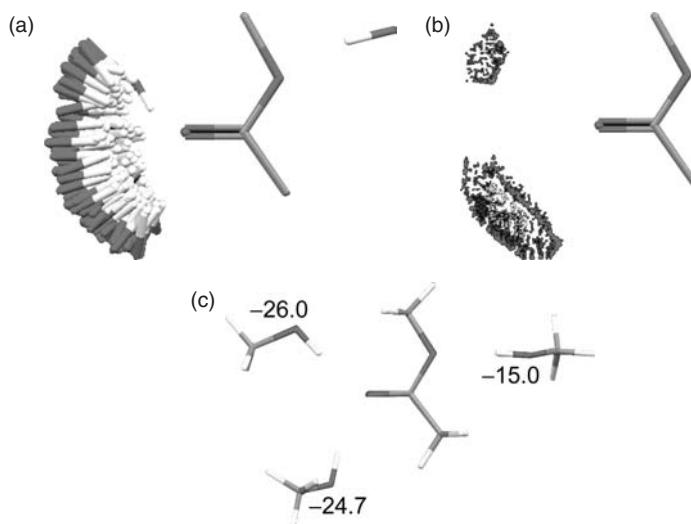
### 1.3 Organisation of Crystallographic Information for Crystal Engineering

Utilisation of crystallographic information for crystal engineering often focuses on features of crystal structures that are not immediately accessible in each individual CSD entry. For example, they might focus on some aspect of molecular symmetry or on some supramolecular feature. Anyone that has ever attempted to search the CSD for supramolecular features using the *ConQuest* software [18] will be aware that searches based on 3-D information are usually vastly slower than those based on 2-D chemical substructures, often to the extent that the 3-D searches become impractical. To some degree, this is a consequence of the construction of the CSD system, but it also reflects the very different nature of molecular and supramolecular features of crystal structures: chemical connectivity of molecules is usually unambiguous and can therefore be coded specifically, while supramolecular geometries and ‘connectivities’ are hugely variable. Thus, 2-D searches of the CSD can be accelerated by well-defined coded shortcuts that cannot readily be applied at the supramolecular level. In this section, we will examine several approaches to retrieve more complex information from the CSD. These include knowledge-based libraries conceptually similar to *Mogul*, and derived databases that allow direct interrogation of information that is not immediately present in each CSD entry.

#### 1.3.1 IsoStar

A knowledge-based library approach for intermolecular interactions has been implemented in the CCDC’s *IsoStar* [19] package by specifying particular chemical groups and examining intermolecular contacts between pairs of these groups. For two chemical groups, A and B, all instances of  $A \cdots B$  contacts in the CSD are identified within specified criteria, and the crystallographic coordinates of each instance are transformed so that the A groups are overlaid by a least-squares approach (usually averaged into a single representation for convenient visualisation). The result is a 3-D scatter plot that illustrates the experimental distribution of the B group with respect to the A group. For example, Figure 1.2 shows such a scatter plot for intermolecular contacts between OH groups and an ester group. In this instance, the scatter plot makes clear that OH groups often form hydrogen bonds to the C=O group of the ester, but seldom to the other O atom. The observed preference is consistent with the optimised intermolecular gas-phase interaction energies (also included in Figure 1.2), which confirm that the  $OH \cdots O=C$  contacts are significantly more stabilising than the alternative  $OH \cdots O$  contacts.

The *IsoStar* library contains over 25 000 pre-generated scatter plots, derived principally from the CSD, but also with contributions from the Protein Data Bank (PDB) [2]. The scatter plots are hyperlinked to both databases so that specific interactions



**Figure 1.2** Scatter plot taken from the knowledge-based library *IsoStar*, showing the distribution of intermolecular contacts between OH groups and an ester group: (a) displayed as overlaid contacts; (b) displayed as a contoured 'interaction density' surface; and (c) showing the optimised gas-phase geometry and interaction energies ( $\text{kJ mol}^{-1}$ ). Reproduced by permission of The Cambridge Crystallographic Data Centre

can be viewed in their native structural environments. Also included are 1500 potential energy minima calculated using distributed multipole analysis and intermolecular perturbation theory [20], permitting an assessment of how far a given supramolecular geometry deviates from its expected energy minimum. The rationale for including three different sources of data is that they are mutually complementary: (1) the crystal structures of small molecules are usually highly precise, but the condensed nature of the crystal lattice can often perturb the 'natural' preferences of intermolecular interactions; (2) a slightly different view might be obtained from crystal structures of protein–ligand complexes, where different constraints might be imposed on the intermolecular interactions. Although the experimental precision of the PDB structures is usually much lower than that of CSD entries, these structures are perhaps more directly relevant to structure-based drug design; (3) finally, theoretical studies refer to the optimised gas-phase situation, where intermolecular interactions do not persist.

The *IsoStar* library is of clear relevance for crystal engineering since it addresses the nature and frequency of observation of supramolecular geometries at the local level. For example, it is well suited to answer fundamental questions regarding the directionality of specific interatomic or inter-group interactions. In this respect, *IsoStar* is able to encompass and present graphically much of the pioneering database work that first established the directional nature of various intermolecular interactions. Some caution must be applied when drawing general conclusions from *IsoStar* since the scatter plots give no indication of the underlying causes of any particular geometrical preference. For example, a clustering of  $\text{H} \cdots \text{O}=\text{C}$  contacts along the directions expected for  $\text{sp}^2$  orbitals on oxygen might reflect some inherent preference associated with orbital interactions, or

it might reflect the fact that most O=C groups are accepting simultaneously two hydrogen bonds and that the directional preferences represent a simple spatial compromise of multiple interactions about the oxygen atom. The theoretical energy minima present in the *IsoStar* library should help to resolve this type of question since they can provide independent evidence that the features seen in the scatter plot reflect inherent properties of the interacting systems. For crystal engineering purposes, the *IsoStar* library may be viewed as a collection of potential supramolecular synthons, with pre-distilled information regarding the robustness of their geometry and the frequency of their formation within crystals.

### 1.3.2 CSDContact

Although *IsoStar* provides convenient facilities for *visualising* intermolecular interactions, it is not immediately suitable for quantitative analysis of sets of interactions, for example to establish the average number of hydrogen bonds formed by a given functional group in all observed intermolecular environments. One approach that has been taken to address this type of question is the creation of the ‘derived database’ *CSDContact* [21]. The derived database approach is subtly different from the knowledge-based libraries *Mogul* and *IsoStar*, which effectively envisage all questions of a particular type that might be posed to the CSD, then pre-prepare answers for rapid retrieval. A derived database consists simply of all CSD entries converted into a different format, which must subsequently be searched in order to obtain specific knowledge. The advantage is that the derived information must be calculated only once, producing a database that forms a more convenient basis for subsequent searches. Several applications of this concept will be discussed throughout this Chapter. The *CSDContact* database includes information concerning all instances of intermolecular contacts between OH<sub>*x*</sub> (*x* = 1–3), NH<sub>*x*</sub> (*x* = 1–4) or SH groups in all relevant structures within the CSD. Contacts are defined on the basis of distances (less than the sum of the van der Waals radii plus 0.1 Å) and each atom involved in each contact is classified as a hydrogen-bond donor, hydrogen-bond acceptor or neither. *CSDContact* also records chemical environment codes for each atom (e.g. O in OH, O in C=O) and a measure of their accessible surface [22]. The information is collated into a relational database using the Microsoft *Access* software [23], which permits flexible interrogation to find occurrences of any particular combination of data values. For example, it can be established (and quantified) immediately that some groups, such as the oxygen atom of the C=O group in a carboxylic acid, show a very high number of contacts in which oxygen is classified as a hydrogen-bond acceptor, while others such as the nitrogen atom of a planar NH<sub>2</sub> group almost never accept hydrogen bonds.

An example where the *CSDContact* methodology has been used to derive knowledge of specific relevance for crystal engineering is the study of competition for specific hydrogen-bond donors and acceptors [21]. For example, in a set of 82 structures containing molecules with one alcoholic OH group and two ketone C=O groups, O–H···O hydrogen bonds were found to occur in 80 cases (98%), of which 13 were intramolecular. Of the intermolecular hydrogen bonds, 50 were formed to the C=O group with the greater accessible surface, while 17 were formed to the C=O group with the smaller accessible surface. Thus, the probability of observing intermolecular O–H···O hydrogen



**Table 1.1** Results of a query from *CSDContact*, showing the frequency of disruption of self-association of acid and amide groups (column A) in the presence of other functional groups (column B). Columns A...A, B...B and A...B refer to the percentages of group-group associations of each possible type, and N is the number of structures in the CSD contributing to each category. Reproduced from [11] by permission of The Royal Society of Chemistry

A	B	A...A (%)	B...B (%)	A...B (%)	N
COOH	C(sp <sup>3</sup> )-OH	12	0	88	49
COOH	CONH	14	28	58	40
COOH	C=O (keto)	62	0	38	86
COOH	C-O-C	92	1	8	26
CONH	C(sp <sup>3</sup> )-OH	35	5	60	30
CONH	COOH	28	14	58	40
CONH	C=O (keto)	89	0	11	26
CONH	C-O-C	97	0	3	35

bonding is established to be 82%, with 61% in favour of the C=O group with the higher accessible surface. This type of analysis can be extended to consider self-association or disruption of intermolecular contacts between particular functional groups [24]. For example, Table 1.1 shows the propensity for self-association of COOH and CONH functional groups in the presence of other functional groups within crystal structures. The results show that a C(sp<sup>3</sup>)-OH group within a molecule has a high probability of disrupting self-association of COOH or CONH groups in the crystal structure of that molecule. Similarly, when both COOH and CONH are present in the same molecule, they are most likely (58%) to be associated with each other. The presence of a C=O (keto) group in a molecule is much more likely to disrupt self-association of COOH groups than it is to disrupt self-association of CONH groups, while the presence of a C-O-C group has very little influence on the propensity for self-association. This type of information is clearly of relevance for crystal engineering strategies since it quantifies the frequency of observation of particular supramolecular synthons, taking into account also the identity of other functional groups that might be present within a molecule of interest.

For these types of studies (which are only representative examples of what might be achieved using the *CSDContact* database), we might consider the extent to which the derived database approach permits derivation of new knowledge. In particular, how do the results differ from those of earlier studies with similar aims? For example, Steiner reported in 2001 the relative ‘success rates’ for 34 types of groups to accept a hydrogen bond from the strong OH donor of a carboxyl group [25]. Amongst Steiner’s results was the observation that planar NH<sub>2</sub> groups have a success rate of zero, which is exactly one of the conclusions present in the initial report of *CSDContact* [21]. However, while Steiner’s study applies only to one specific hydrogen-bond donor group, *CSDContact* can provide immediate access to this type of information for over 100 common chemical groups. Similarly, the global probabilities of formation of 75 intermolecular hydrogen-bonded ring motifs were studied in 1999 by Allen *et al.* [26] It was shown in that study that self-association of carboxylic acid groups occurred with a probability of ca. 33%, and that self-association of acyclic amide groups and mixed acid–amide association both occurred with a probability of ca. 10%. Self-association of cyclic amides (lactams) was

more probable, amounting to ca. 34%. These statistics refer to the number of observations of the motif compared with the number that could have formed given the functional groups present. It is difficult to compare these statistics directly with Table 1.1, since they refer to the global population, while Table 1.1 refers to specific combinations of functional groups within a molecule. However, for the particular cases of the acids and amides, Allen *et al.* went on to subdivide their categories. It was shown that the probability of acid–acid association is ca. 95% for molecules with no competing hydrogen-bond donors or acceptors. This compares with 92% for the COOH/C–O–C combination listed in Table 1.1. Similarly, for primary amides it was shown that the probability of amide–amide association is 83% for monoamides and 95% for diamides, again comparable with the CONH/C–O–C combination in Table 1.1. Thus, it appears that when *comparable* statistics are considered, there is no significant change in the results between 1999 and 2007.

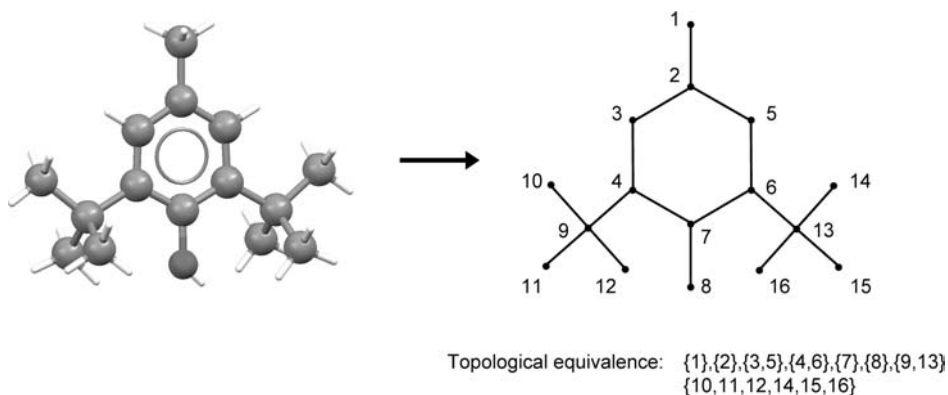
We will encounter similar situations several times in the remainder of this Chapter: when considering comparable statistics, the most recent automated database studies often corroborate conclusions that were obtained much earlier from more limited datasets. However, the benefit of automated analyses is that they permit much more rapid examination of much larger datasets, therefore providing (at least in principle) immediate access to a greater range of structural knowledge. In this particular instance, the information listed in Table 1.1 is considerably more accessible than the global results of Allen *et al.*, illustrating the capacity for *CSDContact* to provide answers to questions that are of immediate relevance for experimental crystal engineering.

### 1.3.3 CSDSymmetry

One of the principal targets that would be of exceptional value for crystal engineering is a robust correlation between the symmetry of molecules and the symmetry of their crystal structures. An obvious approach to develop such a relationship is to gather statistics from the CSD. However, this task is not immediately accessible since the molecular symmetry is not (currently) a part of each standard CSD entry. In cases where a molecule lies on a crystallographic special position and does not exhibit disorder, the molecular symmetry can be deduced directly from the point symmetry of the special position. However, in many cases molecules in crystal structures lie on general positions, with no obvious correlation between molecular symmetry and the positions that the molecules occupy [27]. The derived database *CSDSymmetry* [28] has been developed as a tool to address this issue, comprising a relational database between the perceived point group of each molecule in a CSD entry and crystallographic properties such as space group, occupied Wyckoff position and  $Z'$  values. *CSDSymmetry* permits easy identification of occurrences of any combination of molecular or crystallographic symmetry attributes.

The first requirement for development of *CSDSymmetry* is a robust and automated method for analysing exact and approximate molecular symmetry. The specific methodology that has been employed is a combination of checks for topological equivalence in the 2-D connectivity diagram followed by full 3-D symmetry perception [29]. The first part of the process is based on the fact that atoms related by symmetry within a molecule must have identical chemical environments – that is, symmetry-related atoms





**Figure 1.3** Topological equivalence of symmetry-related atoms used by *CSDSymmetry* for the first step of automated symmetry detection. A full description of the algorithm used to partition the groups is given in Cole et al. [29]

must be topologically equivalent in the 2-D connectivity diagram (Figure 1.3). Topological analysis of the 2-D diagram divides the molecule into groups of atoms that could possibly be related by some symmetry operation in three dimensions. This information is passed to a 3-D symmetry algorithm to test whether the molecule possesses actual symmetry within certain user-defined tolerances. Once all symmetry operations have been identified, their number and specific nature allows the molecular point group to be determined in the standard way. For crystal structures with multiple molecules or ions, symmetry detection is undertaken for each molecule in turn, and each is stored as a separate entry in the *CSDSymmetry* database. As for *CSDContact* [21], *CSDSymmetry* is constructed within a relational database using Microsoft Access [23].

Examples of the types of questions that can be posed to *CSDSymmetry* include the following [28]:

- (1) *How often do centrosymmetric molecules lie on crystallographic centres of inversion?*

This particular question provides another example where an earlier study from a small, hand-edited dataset can be compared with the more extensive, automated study permitted by *CSDSymmetry*. In 1994, Brock and Dunitz suggested a number of correlations between molecular symmetry and space-group symmetry, including the observation that centrosymmetric molecules often occupy crystallographic centres of inversion [30]. Analysis using *CSDSymmetry* (based on the April 2000 release of the CSD, using only structures with one molecular component) identified 18 008 molecules belonging to a point group containing the symmetry element  $\bar{1}$ . Of these, 15 156 (88.4%) lie on a special position of point symmetry  $\bar{1}$ , confirming emphatically the statement of Brock and Dunitz that ‘inversion centres are favourable’ [30].

- (2) *Do high-symmetry space groups prefer highly symmetric molecules?*

*CSDSymmetry* identifies 1420 molecular crystal structures (in which the molecule contains more than five atoms) within space groups that contain both special positions of point symmetry 2 and  $m$ . Of these, 1079 molecules (76.0%) have point symmetry higher than 1,  $\bar{1}$ , 2 or  $m$ . By contrast, there are 2299 structures identified in space group  $P1$ , of which only 167 molecules (7.2%) have point symmetry higher than

1,  $\bar{1}$ , 2 or  $m$ . Therefore, there is a clear preference for molecules with higher point symmetry to be accommodated in space groups with higher symmetry.

(3) Which molecular point groups result most often in the adoption of noncentrosymmetric structures?

For each molecular point group identified in the CSD, it is possible to obtain from *CSDSymmetry* the proportion of molecules that crystallise in noncentrosymmetric space groups. The ten molecular point groups that give rise to the highest proportion of noncentrosymmetric crystal structures are listed in Table 1.2. These results show a clear preference for molecules with point symmetry 3 or 4 to crystallise in noncentrosymmetric space groups. As suggested by the authors of the original literature article [28], ‘is this an indication that the synthesis of molecules with point symmetry 3 or 4 is a promising route to materials that may behave as non-linear optics?’

As is the case for *CSDContact*, some of the results obtained from *CSDSymmetry* might appear to some extent to be ‘intuitive’. For example, it is perhaps not surprising that molecules with internal symmetry elements are unlikely to crystallise in space group  $P1$  since the internal symmetry elements of the molecules can probably be arranged regularly in a crystal in some way that corresponds to one of the higher-symmetry space group types. However, as for *CSDContact*, the great utility of *CSDSymmetry* is that it allows for convenient examination and quantification of the relationship between molecular symmetry and crystallographic symmetry for *all* entries in the CSD. For example, an extensive study using *CSDSymmetry* [31] has established a concrete statistical hierarchy for retention of molecular point symmetry elements in crystal structures: for a molecule containing only one type of symmetry element, it is found that this symmetry element is retained in a crystallographic sense for 99% of molecules with an inversion centre, 66% of molecules with a threefold rotation axis, 59% of molecules with a twofold rotation axis and 26% of molecules with a mirror plane. For molecules with several point symmetry elements, any inversion centre is almost always retained

**Table 1.2** Results of a query from *CSDSymmetry*, showing the ten molecular point groups (with more than 30 occurrences in the CSD) that give rise most often to noncentrosymmetric crystal structures. Reproduced from [28] by permission of the International Union of Crystallography

Molecular point group: Hermann–Mauguin notation (Schoenflies notation)	Percentage of crystal structures in noncentrosymmetric space groups	Number of structures considered
4 ( $C_4$ )	42.9	70
3 ( $C_3$ )	41.1	410
1 ( $C_1$ )	30.7	11 5826
$\bar{4}$ ( $S_4$ )	29.5	1162
32 ( $D_3$ )	26.0	580
$\bar{4}3m$ ( $T_d$ )	24.4	6220
222 ( $D_2$ )	23.6	977
2 ( $C_2$ )	23.1	13 474
5/ $m2m$ ( $D_{5h}$ )	23.1	96
$\bar{4}2m$ ( $D_{2d}$ )	21.9	1420

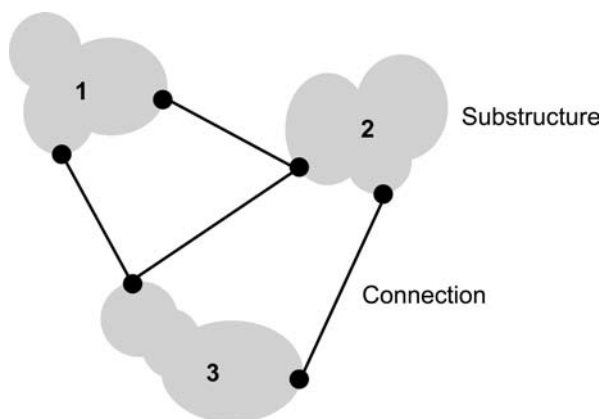
(whether or not in conjunction with other symmetry elements), and other ‘point-acting’ symmetry elements such as  $\bar{4}$  or  $\bar{6}$  axes are also retained frequently. Rotation axes and mirror planes (which act about lines and planes, respectively) are retained less frequently, and the full point group symmetry of any molecule with multiple point symmetry elements is retained in a crystallographic sense quite infrequently. One general challenge in this area seems to be to establish the best way to use the knowledge within *CSDSymmetry* in order to yield genuine enhancements of *experimental* crystal engineering strategies.

## 1.4 New Tools for Database Research

Although the CSD contains the complete body of published crystallographic information, the knowledge that can be derived from it is clearly limited by the scope of the search tools that are available to examine it. For crystal engineering purposes, the standard *ConQuest* software (as it appears in 2007) has several significant limitations for supramolecular searches, not least of which is the practical limitation of search speed. If the full capabilities of the CSD are to be exploited for crystal engineering, the tools to examine its vast information content must be extended. In practice, this means developing new algorithms and new programs for database research. This section of the Chapter focuses on several steps in this direction. The first highlighted example, *3DSEARCH*, is a general tool that promises to provide an effective means for complex supramolecular searches of the CSD. Its potential impact for crystal engineering appears to be significant. Further examples are then described where isolated ad hoc procedures have been developed to address specific questions of relevance to crystal engineering. One of the challenges in this area seems to be to envisage an effective means by which these ad hoc examples can be packaged into a coherent tool that will provide a more general ‘crystal engineering interface’ to the CSD. Finally, we will consider some tools for identifying and visualising trends within sets of crystal structures. These include techniques for automated clustering of CSD search results, methods for detecting and quantifying isostructurality, and methods for defining and identifying consistent structural sub-units within sets of related crystal structures. The key feature of each of these methods is automation, which is required to permit effective interrogation of a large number of structures—clearly, this is a factor that will become increasingly more important as the CSD continues to grow.

### 1.4.1 A General Tool for 3-D Searches of the CSD: 3DSEARCH

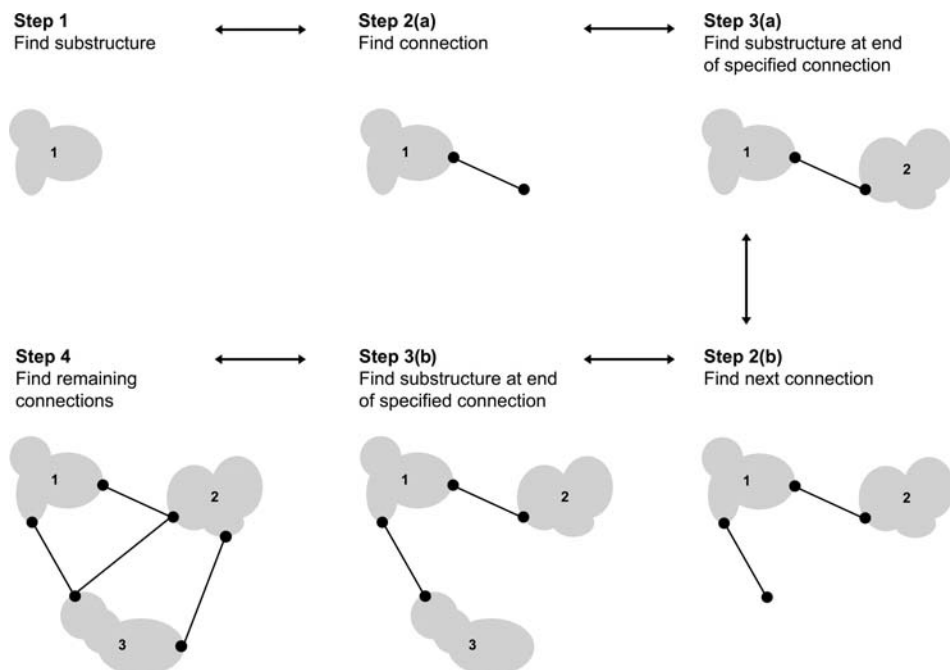
One of the potentially most significant developments for application of the CSD to crystal engineering strategies is the *3DSEARCH* algorithm [32], which has been designed specifically to search 3-D crystal coordinates for extended patterns in an efficient and accurate way. The algorithm permits supramolecular searches to be performed on a reasonable timescale (say, half an hour) where they would have taken many hours or even days using *ConQuest*. A generalised *3DSEARCH* query is viewed as a collection of substructures and connections (Figure 1.4), together with a list of constraints on bond lengths, bond angles or bond torsion angles. The substructures refer to covalently bonded fragments in the usual way, while the connections specify intermolecular geometrical



**Figure 1.4** A generalised 3DSEARCH query containing three substructures (1–3) with four specified connections. The substructures can be of any type and can be specified in the usual way (atom types, bond types, etc.). The connections are defined between specified atoms in each substructure and can be constrained using a range of criteria (intra- or intermolecular, distance, angle, torsion). Reproduced from [32] by permission of the International Union of Crystallography

criteria, such as particular interatomic distances. Within the program, substructures are identified by chemical graphs in a manner similar to that described for *CSDSymmetry* (Figure 1.3) [29]. The query substructures are matched to target CSD entries by comparing node connectivity, which permits efficient matching of all occurrences of a substructure in a particular molecule. To find individual connections, a general search of interatomic distances is applied: an entire unit cell of atomic coordinates is generated, the structure is expanded according to the maximum translations that might be required to encompass the specified interatomic distance, and all atomic positions are checked against the distance criterion. The procedure is general and can identify arbitrarily large interatomic distances which might be either intermolecular or intramolecular.

The 3DSEARCH procedure combines its substructure and connection searches into a ‘depth-first backtracking algorithm’, summarised in Figure 1.5. The algorithm begins (Step 1) by searching for the first substructure amongst the molecules in the crystallographic asymmetric unit of a given CSD entry (there may be more than one match). Each substructure that is found is then examined for the specified connections (Step 2), and each connection that is found is searched for the next specified substructure (Step 3). Each of these steps may provide numerous hits, and all hits are examined to ensure that the search is exhaustive. In a final step (Step 4), a check is made for remaining connections between substructures that have already been found to be connected. This latter step contributes significantly to the speed of the algorithm since it means that a substructure making multiple connections does not necessarily require an extensive substructure search for each connection. If a search step fails at any point (for example, an interatomic connection from an identified substructure does not correspond to a link to the desired substructure in Step 3), the algorithm backtracks to the previous step and continues (for example, it reverts to Step 2 and looks for further connections from the substructure that was initially identified). Each particular path through the search

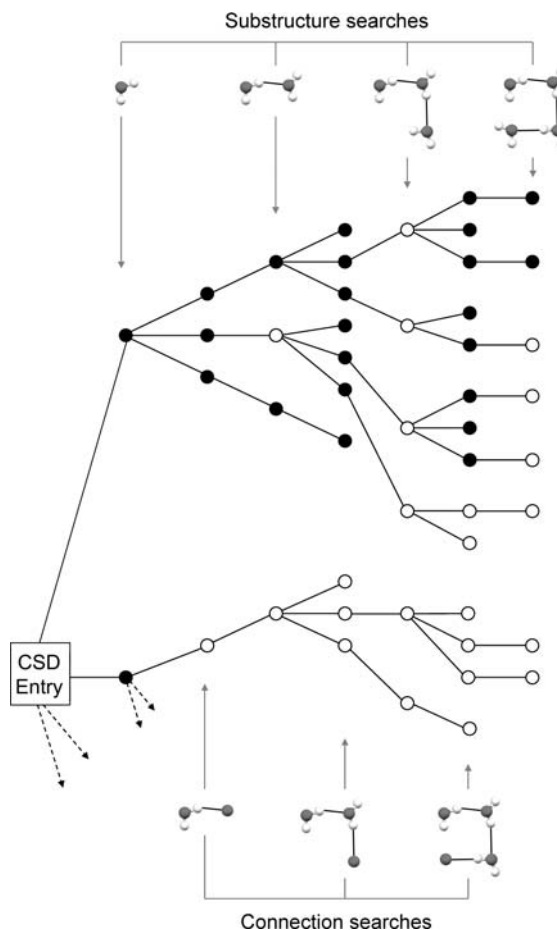


**Figure 1.5** Schematic representation of the search procedure within the 3DSEARCH program. For each substructure identified in Step 1, Steps 2, 3 and 4 are repeated as many times as necessary to identify the specified motif. In this example, two iterations are sufficient. During the search process, failure at any stage causes the process to 'backtrack' to the previous stage, as indicated by the double-headed arrows. Reproduced from [32] by permission of the International Union of Crystallography

procedure can be viewed as one branch of a 'search tree', and the algorithm continues until all possible branches have been examined.

There are several specific ways in which the 3DSEARCH algorithm addresses efficiency. First, the chemical graph representation of the substructures permits rapid assessment of potential substructure matches at the end of each identified connection (Step 3). Potential substructure matches can be ruled out immediately if the node type of the first atom does not match the node type of the desired substructure. If this is the case, there is no need to consider any further the 2-D connectivity of the substructure to be matched. This can be particularly beneficial when searching for relatively large substructures. Secondly, as the complexity of the search query increases, the associated search tree obviously becomes more complex. 3DSEARCH maximises the efficiency by which the search tree is explored by storing and recycling results wherever possible. Each time that the algorithm embarks along a particular branch of the search tree, it checks to see whether any of the required steps might already have been performed so that identical searches are never repeated.

The procedure is illustrated in Figure 1.6, which shows the top half of a search tree corresponding to the identification of a four-membered ring of hydrogen-bonded water molecules. This tree has a 'search depth' of seven: the first level corresponds to



**Figure 1.6** The top section of the search tree explored by 3DSEARCH to identify four water molecules arranged in a hydrogen-bonded ring. Shaded nodes in the search tree denote results obtained from a full substructure or connection search. Unshaded nodes denote results that have been recycled. All nodes in the bottom half of the search tree (not shown) are unshaded. A final rapid step (Step 4 in Figure 1.5) checks for a connection between the first and last identified water molecules. Reproduced from [32] by permission of the International Union of Crystallography

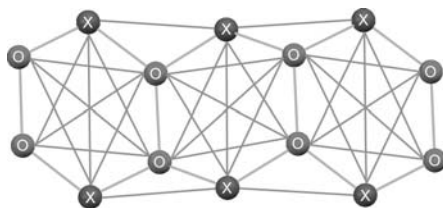
identification of a water molecule, the second level corresponds to identification of an interatomic  $\text{O} \cdots \text{O}$  contact, the third level corresponds to a substructure search to verify that the oxygen atom of the identified contact belongs to a water molecule, and so on. The nodes in the diagram represent successful outcomes at each level. Thus, the nodes in the first column of the diagram represent successful location of a water molecule in a particular CSD entry. The shaded nodes represent steps where full searches have been employed, while the unshaded nodes represent steps where search results have been recycled. For example, the uppermost branch in the search tree consists of four successful steps: (1) a substructure search identifies a water molecule in a CSD entry; (2)

a connection search identifies a suitable  $\text{O} \cdots \text{O}$  contact; (3) a substructure search verifies that the oxygen atom of the contact belongs to a water molecule; (4) a connection search identifies a new  $\text{O} \cdots \text{O}$  connection. At the next step, the algorithm fails—that is, the  $\text{O} \cdots \text{O}$  contact identified in step (4) is not made to another water molecule. Since this branch of the search tree is examined at the very beginning of the search process, each of its substructure and connection matches must be performed in full. However, the results are stored so that it is now *known* that the specific oxygen atom examined at step (3) belongs to a water molecule and the specific oxygen atom examined in the unsuccessful step does not. After failing, the algorithm terminates that particular branch of the search tree and returns to the water molecule identified in step (3), successfully finding a new  $\text{O} \cdots \text{O}$  contact. It happens that this new contact is made to one of the oxygen atoms that have been examined already. Thus, it is known immediately that this oxygen atom belongs to a water molecule and there is no need to perform any substructure search at this step. As the search proceeds, it becomes possible to recycle more and more results, until the search tree in Figure 1.6 comprises only unshaded nodes which can be evaluated very rapidly.

The initial literature report of *3DSEARCH* contains several representative examples of its application to retrieve supramolecular motifs [32]. For example, Figure 1.7 shows a hydrogen-bond motif comprising eight water molecules and six atoms that are allowed to match any element other than C or H. The search is made more complex by specifying contact distances across the hexagons in order to impose geometrical regularity. This query consists of 14 substructures and 47 connections, giving rise to a search tree with a potential depth of 27 levels [in general, the potential depth of the search tree is  $[(2 \times \text{no. substructures}) - 1]$ ]. A search of ca. 270 000 CSD entries was completed in 30 min on a standard PC, identifying this complex motif in 116 structures. Over 50% of the search time is associated with retrieving the coordinate information from the CSD; the actual time spent on the search tree is only 11 min. This performance suggests clearly that the *3DSEARCH* procedure can have a dramatic impact on the use of the CSD for crystal engineering strategies. The new search capabilities have recently been distributed within the *Materials* module of the Mercury CSD system [33].

#### 1.4.2 Ad hoc Software

There are numerous examples in the literature where specific searches of the CSD are achieved through creation of ad hoc search software. Many of these originate from within



**Figure 1.7** Complex hydrogen-bonded motif used as a search query for *3DSEARCH*. The oxygen atoms of the eight water molecules are labelled. The six atoms labelled 'X' are allowed to match any atom except H or D. The lines represent specified connections (47 in total). Reproduced from [32] by permission of the International Union of Crystallography



the CCDC, and some of the tools that are created might eventually find their way into the distributed CSD system. The studies described in this section are grouped under the label ‘ad hoc’ since the search tools were created and employed for a specific purpose. The two selected studies serve as examples of what might be required if we are to utilise the CSD more effectively to answer specific questions in crystal engineering. The second study in particular is illustrative, since it has a direct counterpart in the literature from 15 years earlier. This allows us to examine again the impact that the growth of the CSD and the development of automated search procedures might have for the quantity and quality of crystal engineering knowledge that we are able to obtain from it.

## 1.5 Search for Functional Group Exchanges: *GRX*

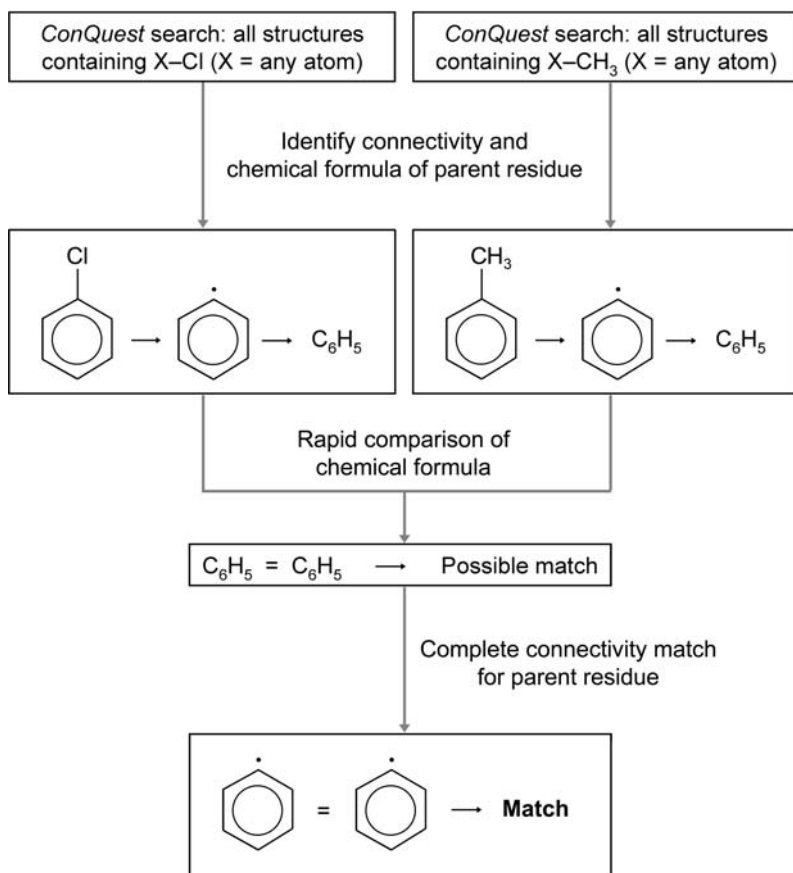
One of the classic paradigms of crystal engineering is the notion of chloro/methyl interchange, where it is observed that pairs of molecules differing only by interchange of a chloro substituent for a methyl group can adopt identical crystal structures [34]. One question that we might reasonably expect to answer by consultation of the CSD is: how often is isostructurality observed for pairs of chloro/methyl substituted molecules? The conceptual steps required to answer such a question are as follows:

- (1) Locate a molecule with a methyl substituent.
- (2) Search for the corresponding molecule with a chloro substituent.
- (3) Where such a pair exists, establish whether they are isostructural or not.
- (4) Repeat for all molecules in the CSD that contain a methyl substituent.
- (5) Establish the total number of relevant molecular pairs in the CSD and then establish the frequency of isostructurality.

Clearly, this is a task that is beyond manual manipulation of the *ConQuest* software, becoming intractable at step (4). To automate this type of process, the *GRX* program has been developed [35].

Input to *GRX* comprises two functional groups that are to be examined for interchange. In the chloro/methyl example, two search queries  $X-\text{Cl}$  and  $X-\text{CH}_3$  would be generated, where  $X$  refers to any atom except hydrogen or deuterium. The steps of the *GRX* search procedure are outlined in Figure 1.8. In the first step, the CSD is searched for all instances of the first functional group, using the chemical connectivities in the usual way. For each entry that is obtained, the functional group in question is subtracted from the total chemical formula to obtain the connectivity and chemical formula of the parent residue. For example, the chlorobenzene molecule  $\text{C}_6\text{H}_5\text{Cl}$  will produce  $\text{C}_6\text{H}_5$  for the parent residue, with a connectivity that resembles a phenyl radical. Where an entry contains more than one molecule type, all specific molecules are searched for the functional group, and a match is stored as a chemical formula of the entire entry with the functional group removed from the appropriate residue (with a note of which residue has its functional group removed). For example, the (hypothetical) two-component crystal  $\text{C}_6\text{H}_5\text{Cl} \cdot \text{C}_6\text{H}_4\text{BrCl}$  would be stored in the list of chemical formulae as ‘1  $\text{C}_6\text{H}_5\text{C}_6\text{H}_4\text{BrCl}$ ’ and ‘2  $\text{C}_6\text{H}_5\text{Cl} \text{C}_6\text{H}_4\text{Br}$ ’. The entire procedure is repeated for the second functional group, producing two sets of parent residues, which must then be screened for matches. Comparison of the connectivity of every entry in the first list with every





**Figure 1.8** Steps in the GRX procedure for the chloro/methyl interchange example. Separate lists of structures containing Cl or CH<sub>3</sub> substituents are obtained using ConQuest, and the connectivity and chemical formula of the parent residue is derived for all structures. The chemical formula is compared in an initial rapid step, then full connectivity matching is undertaken only for pairs that match at the first stage. Reproduced from [35] by permission of the International Union of Crystallography

entry in the second list is a formidable task: for the chloro/methyl example, something in the order of  $1.5 \times 10^{10}$  pairs of entries must be compared for the present size of the CSD. To accelerate the procedure, matching is undertaken initially on the basis of the chemical formula. For entries with multiple molecules, the *total* chemical formula is used, since the procedure aims to identify identical crystal structures rather than identical molecules within different crystal structures. For the chloro/methyl example, this step reduces the set of matches to approximately 22 000. Finally, this set of 22 000 potential matches must be compared according to their complete chemical connectivity. For entries containing more than one molecule or entries where the specified functional groups are found more than once, matching must be applied to every possible combination. For chloro/methyl interchange, the final result from the November 2004 release of the CSD

was 1331 matching pairs, of which approximately 25% were found to be isostructural. According to the initial literature report [35], there are definite plans to integrate the *GRX* procedure into the distributed CSD system.

## 1.6 Search for Solvated and Unsolvated Structures: *Solvates*

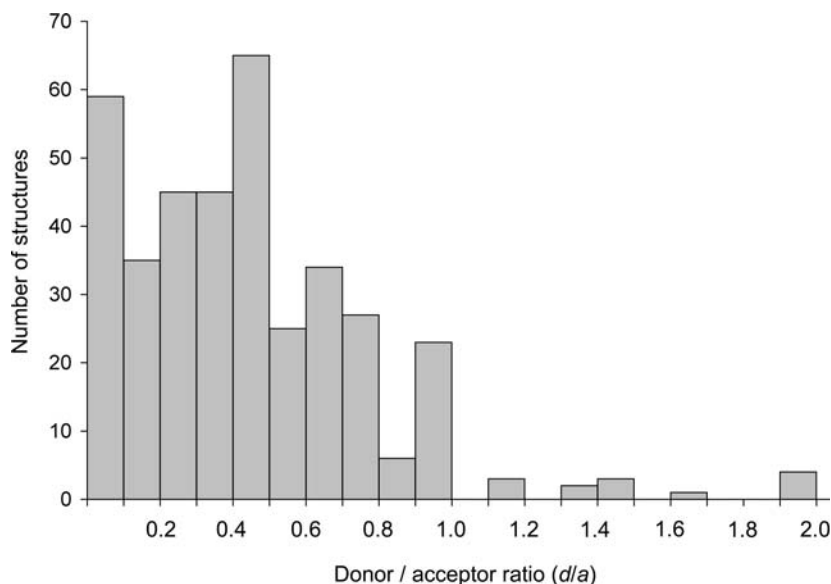
A problem closely related to functional group interchange is the search for sets of crystal structures that contain the same molecule with different solvent molecules or in an unsolvated form. We might wish to ask the question: what role does the solvent molecule play within the crystal structure of a solvated form, and is this correlated in any way with the structures of unsolvated crystal forms? For example, does each unsolvated crystal structure contain any feature that reveals a tendency to incorporate a particular solvent molecule? Can we use this information to avoid certain solvents if we are seeking to crystallise an unsolvated form? To pose this type of question to the CSD, a search tool is required that will identify molecules for which both solvated and unsolvated crystal structures are available. This has been accomplished through the program *Solvates* [36], which performs the task for any specified 'solvent' molecule. The program first locates all CSD entries that contain the solvent molecule, removes that molecule from each entry and the CSD is searched again for the compound that remains. There are no restrictions on the identity of the 'solvent' molecules or on the complexity of the remaining unit, so that all molecular fragments (including counter ions and other solvent molecules) are included in the comparison.

The problems associated with such a task are closely related to those encountered in the *GRX* procedure [35]. For the specific case of hydrates, approximately 36 500 entries can be identified by a coarse search for the word 'hydrate' in the chemical name (which would be expected to provide at least a good approximation to all hydrates in the CSD), each of which must then be used as a search query against the 475 000 CSD entries. This gives a total number of comparisons again in the vicinity of  $1.5 \times 10^{10}$ . The *Solvates* procedure accelerates the search in a different way from the *GRX* program, by reducing the 2-D connectivity of each molecule to its 'topological index', a single number that attempts to encapsulate the 2-D connectivity [37, 38]. Comparison of the full 2-D connectivity is undertaken only for those pairs that are matched in the preliminary stage. For hydrates, this reduces the number of required full comparisons to ca. 5000. The *Solvates* program takes all necessary precautions to ensure the reliability of the structures that are input into the program (for example, applying the common *ConQuest* quality criteria of no disorder, no errors, etc.) then applies some further stereochemical filtering intended to distinguish between enantiomerically pure compounds and racemates. The final result is a list of pairs of crystal structures of anhydrous and hydrated forms of the same molecule. CSD entries containing different amounts of the same solvent molecule are listed with relation to the unsolvated form, so that a dihydrate would not be listed as a 'hydrate of a hydrate', for example.

The question remains as to what should be done with such a list. The *Solvates* procedure generates only a suitable dataset from which we might hope to obtain meaningful answers to the questions that we envisaged at the outset of the study. The particular example of the anhydrous/hydrate dataset is illustrative since it can be compared

directly with a study from 1991 [39]. How does the computationally more sophisticated 2007 study compare with the earlier one? In 1991, Desiraju found 33 886 organic-only entries in the CSD (1988 version), of which 2566 (7.6%) were hydrates [39]. To permit manual inspection, the sample was reduced to 411 structures, first by excluding all salts and cyclodextrins (which have pronounced tendencies for hydrate formation), then by retaining structures on the basis of the journal in which they were published to prevent any bias. Barring a single structure, all of the 411 selected hydrates contain groups capable of participating in hydrogen bonding, and 65% of them have a hydrogen-bond donor:acceptor ratio ( $d/a$ ) < 0.5 (Figure 1.9). Thus, over half of the molecules in hydrated crystal structures have a greater number of hydrogen-bond acceptor groups than hydrogen-bond donor groups. This was explained in the following way: if all good hydrogen-bond donors and acceptors are to be used in hydrogen bonding, molecules with low  $d/a$  ratios must find some means by which to address the  $d/a$  balance. One such means is the adoption of multifurcated interactions in which acceptor groups are associated with more than one hydrogen-bond donor. An alternative is to incorporate water molecules, for which  $d/a = 2$ .

In the 2007 study [36], van de Streek and Motherwell considered 100 864 organic-only structures (after pre-filtering at a strict quality level) of which 5972 (5.9%) were hydrates. Since the 1991 and 2007 studies were based on slightly different pre-filtering conditions, it can reasonably be concluded that the relative proportion of hydrates in the CSD has not changed dramatically over 16 years. For 5146 identified anhydrides, the average  $d/a$  ratio was found to be 0.28. For 5232 hydrates, the average  $d/a$  ratio was found to be 0.63. For ca. 370 molecules where both an anhydrate and a hydrate



**Figure 1.9** Histogram of hydrogen-bond donor/hydrogen-bond acceptor group ( $d/a$ ) ratios for 411 hydrated crystal structures in the 1991 study of Desiraju. The mean value is 0.44. Reproduced from [39] by permission of The Royal Society of Chemistry

are present in the CSD, the average  $d/a$  ratio was found to be 0.65. On the basis of these averages, van de Streek and Motherwell showed that ‘if a compound is soluble in water, it is impossible to predict whether or not it will form a hydrate based on the average number of donors or acceptors or their ratio’, which was interpreted to mean that ‘a low donor/acceptor ratio is not the rationale behind hydrate formation, contradicting Desiraju’s hypothesis’.

So, is this an example where a more comprehensive automated database study negates earlier results? Actually, it seems to be more a matter of interpretation. With the benefit of automated procedures, van de Streek and Motherwell were able to show that the  $d/a$  ratio of a molecule has no absolute predictive power for determining whether it is likely to crystallise as a hydrate. Thus, it seems that the title of Desiraju’s article – ‘Hydration in organic crystals: prediction from molecular structure’ – was not entirely justified. However, the contents of the 1991 article remain perfectly valid. In the 2007 article, van de Streek and Motherwell went on to derive averages for the number of unsatisfied hydrogen-bond donor and acceptor groups: in the anhydrides, there are on average 0.2 unsatisfied donors and 2.8 unsatisfied acceptors, while the hydrates have on average 0.3 unsatisfied donors and 1.8 unsatisfied acceptors. Thus, van de Streek and Motherwell show that ‘the addition of water to a crystal structure has an average effect of reducing the number of unsatisfied acceptors’ [36], which is precisely the conclusion of Desiraju.

This example highlights perhaps one of the most crucial points when we consider the role of the CSD for crystal engineering: the development of automated procedures will provide ever more sophisticated means to explore the vast number of structures in the CSD, including an ever greater variety of statistical measures. However, the chemical conclusions will remain a matter of human interpretation. Desiraju pointed out that an alternative way to satisfy all hydrogen-bond donors in cases where  $d/a < 1$  is by formation of multifurcated interactions. In the case of the hydrates/anhydrides, it might be possible then to show that anhydrides with  $d/a < 1$  are better able to realise multifurcated hydrogen bonds. Thus, the situation should be viewed as a competition between multifurcation and water inclusion, so that prediction of hydrate formation amounts to deciding which of the two possibilities is most favourable for a given molecule. A statistical study of multifurcation seems to be well suited to the *CSDContact* methodology [21]. It would seem then that crystal engineering questions will be addressed most completely and most effectively by *combining* the new search tools to address all conceivable aspects of a given chemical situation.

### 1.6.1 Tools to Examine Structural Similarity

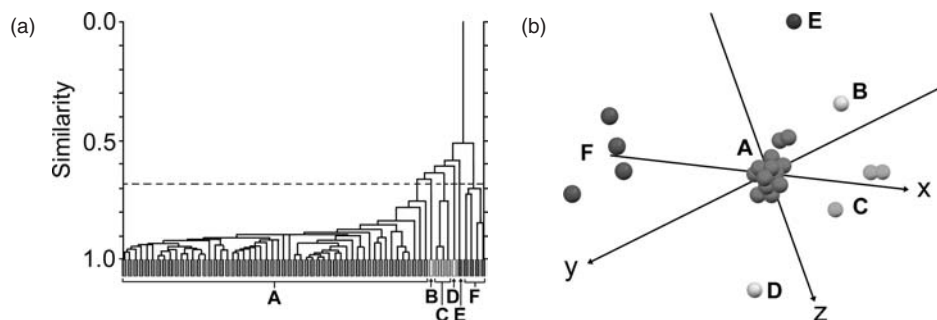
In addition to the tools described for specific searches of the CSD, one of the major requirements for crystal engineering is a capability to compare sets of crystal structures in order to identify degrees of similarity and difference. It is obvious that this task has formed the very basis of the development of crystal engineering, [40] but consistent with the general contemporary trend there is a move towards automated procedures for large-scale structure comparison. The key concept that is required for automation is some unique representation of a crystal structure. Standard items in a CSD entry can cause problems in this respect: for example, unit-cell parameters may differ on account of thermal expansion; the space group may be influenced by a particular choice of origin

or unit-cell setting, and in some cases by missed symmetry or the presence of pseudo-symmetry; the chemical representation may be influenced by subjective assignment of bonds and bond types. This section of the Chapter focuses on three general themes that have been developed recently for the purpose of comparing crystal structures. The first concerns classification and visualisation of large amounts of structural information that might be obtained from a standard CSD search – can the key characteristics of the dataset be extracted and classified in an automated way? The second theme focuses on applications that use the simulated PXRD profile, which can be derived from the primary data items in each CSD entry. These applications include automated determination of isostructurality at the 3-D level, and also techniques for clustering a large number of structures into smaller ‘representative sets’. The last theme to be discussed is more immediately related to traditional concepts of crystal engineering, namely the identification of directly comparable supramolecular motifs within crystal structures. The three themes outlined here are clearly closely related and overlapping. Together, they focus on the common aim of identifying and clarifying trends amongst large sets of crystal structures.

## 1.7 Clustering and Classifying CSD Search Results: *dSNAP*

As the number of entries in the CSD continues to grow, it will become increasingly more difficult to manage even the subsets of data that are produced by specific CSD searches. For example, the ground-breaking 1976 manuscript of Leiserowitz describing the modes of association of carboxylic acids refers to no more than 100 crystal structures, [41] which are certainly amenable to manual analysis. If similar questions are asked today, a search of the CSD for two carboxylic acid groups exhibiting at least one intermolecular contact less than the sum of the van der Waals radii of the atoms involved provides close to 3250 fragments in 2300 structures, [42] which begins to appear a little daunting. In an effort to address this issue, new tools are emerging for classifying and clustering large numbers of structures, with the intention of identifying common features in an automated way. One approach, using techniques of multivariate statistics, is applied in the *dSNAP* program [43].

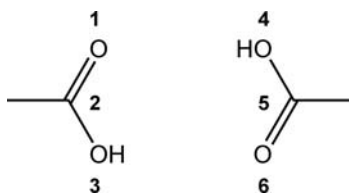
The *dSNAP* methodology is based directly on structural information, employing a distance matrix to describe the geometry of particular fragments extracted from the CSD by means of a standard *ConQuest* search. The fragments might be individual molecules or they might comprise several molecules. For the specific illustration of applications to crystal engineering, the focus here is on intermolecular interactions. The *dSNAP* method involves generation of a matrix whose elements are based on differences between all distances and angles in the structural fragment, not just those that are bonded or specified in some other way. The matrix comprises one row for each fragment to be compared, and each row comprises one column for each geometrical parameter (distance or angle) that has been derived from the fragment. This matrix forms the basis for statistical comparison, the full details of which are outlined elsewhere [43]. The result is a mapping of the dataset onto a 3-D space, producing a metric multi-dimensional scaling (MDS) plot in which each point represents a particular structural fragment and points close together in space represent similar fragments (Figure 1.10). The results can also be represented by a dendrogram, in which each fragment forms a separate



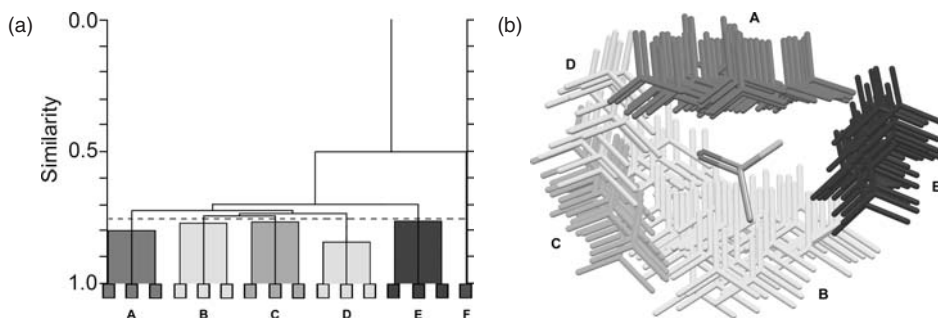
**Figure 1.10** Example of (a) a dendrogram and (b) the accompanying metric multidimensional scaling (MDS) plot produced by the *dSNAP* program. The horizontal dashed line in the dendrogram indicates the cut-off level used to distinguish similar fragments: at the 68% similarity level shown, the complete set of structures is clustered into six distinct groups, labelled A–F. In the MDS plot, each sphere represents one structural fragment, and points close together in space are geometrically similar. Reproduced from [42] by permission of The Royal Society of Chemistry

class at the bottom of the diagram, and the classes are grouped together by tie bars into larger clusters according to specified similarity criteria (Figure 1.10). The height of the tie bar in the dendrogram gives the specific level at which the linked fragments are considered to be similar. For each identified cluster with three or more members, the most representative structure is identified, based on the criterion that it has the minimum average distance from every other member of the cluster in the 3-D space on which the structures are mapped. The same process can be applied to specific distances and angles within the fragment, producing a representation in 3-D space where each point corresponds to a distance or angle, and points that are clustered together correspond to geometrical features that are highly correlated.

The *dSNAP* methodology has been applied to cluster and categorise the intermolecular interactions of carboxylic acid groups [42]. A search of the CSD using the fragment shown in Figure 1.11 provided 3221 fragments from 2313 unique organic-only structures. Using *dSNAP*, the interactions are clustered at the 75% level into five distinct types (plus one erroneous outlier). The majority of the contacts (group A, 2138 structures, 66%) correspond to two carboxylic acid groups approaching each other in a ‘head-on’ fashion. Of these, sub-clustering reveals that 1839 correspond to the ‘ideal’



**Figure 1.11** Carboxylic acid dimer used for the *dSNAP* study described in the text. Contacts (defined as less than the sum of the van der Waals radii) are permitted between any of atoms 1, 2 and 3 and any of atoms 4, 5 and 6. Reproduced from [42] by permission of The Royal Society of Chemistry



**Figure 1.12** (a) Simplified dendrogram and (b) the accompanying Isostar plot obtained from *dSNAP* by clustering the interactions between carboxylic acid groups at the 75% similarity level (indicated by the dashed horizontal line in the dendrogram). Five groups are obtained (labelled **A–E**), plus one erroneous outlier (labelled **F**). The simplified dendrogram provides a clearer indication of the number of groups derived from the dataset, but does not immediately give an indication of the number of structures in each group. Reproduced from [42] by permission of The Royal Society of Chemistry

centrosymmetric carboxylic acid dimer. The recognised ‘*anti*’ interaction that generates catemeric motifs occurs for only 217 fragments (group **C**) confirming the original observation of Leisorowitz that this motif is rare. In fact, contacts between two carbonyl groups are more common – these have been suggested previously to be competitive with hydrogen bonds [44]. One especially helpful feature of *dSNAP* is a synergy with the visualisation tools of *IsoStar*, [19] which enables identified clusters to be viewed as scatter plots, which can reveal their unifying structural features (Figure 1.12).

As with all clustering and classifying procedures, there is a danger (acknowledged openly by its creators) [42] that *dSNAP* might be used to over-classify a dataset. Taken to its extreme, cluster analysis can be made to show that each structure in a dataset is unique. At the other extreme, suitably liberal criteria will cluster all structures into a single group, without any discriminatory power. The most effective clustering level obviously lies somewhere between, and requires some level of input regarding how the results are to be interpreted. An effective procedure usually involves initial clustering into a small group of distinctly different clusters, which can then be sub-clustered to reveal more subtle differences. The clustering procedure is particularly effective for identifying outliers (and therefore provides an excellent way to identify erroneous CSD entries), but chemical and crystallographic knowledge must be applied to distinguish unusual and possibly interesting structures from questionable structures that have been poorly determined. At some stage, it is necessary to acknowledge that there is nothing to be gained by further sub-classification – this can only be assessed by the analyst, with a realistic appraisal of the overall aim of the study.

## 1.8 The PXRD Profile as a Structural Descriptor

One suitable descriptor that can be used to assess global structural similarity is the powder X-ray diffraction (PXRD) profile. PXRD profiles can be simulated from many entries



in the CSD, providing a measure that is influenced only by the periodic arrangement of the atoms in space. Structures that are very similar, perhaps differing only slightly in cell volume or unit-cell shape, or with one atom or group replaced by a similar atom or group, give PXRD profiles that are clearly similar by visual inspection. However, ‘similar by visual inspection’ is not necessarily compatible with automated comparison: in the case of PXRD profiles, the positions of peaks may be very sensitive at a local level to small differences in lattice parameters, which might cause problems for conventional point-by-point comparison (for example, using  $R_{wp}$ ). A more effective similarity measure has been described by de Gelder *et al.*, [45] based upon weighted cross-correlation (WCC) functions which are normalised and can be adapted to account for shifts in peak positions. The WCC procedure employs the concept of a ‘neighbourhood’, by which individual points in the profile are compared not just with exactly equivalent points in the reference profile, but also with points in the local neighbourhood, the scope of which is defined by a triangle function of specified width. Triangle functions that are narrow ignore the neighbourhood of the points in the profile, leading to strict requirements for similarity, while triangle functions that are broad lead often to uniformly high similarity values. The validity and effectiveness of WCC procedures for clustering PXRD profiles has been clearly demonstrated in the literature, [45, 46] and will be illustrated by some specific examples below.

### 1.8.1 IsoQuest

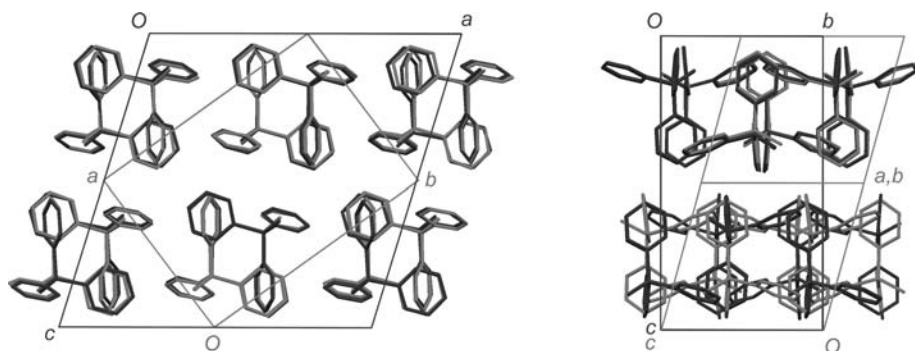
The *IsoQuest* program [47, 48] has been developed on the basis of the WCC method. As a first step, a derived database of PXRD profiles is obtained from CSD entries. This is not necessarily straightforward, since some entries do not contain atomic coordinates, many entries show disorder, contradictions in symmetry, etc., but can be achieved for the majority of entries. The resulting PXRD database can be searched by the *IsoQuest* program, using either a structural model or X-ray diffraction data as input. The ability to input experimental PXRD profiles is important, since it provides a means to assess new data, regardless of whether the pattern can be indexed or the structure solved. A resulting list of identical or closely related structures can be obtained in less than a minute on a modern PC. The *IsoQuest* procedure utilises both the principles of derived databases and knowledge-based libraries. For the latter, a complete database of isostructurality relationships has been derived by comparing the PXRD profile of every structure in the CSD with every other. The resulting *IsoBase* library [47, 48] enables rapid retrieval of isostructural sets, which can be used to augment an *IsoQuest* search, for example providing further hits that might not have been identified in the first instance when using specific search criteria.

As an example, application of *IsoQuest* to the compound  $\text{Ph}_3\text{SiCl}$  [CSD refcode (Nov 2006 version): BARNUD] yields a list of 16 CSD entries, of which 12 are unique structures (Table 1.3) [48]. It is noticeable that the list contains structures determined at different temperatures, illustrating the ability of the WCC procedure to handle temperature-related shifts in peak positions. There are also instances of triclinic structures with unit cells half the volume of the monoclinic structures. These would not have been identified by a standard *ConQuest* search using the reduced cell. Overlaying the monoclinic and triclinic structures (Figure 1.13) shows clearly the structural similarity—the

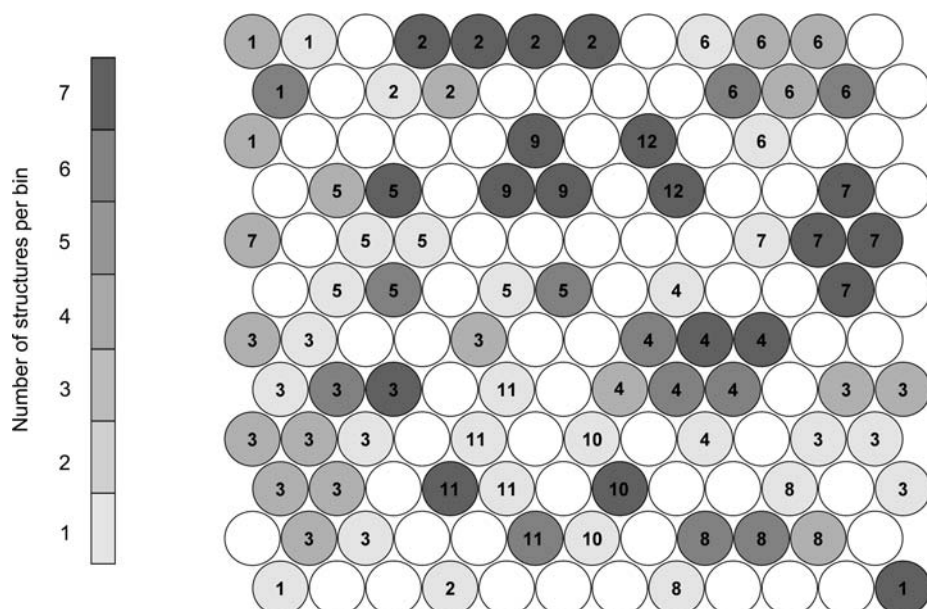


**Table 1.3** Results of an IsoQuest search using the compound  $\text{Ph}_3\text{SiCl}$  [CSD refcode (Nov. 2006 version): BARNUD], listed according to decreasing similarity. The structure of BARNUD is reported in the nonstandard space group  $P2_1/b$ . The triclinic unit cells have half the volume of the monoclinic unit cells

CSD refcode		<i>T</i> (K)	Space group	<i>a</i> (Å)	<i>b</i> (Å)	<i>c</i> (Å)	$\alpha$ (°)	$\beta$ (°)	$\gamma$ (°)
BARNUD	$\text{Ph}_3\text{SiCl}$	298	$P2_1/b$	18.662(7)	18.390(10)	9.667(3)	90	90	107.01(4)
WAKHOF	$\text{Ph}_3\text{SiSH}$	298	$P2_1/c$	18.749(8)	9.663(5)	18.452(9)	90	107.84(4)	90
MTPHEP01	$\text{Ph}_3\text{P}=\text{CH}_2$	100	$P2_1/c$	18.299(2)	9.467(1)	17.971(1)	90	106.90(1)	90
MTPHEP10	$\text{Ph}_3\text{P}=\text{CH}_2$	298	$P2_1/c$	18.455(7)	9.614(7)	18.103(9)	90	106.93(10)	90
TPPOSS	$\text{Ph}_3\text{P}=\text{S}$	298	$P2_1/c$	18.314(4)	9.640(2)	18.013(4)	90	105.99(2)	90
TPPOSS02	$\text{Ph}_3\text{P}=\text{S}$	150	$P2_1/c$	18.162(1)	9.512(1)	17.875(1)	90	106.06(1)	90
TPPOSS03	$\text{Ph}_3\text{P}=\text{S}$	180	$P-1$	9.542(3)	11.207(3)	14.537(4)	88.89(4)	80.39(4)	78.28(3)
TPGEBR	$\text{Ph}_3\text{GeBr}$	298	$P2_1/c$	18.770(4)	9.662(3)	18.467(4)	90	106.84(8)	90
TPPHSE	$\text{Ph}_3\text{P}=\text{Se}$	298	$P2_1/c$	18.471(3)	9.667(1)	18.120(3)	90	106.87(1)	90
TPPHSE01	$\text{Ph}_3\text{P}=\text{Se}$	173	$P-1$	9.613(3)	11.140(3)	14.726(3)	89.23(2)	80.78(2)	78.31(2)
TPASNS	$\text{Ph}_3\text{As}=\text{S}$	298	$P2_1/c$	18.523(8)	9.642(4)	18.140(8)	90	105.93(4)	90
TPASNS01	$\text{Ph}_3\text{As}=\text{S}$	298	$P2_1/c$	18.426(4)	9.750(5)	18.582(4)	90	105.74(6)	90
BONLEV	$\text{Ph}_3\text{PBH}_3$	108	$P-1$	11.021(4)	14.922(5)	9.554(4)	99.28(2)	101.88(1)	89.31(2)
BRTPSN	$\text{Ph}_3\text{SnBr}$	298	$P2_1/c$	19.082(3)	9.763(2)	18.739(3)	90	106.32(8)	90
TPSNCL01	$\text{Ph}_3\text{SnCl}$	298	$P2_1/a$	18.664(4)	9.721(4)	18.983(5)	90	105.60(2)	90
TPSNCL02	$\text{Ph}_3\text{SnCl}$	110	$P2_1/a$	18.410(4)	9.559(1)	18.704(5)	90	105.20(1)	90
QUQCUA	$\text{Ph}_3\text{GeCl}$	173	$P2_1/c$	18.542(5)	9.544(3)	18.305(7)	90	106.18(3)	90



**Figure 1.13** Monoclinic (blue) and triclinic (red) variants of the BARNUD structure. The actual structures shown are MTPHEP01 (space group  $P2_1/c$ ) and BONLEV (after inversion of its published coordinates). The two structures comprise identical layers (in the  $ab$ -plane of the monoclinic unit cell). In the view along the  $a$ -axis of the monoclinic cell (right), the top layers overlay exactly while the bottom layers are offset from each other along the  $b$ -axis of the monoclinic cell. (see colour plate section)



**Figure 1.14** A self-organising map comprising 144 bins. The map has been trained against 205 structures, grouped into 12 distinct classes according to similarity (based on a WCC comparison of their simulated PXRD profiles). The numbers 1–12 in each bin correspond to a structure from a given class mapped to that bin. The 12 classes are separated (i.e. there are no bins that contain structures from more than one class) and the degrees of shading denote the number of structures per bin. For example, the bin at the top left of the map contains two structures from class 1. The edges of the map fold onto each other, and structures from a given class are clearly grouped into neighbouring bins. Reproduced from [49] by permission of the International Union of Crystallography

two structure types contain identical layers, but with different layer stacking sequences. In this case, the WCC procedure is shown even to tolerate variations in the PXRD peak intensities that arise for two slightly different atomic distributions within lattices with similar metric features. Performing a comparable search in *ConQuest*, specifying the unit cell of BARNUD (with the standard tolerances of 1.5%), combined with a search fragment comprising three phenyl rings bound to any atom type yields 23 hits, including only 7 of the 12 unique structures identified by *IsoQuest*. *ConQuest* also identifies a further 16 structures that are similar to BARNUD at a molecular level, but do not exhibit comparable packing arrangements. This example illustrates clearly the ability of *IsoQuest* to provide new crystal engineering knowledge where *ConQuest* cannot.

### 1.8.2 Self-organising maps

A method for *visualising* the extent of similarity amongst a set of crystal structures has been developed from the WCC approach, whereby structures are represented in a self-organising map [49]. The underlying principle is that similar structures are ‘binned’ into subsets so that a ‘continuous space’ of 475 000 unique CSD entries might be reduced

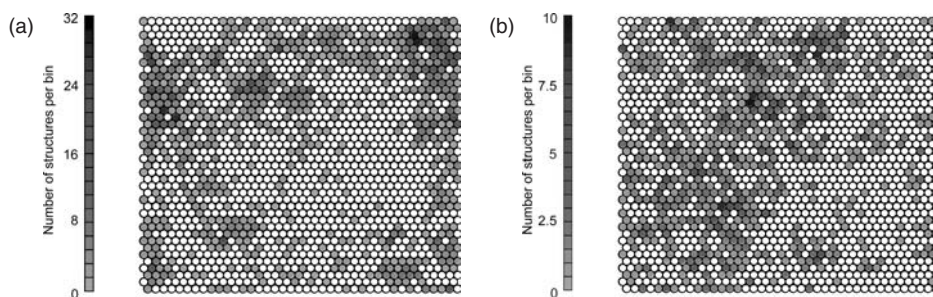
to several thousand bins, each of which contains only similar structures, with a single representative PXRD profile for each bin. The bins may be visualised as a 2-D array (Figure 1.14), where similar structures are mapped to the same bin or to bins that form local clusters in the map. The structure of the map therefore represents the structure of the database in a visually accessible way which might reveal relationships amongst the structures that would otherwise have gone unnoticed. The fact that each bin has a single reference PXRD profile means that subsequent searches for similarity can be made against a small set of reference profiles (equal to the number of bins) rather than against the very large set of PXRD profiles associated with the complete set of CSD entries.

The self-organising property of the map refers to the method of its construction. A number of bins are set up and each is assigned a reference PXRD profile at random. The number of bins to be used depends on the desired level of detail: the more bins that are chosen, the more distinct features will result. The first PXRD profile in the database is compared with the reference PXRD profile in each bin. For the bin that gives the best match, the reference PXRD profile is updated to become more similar to that of the structure presented to it, by an amount determined by a 'learning parameter'. The reference profiles of neighbouring bins in the map are also updated, so that general regions of the map begin to take on comparable features. Each structure in the database is presented in turn to the map, and the whole procedure is repeated over numerous cycles, gradually decreasing the neighbourhood over which the reference profiles are updated. In the early cycles, the reference profiles of *all* bins in the map are updated after every comparison so that the global characteristics of the database are transferred to the map. For example, if there is no PXRD profile in the database that displays any peak below  $2\theta = 10^\circ$ , the reference profiles of all bins will rapidly be updated so that none of them display any such peaks. In the later cycles, the neighbourhood of the matching bin is made progressively smaller until eventually the reference profile is updated only for the specific bin that provides the best match. In this way, similar structures migrate towards specific regions in the map, and individual bins eventually become specialised for the PXRD profiles that are matched to them most frequently.

As an example of the approach [49], a self-organising map comprising 1600 bins was generated from a sample database of 11 165 CSD entries selected from the April 2004 database, and three distinct sets of structures were compared with the trained map:

- (1) A set of 2303 steroid structures were mapped to 654 of the 1600 bins, with a maximum of 32 structures mapped to one particular bin (Figure 1.15a).
- (2) A set of 1262 peptides were mapped to 638 of the 1600 bins, with a maximum of 10 structures mapped to one particular bin (Figure 1.15b).
- (3) All 5789 structures in the July 2004 CSD update were mapped to 1472 of the 1600 bins, covering essentially the whole map.

The confinement of the steroid and peptide datasets to ca. 40% of the bins demonstrates that particular structural types can be related to specific regions of the map. Thus, although the distributions are diffuse, it is possible to speak of the 'steroid' region or the 'peptide' region of the database. The steroid compounds provide another opportunity to compare the performance of the contemporary procedure against established knowledge: in 1993, Kálmán described a small set of steroid structures that display isostructural



**Figure 1.15** Mapping of (a) 2303 steroids structures and (b) the 1262 peptide structures onto a self-organising map comprising 1600 bins trained with 11 165 structures from the April 2004 CSD. The scale on the left-hand side indicates the number of structures mapped to each bin. No compounds are mapped to white bins. In both cases, the structures concentrate in specific (but diffuse) regions of the map. Reproduced from [49] by permission of the International Union of Crystallography. (see colour plate section)

relationships [50]. One of these, digitoxigenin (CSD refcode: DIGTOX), matches five bins in the trained map (to a certain defined tolerance level), which contain in total 59 comparable steroids, including all of those described by Kálmán. The matching procedure is dramatically more rapid than a search of the entire database, taking less than 1 s on a standard PC.

### 1.8.3 Discrimination of polymorphs and redeterminations

The need for clarification of the occurrence of polymorphism in the CSD arises in part from the policy of the CCDC to gather completely and largely uncritically<sup>†</sup> all crystallographic data published in the primary chemical literature. Although completeness is one of the obvious strengths of the CSD, it also means that redeterminations appear frequently. Redeterminations may originate from literature publications that make significant improvements to an otherwise correct structure or they may be corrections to an incorrect structure. In some instances, the same structure may have been published more than once in the literature by the same authors. In any of these cases, the earlier structures are retained within the CSD, and it is up to the user to identify the nature of multiple CSD entries, and to handle them appropriately before attempting to derive conclusions. As a further consequence of the way in which structures are entered into the CSD, specific features such as polymorphism might only be highlighted if they are explicitly mentioned in the primary publication of the crystal structure. If a compound is flagged as polymorphic in the CSD, this does not necessarily mean that the structures of any of the other polymorphs are available. Furthermore, if a compound is flagged as polymorphic, all of its entries will usually be flagged as such, which may include

<sup>†</sup> The CCDC will correct obvious errors if a cause for that anomaly can be found (for example, it is relatively common to see notes of corrections to mistyped atomic coordinates for earlier structures in the database), and they will actively engage in communication with authors where other problems might be suspected. However, the primary policy of the CCDC is to trust in the peer-review process that should be carried out before a structure appears in the primary literature.

one or more redeterminations. For these reasons, it is difficult to draw any meaningful conclusions regarding polymorphism from the *bibliographical* information associated with each CSD entry.

To revert to structural information (which is of course the only sound way to investigate polymorphism), van de Streek and Motherwell considered simulated PXRD profiles in a manner similar to the *IsoQuest* procedure [51]. In an additional effort to address the influence of temperature on the PXRD profile, the unit-cell lengths were normalised for each structure so as to reproduce the unit-cell volumes obtained from summing the average atomic volumes published by Hofmann [52]. The unit-cell angles were assumed to be constant during this normalisation, and no attempt was made to account for anisotropic expansion. The resulting PXRD profiles were then matched using the WCC approach. The normalisation and WCC procedures together provide a very well-defined demarcation that is beneficial for a fully automated procedure: after normalisation of the unit-cell volume, similarities over 0.99 consistently indicate a redetermination while similarities less than 0.97 consistently indicate polymorphism. In the small region between these two values, the algorithm is unable to distinguish reliably between polymorphs and redeterminations.

In an effort to develop a fully automated procedure that can be used to screen the entire CSD, van de Streek and Motherwell considered that some CSD entries are likely to contain errors, which are more likely to be in the atomic coordinates than in the unit-cell parameters. To eliminate the contribution of the unit-cell contents, the structure factor of each reflection in the simulated PXRD profile was set to an arbitrary constant value, thereby producing a 'normalised reduced-cell PXRD profile' that was used for comparison using the WCC procedure. Applying this methodology to the CSD prior to November 2004 produced a list of 2862 unique refcode families which contain reliable pairs of polymorphic structures, amounting to ca. 1% of all unique chemical compounds in the CSD. Numerous strict criteria were applied in the process, so that the list represents a minimum number of reliable observations rather than an exhaustive list.

Some interesting examples are highlighted by the study. Refcodes MEACCU01 and MEACCU02 are described, respectively, in the CSD as discrete molecular units of formula  $[(\text{CH}_3\text{CO}_2)_4(\text{CH}_3\text{O})_4\text{Cu}_4]$  and polymeric units of formula  $[(\text{CH}_3\text{CO}_2)_2(\text{CH}_3\text{O})_2\text{Cu}_2]_n$ . The two compounds actually form identical 2-D polymeric layers, but with different layer stacking sequences (a situation closely comparable with the monoclinic and triclinic structures described for BARNUD) (Figure 1.13). The metric difference is slight, and the atomic arrangements are so closely comparable that differences in the PXRD profiles are also minimal. Thus, this pair is identified to be isostructural by the WCC procedure (regardless of whether the peak intensities are normalised or not) and is a case of subtle polymorphism that would be missed by this automated algorithm. A more positive example which shows the ability of the procedure to overcome possible errors in the atomic coordinates is given by DOTJEB and DOTJEB01. The CSD entry DOTJEB is reported in space group  $P2_1$  while DOTJEB01 is reported in space group  $C222_1$ , with a unit cell twice the volume. The similarity of the full PXRD profiles is only 0.944, suggesting them to be genuine polymorphs. However, matching on the basis of the normalised reduced-cell PXRD profiles suggested the structures to be redeterminations, and closer inspection of DOTJEB01 revealed an error where one coordinate of one chloride atom is missing a



minus sign (the error in the CSD being reproduced from the original publication). After correction, the similarity of the full PXRD profiles is 1.000, and the two structures are shown unambiguously to be identical.

In a manner reminiscent of the self-organising maps, the methodology has been applied to produce a subset of the database that represents the ‘best representative’ of each unique crystal structure [53]. Thus, a compound with 10 CSD entries, comprising one polymorph determined seven times and a second polymorph determined three times will be reduced to two entries, which are considered to represent the two unique structure types. The details of the applied quality tests are extensive [53], but the result is a list of 231 918 structures (derived from 353 666 structures in the November 2005 release) that are considered to be the best representative examples of all unique high-quality structures in the CSD [54]. In this way, the complete contents of the CSD are reduced to a set of representative structures that contain an equivalent amount of structural information, but without any redundancy.<sup>‡</sup> This dataset forms an especially convenient basis for structural searches, since it is free of any duplication.

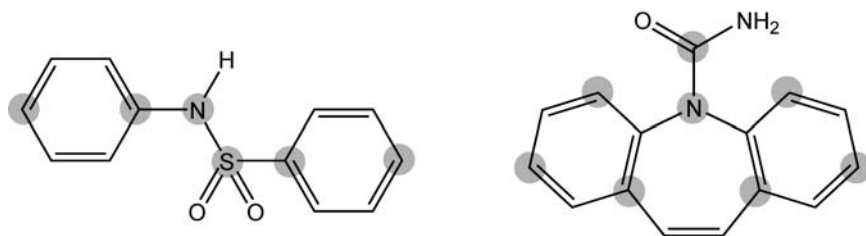
## 1.9 Identifying Supramolecular Constructs: *XPac*

The problem of identifying complete (or almost complete) isostructurality amongst sets of crystal structures is handled effectively by the WCC approach to PXRD profile matching. However, for crystal engineering purposes it is often necessary to identify more limited isostructurality in the form of consistent supramolecular motifs. These might be isolated clusters or they may be larger assemblies that extend in one or two dimensions. Traditional methods to identify such similarity involve manual inspection of crystal structures, which can be hampered by sheer structural complexity or weight of numbers, and might also be biased by preconceived ideas of what constitutes a significant intermolecular interaction. The *XPac* program has been developed to address this task in a systematic and automated way by identifying consistent geometrical features in crystal structures, which have been termed ‘supramolecular constructs’ (SCs) [55]. The distinction between a supramolecular construct and a supramolecular synthon is that the former implies only geometrical similarity, while the latter refers to some specific association of chemical groups.

Like most general methods for assessing structural similarity, *XPac* dispenses with traditional descriptions of unit cells and space groups by considering the environments of molecules [55, 56]. A reference molecule (referred to as the kernel molecule) is defined, and a coordination shell is generated on the basis of some distance criterion. A molecule is considered to form part of the coordination shell if it forms at least one contact to the kernel molecule in which the interatomic distance is less than the sum of the van der Waals radii plus some defined tolerance. The tolerance is usually set to a value of 1.5 Å, which is liberal in the context of conventional views of what constitutes a significant intermolecular interaction. The kernel molecule and all associated shell molecules form

---

<sup>‡</sup> Conceptually, the process is *equivalent* to the generation of self-organised maps. In the case of the best representative set, the conditions for two structures to fall into the same bin include the strict criterion of identical chemical structure. Thus, only redeterminations of the same crystal structure fall into the same bin, and the number of bins in the final map will be equal to the number of unique structures in the CSD.



**Figure 1.16** Ordered sets of points carbamazepine and diarylsulfonamide molecules, suitable for input to the XPac program. The points in the diarylsulfonamide molecule are independent of any rotation of the benzene rings and are therefore more likely to be comparable in a set of structures. Reproduced from [57] by permission of The Royal Society of Chemistry

a representative cluster of the crystal structure. For structures with  $Z' > 1$ , more than one cluster is defined. A set of points is then specified within the structure, based on selected atomic positions. For example, for the comparison of a set of diarylsulfonamide structures (Figure 1.16), the six highlighted atoms that form the molecular core might be an appropriate set of points. In the assessment of structural similarity, the derived differences between two structures will depend on genuine packing differences, which are the object of the study, and also on internal differences in the chosen set of points, which should therefore be minimised. Comparison of structures using XPac is only sensible if genuinely comparable sets of such points can be defined in each structure to be compared.

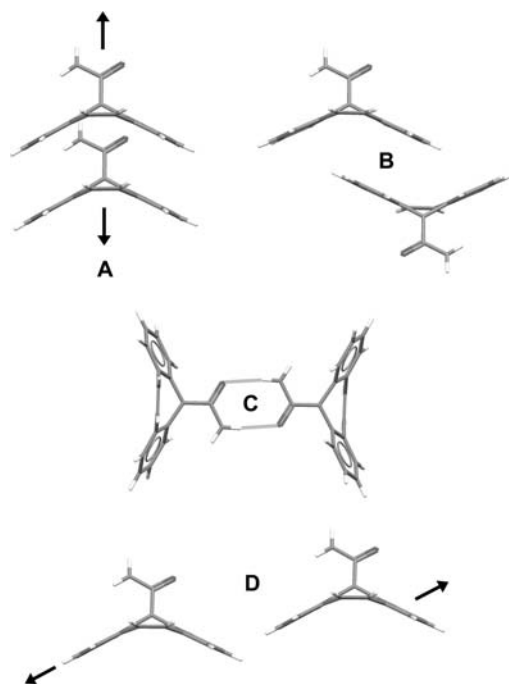
When an appropriate ordered set of points has been established, the arrangement of the various sets of points within the representative cluster can be defined using internal coordinates (which are independent of unit cell or space group). For two structures that are to be compared (referred to as **1** and **2**) the agreement between the ordered sets of points is measured by calculating the mean absolute difference between all corresponding angles. If good agreement is not obtained at this stage, the chosen set of points does not form a sensible basis to probe structural similarity. It is then established whether the representative clusters of structures **1** and **2** contain any pairs of molecules that form comparable arrangements. This assessment is based on all unique angles between groups of three atoms, angles between the normals to planes defined by groups of three atoms, and the torsion angles defined by groups of four atoms. Distance criteria are not taken into account, so that the procedure is not restricted by differences in bond lengths, for example. Absolute values are considered for the torsion angles to account for the fact that they will be inverted for molecules related by inversion centres or mirror planes. For a representative cluster containing  $N$  molecules, there are  $\frac{1}{2}(N^2 - N)$  pairs to consider and the comparison is made for each of these pairs. At this stage, there are three possible outcomes:

- (1) There are no comparable molecular pairs in structures **1** and **2**: the structures exhibit no packing similarity and the analysis is completed.
- (2) There is exactly one comparable molecular pair in structures **1** and **2**: the structures exhibit isolated pairs that are comparable (a 'zero-dimensional' SC) and the analysis is completed.

- (3) There are several comparable molecular pairs in structures **1** and **2**: it is necessary to assess whether the pairs form part of a larger comparable motif.

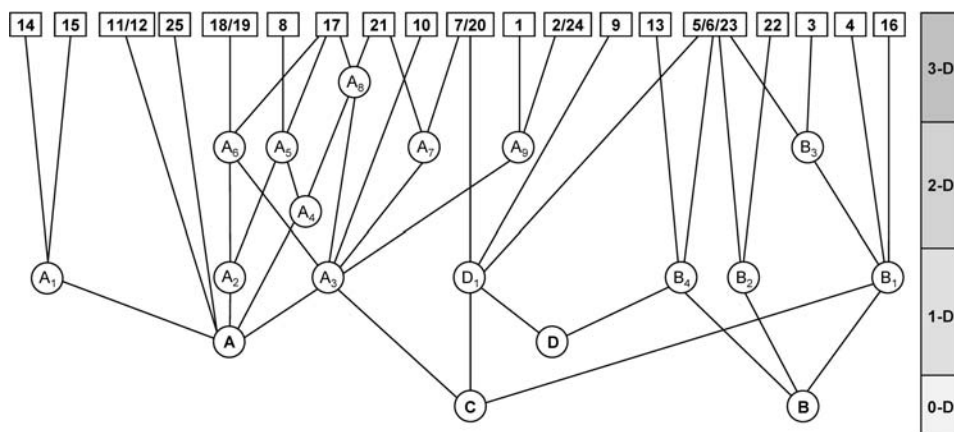
For the latter case, the process is applied again to all possible combinations of pairs containing a common molecule. Thus, the comparison is applied over three molecules rather than two. After this step, any similar molecular triples are combined to form the ‘seed’ of an SC. Any comparable molecular triples that share a common molecule must form part of the same seed. In some cases, there may be more than one distinct seed. Finally, the unique seeds are expanded by the symmetry operators of the structure to yield the SC(s). At this point, information concerning the dimensionality and the base vectors of the SC can be extracted.

An example application of the *XPac* method is a study of 25 crystal structures containing the carbamazepine molecule (CBZ) (Figure 1.16) [57]. The group comprises four polymorphs of CBZ, 19 multi-component crystals and two molecules that are close analogues of CBZ. Analysis using *XPac* reveals four primary SCs (Figure 1.17): two different types of one-dimensional chain (**A** and **D**) and two different types of centrosymmetric dimer (**B** and **C**). These can be combined in various ways to yield a complex set of structural relationships, which are most conveniently represented in the form of a ‘tree diagram’ (Figure 1.18). The diagram consists of nodes that symbolise either crystal structures or SCs, joined by lines that represent relationships between them. Nodes positioned lower in the diagram are subsets of nodes towards the top of the diagram: an



**Figure 1.17** The four primary supramolecular constructs (**A–D**) in the CBZ family, identified by the *XPac* program. Reproduced from [57] by permission of The Royal Society of Chemistry





**Figure 1.18** Diagram showing the relationships between 25 structures of the carbamazepine family, generated by the XPac program. The 25 crystal structures are represented in the top row. The remaining nodes represent supramolecular constructs (SCs). The dimensionality of the SCs is shown in the scale on the right-hand side. Related structures are connected by lines. Reproduced from [57] by permission of The Royal Society of Chemistry

SC is a subset of crystal structure, a 2-D SC can be a subset of a 3-D SC, and so on. Thus, the crystal structures to be compared are represented along a horizontal row at the top of the diagram and the dimensionality of the SCs decreases down the diagram. Isostructural crystals are represented by a single node. All structures that contain a certain SC can be identified by following branches that emerge upwards from that SC. Similarly, common SCs can be identified by following branches downwards from each crystal structure until they meet at a common node. If their branches never meet, the two structures do not possess any common fragment.

For the set of CBZ derivatives (Figure 1.18), each structure has at least one indirect connection to the four primary SCs (A–D). It can be seen that 24 of the 25 structures possess a downward path to either **A** or **B** (but never both), showing that these primary SCs are present in all but one structure and that they never occur simultaneously in a structure. Both of these SCs are based on close-packing of the V-shaped CBZ molecule (Figure 1.17): in **A**, the two molecules are related by translation, which can be extended infinitely to yield a 1-D SC, while in **B** they are related by inversion, which yields an isolated (0-D) SC. Structure **9** (which has no *downward* path to either **A** or **B** in the diagram) is a two-component crystal in which a saccharine molecule intervenes to preclude formation of either **A** or **B**. The SC **C** is based on a centrosymmetric dimer formed by hydrogen bonds between the amide functionalities in two CBZ molecules (Figure 1.17). It can be seen from Figure 1.18 (by following all possible routes upwards from **C**) that 17 structures contain this SC. The SC **D** is associated with ‘side-by-side’ translational arrangement of CBZ molecules (Figure 1.17), which occurs in seven structures. A vast number of structural relationships can be deduced from Figure 1.18. For example, SCs **A** and **C** are present simultaneously in the secondary 1-D SC denoted **A<sub>3</sub>** (found in 10 structures), which is in turn present in the 2-D SCs **A<sub>6</sub>**, **A<sub>7</sub>** and **A<sub>9</sub>**, and also in the 3-D SC **A<sub>8</sub>**.

As for all of the examples concerned with structural similarity, *XPac* is capable of providing a complex set of structural relationships that can appear overwhelming – one vast set of crystal structures can be transformed into an equally vast set of relationships that exist between them. How are we to make sense of this information and apply it to future crystal engineering studies? In the CBZ example, [57] the most important result is probably the identification of the mutually exclusive primary SCs **A** and **B**. It can be noted that **A** and **B** do not contain directional interactions between specific functional groups, which means that they might not be considered within a ‘conventional’ assessment of structural relationships. Conventional decomposition of the CBZ structures would probably arrive at SC **C** as the most notable motif since it is based on the hydrogen-bonded amide–amide supramolecular synthon. However, strategies to construct (or to predict) crystals containing the CBZ unit would probably be better based on the assembly of SCs **A** or **B**, which are identified by *XPac* to be most robust. This idea mirrors the approach taken by Coquerel and co-workers, who have considered 1- or 2-D periodic fragments in crystal structures to describe cases of polymorphism, crystal twinning and epitaxy [58].

### 1.10 Concluding Remarks: The Future Role of Crystallographic Databases

It is clear that the CSD contains a vast amount of crystallographic information that constitutes the very core of crystal engineering. It is also clear that the potential value of this knowledge is only realised when it is gathered into a coherent collection with appropriate tools to manipulate it. At present, this means a single consistent structural format (rather than numerous formats with ‘on-the-fly’ conversion) managed at a single location (rather than any distributed database that is united only in a virtual sense). As computing and global networking capabilities become more powerful, alternative models might become viable, but at the present time the single collated CSD system is the way in which crystallographic information is organised and distributed. It is clear that the CSD does not constitute the *complete* body of crystallographic information. A conservative estimate of the actual number of crystal structures that have been determined worldwide is at least three or four times the number of entries in the CSD. The discrepancy is accounted for by structures in academic laboratories that have not yet been made public, and structures in commercial laboratories that are unlikely ever to be made public. Having seen the scientific benefits gleaned from 475 000 crystal structures, how much more could be gained from a unified collection of 2 000 000 structures or more? One of the principal challenges to ensure effective progress in crystal engineering is to ensure not only that crystallographic information is maintained and managed appropriately, but also that it is augmented to the fullest extent possible.

With the advent of rapid instrumentation for X-ray data collection and powerful software for structure solution, refinement and validation, it is becoming increasingly apparent that traditional methods for publication of crystallographic data are unable to handle the vast number of crystal structures that are being determined [59]. One step that might be taken to address this problem would be for the CCDC to acknowledge itself as the *primary* repository for all crystallographic data – that is, for its role to change from a passive collector of published crystallographic data into a primary publication source.

As the automated checking of crystallographic data becomes more sophisticated (not least through application of *Mogul*), the CSD could effectively regulate its own content, becoming the primary referee of *all* crystallographic data, which can *subsequently* be made available for discussion in the chemical and crystal engineering literature. At the heart of our academic system is the concept of peer review, which means (at least in principle) that somebody ensures the integrity of every structure that eventually finds its way into the CSD. It would seem sensible and most efficient for this review to be performed in a consistent manner by an organisation that specialises in the maintenance of crystallographic data. With a sufficiently flexible and open outlook, which must certainly include clearly documented and justifiable assessment criteria, and which might also include sending some proportion of structures for external review, the CCDC need not be viewed as the 'crystallography police', but rather as the key managers of all crystallographic data.

With specific reference to crystal engineering, this Chapter has illustrated the requirement for development and distribution of more effective tools to explore crystal structures beyond the molecular level. The tools that have been described herein address various aspects of this problem, apparently solving several of them more or less completely. One principal challenge that seems to be immediate is to ensure that this variety of tools, which originate both from within the CCDC and from outside of it, are collected together into a coherent package that can form a 'crystal engineering interface' to the CSD, in the way that *ConQuest* forms an effective 'chemical and bibliographical' interface. The great success of *ConQuest* (even more so than the *Quest* program that came before it) is largely due to its coherence and ease of use, which allows it to be immediately accessible to scientists that might have little or no crystallographic knowledge. The same accessibility must be the target of a crystal engineering interface, so that any scientist can perform effective searches of the CSD at a supramolecular level. This type of package (together with a more realistic approach to what constitutes novelty) will help to limit the number of misleading structural reports that are beginning to become apparent in the primary crystal engineering literature.

Finally, what about the results that are being obtained from contemporary database studies? In the midst of all of this '*in silico* crystal engineering', we should not lose sight of the fact that crystals are physical objects, and we should consider for a moment the extent to which knowledge obtained from the CSD really does facilitate new experimental crystal engineering strategies. The question has been raised at several points through the Chapter, but it is not one that is easy to answer. It is clear that the vast majority of crystal engineering studies are heavily reliant on crystallographic knowledge extracted from the CSD, but it is difficult to pinpoint specific examples where it can be said *absolutely* that experimental results would not have been obtained in the absence of the CSD. For 'traditional' types of database studies, such as investigation of the frequency of formation of hydrogen bonds, it seems that contemporary approaches yield datasets and conclusions that are largely in agreement with earlier studies based on more limited datasets. The strength of the newer automated procedures is that they consider very much larger datasets and usually produce a much greater variety of statistical measures that are more immediately accessible. There does not so far seem to be any genuine instances where a large-scale automated search has negated the knowledge obtained from any well-considered earlier analysis. Contemporary methods are likely to have their greatest

impact where they are able to identify structural trends that might never have been imagined. The automated tools for assessing structural similarity are promising in this respect. It is noticeable (and necessary) that much of the progress in this area has been made by scientists that are effective computer programmers. Amongst such a group, is there still a place for 'old-fashioned' chemical intuition? Curiously, it seems that this may be the key to taking the next big steps in this research area. The advent of the CSD and the original tools for exploring its contents laid many of the key foundations that were developed by intuitive chemists into the research area that we now call crystal engineering [40]. The emergence of more sophisticated search tools should present opportunities in the coming decades for similarly creative chemists to expand the crystal engineering frontiers.

## References

1. F. H. Allen, The Cambridge Structural Database: a quarter of a million structures and rising, *Acta Cryst.*, **B58**, 380–388 (2002).
2. RCSB Protein Data Bank: <http://www.rcsb.org/pdb>
3. Inorganic Crystal Structure Database: <http://www.fiz-karlsruhe.de/icsd.html>
4. Nucleic Acid Database: <http://www.ndbserver.rutgers.edu>
5. Crystallography Open Database: <http://www.crystallography.net>
6. CSD Version 5.30 (November 2008) plus November 2008 and February 2009 updates. 472 200 entries in total.
7. F. H. Allen and W. D. S. Motherwell, Applications of the Cambridge Structural Database in organic chemistry and crystal chemistry, *Acta Cryst.*, **B58**, 407–422 (2002).
8. A. G. Orpen, Applications of the Cambridge Structural Database to molecular inorganic chemistry, *Acta Cryst.*, **B58**, 398–406 (2002).
9. A. Nangia, Database research in crystal engineering, *CrystEngComm*, **4**, 93–101 (2002).
10. WebCite [http://www.ccdc.cam.ac.uk/free\\_services/webcite](http://www.ccdc.cam.ac.uk/free_services/webcite)
11. J. A. Chisholm, E. Pidcock, J. van de Streek, L. Infantes, W. D. S. Motherwell and F. H. Allen, Knowledge-based approaches to crystal design, *CrystEngComm*, **8**, 11–28 (2006).
12. S. R. Hall, F. D. Allen and I. D. Brown, The Crystallographic Information File (CIF): a new standard archive file for crystallography, *Acta Cryst.*, **A47**, 655–685 (1991).
13. International Union of Crystallography <http://www.iucr.org>
14. I. J. Bruno, J. C. Cole, M. Kessler, J. Luo, W. D. S. Motherwell, L. H. Purkis, B. R. Smith, R. Taylor, R. I. Cooper, S. E. Harris and A. G. Orpen, Retrieval of crystallographically-derived molecular geometry information, *J. Chem. Inf. Comput. Sci.*, **44**, 2133–2144 (2004).
15. P. W. Betteridge, J. R. Carruthers, R. I. Cooper, K. Prout and D. J. Watkin, *CRYSTALS* version 12: software for guided crystal structure analysis, *J. Appl. Cryst.*, **36**, 1487 (2003).

16. W. I. F. David, K. Shankland, J. van de Streek, E. Pidcock, W. D. S. Motherwell and J. C. Cole, *DASH*: a program for crystal structure determination from powder diffraction data, *J. Appl. Cryst.*, **39**, 910–915 (2006).
17. M. Mascal, L. Infantes and J. A. Chisholm, Water oligomers in crystal hydrates – what’s news and what isn’t? *Angew. Chem. Int. Ed.*, **45**, 32–36 (2006).
18. I. J. Bruno, J. C. Cole, P. R. Edgington, M. Kessler, C. F. McCrae, P. McCabe, J. Pearson and R. Taylor, New software for searching the Cambridge Structural Database and visualising crystal structures, *Acta Cryst.*, **B58**, 389–397 (2002).
19. I. J. Bruno, J. C. Cole, J. P. M. Lommerse, R. S. Rowland, R. Taylor and M. L. Verdunck, *IsoStar*: a library of information about nonbonded interactions, *J. Comput. Aided Mol. Des.*, **11**, 525–537 (1997).
20. I. C. Hayes and A. J. Stone, An intermolecular perturbation theory for the region of moderate overlap, *Mol. Phys.*, **53**, 83–105 (1984).
21. L. Infantes and W. D. S. Motherwell, The probable number of hydrogen-bonded contacts for chemical groups in organic crystal structures, *Chem. Commun.*, 1166–1167 (2004).
22. L. Infantes and W. D. S. Motherwell, Prediction of H-bonding motifs for pyrazoles and oximes using the Cambridge Structural Database, *Struct. Chem.*, **15**, 173–184 (2004).
23. Microsoft Office Access, Microsoft Corporation (1992–2003).
24. L. Infantes and W. D. S. Motherwell, Hydrogen bond competition between chemical groups: new methodology and the Cambridge Structural Database, *Z. Kristallogr.*, **220**, 333–339 (2005).
25. T. Steiner, Competition of hydrogen-bond acceptors for the strong carboxyl donor, *Acta Cryst.*, **B57**, 103–106 (2001).
26. F. H. Allen, W. D. S. Motherwell, P. R. Raithby, G. P. Shields and R. Taylor, Systematic analysis of the probabilities of formation of bimolecular hydrogen-bonded ring motifs in organic crystal structures, *New J. Chem.*, **23**, 25–34 (1999).
27. W. D. S. Motherwell, Distribution of molecular centres in crystallographic unit cells, *Acta Cryst.*, **B53**, 726–736 (1997).
28. J. W. Yao, J. C. Cole, E. Pidcock, F. H. Allen, J. A. K. Howard and W. D. S. Motherwell, *CSDSymmetry*: the definitive database of point-group and space-group symmetry relationships in small-molecule crystal structures, *Acta Cryst.*, **B58**, 640–646 (2002).
29. J. C. Cole, J. W. Yao, G. P. Shields, W. D. S. Motherwell, F. H. Allen and J. A. K. Howard, Automatic detection of molecular symmetry in the Cambridge Structural Database, *Acta Cryst.*, **B57**, 88–94 (2001).
30. C. P. Brock and J. D. Dunitz, Towards a grammar of crystal packing, *Chem. Mater.*, **6**, 1118–1127 (1994).
31. E. Pidcock, W. D. S. Motherwell and J. C. Cole, A database survey of molecular and crystallographic symmetry, *Acta Cryst.*, **B59**, 634–640 (2003).
32. J. A. Chisholm and W. D. S. Motherwell, A new algorithm for performing three-dimensional searches of the Cambridge Structural Database, *J. Appl. Cryst.*, **37**, 331–334 (2004).
33. Information is available from the CCDC’s website: [http://www.ccdc.cam.ac.uk/products/csd\\_system/materials\\_module/](http://www.ccdc.cam.ac.uk/products/csd_system/materials_module/)

34. M. R. Edwards, W. Jones, W. D. S. Motherwell and G. P. Shields, Crystal engineering and chloro-methyl interchange: a CSD analysis, *Mol. Cryst. Liq. Cryst.*, **356**, 337–353 (2001).
35. J. van de Streek and W. D. S. Motherwell, *GRX*: a program to search the CSD for functional group exchanges, *J. Appl. Cryst.*, **38**, 694–696 (2005).
36. J. van de Streek and W. D. S. Motherwell, New software for searching the Cambridge Structural Database for solvated and unsolvated crystal structures applied to hydrates, *CrystEngComm*, **9**, 55–64 (2007).
37. L. A. Evans, M. F. Lynch and P. Willett, Structural search codes for online compound registration, *J. Chem. Inf. Comput. Sci.*, **18**, 146–149 (1978).
38. D. Bawden, J. T. Catlow, T. K. Devon, J. M. Dalton, M. F. Lynch and P. Willett, Evaluation and implementation of topological codes for online compound search and registration, *J. Chem. Inf. Comput. Sci.*, **21**, 83–86 (1981).
39. G. R. Desiraju, Hydration in organic crystals: prediction from molecular structure, *J. Chem. Soc., Chem. Commun.*, 426–428 (1991).
40. G. R. Desiraju, Midsummer madness, *Nature*, **431**, 25 (2004).
41. L. Leiserowitz, Molecular packing modes: carboxylic acids, *Acta Cryst.*, **B32**, 775–802 (1976).
42. A. Parkin, G. Barr, W. Dong, C. J. Gilmore and C. C. Wilson, Identifying structural motifs in intermolecular contacts using cluster analysis, Part 1: interactions of carboxylic acids with primary amides and with other carboxylic acid groups, *CrystEngComm*, **8**, 257–264 (2006).
43. G. Barr, W. Dong, C. J. Gilmore, A. Parkin and C. C. Wilson, *dSNAP*: a computer program to cluster and classify Cambridge Structural Database searches, *J. Appl. Cryst.*, **38**, 833–841 (2005).
44. F. H. Allen, C. A. Baalham, J. P. M. Lommerse and P. R. Raithby, Carbonyl–carbonyl interactions can be competitive with hydrogen bonds, *Acta Cryst.*, **B54**, 320–329 (1998).
45. R. de Gelder, R. Wehrens and J. A. Hageman, A generalised expression for the similarity of spectra: application to powder diffraction pattern classification, *J. Comp. Chem.*, **22**, 273–289 (2001).
46. E. L. Willighagen, R. Wehrens, P. Verwer, R. de Gelder and L. M. C. Buydens, Method for the computational comparison of crystal structures, *Acta Cryst.*, **B61**, 29–36 (2005).
47. R. de Gelder and J. M. M. Smits, *SYSTER* and *ISOQUEST*: how good and how unique are your data and structure? *Acta Cryst.*, **A60**, s78 (2004).
48. R. de Gelder, Quantifying the similarity of crystal structures, *International Union of Crystallography Commission on Crystallographic Computing Newsletter*, No. 7, 59–69 (Nov. 2006).
49. R. Wehrens, W. Melssen, L. Buydens and R. de Gelder, Representing structural databases in a self-organising map, *Acta Cryst.*, **B61**, 548–557 (2005).
50. A. Kálmán, L. Párkányi and G. Argay, Classification of the isostructurality of organic molecules in the crystalline state, *Acta Cryst.*, **B49**, 1039–1049 (1993).
51. J. van de Streek and W. D. S. Motherwell, Searching the Cambridge Structural Database for polymorphs, *Acta Cryst.*, **B62**, 504–510 (2006).



52. D. W. M. Hofmann, Fast estimation of crystal densities, *Acta Cryst.*, **B58**, 489–493 (2002).
53. J. van de Streek, Searching the Cambridge Structural Database for the ‘best’ representative of each unique polymorph, *Acta Cryst.*, **B62**, 567–579 (2006).
54. The best representative list is updated for each major CSD release and is available from: [http://www.ccdc.cam.ac.uk/free\\_services/best\\_representative/](http://www.ccdc.cam.ac.uk/free_services/best_representative/)
55. T. Gelbrich and M. B. Hursthouse, A versatile procedure for the identification, description and quantification of structural similarity in molecular crystals, *CrystEngComm*, **7**, 324–336 (2005).
56. T. Gelbrich, The *Xpac* program for comparing molecular packing, *International Union of Crystallography Commission on Crystallographic Computing Newsletter*, No. 7, 39–44 (Nov. 2006).
57. T. Gelbrich and M. B. Hursthouse, Systematic investigation of the relationships between 25 crystal structures containing the carbamazepine molecule or a close analogue: a case study of the *XPac* method, *CrystEngComm*, **8**, 448–460 (2006).
58. C. Gervais and G. Coquerel, Simple model designed to generate new crystal structures derived from a mother phase: application to molecular compounds, *Acta Cryst.*, **B58**, 662–672 (2002).
59. A. D. Bond and J. E. Davies, *Chem. Br.*, **39**, 44 (2003).





# 2

## Computational Crystal Structure Prediction: Towards *In Silico* Solid Form Screening

Graeme M. Day

### 2.1 Introduction

The holy grail of computational modelling, as applied to molecular materials, is the prediction of structure and properties from first principles, from an understanding of the intermolecular interactions that are at work and of how they guide the packing of molecules in crystals. There is a wealth of information about the interactions between molecules in the rapidly expanding Cambridge Structural Database (CSD), which now holds the crystal structures of over 400 000 compounds [1]. Each of these structures is a delicate balance of a variety of interactions with a wide range of strengths. Therefore, crystal structure modelling is a stringent test of our understanding of these interactions and our ability to describe them in a computationally useful form. Until recently, the main aim of the computer modelling of molecular crystal structures has been to understand and explain, at the level of intermolecular interactions, the properties of these materials. However, as methods have improved and computing power continues to increase, the goals have become more ambitious, with the *ab initio* prediction of crystal packing now a target, and the hope that computational (*in silico*) studies will progress to the level of guiding experimental investigations in crystal engineering.

That we even consider crystal structure prediction as a possibility is a measure of the impact of improvements in our understanding of intermolecular interactions, and

the computational advances that allow the theories to be applied to real systems. The quantification of crystal packing forces traces back to Kitaigorodskii's arguments for close packing as the main driving force that determines crystalline structure [2, 3]. It is interesting that, while describing the many possibilities for the close packing of molecules, he proposed that it 'would not, of course, be feasible to carry out one-by-one examination of all possible packing patterns of molecules for unit cells of different symmetries and different dimensions' [4]. However, such studies are now, in a way, possible, with the 'examination' normally limited to a computation of some scoring function (usually the lattice energy) and it is the methods, reliability and application of such studies that will be discussed in Sections 2.2, 2.3 and 2.4, respectively.

Before embarking on a description of the computational methods involved, and how well they perform, we should address the goals of crystal structure prediction. At its most ambitious level, the aim is to start from nothing more than the structural formula of a molecule and to predict, with perfect reliability, the structure of the resulting solid, with no input from experimental observations. (Here, by structure, we mean the space group, unit cell parameters and a full specification of all atomic positions.) This goal is, of course, unrealistic; polymorphism in molecular crystals tells us that there is often not just one crystal structure for a molecule and we know that the crystal that is produced in an experiment depends on a variety of factors, from thermodynamic descriptors of the system (temperature and pressure) to the method of crystallization, solvent used and the presence of impurities. Without a detailed description of the crystallization conditions, prediction of *the* resulting structure cannot be the aim. Furthermore, many of these factors are not sufficiently well understood to be represented in a computational procedure for crystal structure prediction.

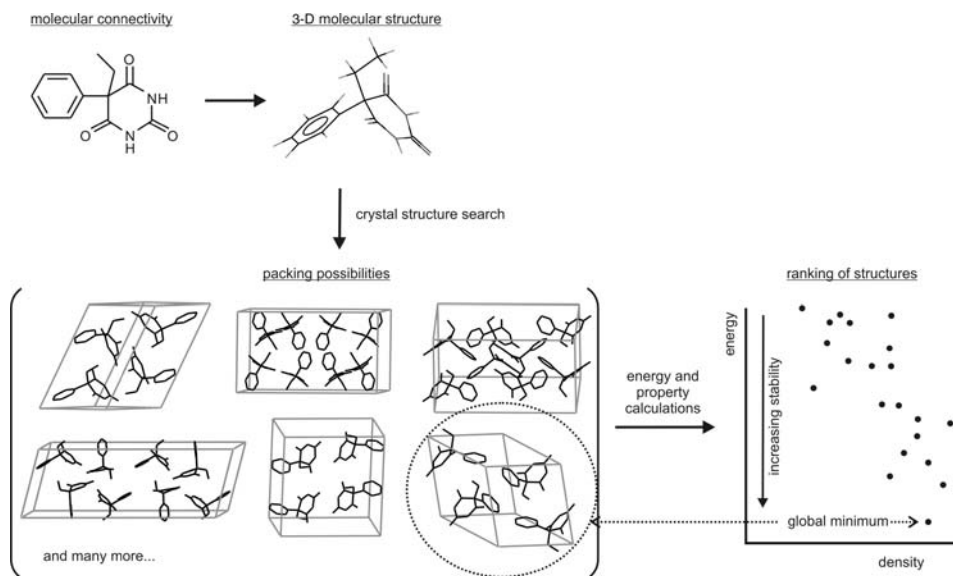
A less ambitious, more realistic goal is the enumeration of the most likely structures – the 'inherent' crystal structures, or putative polymorphs, of a molecule. A list of these, along with a measure of their relative stabilities, should already be useful to the experimental crystallographer, particularly if the calculations can be performed on a relatively fast timescale. This is the goal towards which most methods strive: producing a ranked list of crystal structures within which all possible polymorphs are at, or near, the top. From the crystal engineering point of view, *in silico* experiments could investigate the influence of changes in molecular structure on the likely crystal structures and help to guide the placement of functional groups on a molecule when aiming for certain structural characteristics in the resulting crystal. Another application is structure solution from limited experimental data (e.g. powder X-ray diffraction), where predicted structures might provide the necessary starting point for determining the structure.

The aim of this chapter is to give a brief description of the methods currently used for the *ab initio* prediction of crystal structures, to provide an examination of the reliability of these methods and to show how such *in silico* examinations of crystal packing can complement experimental studies in developing our understanding of the crystal packing of small organic molecules. The examples of applications are taken from pharmaceutical materials science, which is one of the main industrial interests in crystal structure prediction.

## 2.2 Methods used to Predict Crystal Structures

The most general approach to the prediction of crystal structures, and that which will be focused on in this chapter, consists of three general steps, as summarized in Figure 2.1: (i) obtaining a 3-D model of the molecule from the connectivity of its atoms; (ii) a *search procedure* to sample phase space for all packing possibilities for the molecule; followed by (iii) a *ranking method* to assess which of these putative structures are most likely to be experimentally realizable. By far the most commonly used method for assessing the computer-generated crystal structures is to calculate their lattice energies and assume that those with lowest energy are most likely to be observed [5]. The global minimum (i.e. the lowest in energy) is taken to be the most likely crystal structure for that molecule.

Each step in this ‘lattice energy minimization’ approach to crystal structure prediction poses its own challenges, which have had to be addressed in developing existing computational methods. Calculating the molecular geometry from the structural formula is now, at least for small molecules, rather routine. Quantum mechanics calculations can provide accurate gas phase geometries, which are often used as the starting point for predictions. Therefore, the two main areas for development have been in algorithms for searching the multi-dimensional phase space for all possible crystal structures and in developing the methods used to accurately predict the relative energies of the resulting crystal structures.



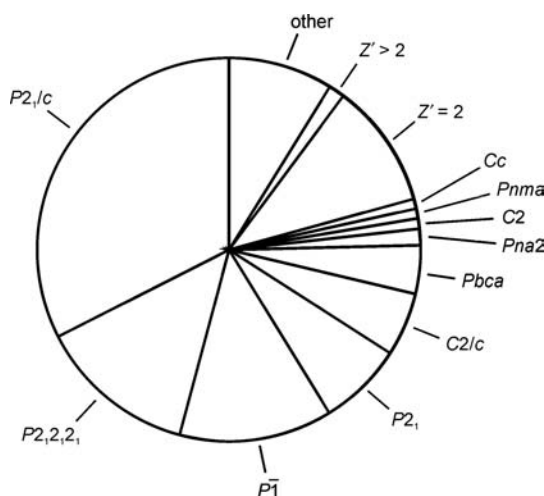
**Figure 2.1** Schematic representation of the lattice energy minimization approach to crystal structure prediction

### 2.2.1 Search Methods

Once a molecular model has been calculated, this is used as a building block to generate crystal structures. This step in crystal structure prediction is a demanding mathematical problem of finding all minima on a highly dimensional potential energy surface. Here, space group symmetry is normally used to reduce the complexity of the problem. Within a given space group, the search method must sample all possible positions and orientations of molecules in the asymmetric unit (up to  $6 \times Z'$  degrees of freedom), as well as up to 6 degrees of freedom describing the shape and size of the unit cell, locating all resulting energy minima for further analysis.

As space group symmetry is generally used while searching for structures, there is a need to decide which space groups should be considered. Here, we can take advantage of the fact that space groups are unequally populated by molecular crystals, with the majority of known crystal structures occupying a small subset of the 230 space groups. A search of the CSD reveals that the six most common space groups with  $Z' \leq 1$  cover over 75% of observed homomolecular organic crystal structures (Figure 2.2). The popularity of space groups in the CSD should reflect a molecule's ability to close pack and satisfy important intermolecular interactions in each space group. Therefore, it is assumed that their relative popularities are also related to the likelihood of finding a relevant low energy crystal structure in a lattice energy search. Therefore, calculations are normally restricted to a set of the most commonly observed space groups. In practice, the number of space groups considered is determined by the time available for the study and the importance placed on finding every possible crystal structure; the minimum is normally 9–10 space groups, covering 80% of the space group statistics (Figure 2.2). Ideally, where computing resources are available, a larger set is considered, also including  $Z' > 1$ .

Besides the use of space group statistics, which is common across most methods, there is a great variety of search algorithms, including simple random sampling of phase



**Figure 2.2** Distribution of common space groups for molecular organic crystal structures in the Cambridge Structural Database [6].  $Z' \leq 1$  unless specified

space, Monte Carlo energy-guided stochastic searches and genetic algorithms, amongst many others. The discussion in this chapter focuses mainly on analysis and evaluation of the resulting structures; in-depth descriptions and comparisons of the search methods can be found elsewhere [7–10]. It is sufficient to know that all methods aim to produce a comprehensive list of plausible crystal structures.

## 2.2.2 Evaluating the Computer-generated Crystal Structures

Once the set of crystal structures has been produced, the challenge remains to rank them in order of likelihood. In the lattice energy approach this means evaluating the relative energies of the full set of structures. The distribution of putative crystal structures resulting from a computational study is often visualized via a scatter plot (Figure 2.1, lower right) of the relative energies of the computer-generated structures, plotted against a second property (often the crystals' densities) – this is a convenient means of visualizing the number of distinct structures in the relevant energy range, and the variability amongst the low-energy structures of the second property (here, the density). The lattice energy minimization approach to crystal structure prediction is a purely thermodynamic approach and ignores any influence of kinetics during nucleation and crystal growth; crystallization is not generally an equilibrium process and the structure resulting from a particular crystallization might represent that structure which crystallizes most rapidly or easily. Furthermore, the thermodynamic model itself is very simplified – free energies should be used to evaluate relative stabilities, while lattice energies ignore all thermal contributions (e.g. entropy). Nevertheless, we will explore how well this basic approach to crystal structure prediction performs, despite the important factors that are ignored.

Search procedures often produce lists of thousands of crystal structures and experience has shown that, for most molecules, a surprisingly large number of these are energetically feasible. A rule of thumb for the maximum energy differences between polymorphs is  $10 \text{ kJ mol}^{-1}$  [4, 11], so we should consider all computer-generated crystal structures in this range above the global minimum to be energetically plausible. For many molecules, lattice energy searches produce tens or even hundreds of distinct crystal structures in this range [12, 13]. Therefore, correctly ordering the relative stabilities of the set of crystal structures is a great challenge: when energy differences are on the order of  $0.1\text{--}1 \text{ kJ mol}^{-1}$ , the required accuracy is a stringent test of whichever method is used to calculate the crystal structures' energies. A further requirement is that the energy calculation must be reasonably fast, considering that the energies of thousands of computer-generated crystal structures must typically be evaluated.

The most common approach used to model intermolecular interactions in crystals is the atom–atom method, where the interaction between molecules is approximated as a sum of atom–atom contributions. The approach was pioneered by Kitaigorodskii [14, 15] and has been an invaluable tool for the modelling of molecular crystals. In its most common form, the interaction energy between two molecules is expressed as:

$$\begin{aligned}
 U_{MN}(R_{MN}, \Omega_{MN}) &= \sum_{i \in M, k \in N} u_{ik}(R_{ik}) \\
 &= \sum_{i \in M, k \in N} A_{ik} \exp(-B_{ik} R_{ik}) - C_{ik} R_{ik}^{-6} + \frac{q_i q_k}{4\pi \epsilon_0 R_{ik}} \quad (2.1)
 \end{aligned}$$

Here,  $R$  and  $\Omega$  represent the separation and relative orientation of the molecules or atoms, and  $i$  and  $k$  are constituent atoms of molecules  $M$  and  $N$ , of atom type  $\iota$  and  $\kappa$ , respectively. Atom types are usually defined by element and sometimes further categorized by bonding environment, e.g. ‘hydroxyl hydrogen’ or ‘sp<sup>3</sup> carbon’. The parameters  $A$ ,  $B$  and  $C$  are typically empirically parameterized, for example to reproduce crystal structures and sublimation enthalpies of molecular crystals, while atomic partial charges ( $q_i$ ) are normally derived from quantum mechanical electronic structure calculations on the isolated molecule. In Equation (2.1), the interaction between a pair of atoms only depends on their separation. Until fairly recently, almost all molecular modelling has used this isotropic atom approximation, largely because of the added complexity of anisotropic models, where atom–atom interactions are also dependent on the relative orientation of atoms. Unsurprisingly, isotropic atom–atom models sometimes fail to reproduce structures and properties satisfactorily; we know that certain interactions are structure-directing due to their orientational preferences. For example, hydrogen bonds are usually aligned to lone pairs on the acceptor atom. The dominant feature here is the electrostatic contribution to the interaction energy and electrostatics are generally the most orientation-dependent contribution to intermolecular forces [16]. Therefore, most of the effort in improving beyond the isotropic atom approximation has been in developing more elaborate descriptions of electrostatic interactions. One approach that has been demonstrated to improve the modelling of molecular crystals is to replace the atomic charges by multipole expansions (charge, dipole, quadrupole, etc.) on each atom in the molecule [17]. Such distributed multipole models are able to describe important details of the charge density distribution in molecules, such as lone pairs and  $\pi$ -electron density, which are often not adequately described by atomic charges alone. The effects of using the more elaborate electrostatic models are explored in the following section, where two appraisals of current capabilities in crystal structure prediction are discussed: the blind tests of crystal structure prediction and an assessment based on a large test set of small molecules.

## 2.3 Current Capabilities of Crystal Structure Prediction

### 2.3.1 The Blind Tests

A survey of the literature shows that, by the late 1990s, somewhere on the order of 60–70 molecules had been studied by crystal structure prediction methods using the lattice energy minimization approach outlined in the previous section [5, 18]. Amongst these studies there were varying claims for success of the methods and, in 1999, researchers at the Cambridge Crystallographic Data Centre decided that an objective test of the capabilities of the various methods should be performed. The resulting ‘blind test’ of crystal structure prediction was published in 1999 [7] and, following the interest generated by the first blind test, three more followed in 2001, 2004 and 2007, to investigate developments in the methods [8, 10, 19]. For each blind test, active researchers were invited to test their methodologies; the number of participating groups grew from 11 in 1999 to 17 in 2001, 18 in 2004 and 14 in 2007. The results of these four international collaborations are described below and can be taken as one measure of the current capabilities and limitations of the computational methods.

For each blind test, recently determined, but unpublished, crystal structures were collected from crystallographers kind enough to delay publication of their structures. From these, a set of molecules were chosen for the test, and their structural diagrams (Table 2.1) were circulated to the participants. The experimentally determined crystal structures were withheld for 6 months by an independent referee, during which time predictions had to be submitted: each participant was asked to propose three crystal structures, in order of preference, for each molecule. A prediction was considered to be successful if one of the three submitted structures was an adequate representation of the experimentally determined crystal structure [7].



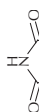

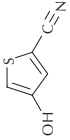
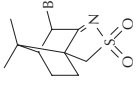
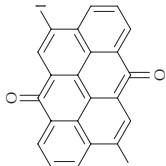
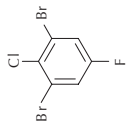
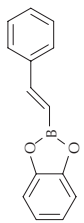
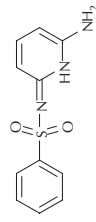
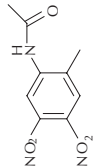
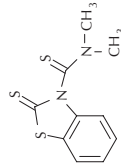

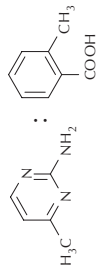
Three categories of molecule have been considered in each of the blind tests: category 1 – small and rigid molecules (up to about 25 atoms) with only the elements C, N, O and H; category 2 – rigid molecules with more challenging elements or functional groups that are intended to present a challenge for parameterization of modelling methods; and category 3 – moderately flexible molecules (usually with two or three flexible acyclic torsion angles). In the first two blind tests, the crystal structures were guaranteed to be in a ‘common’ space group with  $Z' \leq 1$ , while no restriction was placed on space group in the third and fourth rounds (except  $Z' \leq 2$  in 2004 and  $Z' \leq 1$  in 2007). The 2007 blind test also included a co-crystal to test methods of dealing with multiple independent molecules. In total, 15 molecules have been studied in the four blind tests; these are numbered I–XV (Table 2.1).

A few things can be said about the overall results of the four blind tests (Table 2.2): there has been some success for rigid molecules, although the predictability of the different category 1 and 2 crystal structures is variable. Molecular flexibility is a serious obstacle for current methods of crystal structure prediction; only one ‘category 3’ success was achieved in the first three blind tests and, while more success was achieved for the flexible molecule in the latest test, this was partly due to the more restricted molecular flexibility of the molecule chosen that year [19].

Because all participating groups followed the general approach of (i) searching the energy surface to generate a set of likely structures, then (ii) ranking the resulting structures to choose their three predictions (usually based on calculated energies), the success rates are a combination of the successes and failures of these steps. However, an analysis based solely on the top three predictions from each group cannot distinguish between methods that fail because of shortcomings of the search methods or unsuccessful ranking. In the first blind test, only the three ‘official’ predictions were collected from each participant, while for the second, third and fourth blind tests participants were asked to also submit extended lists of the crystal structures generated by their search method. Therefore, for the latest three tests, we can separate the performance of the search methods (i.e. what proportion of participants located the correct structure *somewhere* in their list of structures, not just in their top three) and the success in the ranking of the structures.

Success rates of the search methods generating the experimental structure somewhere in the list of computer-generated crystal structures are given in Table 2.2. Of all of the rigid molecules, molecule XI stands out – this molecule crystallizes with  $Z' = 2$  and appeared as a target because the  $Z' \leq 1$  restriction had been lifted for the third blind test. Dealing with multiple independent molecules is clearly a limitation of too many of the methods. For the other rigid molecules, the 56–79% success rates of locating

**Table 2.1** Diagrams of the molecules included as targets in the four blind tests of crystal structure prediction

Category	1999 [7]	2001 [8]	2004 [10]	2007 [19]
1	 (molecule I)	 (molecule IV)	 (molecule VIII) <sup>a</sup>	 (molecule XII)
2	 (molecule II)	 (molecule V)	 (molecule IX)	 (molecule XIII)
3	 (molecule III)	 (molecule VI)	 (molecule X)	 (molecule XIV)
Extra	 (molecule VII) <sup>b</sup>			 1:1 co-crystal (molecule XV) <sup>c</sup>

<sup>a</sup> Molecule XI was included as an additional category 1 target midway through the 2004 blind test, after it was found that some information on the crystal structure of molecule VIII had been reported. Although most participants continued their predictions without using this experimental information, this molecule might not be considered a true blind test.

<sup>b</sup> Molecule VII was added to the main three 1999 blind test molecules as an additional challenge.

<sup>c</sup> The category of two-component crystals was added to the fourth blind test.



**Table 2.2** Success rates in the blind tests of crystal structure prediction

Molecule	Number of participants	Participants who generated the correct structure in their list	Successful predictions (in the top 3)
<b>Category 1 – Small rigid molecules</b>			
I (form I)	11	– <sup>b</sup>	0
I (form 2) <sup>a</sup>			4
IV <sup>c</sup>	16	10 of 14 <sup>d</sup> (71%)	3
VIII	14	11 (79%)	5
XI	18	4 (22%)	0
XII	13	10 (77%)	4
<b>Category 2 – Rigid molecules with challenging functional groups</b>			
II	8	– <sup>b</sup>	1
V	15	11 of 14 <sup>d</sup> (79%)	4
IX	16	9 (56%)	1
XIII	14	9 (64%)	4
<b>Category 3 – Flexible molecules</b>			
III	11	– <sup>b</sup>	1
VI	11	4 of 10 <sup>d</sup> (40%)	0
X	15	7 (47%)	0
XIV	12	9 (75%)	3
<b>Category 4 – Multi-component crystals</b>			
XV	12	5 (42%)	2
<b>Extra</b>			
VII	6	– <sup>b</sup>	1

<sup>a</sup>Molecule I was polymorphic, but all successful predictions were for the metastable form 2.

<sup>b</sup>The publication of the first blind test does not report enough detail to know the overall success rate of the search methods.

<sup>c</sup>The strategy and results for molecule IV from one of the participants were published separately from the main blind test publication [20].

<sup>d</sup>Some participants in the second blind test borrowed the candidate structures from another group for re-ranking by another scoring method or energy model. There were, therefore, only 14 independent searches performed for molecules IV and V, and 10 for molecule VI.

the observed structure are lower than what would be needed for reliable predictions. However, it is promising that some methods that have been used over more than one blind test are consistently successful for this step in the prediction process. The methods are clearly less robust for flexible molecules: for molecules VI and X, under half of the participants explored packing space sufficiently to locate the correct structure, which partly explains the low overall success rates for these molecules. There is clearly a need to develop reliable search methods for flexible molecules and systems with more than one independent molecule – both capabilities would be necessary for many applications of crystal structure prediction.

In judging the success of ranking methods, we should discard those entries that did not produce the structure in the first step. Instead, we ask: how many of those that did generate the correct structure ranked it within their top three? Category 1 success rates were: 3 of 10 = 30% (IV), 5 of 11 = 45% (VIII), 0 of 4 = 0% (XI) and 4 of 10 = 40%

(XII); those for category 2 molecules are 4 of 11 = 36% (V), 1 of 9 = 11% (IX) and 4 of 9 = 44% (XIII). There were no successes in category 3 for the second and third blind tests (0 of 4 for VI and 0 of 7 for X), and an increased success rate (3 of 9 = 33%) for the flexible molecule (XIV) in the latest blind test. For four of the molecules, there is a reasonable (ca. 1 in 3) rate of success, as long as the search method for generating structures does not fail. Molecules XI and IX were more problematic and illustrate one problem with over-interpreting the blind test results: the difficulty of molecules varies too much (even within one of the categories of molecule) to give a reliable measure of developments in the methods, given the necessarily small number of molecules that can be included in these projects. It is usually only in post-analysis that the factors leading to poor predictability can be assessed and both of these molecules were deemed to be particularly problematic [10].

Although the small number of molecules studied does limit the conclusions that can be drawn from the blind tests, they are a useful exercise for highlighting the most important limitations of existing methods; successes and failures over the 15 molecules in the four blind tests point to the areas where developments are most urgent. Those methods that are showing promise for becoming robust for small rigid molecules must be extended to meet the challenges posed by more complex systems.

### 2.3.2 Further Assessments of Crystal Structure Prediction

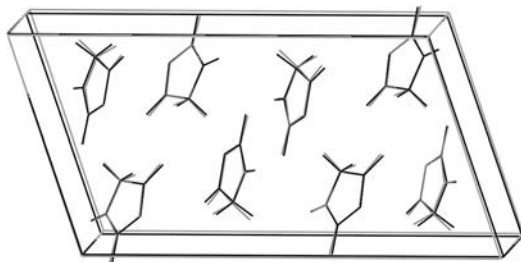
Another approach to assessing lattice energy minimization as a method for crystal structure prediction is by, instead of applying a wide range of methods to a small set of molecules, applying a given protocol to a much larger set of molecules [13, 21, 22]. An advantage of such an approach is that, if the set of molecules is large and varied enough, it should contain both the difficult and the more easily predictable cases and, so, is less subject to the ‘luck of the draw’ than the blind tests, where one particularly difficult molecule can give the impression that no progress has been made.

In a recent study, lattice energy searches were performed for 50 rigid molecules of varying shapes and sizes (molecular weights ranging from 28 to 111 amu), containing a range of common functional groups [13, 21]. Each of these molecules has at least one crystal structure in the CSD, in one of the common space groups. Several are polymorphic, giving a total of 64 known crystal structures in the test set. Following crystal structure searches and lattice energy minimizations for all 50 molecules, the lists of computer-generated crystal structures for each molecule were compared with the experimentally determined structures. While the performance of the search methods was not the focus of this study, it is notable that all but four of the crystal structures were located using a consistent search protocol [13]. The two main questions that were asked of the results were: (i) how often does the experimentally observed crystal structure for a given molecule correspond to the global minimum in calculated lattice energy? (Or, for polymorphic systems, how often are the known polymorphs the lowest energy possibilities?); and (ii) how much variability is there between types of molecules: are there certain molecular characteristics that lead to more or less reliable predictions based on lattice energy minimization?

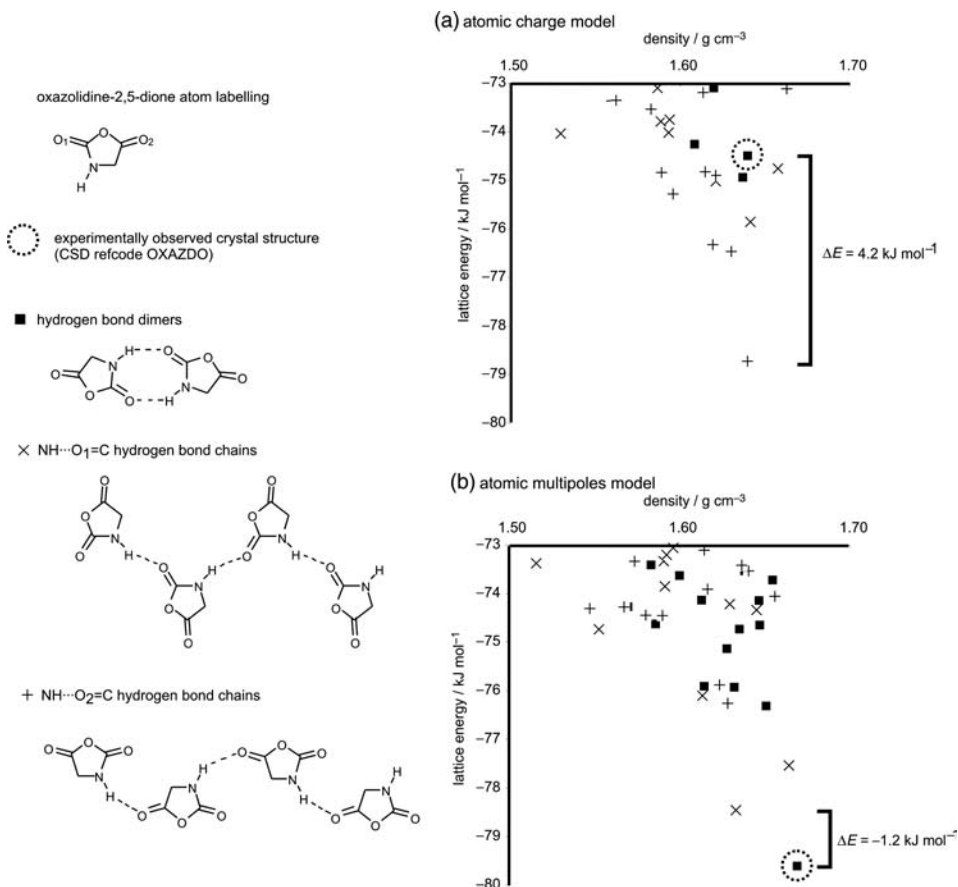
### 2.3.2.1 The Likelihood of Finding the Observed Structure at the Global Minimum

The rate of finding the observed crystal structures at, or near, the global minimum in calculated energy depends on at least two issues: (i) how often crystallization actually leads to the thermodynamic product; and (ii) the quality of the method used to calculate the relative energies of the computer-generated crystal structures. The two effects can only be separated by systematically improving the energy model until no overall improvement in ranking of the structures can be attained. At this point, we could confidently assume that high-energy structures represent kinetically determined crystallization products. As a first step in addressing the importance of the quality of the energy model, two models were used to evaluate the lattice energies of the computer-generated crystal structures of all 50 molecules: (i) a common atom–atom potential of the form given in Section 2.2.2, using atomic charges calculated for each molecule; and (ii) the same model potential with these atomic charges replaced by atomic multipoles.

Assessing the overall results for such a large set of molecules requires a measure of how successful each lattice energy prediction has been. One measure for summarizing results is the relative energy  $\Delta E$ , defined as the calculated lattice energy difference between the prediction corresponding to the experimentally observed crystal structure and the lowest energy computer-generated, but unobserved, crystal structure. Therefore, when  $\Delta E < 0$ , the observed structure is more stable than any other computer-generated alternative whilst, when  $\Delta E > 0$ , its magnitude measures how far above the lowest energy predicted structure the observed crystal structure is found. As an example, we examine the results for one of the 50 molecules: oxazolidine-2,5-dione. One of the computer-generated crystal structures is a very good match to the structure reported in the CSD (refcode OXAZDO) – an overlay of the X-ray determined structure and the lattice energy minimum (Figure 2.3) demonstrates the accuracy to which the crystal structure is predicted. With the atomic charge model of electrostatics, 11 computer-generated crystal structures are calculated to be more stable than the observed structure (Figure 2.4a) and the global minimum of the search is  $\Delta E = 4.2 \text{ kJ mol}^{-1}$  lower in energy than the true crystal structure. Most of these lower energy crystal structures

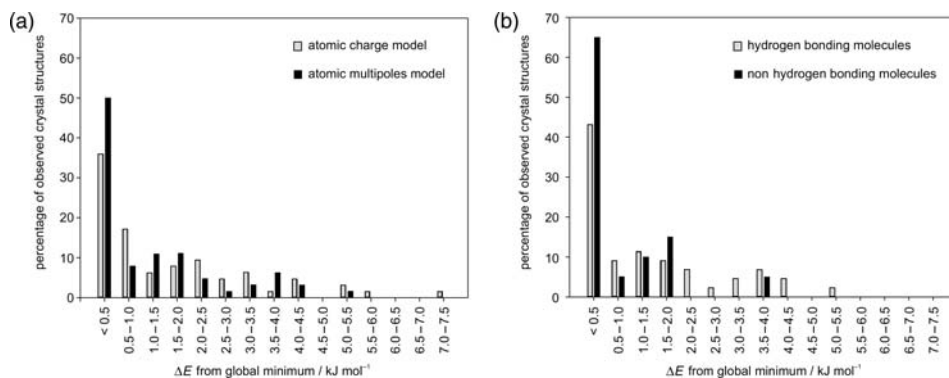


**Figure 2.3** Overlay of the X-ray determined crystal structure (black) [23] of oxazolidine-2,5-dione (CSD refcode OXAZDO) and the lowest energy predicted crystal structure after lattice energy minimization using an atomic multipole-based model potential (grey)



**Figure 2.4** Lattice energy versus density plots of the crystal structure prediction results for oxazolidine-2,5-dione, after lattice energy minimization using (a) atomic point charge and (b) atomic multipole electrostatic models. Predicted structures corresponding to the experimentally determined structure are circled. Data from [13, 21]

have different hydrogen bonding than the dimers in the observed structure, with two types of hydrogen bond chains prevalent in the lowest energy predictions (Figure 2.4). From these calculations, it seems that the observed crystal structure is thermodynamically metastable and that lattice energy minimization is unable to predict the correct structure. However, when the lattice energies are calculated using the atomic multipoles electrostatic model, there is a striking rearrangement of the relative stabilities (Figure 2.4b). Many more dimer-based structures appear in the low-energy region and the chain structures are disfavoured, especially the N-H...O<sub>2</sub>=C chains, which were present in the three lowest energy structures from the atomic charge calculations. Most importantly, upon re-ranking, the observed crystal structure appears as the global minimum, 1.2 kJ mol<sup>-1</sup> more stable than the next lowest energy structure ( $\Delta E = -1.2$  kJ mol<sup>-1</sup>): relative energies calculated using the more elaborate description of electrostatics suggest

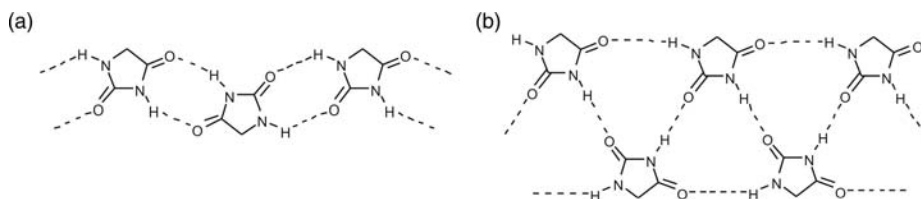


**Figure 2.5** Distributions of the energy difference,  $\Delta E$ , between observed and global minimum energy crystal structures for (a) atomic point charge and atomic multipoles models and (b) hydrogen bonding and nonhydrogen bonding molecules (using the atomic multipoles-based model potential). Adapted from [21] with permission. Copyright 2005 American Chemical Society

that the observed crystal structure is, in fact, the thermodynamically favoured product of crystallization.

The influence of changes in the electrostatic model is not always so pronounced and, for some of the molecules in the test set, the ranking of the ‘true’ crystal structure in the list of predictions is worsened when charges are replaced by atomic multipoles. It is the overall influence on the predictions across the entire set of molecules that reveals how important the differences are for the reliability of crystal structure prediction. The distributions of  $\Delta E$  for the 64 crystal structures with both energy models (Figure 2.5a) demonstrate a significant shift of the distribution towards lower  $\Delta E$  with the atomic multipoles model. With the multipoles model of electrostatics, 50% of the experimentally observed crystal structures are located either at the global minimum (33% as the global minimum itself), or within  $0.5 \text{ kJ mol}^{-1}$ , an increase from 36% with the atomic charge model. Furthermore, with the multipoles model the observed crystal structure is never higher than  $5.1 \text{ kJ mol}^{-1}$  from the global minimum. In terms of the absolute ranking, this translates to the true crystal structure being found amongst the best three computer-generated possibilities in 55% of cases [21]. If the list is increased from the best three to the best 10 structures, the observed crystal structure is present in 80% of cases. While the lattice energy minimization approach clearly does not always find the observed crystal structure as the lowest energy possibility, results of this quality are promising; given a small, rigid molecule, a short list of possible crystal structures can be produced, within which the observed crystal structure(s) should be found with a chosen level of confidence, e.g. a confidence level of about 55% for a short-list of three predicted crystal structures. This is reasonably successful, given that no account has been taken of growth conditions, and that the thermodynamic model is still fairly crude.

The improvement in results with more accurate electrostatic models can also be seen by studying the blind test results in more detail. For example, the observed crystal structure of molecule VIII contains molecular tapes based on  $R_2^2(8)$  hydrogen bond dimers

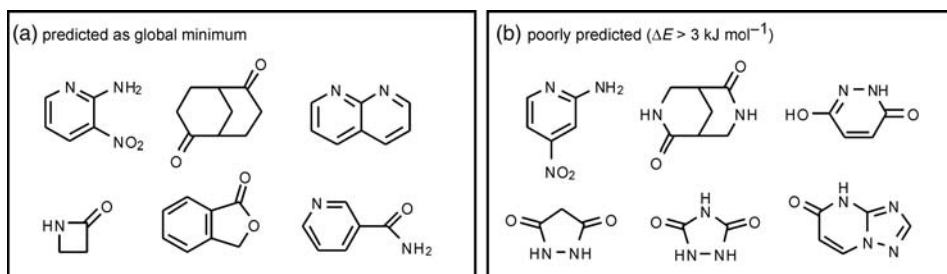


**Figure 2.6** Hydrogen bond patterns in the predicted crystal structures of blind test molecule VIII [10]

(Figure 2.6a). Eleven of 14 predictions by participants using detailed descriptions of electrostatics (i.e. atomic multipoles, off-nuclear charge sites or Gavezzotti's PIXEL method [24]) contained this correct hydrogen bonding. However, the nine participants using atomic charge models overwhelmingly favoured alternative hydrogen bonding schemes: 14 of 27 predictions (3 predictions  $\times$  9 participants) contained alternative tapes (Figure 2.6b), only five had the  $R_2^2(8)$  motif (Figure 2.6a) and eight contained other hydrogen bond patterns (structures were taken from the supplementary information of reference [10]). There does seem to be convincing evidence that, for polar or hydrogen bonding molecules, the description of electrostatic interactions must be more detailed than the traditional atomic partial charges, especially in distinguishing between competing hydrogen bond motifs.

### 2.3.2.2 The Influence of Hydrogen Bonding on Crystal Structure Predictability

Another question to be asked of the study of 50 molecules is whether certain molecular properties lead to highly predictable or poorly predicted crystal structures based on the lattice energy minimization approach. Some of the 50 molecules that were predicted successfully (as the global minimum) and relatively unsuccessfully (at  $\Delta E > 3 \text{ kJ mol}^{-1}$ ) are shown in Figure 2.7. The results of this study suggest that the most important differences are between hydrogen bonding and nonhydrogen bonding molecules (Figure 2.5b). The observed crystal structures of molecules *without* the capacity to form strong hydrogen bonds are almost always found as, or very close to, the global minimum in lattice energy. By contrast, there is an important proportion of crystal structures of molecules that *do*



**Figure 2.7** Examples of molecules whose known crystal structures (a) were found as global minimum in lattice energy and (b) were poorly predicted in terms of lattice energy ( $\Delta E > 3 \text{ kJ mol}^{-1}$ ), based on energy rankings using the atomic multipoles electrostatic model

hydrogen bond that are found at high relative energies. Using the atomic multipoles electrostatics model, 95% of the observed crystal structures of nonhydrogen bonding molecules are found within  $2\text{ kJ mol}^{-1}$  of the global minimum, while only 73% of observed crystal structures of hydrogen bonding molecules are located within the same energy range (Figure 2.5b). Therefore, almost all experimentally observed crystal structures that are farther than  $2\text{ kJ mol}^{-1}$  from the global minimum on their crystal packing energy surface are crystal structures of hydrogen bonding molecules [13].

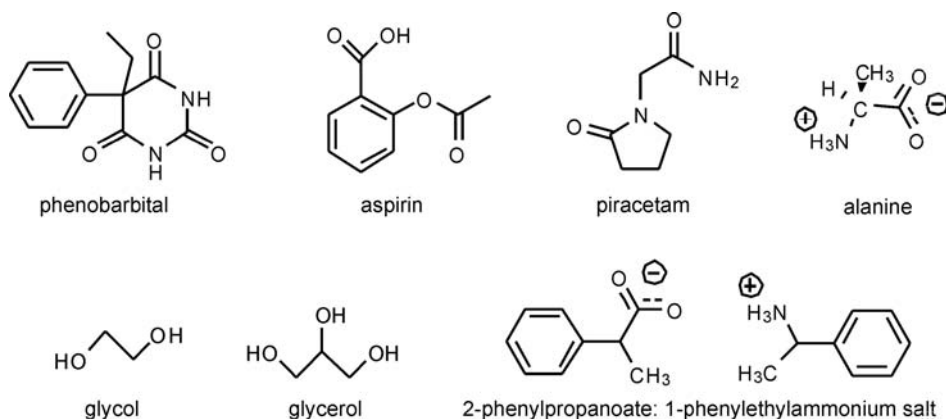
The differences between hydrogen bonding and nonhydrogen bonding molecules are less significant with the more elaborate electrostatic model than with the simpler model. This suggests that some of the differences between the two classes of molecules result from difficulties in accurately modelling hydrogen bond interactions to get the relative energies of competing hydrogen bond motifs correct. However, the differences are large enough, even with the accurate electrostatics, to also suggest that hydrogen bonding molecules are more often kinetically trapped in metastable crystal structures than molecules with no hydrogen bonding capability. This can be rationalized by the premise that strong directional interactions, such as hydrogen bonds, lead to high energy barriers between structures. Therefore, preferred molecular arrangements in solution and the early stages of nucleation are likely to persist in the final crystal structure, even if they do not lead to the thermodynamic global minimum. Molecules that only interact via weak interactions, however, are relatively free to explore packing possibilities in search of the deepest minimum and, according to this study, appear to be very efficient at finding the global minimum energy crystal structure.

### 2.3.3 Flexible Molecules

While much of the development of methods has necessarily focussed on the simplest small, rigid molecules, the focus of recent crystal structure prediction studies has noticeably shifted to more complex molecular systems. This shift in the nature of targets for structure prediction somewhat reflects the feeling that, for small rigid molecules, current methods are already useful and, while they still require further refinement, there is a pressing need to extend the applicability of what has been learnt from rigid molecules to more general, flexible molecules.

Many of the developments for conformationally flexible molecules have been very recent. Therefore, outside of the blind tests, there has not been any systematic evaluation of methods on the same scale as has been done for rigid molecules. However, several recent investigations of individual systems have demonstrated that successful predictions are now possible for larger, more flexible molecules. A few examples are (Figure 2.8): phenobarbital, whose thermodynamically stable polymorph was found as the global minimum in lattice energy [25]; aspirin, for which the calculations anticipated a new polymorph [26] which was later discovered during attempted co-crystallization [27]; piracetam, for which the structure of a new polymorph was predicted under informal blind test conditions [28]; the L- and DL-alanine crystal structures were located as global minima amongst chiral and racemic space groups, respectively [29]; the known structures of glycol and glycerol were ranked amongst the lowest energy calculated structures [30], and the known n- and p-salt structures of 2-phenylpropanoate:1-phenylethylammonium were found amongst the low energy computer-generated structures, the p-salt as the





**Figure 2.8** Examples of more flexible molecules recently studied by crystal structure prediction methods

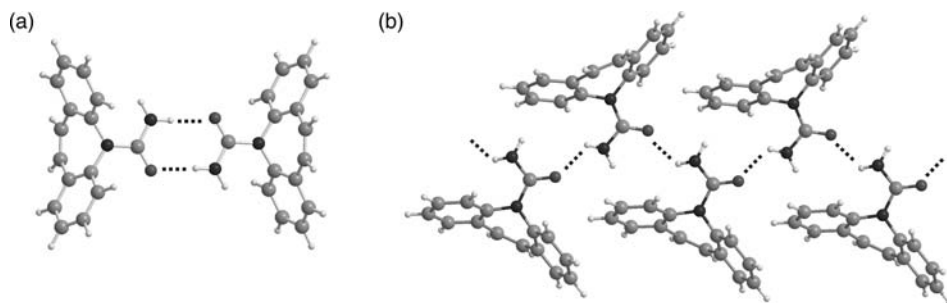
global minimum [31]. In all of these studies, the accurate models of intermolecular interactions have been combined with quantum mechanically calculated conformational energies, sometimes with Cambridge Structural Database guidance on the choice of molecular conformations. These reports demonstrate promising progress in the flexibility and size of molecules that can be confidently studied by computational methods.

## 2.4 Exploration of Crystal Forms. A Case Study: Carbamazepine

The results presented in the preceding section demonstrate that, while a lattice energy search will not always locate the observed crystal structure(s) of a molecule as the best ranked structure(s), a set of low-energy crystal structures can be produced, with a high likelihood that experimentally realizable crystal structures will be found within a short list of the best ranked possibilities. While this does not represent perfectly reliable prediction, this level of results should already be useful for exploring the likely crystallization behaviour of small organic molecules.

One of the main industrial interests in crystal structure prediction is the pharmaceutical industry, due to the potential consequences of unanticipated polymorphism of a drug molecule. The appearance of a new polymorph during manufacture could have serious consequences, due to differences in physical properties between polymorphs. One well-documented case was the appearance, after its launch onto the market, of a second polymorph of the protease inhibitor ritonavir [32, 33]. The scientists who dealt with the impact of the new polymorph concluded that ‘it is highly advisable to put enough resources to carry on exhaustive research to identify the most stable and all possible polymorphs.’ [32] High-throughput screening of crystallization conditions is one approach to finding all possible solid forms [34–36]. However, polymorphs can stay hidden, even when crystallized on large scales and under wide-ranging conditions (e.g. 1,3,5-trinitrobenzene [37] and maleic acid [38]). Crystal structure prediction calculations could be a valuable complement to experimental polymorph screening methodologies,





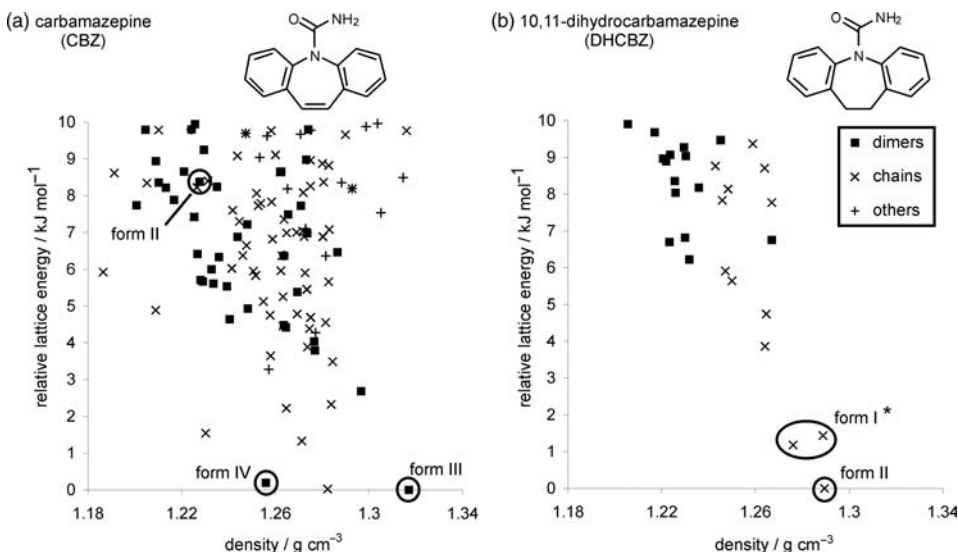
**Figure 2.9** Hydrogen bond (a) dimers and (b) chains of carbamazepine.  $N-H \cdots O$  hydrogen bonds are indicated as dotted lines

both in identifying all possible polymorphic forms for a given molecule and in judging their relative stabilities – an examination and evaluation of all potential crystal structures of a given compound is a step towards an understanding of its polymorphism.

Because of the importance of knowing all polymorphic forms during pharmaceutical solid form development, this application has been a driving force for making the necessary developments to deal with increasingly complex molecules. Furthermore, computational studies are not limited to searching for polymorphs of a molecule, but can be used to investigate the formation of co-crystals, salt forms and solvates without need for any material, opening the possibility of gaining a comprehensive picture of solid form diversity, perhaps even prior to synthesis of the molecule. One widely studied pharmaceutical molecule is the anticonvulsant carbamazepine (CBZ), which has attracted interest because of its propensity for polymorphism [39], as well as solvate and co-crystal formation [40]. A series of recent computational studies of this molecule and its derivative 10,11-dihydrocarbamazepine (DHCBZ) [41–44] (Figure 2.9) will be used to illustrate computational capabilities and the insight that can be gained from computational methods in crystal structure modelling and prediction.

#### 2.4.1 Polymorphism and the Influence of Small Molecular Changes on Packing

A reason for the interest in CBZ is its extensive polymorphism. To date, four polymorphs are known, all comprising the same  $R_2^2(8)$  dimer hydrogen bonding (Figure 2.9a), with the differences being in the packing of the dimer units. A computational study was performed to assess the predictability of the polymorphs of CBZ and assess the likelihood of further polymorphs. Crystal structures were generated in 10 of the most common space groups with  $Z' = 1$  and these were lattice energy minimized using an atomic multipole-based description of electrostatic interactions, also taking into account molecular flexibility around the amide nitrogen [41]. Three of the four polymorphs (forms I, II and IV) are in the list of low energy structures from this search (Figure 2.10a) – the fourth, form I, is of unusually low symmetry ( $Z' = 4$ ) which would generally not be predicted. The global minimum energy structure from the search corresponds to form III, the thermodynamically stable polymorph at ambient conditions, and two other structures are found within  $0.2 \text{ kJ mol}^{-1}$ : form IV and an experimentally unobserved crystal structure containing hydrogen bond chains (Figure 2.9b). Forms III and IV clearly would



**Figure 2.10** Lattice energy versus density plots of the low energy predicted crystal structures of (a) carbamazepine and (b) 10,11-dihydrocarbamazepine. Observed crystal structures are circled. \*Two structures are labelled DHCBZ form I: here, two lattice energy minima are structurally very similar to the reported crystal structure and these probably converge to a single free energy minimum. Adapted from [40] with permission. Copyright 2003 American Chemical Society

have been predictable by lattice energy. However, form II was calculated to be higher in energy than many hypothetical structures (more than 8 kJ mol<sup>-1</sup> above the other two polymorphs).

One observation is that no lower energy polymorphs than form III seem possible; this is an important conclusion for a pharmaceutical molecule because it rules out phase transformations to a new, more stable structure, at least under ambient conditions. However, the calculations raised two questions: (i) why does form II crystallize when it is apparently so high in energy relative to many alternative crystal structures; and (ii) does the crystal structure based on hydrogen bond chains, calculated to be nearly equi-energetic with forms III and IV, represent an experimentally obtainable polymorph? The existence of possible hydrogen bond chain-based polymorphs of carbamazepine was the motivation for an experimental high-throughput screen [35]. However, no new polymorphs were found and the dimer hydrogen bond motif appears to be robust. The explanation could be that, while both possibilities for hydrogen bonding (dimers and chains) can lead to equi-energetic crystal structures, the dimers are preferred in solution, so favour dimer-based structures during nucleation. To back this up, further calculations showed that the hydrogen bonds in CBZ dimers are stronger than those in CBZ chains [41]; if the strongest hydrogen bonds exist in solution prior to nucleation, then a chain polymorph will only form if there is a strong thermodynamic driving force during crystal growth. Here, the best chain structure is just above the dimer-based polymorphs in energy, so there is no such driving force to formation.

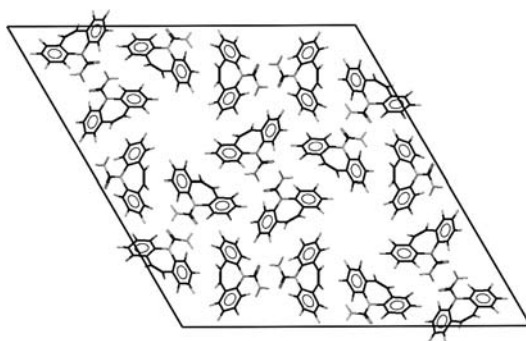
Interestingly, despite the apparently small difference in molecular structure (Figure 2.10) DHCBZ does crystallize to form hydrogen bond chains [45], rather than the dimers favoured by CBZ. Can calculations reproduce and help explain the differences in crystallization behaviour of the two related molecules? A lattice energy search for low energy crystal structures of DHCBZ does indeed agree with the experimental observation: all structures up to about  $6\text{ kJ mol}^{-1}$  above the global minimum contained chains rather than dimers (Figure 2.10b) [43]. Here, even if dimers exist in solution, there is a substantial driving force towards chains in the final crystal structure. The crystal structure prediction calculations for these two molecules provide an opportunity to analyse how these molecules would pack if they adopted a different hydrogen bonding in their crystal structures: the lowest energy possible crystal structures of DHCBZ based on a dimer motif and the best chain-based CBZ structures can be investigated to try to understand why they have not (as yet) been observed. An analysis of the packing in these hypothetical crystal structures gives insight as to why the two similar molecules crystallize with different hydrogen bonding. Here, a comparison of the intermolecular contacts in the observed and hypothetical crystal structures revealed that it is the subtle change in shape of the fused ring system that allows improved packing of hydrogen bond chains in DHCBZ, favouring their formation over dimers [43]. A complete understanding of the systems might not have been possible based solely on the experimentally determined crystal structures of both molecules.

A further observation was that the known crystal structure of DHCBZ (labelled form I in Figure 2.10b) did not correspond to the global minimum of the search – one lower energy structure was suggested by the calculations. In fact, a new polymorph of DHCBZ was reported while these calculations were being analyzed [46] and it is satisfying from the crystal structure prediction viewpoint to find that the structure of this new polymorph does correspond to the lowest energy structure from the predictions (form II in Figure 2.10b) [43].

## 2.4.2 Solvate and Co-crystal Formation

One of the conclusions of the carbamazepine study was that, while form II is present in the list of computer-generated crystal structures, it is far higher than the other polymorphs in lattice energy. The lattice energy difference between form II and the most stable crystal structures (including forms III and IV) was calculated to be  $8.4\text{ kJ mol}^{-1}$ . This is out of the range where we normally find observable crystal structures of small molecules (Section 2.3.2), so prompted a further investigation of the stability of this polymorph.

What is striking about the form II crystal structure is the presence of channels running along the *c*-axis (Figure 2.11), which contribute to its relatively low density. A possible explanation for the existence of form II is the inclusion of solvent molecules in these channels and a modelling study demonstrated that the structure could be stabilized by up to 3–4.5% by weight of toluene in the channels [44]. Indeed, subsequent studies of crystals grown from toluene confirmed between 2.7 and 4.2% by mass (2.7–3.7% by thermal gravimetric analysis, 4.2% by solution NMR) of solvent in the crystals [44]. While the presence of solvent in the channels might provide the crucial stabilization needed for the growth of form II, it is important that the structure was located amongst



**Figure 2.11** Carbamazepine form II, viewed down the *c*-axis

the structures proposed by crystal structure prediction of pure CBZ. This observation suggests that certain inclusion solvates can be anticipated by analysis of solvent-sized voids in computer-generated crystal structures of the *pure* compound [44], a computationally less expensive route than explicitly including solvent molecules from the start of the calculations.

In general, the prediction of multi-component crystal structures (co-crystals, solvates and salts) would require both molecules to be included in the calculations from the start; the search for possible crystal structures involves consideration of the relative positions and orientations of both component molecules, in addition to the usual set of variables. The increased complexity means that few co-crystal, salt or solvate systems have been subjected to crystal structure prediction calculations. To test current methods for multi-component crystal structures, and further explore the solid form diversity of CBZ, which is known to form numerous co-crystals and solvates [35, 40], lattice energy searches were performed for CBZ with acetic acid (AcOH) [42], one of the smaller molecules with which it is known to co-crystallize. The results of these calculations were pleasing – the known CBZ:AcOH crystal structure was not only located amongst the computer-generated crystal structures, but was found as the global minimum in calculated lattice energy. Following this promising result, calculations were performed for DHCBBZ with AcOH, where no crystal structure was known. Several observations were made from the computational results. First, the relative energies of pure DHCBBZ and the computer-generated two-component crystal structures suggested that a DHCBBZ:AcOH crystal could be more stable than the individual components and should form. Furthermore, unlike the pure crystals, where the slight changes in molecular shape caused dramatic differences in the low-energy possible crystal structures, there is striking similarity in the crystal packing possibilities for CBZ:AcOH and DHCBBZ:AcOH. In fact, the third most stable predicted DHCBBZ:AcOH crystal structure is isostructural with the known CBZ:AcOH structure. Following the calculations, DHCBBZ:AcOH crystals were grown, the structure was determined, and was found to be this third lowest energy predicted structure, isostructural with CBZ:AcOH. It is interesting that, despite contrasting hydrogen bond behaviour in their pure crystal structures, the presence of acetic acid promotes isostructurality. It is also noteworthy that the crystal structure prediction calculations suggested that such behaviour was likely. The study demonstrates

the utility of existing methods for exploring and predicting the crystal structures of multi-component crystal structures; such calculations should be similarly applicable to molecular salts.

### 2.4.3 Computational Solid Form Screening?

Although CBZ is a relatively small, rigid molecule in the world of pharmaceuticals, these studies of its polymorphism, solvate and co-crystal formation suggest that computational methods for crystal structure prediction are approaching the point where they can provide useful information in the solid form selection process of a drug molecule. Although there is still considerable progress required in dealing confidently with the larger, conformationally flexible molecules (i.e. more typical of drug molecules), one can imagine the design of an *in silico* solid form screen that could accompany and inform experimental screening for crystal forms. What would be involved in such a procedure and what information might it provide? At minimum, it would involve calculations to determine the lowest energy crystal structures of the pure drug molecule, ideally within a large set of the common space groups, including  $Z' = 2$ . This could provide a list of the lowest energy crystal structures of the molecule and, if there is already a known crystal structure for the molecule, the results might provide guidance as to whether it is the thermodynamically optimal form.

Recent forays into the prediction of multi-component crystal structures, such as the CBZ and DHCBZ:AcOH structures, as well as work on molecular salts [31, 47] and the co-crystal in the 2007 blind test, indicate that, while challenging, the prediction of such structures is possible. Therefore, should it be desirable to formulate a salt or co-crystal instead of the pure molecule, crystal structure prediction methods could be used to assess the low energy crystal structures of the pharmaceutical molecule with selected co-crystal or salt formers. The relative stabilities of the pure compound versus its putative salts or co-crystals could help choose the best candidates for experimental investigations. At the moment, the time required for such calculations would be longer than required for an experimental screen, but the advantage is that no material is required. Similarly, the use of computational methods for anticipating solvate formation could be incorporated: from calculated energies of putative solvate crystal structures, and the corresponding pure crystals of the compound, the occurrence of undesirable solvates might be anticipated. Such calculations could inform the crystallization process of what solvents to avoid for solution crystallization.

Computational methods could, therefore, potentially impact on polymorph screening, salt or co-crystal selection, as well as the avoidance of solvates. The outline of such a possibility should not be viewed as an over-optimistic assessment of current capabilities, but rather a goal towards which developments should aim. Given that progress is being made on flexible molecules and multi-component crystal structures, the methods that are necessary for such *in silico* screens are attainable, although it is difficult to predict the rate of progress and, therefore, when such calculations will be practical for the typical pharmaceutical molecule. Moreover, results will always need to be interpreted with care and a realistic view of the approximations and limitations of the methods. This is where continued assessment of methods on well-characterized systems is needed, to inform our level of confidence in the calculations.

## 2.5 Summary

The purpose of this chapter has been to present a perspective on computational crystal structure prediction as a maturing tool for understanding and predicting the likely crystal structures of organic molecules. The reliability of existing methods has been discussed and it is clear that (as was expressed in the conclusions on the publication of the second blind test) ‘crystal structure prediction, although beset by fundamental and technical difficulties, is no longer scandalously hopeless.’ [8] Further developments in the methods are clearly needed, to cope with simple molecules with greater reliability and to continue to extend their applicability to more complex systems. However, the computational methods have developed to a stage of providing a useful tool for exploring packing possibilities and understanding the crystallization behaviour of small organic molecules, with important applications in industry as well as much promise for the modern crystal engineer.

## Acknowledgements

I am grateful to Dr Aileen Day for helpful comments on the manuscript and to all colleagues who were involved in the work presented here, especially Aurora Cruz Cabeza for her thorough investigations of carbamazepine.

## References

1. F.H. Allen, *Acta Cryst.*, **B58**, 380–388 (2002).
2. A.I. Kitaigorodskii, *Organic Chemical Crystallography*, Consultants Bureau, New York, 1961.
3. A.I. Kitaigorodskii, *Acta Cryst.*, **18**, 585–590 (1965).
4. A.I. Kitaigorodskii, *Molecular Crystals and Molecules*. Physical Chemistry: A Series of Monographs, Vol. 29, Academic Press, New York and London, 1973.
5. T. Beyer, T. Lewis and S.L. Price, *CrystEngComm*, **3**, 178–212 (2001).
6. The search was performed on version 5.28 of the CSD, including January 2007 updates. The search criteria were: organic structures with no errors and no polymers. The search was furthermore restricted to crystal structures containing one type of molecule (RES = 1).
7. J.P.M. Lommerse, W.D.S. Motherwell, H.L. Ammon, J.D. Dunitz, A. Gavezzotti, D.W.M. Hofmann, F.J.J. Leusen, W.T.M. Mooij, S.L. Price, B. Schweizer, M.U. Schmidt, B.P. van Eijck, P. Verwer and D.E. Williams, *Acta Cryst.*, **B56**, 697–714 (2000).
8. W.D.S. Motherwell, H.L. Ammon, J.D. Dunitz, A. Dzyabchenko, P. Erk, A. Gavezzotti, D.W.M. Hofmann, F.J.J. Leusen, J.P.M. Lommerse, W.T.M. Mooij, S.L. Price, H. Scheraga, B. Schweizer, M.U. Schmidt, B.P. van Eijck, P. Verwer and D.E. Williams, *Acta Cryst.*, **B58**, 647–661 (2002).
9. P. Verwer and F.J.J. Leusen, *Rev. Comput. Chem.*, **12**, 327–365 (1998).



10. G.M. Day, W.D.S. Motherwell, H.L. Ammon, S.X.M. Boerrigter, R.G. Della Valle, E. Venuti, A. Dzyabchenko, J.D. Dunitz, B. Schweizer, B.P. van Eijck, P. Erk, J.C. Facelli, V.E. Bazterra, M.B. Ferraro, D.W.M. Hofmann, F.J.J. Leusen, C. Liang, C.C. Pantelides, P.G. Karamertzanis, S.L. Price, T.C. Lewis, H. Nowell, A. Torrisi, H.A. Scheraga, Y.A. Arnautova, M.U. Schmidt and P. Verwer, *Acta Cryst.*, **B61**, 511–527 (2005).
11. J.S. Chickos, C.M. Braton, D.G. Hesse and J.F. Liebman, *J. Org. Chem.*, **56**, 927–938 (1991).
12. W.T.M. Mooij, B.P. van Eijck, S.L. Price, P. Verwer and J. Kroon, *J. Comput. Chem.*, **19**, 459–474 (1998).
13. G.M. Day, J. Chisholm, N. Shan, W.D.S. Motherwell and W. Jones, *Cryst. Growth Des.*, **4**, 1327–1340 (2004).
14. A.J. Pertsin and A.I. Kitaigorodskii, *The Atom-Atom Potential Method. Applications to Organic Molecular Solids*. F.P. Schäfer and J.P. Toennies (Eds), Springer Series in Chemical Physics, Vol. 43, Springer-Verlag, Berlin, 1987.
15. A.I. Kitaigorodskii, *Tetrahedron*, **14**, 230–236 (1961).
16. A.J. Stone, *The Theory of Intermolecular Forces*. J.S. Rowlinson (Ed.), International Series of Monographs on Chemistry, Vol. 32, Oxford University Press, Oxford, 1996.
17. D.S. Coombes, S.L. Price, D.J. Willock and M. Leslie, *J. Phys. Chem.*, **100**, 7352–7360 (1996).
18. Basic Technology Programme: Control and Prediction of the Organic Solid State website. <http://www.chem.ucl.ac.uk/basictechorg/survey/survey.shtml>.
19. G.M. Day, T.G. Cooper, A.J. Cruz-Cabeza, K.E. Hejczyk, H.L. Ammon, S.X.M. Boerrigter, J.S. Tan, R.G.D. Valle, E. Venuti, J. Jose, S.R. Gadre, G.R. Desiraju, T.S. Thakur, B.P. van Eijck, J.C. Facelli, V.E. Bazterra, M.B. Ferraro, D.W.M. Hofmann, M.A. Neumann, F.J.J. Leusen, J. Kendrick, S.L. Price, A.J. Misquitta, P.G. Karamertzanis, G.W.A. Welch, H.A. Scheraga, Y.A. Arnautova, M.U. Schmidt, J.v.d. Streek, A.K. Wolf and B. Schweizer, *Acta Cryst.*, **B65**, 107–125 (2009).
20. J.A.R.P. Sarma and G.R. Desiraju, *Cryst. Growth Des.*, **2**, 93–100 (2002).
21. G.M. Day, W.D.S. Motherwell and W. Jones, *Cryst. Growth Des.*, **5**, 1023–1033 (2005).
22. B.M. Rice and D.C. Sorescu, *J. Phys. Chem. B*, **108**, 17730–17739 (2004).
23. H. Kanazawa, Y. Matsuura, N. Tanaka, M. Kakudo, T. Komoto and T. Kawai, *Bull. Chem. Soc. Jpn*, **49**, 954–956 (1976).
24. A. Gavezzotti, *CrystEngComm*, **5**, 429–438 (2003).
25. G.M. Day, W.D.S. Motherwell and W. Jones, *Phys. Chem. Chem. Phys.*, **9**, 1693–1704 (2007).
26. C. Ouvrard and S.L. Price, *Cryst. Growth Des.*, **4**, 1119–1127 (2004).
27. P. Vishweshwar, J.A. McMahon, M. Oliveira, M.L. Peterson and M.J. Zaworotko, *J. Am. Chem. Soc.*, **127**, 16802–16803 (2005).
28. H. Nowell and S.L. Price, *Acta Cryst.*, **B61**, 558–568 (2005).
29. T.G. Cooper, W. Jones, W.D.S. Motherwell and G.M. Day, *CrystEngComm*, **9**, 595–602 (2007).
30. B.P. van Eijck, W.T.M. Mooij and J. Kroon, *J. Comput. Chem.*, **22**, 805–815 (2001).

31. P.G. Karamertzanis and S.L. Price, *J. Phys. Chem. B*, **109**, 17 134–17 150 (2005).
32. S.R. Chemburkar, J. Bauer, K. Deming, H. Spiwek, K. Patel, J. Morris, R. Henry, S. Spanton, W. Dziki, W. Porter, J. Quick, P. Bauer, J. Donaubauer, B.A. Narayanan, M. Soldani, D. Riley and K. McFarland, *Org. Proc. Res. Dev.*, **4**, 413–417 (2000).
33. J. Bauer, S. Spanton, R. Henry, J. Quick, W. Dziki, W. Porter and J. Morris, *Pharm. Res.*, **18**, 859–866 (2001).
34. S.L. Morissette, Ö. Almarsson, M.L. Peterson, J.F. Remenar, M.J. Read, A.V. Lemmo, S. Ellis, M.J. Cima and C.R. Gardner, *Adv. Drug Del. Rev.*, **56**, 275–300 (2004).
35. A.J. Florence, A. Johnston, S.L. Price, H. Nowell, A.R. Kennedy and N. Shankland, *J. Pharm. Sci.*, **95**, 1918–1930 (2006).
36. J.M. Miller, B.M. Collman, L.R. Greene, D.J.W. Grant and A.C. Blackburn, *Pharm. Dev. Tech.*, **10**, 291–297 (2005).
37. P.K. Thallapally, R.K.R. Jetti, A.K. Katz, H.L. Carrell, K. Singh, K. Lahiri, S. Kotha, R. Boese and G.R. Desiraju, *Angew. Chem. Int. Ed. Engl.*, **43**, 1149–1155 (2004).
38. G.M. Day, A.V. Trask, W.D.S. Motherwell and W. Jones, *Chem. Commun.*, 54–56 (2006).
39. S.J. Clark, M.D. Segall, C.J. Pickard, P.J. Hasnip, M.J. Probert, K. Refson and M.C. Payne, *Z. Kristallogr.*, **220**, 567–570 (2005).
40. S.G. Fleischman, S.S. Kuduva, J.A. McMahon, B. Moulton, R.D. Bailey Walsh, N. Rodriguez-Hornedo and M.J. Zaworotko, *Cryst. Growth Des.*, **3**, 909–919 (2003).
41. A.J. Cruz Cabeza, G.M. Day, W.D.S. Motherwell and W. Jones, *Cryst. Growth Des.*, **6**, 1858–1866 (2006).
42. A.J. Cruz Cabeza, G.M. Day, W.D.S. Motherwell and W. Jones, *J. Am. Chem. Soc.*, **128**, 14466–14467 (2006).
43. A.J. Cruz Cabeza, G.M. Day, W.D.S. Motherwell and W. Jones, *Cryst. Growth Des.*, **7**, 100–107 (2007).
44. A.J. Cruz Cabeza, G.M. Day, W.D.S. Motherwell and W. Jones, *Chem. Comm.*, 1600–1602 (2007).
45. G. Bandoli, M. Nicolini, A. Ongaro, G. Volpe and A. Rubello, *J. Cryst. Spectrosc. Res.*, **22**, 177–183 (1992).
46. W.T.A. Harrison, H.S. Yathirajan and H.G. Anilkumar, *Acta Cryst.*, **C62**, o240–o242 (2006).
47. F.J.J. Leusen, *Cryst. Growth Des.*, **3**, 189–192 (2003).



# 3

## Multi-component Pharmaceutical Crystalline Phases: Engineering for Performance

*Matthew L. Peterson, Edwin A. Collier, Magali B. Hickey, Hector Guzman and Örn Almarsson*

### 3.1 Introduction

Pharmaceuticals continue to be an essential part of health care and everyday life. In this first decade of the 21st century, the conversation about health care and pharmaceuticals revolves around value to patients. In short, there is an ever growing need to:

- (i) develop novel medicines with improved performance to satisfy important unmet medical needs in cancer, immunology, infectious diseases, etc.;
- (ii) manage the cost of drugs and overall health care in the face of increasing demand by a generally ageing population;
- (iii) expand access for patients who cannot afford even the most inexpensive drugs; and
- (iv) enable cost-effective delivery of life-saving medications (including vaccines and other products possessing limited thermal stability) in emerging economies and ensure access in the developing world.

Meanwhile, the pharmaceutical industry's productivity has been losing ground, despite continued investments [1].\* On a positive note, significant impact is emerging

\* Tufts CSDD Impact Report Sept./Oct. 2006.



**Scheme 3.1** *Three-step process encompassing API discovery and development*

from multi-disciplinary efforts in pharmaceutical materials science, a discipline which has become increasingly prominent within recent years [2]. The idea of transforming molecules to useful medicines via materials science is gaining considerable interest as one means among many to tackle the productivity challenge and to increase quality in the pharmaceutical arena. This chapter will deal with questions regarding the performance of drug products and how to approach the engineering of product attributes, as illustrated by current themes and examples. A clear focus is placed on the engineering of multi-component crystalline phases of active pharmaceutical ingredients (APIs), which frequently form the basis of the dosage forms used by patients.

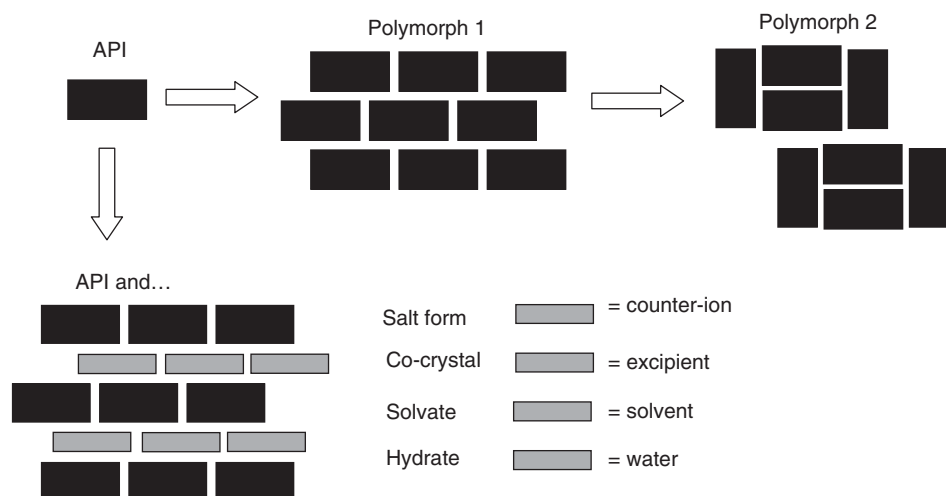
In the drive to optimize performance of an API and dosage forms derived from it, physical chemistry and engineering principles can be applied to connect the discovery of a molecular structure (the active compound) to the study of the new medicine (effective formulation), as shown in Scheme 3.1.

As indicated in Scheme 3.1, the crystallization of an API into an optimal solid form can be viewed as an essential part of the drug development process. The identification of a crystalline form of a compound allows manufacturers to produce bulk supplies in high purity and possessing stability consistent with long-term storage requirements. Crystalline materials are usually significantly more resistant to chemical degradation than amorphous (noncrystalline) forms. Although amorphous forms occasionally find use, the properties of the available crystal forms of a given substance are typically studied extensively to find one with optimal physicochemical properties to achieve the delivery objective. However, the use of structural information to guide development through predictive evaluation of the materials properties of the crystalline forms being used is still very much in its infancy.

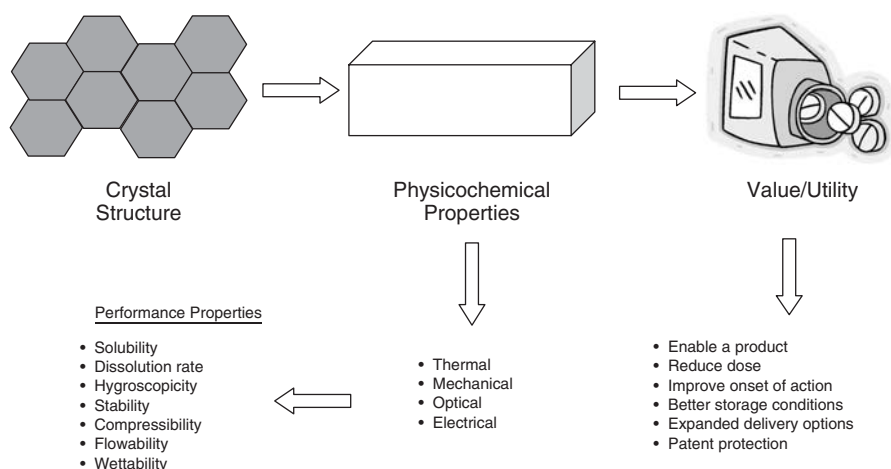
### 3.2 Exploring Crystal Form Diversity

The development of an API into a viable dosage form requires a vast amount of information to be gathered at the pre-formulation stage. Gathering knowledge of the crystal form diversity exhibited and evaluation of physicochemical profiles of thermodynamically stable forms are activities that facilitate form selection for development of a product. Crystal form diversity of the API is often thought of as its propensity to exhibit polymorphism, but it can also be described further by multi-component phases such as salt forms, solvates, hydrates and co-crystals, all of which may, too, be polymorphic (see Figure 3.1) [3–5].

Ultimately the various crystal form types give rise to solids that can exhibit markedly different ‘performance properties’, as evident by changes in solubility and dissolution or differences in chemical and physical stability. The solubility, dissolution rate and overall bioavailability of a drug will be greatly influenced by the crystal packing adopted and the surface chemistry this exposes. The crystal form can also affect the hygroscopicity



**Figure 3.1** Schematic of arrangements of API molecules to form both polymorph and multi-component crystalline forms



**Figure 3.2** The relevance of crystal structure and corresponding physicochemical properties to overall API value and utility

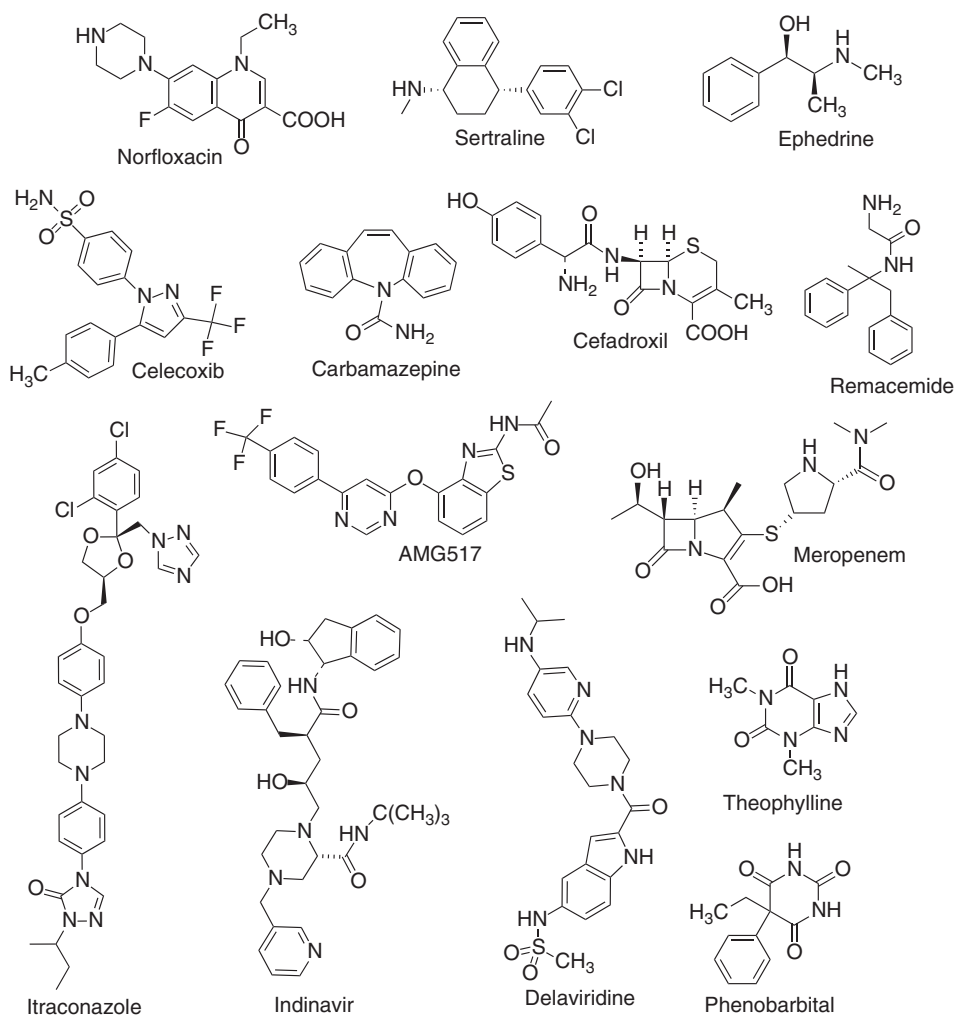
(capacity to adsorb moisture from the air) and stability profile of the drug when exposed to ambient conditions. Selection of the crystal form based on these properties may improve onset of action, lower the overall dose amount required or enhance long-term storage stability [6]. In practice, a balance must usually be struck between those properties deemed most critical for the dosage form required whether it is for oral, intravenous, inhalation or transdermal application, for example (Figure 3.2).

Because the primary focus in early development is often on solubility and dissolution rate along with chemical and physical stability, the examples chosen emphasize systems

where such properties are impacted by the choice of form. The material properties related to handling, flow, compression, etc. are often considered secondary to the stability and solubility properties and can often be optimized using external excipients. The properties of interest that affect manufacturing and performance of solid dosage forms are listed in Table 3.1. The properties are presented here as they relate to specific model API

**Table 3.1** *API system properties – problems and solutions*

Critical property	System	Problem	Solutions/identification of new form
Solubility	Indinavir	Low solubility – limiting bioavailability	Indinavir sulfate ethanolate: salt with acceptable performance characteristics
	Norfloxacin	Formation of low solubility salts with alkali metal ions	Norfloxacin succinate: salt with improved solubility
	Celecoxib	Low solubility, incomplete oral bioavailability, slow onset of action	Celecoxib sodium and celecoxib nicotinamide: salt or co-crystal plus formulation with improved dissolution profiles
	AMG517	Low solubility, incomplete oral bioavailability	AMG517 sorbic acid co-crystal with improved dissolution profile
Chemical stability	$\beta$ -Lactams	Chemical stability – hydrolysis of the $\beta$ -lactam ring by water	Crystallization – hydrates of $\beta$ -lactams are often stable requiring only limited environmental controls on storage and handling
Form diversity	Carbamazepine	Polymorph/solvate formation leading to low solubility form	Pharmaceutical salts or co-crystals that are less prone to form conversion
Physical interactions in dosage unit	Theophylline: phenobarbital	Slower dissolution from combination	Identification of a co-crystal phase with slower dissolution than pure phase. Avoidance of co-crystal formation during manufacture and storage
	Delaviridine mesylate	Storage in humid environments results in ion exchange with croscarmellose sodium	Package and store in humidity controlled environment



**Figure 3.3** Molecular structures of the active pharmaceuticals used as examples

systems (Figure 3.3), each exhibiting its own individual set of problems that need to be overcome via discovery of an optimal solid form.

A key issue in current drug development is the poor water-solubility of an API and the corresponding lack of dissolution into aqueous media.<sup>†</sup> This deficiency is occasionally a major limitation for injection formulations and oral delivery. The chemical structure of

<sup>†</sup> It is not uncommon for metastable polymorphs of a compound to be obtained from initial crystallization experiments. Sometimes these metastable forms are used in initial animal pharmacokinetics experiments with good results. After subsequently identifying and testing the stable form it is sometimes found that the solubility is not adequate to achieve an efficacious dose by oral administration of the compound. For a review of the differences in the solubilities of some pharmaceuticals, see [7].

a new active compound is conceived for desired biological properties, and is often based on a structure–activity relationship (SAR) that is developed for a given biological target. Given the affinity and selectivity of an active molecule to its target, there is a reasonable chance that aqueous solubility may be very low. Though such intuition is often borne out, the fundamentals of the situation are often only realized after a candidate has already been selected. The thermodynamic definition of solubility is the activity of the compound in solution in equilibrium with a crystalline state. However, in an attempt to remedy poor solubility, one may engender new problems that must be addressed. For example, in the development of the Pfizer drug setraline (Zoloft®), a selective serotonin reuptake inhibitor used in the treatment of depression, the HCl salt of the API was selected for development. Incidentally, HCl is among the most commonly used salt formers in the industry [8]. Unfortunately, along with the increased solubility offered by the HCl salt came the heightened propensity for crystal form diversity, including polymorphism and solvate formation [9, 10]. Subsequent salt screens on the free base form revealed that changing the counter ion to either HBr or benzoic acid increased aqueous solubility without the added complication of a large number of crystal forms.

Many criteria are in play when making a decision on a form for drug development, and some may be weighted more heavily in considering the options for a particular product. The advent of rapid form screening technology enables the discovery of more stable crystalline forms earlier in the development process. In some cases several forms are found with similar first line performance (solubility, dissolution, stability, hygroscopicity) properties. Additional selection criteria can then be applied, focusing on properties that will best allow for the desired product image to be met.

### 3.3 High-throughput Experimentation

Classically, bench top solvent-mediated experiments have been the method of choice within the pharmaceutical industry to explore potential crystalline polymorphs and/or salts of an API on an iterative basis. Due to the legal ramifications and secrecy of development filings<sup>‡</sup> that exist when publishing on pharmaceutical compounds, limited literature evidence can be found describing these lower throughput screens and their resulting datasets. Therefore, indication of the exact rationale behind why specific forms were selected or discounted for development is often unavailable. Also, for the crystal engineer aiming to understand potential structure–property relationships within these datasets [11], crystal structures and corresponding analytical data for polymorphs or indeed, multiple salts or co-crystals can be found for no more than a handful of APIs found. Although exploration of the Cambridge Structural Database [12] allows many useful trends to be drawn between similar structures or indeed, those with specific interactions, most researchers combine this database analysis with the actual laboratory work required to generate their own systems of interest such as those for both Remacemide [13] and Ephedrine salts [14]. Along with solution experiments, alternative methods to explore form diversity include both co-crystal and/or salt formation

---

<sup>‡</sup> FOI NDA reviews are publically available. However, while safety and clinical data are completely untouched, often form and formulation data (CMC) are redacted.

via the solid-state ‘dry’ or ‘solvent-drop’ grinding approach [15–17]. These methods have shown great potential to complement and, in some cases, replace the need for solution-based crystallizations to be performed when searching for novel pharmaceutical complexes [18].

Today the application of high-throughput (HT) experimentation to aid in the exploration of crystal form diversity has become far more commonplace within pharmaceutical research and development [2]. This technology has been pioneered to complement the aforementioned laboratory bench work approach. This offers the pre-formulation scientist a markedly increased choice in the discovery and characterization of multiple crystal forms of an API that may be taken forward into the drug development arena. The primary advantage offered by HT experimentation lies in the ability to explore not just multiple composition types, but also the large number of variables and processes that govern the overall crystallization mechanism. From a processing standpoint, these may include first performing true recrystallization experiments via thermal cycling to achieve supersaturation, followed by a more classical evaporative method or indeed an anti-solvent addition where required. In parallel, slurry conversions can be conducted in which the most stable crystal form is often discovered [19]. In order to obtain maximum benefit from the technology available, HT experiment design must build on a thorough understanding of the fundamental science and an appreciation of the individual API’s characteristics [5]. In addition, the large number of ‘hits’ often gathered within the screen dictates that adequate characterization techniques are in place offering a throughput level similar in magnitude to the actual screen. This helps to avoid subsequent analysis bottlenecks and hence, achieve maximum benefit and understanding from the larger scale experiments [20].

From a crystal engineering perspective, the design of molecular pharmaceutical solids with useful physicochemical properties encompasses an appreciation of the underlying intermolecular interactions, along with how these interactions may arise or indeed, be ‘promoted’ within the crystallizing medium. It has long been possible to postulate the interaction of solvent or additive molecules in the promotion of specific crystal morphologies and ultimately, novel polymorphs [21–24]. The HT approach further opens the experimental space with respect to composition types, allowing almost limitless solvent combinations using relatively small quantities of precious API compound in the process. It is in the very nature of HT that often an unexpected positive result may be obtained quite serendipitously, owing mainly to the relative ease in which vast numbers of experiments can be designed and a multitude of variables addressed. Furthermore, through 3-D molecular modelling and the concept of supramolecular synthons, or more simply through consideration of acid–base chemistry and corresponding  $\Delta pK_a$ , one can conceivably predict the formation of a salt or co-crystal of an API. For example, in a salt or co-crystal screen where an acid–amide dimer is targeted, HT processes enable multiple acids and/or bases to be used with varying stoichiometries, again in many different unary, binary or ternary and above solvent systems. However, unlike salt formation where there are also corresponding electrostatic forces at play helping to drive the process, the success of co-crystal screening from solution often requires the individual components to possess relatively similar activities in the solvent system used. Finding a novel co-crystal is potentially much more likely when one is able to explore many different conditions in parallel. Table 3.2 delineates the HT workflow in terms of starting

**Table 3.2** Relative composition and processing variables, along with high-throughput (HT) *in situ* and low-throughput (LT) *ex situ* analysis techniques available

Composition type		Process variables				Analysis methods	
Polymorphs and solvates	Salts and co-crystals	Thermal	Evaporation	Anti-solvent	Slurry conversion	Other variables	
Solvent type	Counter-ion type	Heating rate	Rate of evaporation	Anti-solvent type	Solvent type	Mixing rate	DSC
Degree of supersaturation	Stoichiometry	Cooling rate	Surface-volume ratio	Rate of anti-solvent addition	Incubation temperature	Impeller design	TGA
	Ionic strength	Maximum temperature			Incubation time	Crystallization vessel design (including capillaries, etc.)	HPLC
Additive type	$\Delta pK_a$	Incubation temperature(s)	Carrier gas	Temperature of anti-solvent addition	Thermal cycling and gradients		DVS
Additive concentration	pH					FTIR	Single crystal XRD
	Solvent type	Incubation time		Time of anti-solvent addition			
	Degree of supersaturation						



compositions, processing conditions and subsequent experiment by experiment analyses that can be performed.

Recent examples have highlighted the importance of HT experimentation to help promote a thorough understanding of multiple crystal forms that may exist, whether they are in fact actual polymorphs, salts or solvates of the parent API [5, 10, 20]. However, it is well understood that even HT experimentation is not always adequate to uncover every possible form of a given compound. The timely gathering of this knowledge can ensure development proceeds with the most suitable form, whether that is the most stable polymorph of a base compound or optimal salt candidate. In addition, options for patent protection required for later filing can also be elucidated. In the sections that follow, we consider some of the properties of materials that affect development of products with the desired performance. While the concept outlined in Scheme 3.1 is operable, it is important to stress that ‘design’ of materials is still a retrospective activity. Property prediction from 2-D concomitancy and 3-D structures of molecules can at best help guide experimental determination of performance. Hence, we need to generate materials for characterisation and base performance evaluation and refine predictive models on good experimental data.

### 3.4 Examples of ‘Form and Formulation’

#### 3.4.1 Indinavir sulfate ethanolate – a case of a salt hydrate/solvate

The HIV protease inhibitor Indinavir, marketed as Crixivan<sup>®</sup> by Merck, represents a remarkable case where an unusual crystal form enabled the development of a breakthrough pharmaceutical [25]. One may ask the question, is a salt solvate, such as a sulfate ethanolate of Indinavir, unusual? In searching the literature, besides finding numerous hydrates of pharmaceuticals, very few examples can be identified of actual solvates in approved products. In general, the solvent involved is ethanol due to its wide acceptance and use in pharmaceutical formulations. Another ethanol-solvated protease inhibitor besides Indinavir was recently approved; Darunavir ethanolate.<sup>§</sup> In addition, the antibiotic Doxycycline HCl also exists as an ethanol solvate.<sup>§</sup> The limited number of precedents indicates an aversion to take on the physical challenge of making and maintaining a solvate of an active compound. Indeed, solvates can be prone to solvent loss on standing to varying degrees. In several cases, solvent loss from the crystal form leads to undesirable changes (increasing disorder, loss of chemical stability, deliquescence in humid conditions, etc.). Solvates are therefore often regarded as intermediates rather than candidates for use in a final product.

Given that reliable performance is a critical feature of an antiviral compound, a key issue in the development of Indinavir was managing the problem of consistent oral bioavailability [25]. Indinavir sulfate ethanolate is a crystalline compound, which was selected for the drug product over the crystalline free base monohydrate form of Indinavir on the basis of superior oral pharmacokinetics in animals including humans. The base form, a material with less than 0.02 mg ml<sup>-1</sup> solubility in water at 22 °C, is only

---

<sup>§</sup> Physician’s Desk Reference, electronic library version, MicroMedex (Thompson Healthcare), 2003; The Merck Index, electronic version, 2001, Monograph number 3474.

**Table 3.3** Solubility values of Indinavir free base in aqueous buffer at 22 °C

pH	mg ml <sup>-1</sup>
6.9	0.019
5.9	0.037
3.4	61.2
Sulfate ethanolate (<3)	>500

**Table 3.4** Crystal form dependent oral bioavailability of 200 mg (base equivalent) from capsules in humans

Crystal form	C <sub>max</sub> (mM)	AUC (mM h)	T <sub>max</sub> (h)
Free base monohydrate	0.80 (69%)	1.04 (75%)	0.9 ± 0.5
Sulfate ethanolate	1.30 (28%)	1.83 (31%)	0.8 ± 0.3

Data from Yeh *et al.* [27]. C<sub>max</sub> refers to the maximum plasma concentration achieved, AUC is the calculated area under the plasma–time curve, and T<sub>max</sub> is the time at which C<sub>max</sub> occurs. The numbers in parentheses are the percentage variability on the means. The salt form has lower variability between subjects than the free base monohydrate.

well absorbed orally from acidic solutions at a pH where the protonated drug is soluble. In contrast, the intrinsically soluble acidic sulfate salt exhibits extraordinary solubility in water [26] (>500 mg ml<sup>-1</sup> at pH < 3; Table 3.3) and excellent oral pharmacokinetics in humans (Table 3.4) [27]. Importantly, the salt form achieves high plasma concentrations after oral administration from a solid dosage form. In other words, a pre-made solution dosage was not required to get consistent oral bioavailability and high levels of drug in the circulation. In the case of Indinavir, a solution product of the base form made by acidification was not feasible from patient acceptability and chemical stability considerations. The crystalline, highly soluble ethanol solvate of the sulfate salt allowed the preparation of a solid dosage form, which has room temperature stability for 2 years and high drug content (400 mg per capsule).

The physical property that created some challenge was the overt hygroscopicity of the salt at relative humidities >50% (25 °C). This means that on standing in humid conditions, the salt can absorb significant moisture from the air to become deliquescent and chemically unstable. The solvate is initially lost to form a transient hydrate, which then with further water yields an amorphous, hydrous deliquescent mass. A solution-mediated mechanism of degradation ensues, through intramolecular lactonization with loss of the *cis*-aminoindanol moiety, and the product loses its integrity over relatively short periods of time (days). Hence, the moisture-adsorbing property of Indinavir sulfate ethanolate in the pure drug substance, formulation processing and capsule needed to be managed to achieve success of the drug product.

The lesson from developing the drug product for Crixivan® was one of integration of medicinal chemistry design, detailed metabolism research, understanding of the oral absorption problem and application of sound pharmaceutical chemistry. The discovery of the sulfate salt ethanolate was enabling for an important product, and its success illustrates that a slightly unusual form with challenging but manageable physical properties can be developed if the determination is there. Bioavailability from an oral dosage form

was the key criterion for Crixivan – when this situation is encountered in pharmaceutical research and development, concessions regarding the physicochemical properties of a crystal form may need to be made.

### 3.4.2 Norfloxacin – polymorphs, solvates, salts, complexes and co-crystals

Norfloxacin is a poorly soluble zwitterionic fluoroquinolone oral antibacterial originally marketed as Noroxin® by Merck. The compound is noted for its use in treating common and complicated urinary tract infections, gonorrhoea, as well as acute and chronic prostatitis, an inflammation of the prostate gland. The zwitterionic form of norfloxacin is about 40% bioavailable when taken orally [28]. Higher solubility forms of the drug may allow for alternative routes of administration such as ocular solutions, or they could increase the bioavailability of the oral product, although the drivers for improving the current oral product are, at best, limited [29]. Norfloxacin is known to exist as a pair of enantiotropically related polymorphs of the free zwitterionic form [30, 31], a dihydrate [32], a hydrate ethanolate [33], metal ion complexes with di- and trivalent metal ions [28], organic salts [34–37] and as co-crystals [34]. In fact, one caveat for the use of norfloxacin and several other fluoroquinolone antibacterial agents is their affinity to form metal ion complexes with ions found in common antacids, supplements and dairy products [28, 38]. Invariably, for norfloxacin, these metal ion complexes have lower aqueous solubility than the uncomplexed compound. The decreased solubility translates into lower oral bioavailability in these cases.

Norfloxacin forms metal ion complexes with many di- and trivalent ions including  $\text{Ca}^{2+}$ ,  $\text{Mg}^{2+}$ ,  $\text{Zn}^{2+}$ ,  $\text{Fe}^{3+}$  and  $\text{Al}^{3+}$  [28]. Formation constants for these complexes have been measured using potentiometric titrations between pH 2 and 11 ( $\text{pK}_a$ s are 6.3 and 8.5 for the carboxylic acid and piperazinyl amine, respectively). Complexation occurs at the carbonyl-carboxylate site initially and subsequently at the piperazinyl amine as the pH increases, with these sites of interaction elucidated by single crystal structure determination [39–41]. At pH 6.5 the largest formation constant is found with  $\text{Al}^{3+}$  and decreases according to the following trend:  $\text{Al}^{3+} > \text{Fe}^{2+} \approx \text{Zn}^{2+} > \text{Mg}^{2+} > \text{Ca}^{2+}$ .

In an open, randomized, crossover study in mixed breed dogs,  $12 \text{ mg kg}^{-1}$  norfloxacin in 0.01 M HCl was given alone or co-administered with equimolar amounts of a metal chloride salt. Differences in the  $C_{\text{max}}$  and AUC were observed and found to follow the same rank order as the complex formation constants. A follow up study looking at the effect of five different doses of  $\text{Mg}^{2+}$  (0.25, 0.5, 1, 2 and 4:1  $\text{Mg}^{2+}$ :norfloxacin), from  $\text{MgCl}_2$ , showed that increasing the available  $\text{Mg}^{2+}$  resulted in a rank order decrease in the oral absorption of norfloxacin, most likely because more of the active compound is tied up as the metal ion complex.

In contrast to these studies where decreases in oral bioavailability are correlated with complex formation, increasing the solubility of norfloxacin using specific solubility enhancers has been shown to increase the oral bioavailability [29, 42, 43]. Both EDTA ( $\text{pK}_{a1}$  1.7) and sodium caprate (caprylic acid  $\text{pK}_a \sim 5$ ) were thought to act as chelating agents to increase the dissolution of norfloxacin into aqueous media [42]. The combinations of norfloxacin and either EDTA or sodium caprate were dosed as mixed powders in capsules. While both excipients increased the oral bioavailability of norfloxacin in rabbits, sodium caprate was more effective than EDTA (sodium

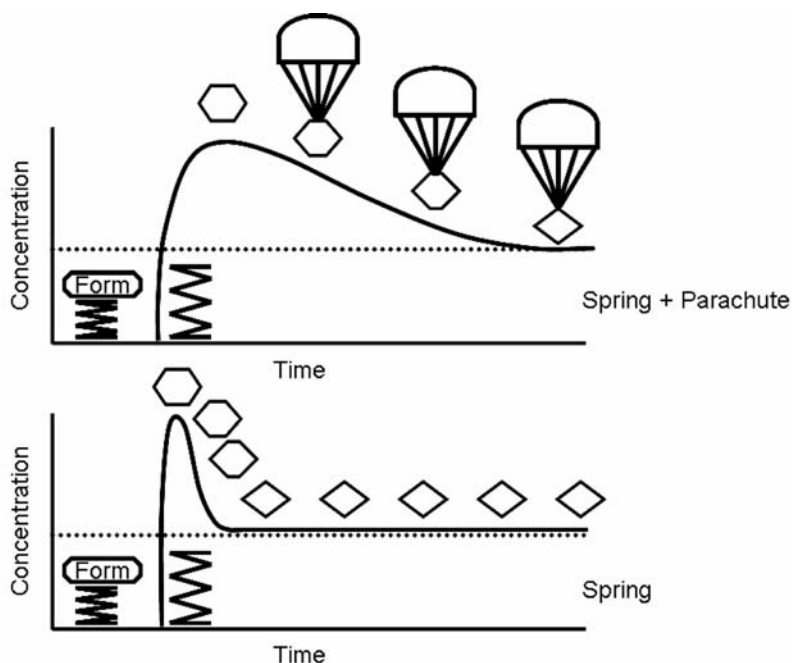
salt?) when five molar equivalents were used. However, with one mole equivalent the effect is reversed. Differences in the solubilizing ability of the two excipients and relatively low solubility of sodium caprate in gastric media were discussed as possible reasons for the change in increasing enhancement with increasing amounts of sodium caprate.

Recently, several organic salts and a co-crystal of norfloxacin were prepared. Salts were formed with ascorbic [35] ( $pK_a = 4.17$ ), lactic [36] ( $pK_a = 3.89$ ), succinic [34] ( $pK_{a1} = 4.20$ ), malonic ( $pK_{a1} = 2.9$ ) and maleic ( $pK_{a1} = 2.0$ ) acids. Considering the low  $pK_{a1}$  of EDTA (1.7), discussed above, it is possible that norfloxacin EDTA salts have also been prepared in solution. The succinate, malonate and maleate salts were determined to be 18–46 times more soluble than the free form of norfloxacin using a modified Higuchi–Connors method [34, 44]. While not tested *in vivo*, it is reasonable to speculate that the higher solubility of the salts could translate into higher bioavailability if rapid precipitation of other forms does not limit the bioavailability, especially in light of the positive effect seen with EDTA. In addition to the salts, an isonicotinamide co-crystal chloroform solvate has also been prepared and crystallographically characterized. Although not pharmaceutically acceptable as a chloroform solvate, the co-crystal represents the fifth type of crystalline form characterized for norfloxacin, following the free form, solvates, metal ion complexes and organic salts. The rich solid form diversity found with norfloxacin exemplified the breadth of possibilities associated with crystal form studies on pharmaceuticals. It also showcases how a simple polymorph of a compound can often perform well enough, even in cases where specific undesirable chemical interactions result in poor performing forms, to facilitate the more than adequate dosing of an important drug. In a world where even the smallest advantage can be marketed aggressively, it also should stand as an example of the extent to which innovator companies may be pushed to ensure they are not surprised by ‘better’ alternatives to their own drugs.

### 3.5 AMG517 and Celecoxib – ‘Spring and Parachute’ Approach

AMG517 is a vanilloid receptor 1 inhibitor that was being evaluated by Amgen, Inc. as a treatment for acute and chronic pain [45–47]. The compound had low aqueous solubility ( $0.02 \text{ mg ml}^{-1}$  at pH 6.8) and slow dissolution and a  $pK_a$  below 1 associated with the aminobenzothiozale group, making traditional salt formation challenging. One of the preclinical formulations was prepared in OraPlus®, an oral suspending vehicle that contains sorbic acid as a preservative. The formulation was a suspension and showed increased exposure, relative to previous suspensions, when dosed in rats. After further characterization of the suspension it was determined that the solids had converted into a new form, the AMG517 sorbic acid co-crystal.

*In vitro* dissolution testing of the sorbic acid co-crystal showed that this form supersaturated with respect to the equilibrium solubility of AMG517 in simulated gastric fluid (SGF) and simulated intestinal fluid (SIF). The kinetic solubility was such that when dosed as a suspension the exposures achieved using the co-crystal were nine times higher in terms of AUC(0-inf) and almost eight times higher in terms of  $C_{max}$  when compared with exposures from an equivalent dose of the free form [45]. This supersaturating effect



**Figure 3.4** 'Spring and parachute' approach to promote and maintain supersaturation of celecoxib in solution. A high energy form of the drug (the spring) provides the driving force to solubilize the drug at a concentration greater than its equilibrium solubility level. A combination of excipients (the parachute) inhibits and/or retards precipitation

of the co-crystal, often discussed in terms of kinetic solubility or apparent solubility, was significant enough ( $\sim 10$  times at 1 h), and precipitation of the free form slow enough that no further formulation was required to take advantage of the form. It was sufficient for the form to act as a spring generating relatively high concentration solutions *in vivo* for the increase in exposures to be realized (Figure 3.4).

Subsequent to this work, twenty-one other co-crystals with AMG517 were reported [46, 47]. All were formed using carboxylic acids or primary amides. The hydrogen bonding observed in the crystal structures was between the benzothiazole on AMG517 and either the carboxylic acid or primary amide on the co-crystal former. The stoichiometry of the co-crystals was either 1:1 or 2:1 where a single molecule of the co-crystal former interacted and two molecules of AMG517 comprise the co-crystal. Melting points were found to fall, for the most part, between that of the co-crystal former and AMG517 free base. The only exception to this was for the AMG517 succinic acid co-crystal. During *in vitro* dissolution testing of the co-crystals in SIF all of the co-crystals were found to supersaturate relative to the free form of AMG517.

Celecoxib is a COX-2 inhibitor marketed by Pfizer as Celebrex® to treat pain of osteoarthritis. The compound is poorly soluble in aqueous media. Although the compound is fairly well absorbed when administered orally, the poor solubility translates into modest bioavailability in animals [48]. Traditionally, salt forms of compounds with poor

aqueous solubility would be prepared and tested in order to determine if any advantage could be gained by increasing the solubility. The compound is capable of forming salts ( $pK_a = 11.1$  [48]) although the relatively high  $pK_a$  limits the selection of possible bases. Nevertheless, formulations containing sodium salts were prepared and were shown to increase the systemic exposure relative to a fixed dose and to improve the rate of uptake of the compound in dogs [49].

Two forms of the sodium salt were described, one hydrated form and one propylene glycol (PG) solvate trihydrate. Details of the structures have not yet been elaborated, but it is certain that both are 1:1 sodium salts with the deprotonation occurring at the sulfonamide group. The challenge in this case was not in preparing a salt form with improved kinetic solubility, rather it was in delaying the precipitation of the crystalline free acid under physiological conditions. Dissolution of the sodium salts in SGF ( $pH = 2$ ) [50] resulted in high initial concentrations. Not unexpectedly, there was rapid conversion to the free acid and therefore, a concurrent decrease in the concentration of celecoxib in solution with respect to the equilibrium concentration of the free acid. For the PG solvate it was demonstrated that the precipitation of the free acid could be delayed by about 30 min using a combination of the nonionic surfactant  $\alpha$ -tocopherol polyethylene glycol succinate (TPGS) or Poloxamer 407 (Pluronic F127, BASF) and hydroxypropylcellulose (HPC) [49].

The combination of a crystalline form with high kinetic solubility (spring) and excipients shown to delay the precipitation (parachute) of a related, lower solubility crystalline form leads to increased solution concentrations of the compound over a prolonged period of time, as shown schematically in Figure 3.4. Animal testing of celecoxib sodium, formulated with TPGS/HPC or F127/HPC, resulted in nearly complete absorption (100% oral bioavailability) of the compound over three different dosing levels, 2.5, 5.0 and 7.5 mg kg<sup>-1</sup>. In comparison, Celebrex<sup>®</sup> achieved 40% oral bioavailability and the free acid, celecoxib, formulated in TPGS/HPC achieved 30% oral bioavailability when dosed at 5 mg kg<sup>-1</sup>. The relative oral bioavailability of Celebrex<sup>®</sup> has been reported to decrease at higher doses.\*\*

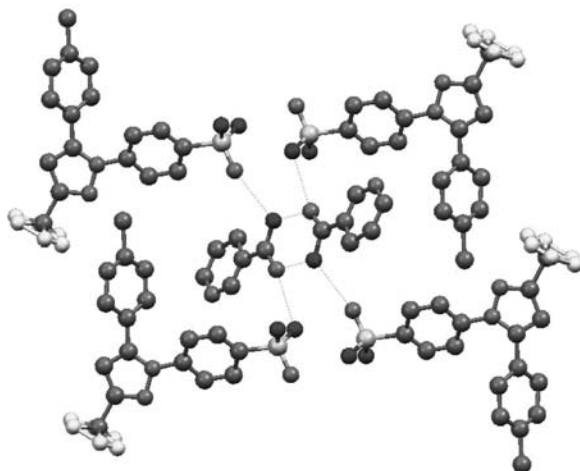
A co-crystal of celecoxib with nicotinamide (Vitamin B3) was identified [49]. The co-crystal has a 1:1 stoichiometry and is comprised of hydrogen bonded nicotinamide dimers that sit within channels formed by celecoxib molecules, as elucidated by powder diffraction structure determination (Figure 3.5). The form rapidly dissociates when dispersed in aqueous solutions and the parent, celecoxib, precipitates in one of several forms depending on the excipients present. Although complex, the dissolution behaviour of the co-crystal has been evaluated and through the use of a spring and parachute approach, the *in vitro* performance of celecoxib:nicotinamide was shown to be superior to that of the parent compound.

The celecoxib:nicotinamide co-crystal was first prepared using a solvent drop grinding approach, where a slight excess of nicotinamide was ground together with celecoxib using a small amount of THF in a ball mill followed by thorough drying at elevated temperatures. Several iterations of the grinding–drying cycle were necessary to complete the transformation to the co-crystal. Solution crystallization of this form was

---

\*\* Pharmacia Corporation. Celecoxib New Drug Application. Pharmacia Corp. (No. 20-998). 1998. Rockville, MD, USA.





**Figure 3.5** Supramolecular association in the co-crystal celecoxib:nicotinamide showing a  $R_2^2(8)$  nicotinamide dimer and its four hydrogen bonded nearest neighbour celecoxib molecules. Hydrogen atoms have been omitted for clarity

successful from methylene chloride using a cooling procedure followed by drying in a vacuum oven. Somewhat larger crystals were prepared by incubating methylene chloride solutions containing small amounts of dimethoxyethane at 40 °C.

Rapid form dissociation and precipitation of known forms of the parent celecoxib was observed in 0.1 N HCl, SGF, 0.02 N HCl, pH 6.5 phosphate buffer and pH 6.5 phosphate buffer with sodium dodecyl sulfate (SDS). The initial dissolution rate was found to be superior for the co-crystal versus the stable form of the parent compound, but rapid conversion to aggregated crystalline celecoxib made this co-crystal a poor candidate for direct use as a dosage form. This initial dissolution advantage was exploited further by formulating the co-crystal with PVP-K30 and the ionic surfactant SDS. Using this combination of excipients, the co-crystal was repeatedly observed to precipitate as a poorly crystalline mixture of metastable celecoxib form 4 and amorphous material. Although the co-crystal was not tested *in vivo*, the *in vitro* performance improvements are consistent with improvements observed in other celecoxib formulations with demonstrated *in vivo* advantages over Celebrex®.

Celecoxib:nicotinamide is not the first co-crystal to be formulated using the spring and parachute approach. In fact, it is a progression of the spring and parachute concept as it was employed to enhance the *in vivo* performance of itraconazole co-crystals [51]. This system was recently reviewed in detail [52]. In short, several co-crystals of itraconazole were identified during a HT crystal form screen [53]. These forms were evaluated and shown to have improved kinetic solubility versus the parent form. Through insightful formulation development, the timeframe of the supersaturation relative to the free base was extended and ultimately led to improved bioavailability over the crystalline free base. This is an early example of formulating a pharmaceutical co-crystal to take advantage of performance improvements, by first exercising control over the solid form and therefore, delaying the onset of subsequent crystallization.

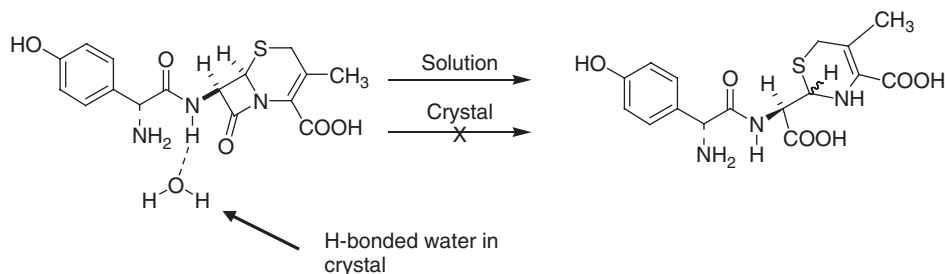
### 3.5.1 $\beta$ -Lactam antibiotics and hydrates – the importance of crystallinity

A survey of crystalline hydrates of  $\beta$ -lactams was recently reported in order to probe the significance of crystallinity on the stability of reactive APIs [54]. Compounds which contain the  $\beta$ -lactam core, including penicillin derivatives, carbapenems and cephalosporins, are known to undergo hydrolysis both in solution and in the solid state [55, 56]. Solid-state stability has been shown to be dependent on storage conditions and degree of crystallinity [57–59]. It has also been shown that even hydrates of  $\beta$ -lactams, where the water is regarded as a potential reactant, can be stable in the solid state for years.

An example highlighting the impact of crystallinity on stability is that of meropenem (MERREM®), a trihydrate of a carbapenem antibiotic. Meropenem batches of different hydration states were examined for solid-state stability at 40 °C over a 1 month period [59]. An increase in amorphous content was shown to lead to a higher extent of and more rapid degradation in the solid state. However, the crystalline trihydrate exhibited a nearly 300-fold improvement in stability compared with a sample which was 80% amorphous. Similar observations have also been made on cephalosporin antibiotics as reported by Pikal and co-workers [58].

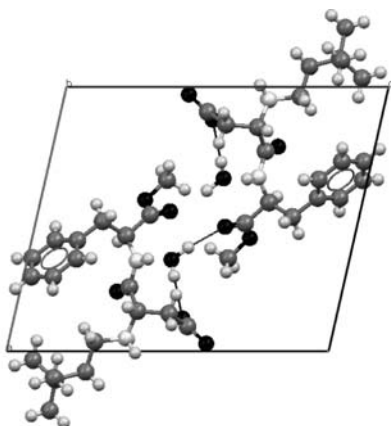
An analysis of all crystal structures available in the Cambridge Structural Database for hydrates of  $\beta$ -lactams reveals that the water molecules in the crystal structures are not in sufficient proximity (within the sum of van der Waals distance) to the reactive center on the  $\beta$ -lactam core to bring about reaction. For cases where water molecules are in close proximity to the reactive centre, further rationalization can be made in terms of the relative orientation of reactants and the steric hindrance of the initial approach of the water, or the low degree of conformational flexibility in the crystal. This type of calculation on cefadroxil monohydrate (reaction scheme shown in Figure 3.6) which contains a water molecule within van der Waals distance of the reactive  $\beta$ -lactam carbonyl group, has shown that several atoms would lie outside of the molecular surface formed by the rest of the crystal, indicating steric clashes.

Crystalline hydrates of other electrophilic drugs other than  $\beta$ -lactams have also been studied. The cases of neotame and aspartame illustrate a similar principle to the one exemplified by  $\beta$ -lactams. The methyl esters are relatively stable in the solid state unless they are rendered amorphous [60, 61]. It is also true that in the case of neotame



**Figure 3.6** Molecular structures of the native and ring open hydrolysed degradant of cefadroxil. Arrows highlight the hydrogen bonded water that is present in the crystal. The hydrolysis proceeds readily in aqueous solution but not in the crystalline hydrate





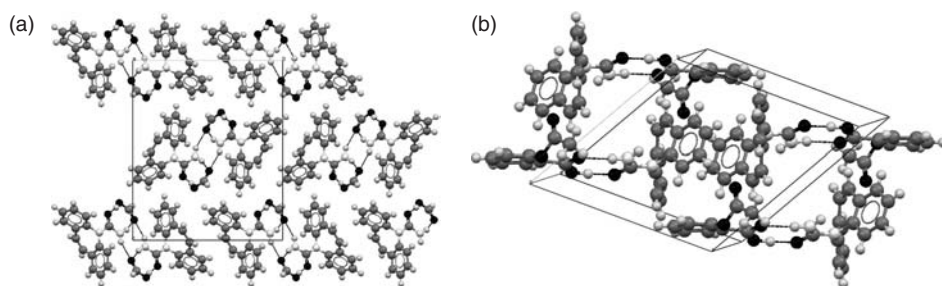
**Figure 3.7** Partial unit cell contents for neotame monohydrate

monohydrate, the water molecule is also in close proximity to the reactive ester group, as shown in Figure 3.7.

### 3.6 Carbamazepine – Stabilization Against a Hydrate

Another example of water absorption reducing the solubility and dissolution rate of a formulated drug is the drug carbamazepine, marketed as Tegretol® by Novartis and also by generic companies. Carbamazepine is an anti-epileptic drug that has been in use for decades. The drug must be titrated in patients because the therapeutic window is somewhat narrow and varies from patient to patient. Moreover, therapeutic doses increase over time for individual patients due to metabolic induction (a feature that is not affected by formulation or form). Carbamazepine is known to exist in at least four polymorphic modifications, several solvated forms, including a dihydrate that is usually the stable form in water, and also in several co-crystal forms [62–65]. The relationship between the relative hygroscopicities and dissolution behaviour of the anhydrous forms of carbamazepine and the dihydrate has been determined [66, 67]. In aqueous solvents, several of the anhydrous polymorphs also show propensity for converting to the dihydrate form.

A common packing motif observed in all of the polymorphs, the dihydrate, several solvates and the co-crystals of carbamazepine is the  $R_2^2(8)$  dimer [3, 68, 69]. The robust nature of this packing element is, in some respects, remarkable. For the hydrate, solvates and co-crystals that maintain the  $R_2^2(8)$  dimer, additional hydrogen bonding interactions are formed with the N-H donor and C=O on the periphery of the dimer. The only structures where the carbamazepine dimer is not present are those in which carbamazepine is co-crystallized with a second component containing a carboxylic acid group, such as the formic acid solvate and the acetylsalicylic acid co-crystal, as well as in the recently discovered metastable carbamazepine:saccharin co-crystal [70]. In the majority of these examples, pairs of acid-amide heterodimers form discrete hydrogen bonded packing units, as shown in Figure 3.8.



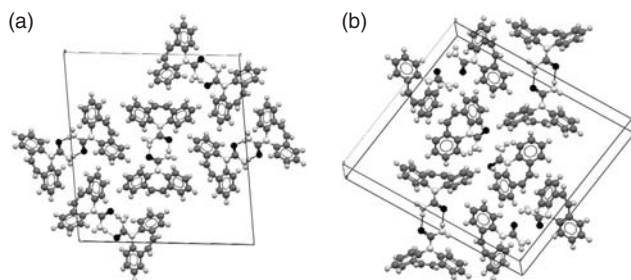
**Figure 3.8** (a) Packing diagram of carbamazepine:formic acid solvate and (b) packing diagram of carbamazepine:acetylsalicylic acid co-crystal

There has been some discussion in the literature about the impact of inadvertent dihydrate formation on the performance of some carbamazepine formulations by reducing the solubility and dissolution rate. The importance of solubility and dissolution rate in relation to pharmacokinetic performance in dogs has been shown using cyclodextrin [71] and poly(ethylene glycol) (PEG) [66, 67] solutions, anhydrous polymorphs, the dihydrate and the saccharin co-crystal [62] of carbamazepine. During several *in vitro* dissolution experiments in water carried out on the stable form of carbamazepine (form 3) and the saccharin and nicotinamide co-crystals, it was found that the saccharin co-crystal exhibited a nearly twofold increase in kinetic solubility compared with the dihydrate of carbamazepine.<sup>††</sup> Both the nicotinamide co-crystal and anhydrous stable form of carbamazepine display an initial increase in the rate of dissolution, but convert to the less soluble dihydrate within 1 h. Interestingly, the saccharin co-crystal showed sustained solubility over a long period of time, indicating that conversion to the less soluble dihydrate did not occur under those conditions.

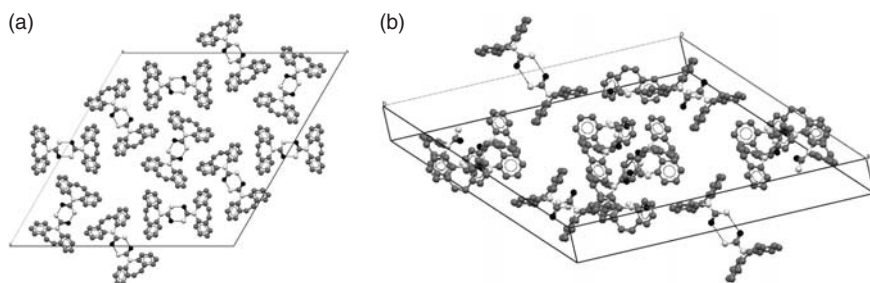
It has been shown that a saccharin co-crystal of carbamazepine is stable against formation of the dihydrate in water slurry for several days. In addition, only one form of the co-crystal was found during extensive screening [72]. In a small probe study the pharmacokinetics performance of the carbamazepine saccharin co-crystal was found to be similar to that of the marketed immediate release product, Tegretol®, but there was some indication that the variability in  $C_{\max}$  and AUC were reduced.

The lack of conversion of the carbamazepine:saccharin co-crystal to the dihydrate when slurried in water can be rationalised from examination of the crystal structures of the anhydrous polymorphs, and the nicotinamide and saccharin co-crystals. The packing diagrams of the four anhydrous polymorphs of carbamazepine are shown in Figures 3.9–3.12. The differences in the crystal structures have been described and are only apparent in the packing of the carboxamide  $R_2^2(8)$  dimer [73, 74]. Weak C-H $\cdots$ O interactions from a vinylic hydrogen atom on the azepine ring to an oxygen acceptor are present in all four structures. In triclinic form 1 and trigonal form 2, shown in Figures 3.9 and 3.10, respectively, no  $\pi$ -stacking was observed. The structures of the two monoclinic polymorphs (forms 3 and 4) also indicate that the systems contain discrete sets of dimers which are  $\pi$ -stacked, as shown in Figures 3.11 and 3.12.

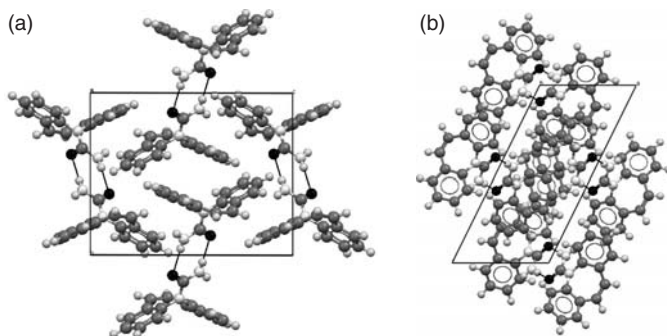
<sup>††</sup> N. Rodríguez-Hornedo, personal communication.



**Figure 3.9** Views of the crystal packing of carbamazepine, form 1

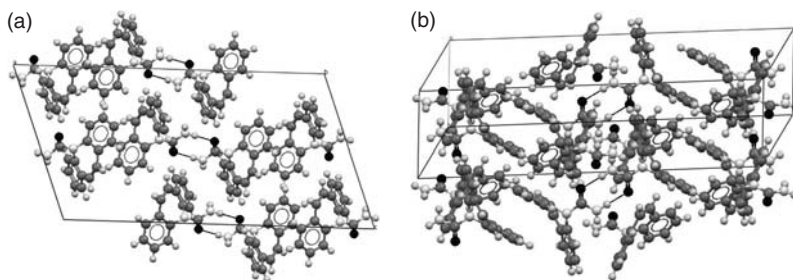


**Figure 3.10** Views of the crystal packing of carbamazepine, form 2

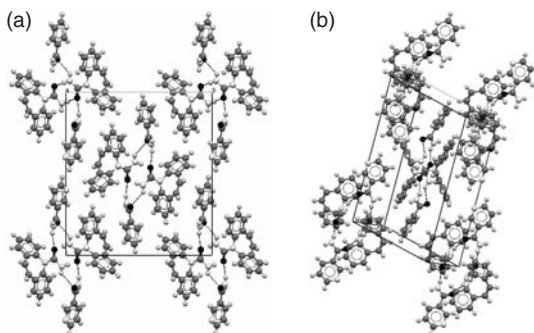


**Figure 3.11** Views of the crystal packing of carbamazepine, form 3. This is the stable polymorph of carbamazepine at room temperature

The packing diagrams of the nicotinamide and saccharin co-crystals of carbamazepine are shown in Figure 3.13. The  $R_2^2(8)$  carboxamide dimer is maintained in the carbamazepine:nicotinamide co-crystal with additional hydrogen bonding from the amide group of nicotinamide to the periphery of the carbamazepine dimer. There is also evidence of  $\pi$ -stacking between the aromatic nicotinamide ring and one of the rings on carbamazepine, as shown in Figure 3.13. Like the anhydrous polymorphs of carbamazepine, the  $\pi$ -stacking is only present in discrete units. Analysis of the packing



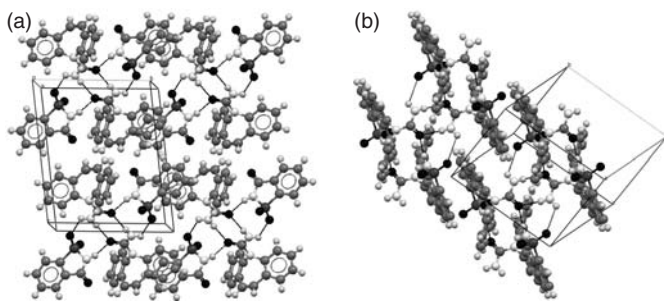
**Figure 3.12** Views of the crystal packing of carbamazepine, form 4



**Figure 3.13** Packing diagram of the carbamazepine:nicotinamide co-crystal and rotated version to emphasize the  $\pi$ -stacking

diagram of the carbamazepine:saccharin co-crystal reveals that the structure is made up of hydrogen bonded tetramers. The hydrogen bonding motif is described by the common  $R_2^2(8)$  carboxamide dimer with additional  $N-H \cdots O$  hydrogen bonding from the carbamazepine nitrogen to one of the sulfonamide oxygen atoms on saccharin and from the nitrogen atom on saccharin to the oxygen atom on the carbamazepine amide group. Of the many solvates and co-crystals of carbamazepine, the carbamazepine saccharin co-crystal is the only one that has an extended  $\pi$ -stacking network, as shown in Figure 3.14. This network is unique to the saccharin co-crystal, form 1, and is not observed in any of the other structures that have been analysed. However, a recent paper describes the discovery of a second polymorph of carbamazepine with saccharin in the presence of polymer heteronuclei [70]. This form was also found to break up the carbamazepine homodimer in favour of a carbamazepine saccharin heterodimer, to convert to the more stable tetramer in water and therefore is believed to be a metastable form relative to the co-crystal polymorph characterized initially.

While much of the literature descriptions of the crystal structures of various carbamazepine forms have focused on differences in hydrogen bonding, it is possible that saccharin disrupts the  $\pi$ -stacking observed in the anhydrous forms and the dihydrate inhibiting the precipitation of those forms from solution. In this case the co-crystal former, saccharin, may also act as a very effective crystal growth inhibitor for certain, less



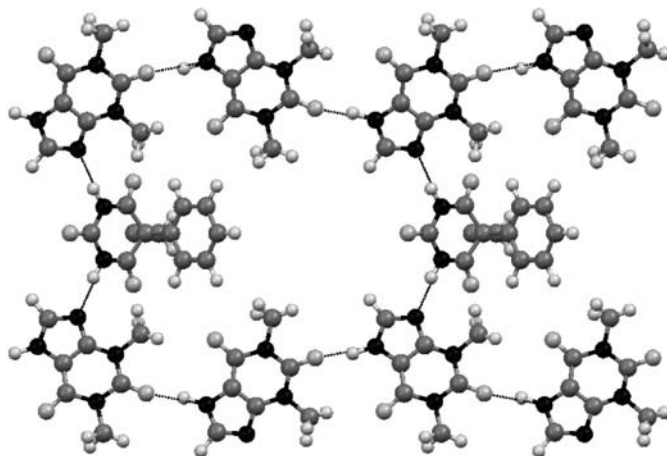
**Figure 3.14** Packing diagram of the carbamazepine:saccharin co-crystal and rotated version to emphasize the  $\pi$ -stacking

desirable forms [75, 76]. While no particular attention has been paid to the morphology of crystals grown during co-crystallization experiments, it would not be surprising if significant changes in morphology would be observed in many cases where only the starting materials crystallized.

### 3.7 Theophylline:Phenobarbital – Two is Better Than One

Combination therapies have been in use for several decades [77]. Often, these products are developed to allow for more convenient dosing of two or more drugs that are used in combination, or to take advantage of the synergies that may exist between two drug compounds upon co-administration. The benefits of combination products include reduced pill burden for patients and may improve patient compliance with the intended therapy. Caution must be used when developing combination products to avoid undesirable interactions, including physical and chemical ones, in the dosage forms. Theophylline is a methylxanthine compound provided under a variety of trade names for the treatment of respiratory diseases such as asthma and chronic obstructive pulmonary disease. In use for nearly a century, the free acid has an aqueous solubility of about  $8 \text{ mg ml}^{-1}$  at low and neutral pH [78]. At higher pH the solubility of theophylline increases. If rapid onset of action is needed from the oral product, the ethylenediamine salt is preferred over the less soluble free acid [79]. One of the side-effects associated with theophylline is increased heart rate [80]. In fact, theophylline has a greater stimulant effect than caffeine.

Phenobarbital is an anticonvulsant in the barbiturate class with sedative and hypnotic effects [78]. In the past, combinations between barbiturates and analgesics were used because barbiturates tend to increase the action of these drugs. Phenobarbital has also been co-administered with theophylline to partially offset the increased heart rate associated with the compound. A combination product containing theophylline, phenobarbital and ephedrine HCl is currently in use for the treatment of asthma, emphysema and bronchitis. Several years ago, variation in the performance of the brand name combination product Tedral, marketed by Parke-Davis Canada Inc., compared with generic versions were noted [81]. Over time, and possibly guided by a clinical observation of performance when co-administering a combination of theophylline and phenobarbital, it was



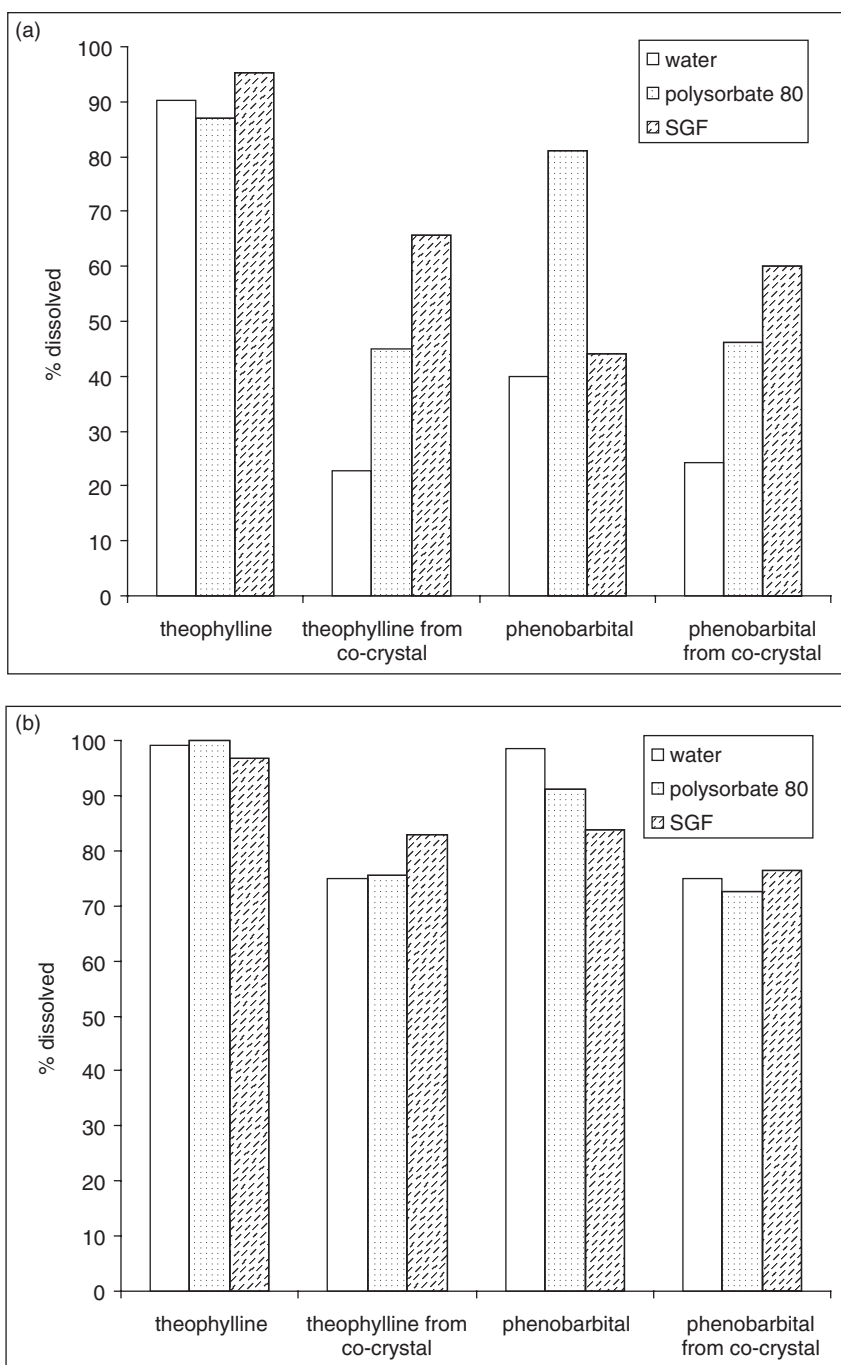
**Figure 3.15** Supramolecular association in the theophylline:phenobarbital co-crystal showing a layer of the hydrogen bonded network and the 2:1 stoichiometry

discovered that a co-crystal between the two compounds existed [82]. The co-crystal has a 2:1 stoichiometry, as elucidated by single crystal structure determination [83]. Theophylline molecules accept hydrogen bonds from the ring amido nitrogen on a phenobarbital molecule and from the imidazole ring of a neighbouring theophylline molecule (Figure 3.15). Several chain (C2) interactions come together to form a  $R_g^8(43)$  ring.

Diffusion of theophylline and phenobarbital through a cellophane membrane at 37 °C in pH 1.2 buffer was examined using the pure starting materials and the recrystallized co-crystal [84]. The steady state diffusion value starting from the pure components was three times larger for theophylline than for phenobarbital. When starting with the co-crystal at the same nominal concentration as used for the pure components, the ratio of steady state diffusion values was 2:1 theophylline:phenobarbital, indicating that the complex moving through the membrane is, in fact, weakly bound. However, UV spectroscopy did not indicate any difference in absorption maxima between the peaks observed for the individual species and for the complex. The nature of any complex that exists between theophylline and phenobarbital in solution is unclear, although it is clear that the association is a weak one.

The dissolution profiles for theophylline, phenobarbital and the co-crystal were measured under sink conditions on powders in stirred water, 0.02% (v/v) polysorbate 80 in water and SGF, all at 37 °C. In all media it was found that theophylline dissolved faster from the pure powder than from the co-crystal (Figure 3.16). Phenobarbital dissolved faster from the pure form in water and polysorbate 80, but more slowly in SGF relative to the co-crystal. At 5 min nearly 90% of the available theophylline is dissolved from the free powder but only between 25% and 70%, depending on the media, when starting with the co-crystal. At 30 min theophylline is nearly completely dissolved starting from pure theophylline and about 80% starting from the co-crystal. The range is broader between the three dissolution media for phenobarbital, but in all cases the percentage of phenobarbital dissolved is lower when starting with the co-crystal than when starting





**Figure 3.16** Bar graph of percentage dissolved at 5 (a) and 30 (b) min for theophylline, theophylline from the co-crystal, phenobarbital and phenobarbital from the co-crystal in water, 0.02% (v/v) polysorbate 80 and SGF. Plotted using data from [81]

with the pure powder. Dissolution from generic tablets also showed dramatically slower dissolution of theophylline than the marketed product of comparison, Tedral.

A pharmacokinetic comparison of theophylline, the theophylline:phenobarbital co-crystal, Tedral or one of two generic tablets was conducted in nine healthy human volunteers [81]. All subjects were fed cereal and milk 30 min prior to dosing and blood samples were collected at 30, 60, 120 and 240 min time points. In seven of the subjects the serum concentrations of theophylline increased more rapidly when dosed as the pure drug than with the co-crystal. These results supported the hypothesis that decreased dissolution rate observed *in vitro* for the co-crystal would lead to slower absorption of theophylline in humans. This study also highlighted the importance of characterizing the amount of co-crystal obtained when new manufacturing processes were evaluated [85, 86]. Because theophylline and phenobarbital are used in combination, but the specific co-crystal between the two is undesirable from a pharmacokinetics standpoint, one can speculate that there would be some benefit to preparing combinations using alternative forms of theophylline, phenobarbital or both.

Theophylline has also been co-crystallized with urea [87] *N*-(2-ammonioethyl)carbamate [88], chlorosalicylic acid [89], sulfathiazole, 5-fluorouracil [90], *p*-nitroaniline [91], succinic acid, malonic acid, maleic acid and oxalic acid [92]. There is evidence that at least one of these co-crystals, between theophylline and oxalic acid, can improve the physical stability of theophylline by protecting it from converting to the less soluble hydrate at high humidities. In this case, oxalic acid and water both hydrogen bond with theophylline at the same site. It is unclear whether or not the occupancy of the hydrogen bonding site by another molecule, in this case oxalic acid prevents the conversion to the hydrate or, if a general decrease in solubility of the oxalic acid co-crystal versus the hydrate is responsible for the protective effects.

Similar protective effects may be exploitable when attempting to prevent the formation of an undesirable co-crystal between two compounds in the same formulation, as is the case in 2(theophylline):phenobarbital. Because phenobarbital forms hydrogen bonds with theophylline by donating a hydrogen bond to the free nitrogen on the imidazole ring in the co-crystal, any compound that has stronger interactions with that site could protect against formation of theophylline:phenobarbital. Of course, the dissolution behaviour of the alternative co-crystal used would have to have similar, or superior, dissolution properties in comparison with the free form of theophylline for this approach to be useful. The rather benign nature of succinic acid and urea could make them interesting candidates in this regard. Of the other known co-crystals with theophylline, only those with 5-fluorouracil and *p*-nitroaniline do not form hydrogen bonding interactions with the free nitrogen atom of the imidazole ring.

### 3.8 Delaviridine Mesylate – Material Misbehaviour

To reiterate a common theme, characterization of the crystalline form selected should be extensive in order to assure a sound scientific understanding of the material properties of the substance is in hand during the development process. An appreciation of the chemical and physical compatibility of the solid form with the excipients to be used in the final formulation is required. Often called excipient compatibility screening,



this process is usually started before entering into phase 1 clinical trials [79, 93–98]. One of the challenging aspects of these studies is that long-term incompatibilities and manufacturing impact are sometimes difficult to quantify under laboratory conditions and development timelines. It is especially important to uncover such incompatibilities when the solid form selected for development can chemically or physically convert to a less desirable form.

Delaviridine, a reverse transcriptase inhibitor used to treat HIV infection, is chosen here as an example because of the thorough understanding of the chemistry surrounding the effect of storage conditions on the incompatibility of the solid form and the excipients utilized for formulation purposes. Delaviridine is an example of a compound with rich form diversity, but the lack of crystallographic structure determination in the open literature prevents the discussion of potential interactions from a more traditional crystal engineering perspective [99]. Nonetheless, the importance of understanding the subtleties of how a crystalline form can be affected by another compound incorporated into the matrix of the dosage form warrants its inclusion.

In this example water vapour was shown to facilitate acid–base chemistry between delaviridine mesylate and croscarmellose sodium (a carboxylate-functionalized polysaccharide, a commonly used functional excipient in tablets) [100]. The product, Rescriptor<sup>®</sup>, was formulated using anhydrous form 11 as a tablet containing lactose, microcrystalline cellulose, croscarmellose sodium, magnesium stearate, colloidal silicon dioxide and carnauba wax.<sup>‡‡</sup> During the development of the product it was shown that exposure of the tablets to high humidity reduced the rate and extent of delaviridine mesylate dissolution [100]. Using the combined techniques of IR spectroscopy and solid-state NMR it was elegantly shown that delaviridine mesylate underwent an acid–base reaction with the croscarmellose sodium upon exposure to high humidity. This resulted in the formation of the less-soluble delaviridine free base. Ultimately this route for potential product degradation was managed using packaging and desiccant. Given the rich form diversity of delaviridine, alternative forms possessing these similar advantages in solubility and dissolution as the mesylate salt but without the potential for acid–base reaction, could possibly have avoided the resulting necessity for special packaging and storage. In fact, an understanding of the potential for co-crystals to provide chemical and physical stabilization against water mediated instabilities had not been appreciated until recently [92].

### 3.9 Summary and Outlook

Engineering for performance broadly encompasses ‘traditional’ crystal engineering, where an understanding of hydrogen bonding interactions is used to ‘design’ crystal forms. This very principle was utilized by Schmidt [101] in making a co-crystal between methylthymine and methyladenine, where he exploited the intermolecular hydrogen bonding interactions to form a new multi-component crystalline material. The use of hydrogen bonds as critical design elements in co-crystals was championed further by Etter [102] and later by Desiraju [103] among others, with some further work helping to

---

<sup>‡‡</sup> This is according to the Rescriptor<sup>®</sup> product label.

conceptualize the possible limits of this methodology [104–108]. The re-emergence of pharmaceutical co-crystals was recently led by Zaworotko and co-workers [62]. From this point of view technology has made the search for new, engineered crystal forms easier [5, 15, 17]. However, engineering for performance extends beyond the approach of using often observed and well described hydrogen bonding patterns in an attempt to stack the deck in favour of finding a new multi-component crystalline material from often complex pharmaceutical compounds. In preparing pharmaceutical materials, it is necessary to include the utilization of physical properties and behaviour of the forms, the exertion of control over form interconversion and the ultimate protection against the formation of less desirable forms. In this paradigm, identification of a new form is just the beginning. For the full advantage of a new form to be exploited, an understanding of how to control its physical behaviour during processing and use is as important as the form itself.

To capture the essence from the above examples, one of the key technical challenges is to improve the solubility or dissolution profile of a compound to facilitate adequate oral absorption for a given study or product. It is evident that an understanding of what hydrogen bonding interactions are actually present in important crystalline solids is easily obtained after the crystal structures have been determined. However, developing an understanding of how the dissolution profile of a given form can be manipulated using external excipients is of paramount importance. Also of great importance is developing an understanding of how to control the crystallization of related species, for example by delaying the precipitation of a stable form of celecoxib free acid. The examples presented herein highlight the idea of engineering for performance in the context of maximizing the advantage gained by increased kinetic solubility, regardless of the type of crystalline form used to achieve this goal.

Opportunities to exploit the principles of crystal engineering exist in addition to delaying precipitation of less desirable crystalline forms. Simple improvements, such as increasing melting point or decreasing hygroscopicity, are easy to envision. The improvements in the primary physical properties of compounds that can be associated with specific crystalline forms are not the end of the story; other protection may also be exploited. As highlighted for the  $\beta$ -lactam class of antibiotics, protection against hydrolysis is conferred simply by utilizing crystalline forms, even when in many cases such forms are crystalline hydrates. The fact that two reactants can be combined to form a stable crystalline material highlights the controlling power that crystallinity can exert over chemical reactivity. For theophylline, the oxalic acid co-crystal was demonstrated to resist formation of the hydrated form of theophylline [92]. It is possible that this or other known co-crystals of theophylline could resist the formation of complexes with phenobarbital and in doing so, may allow for more controlled dosing when combinations of these are used. Co-crystals or organic salts of norfloxacin may be resistant to, or delay, the formation of metal ion complexes *in vivo*. Such a situation could be advantageous in allowing for less restrictive product labelling or dosing instructions by avoiding undesirable chemical interactions. Finally, the example of delaviridine mesylate highlights the importance of understanding how the constituents of a multi-component crystalline solid can interact and react with excipients used in formulations. All dosage form design studies involve the delicate balance of *in vivo* performance and physical and chemical stability. Sometimes, deleterious interactions are difficult to identify using

standard laboratory conditions and analytical methods, but with an open mind, good analytical capabilities and the good fortune of form diversity, more options should be available for development of a well performing drug product.

## References

1. S. Frantz, Study reveals secrets to faster drug development, *Nat. Rev. Drug Discovery*, 5, 883 (2006).
2. C. Gardner, C. Walsh, and Ö. Almarsson, Drugs as materials: valuing physical form in drug discovery, *Nat. Rev. Drug Discovery*, 3, 926–934 (2004).
3. J. Bernstein, *Polymorphism in Molecular Crystals*, Oxford Science Publications, Oxford (2002).
4. H. Brittain (Ed.), *Polymorphism in Pharmaceutical Solids*, Marcel Dekker, New York (1999).
5. S. Morissette, Ö. Almarsson, M. Peterson, J. Remenar, M. Read, A. Lemmo, S. Ellis, M. Cima, and C. Gardner, High-throughput crystallization: polymorphs, salts, co-crystals and solvates of pharmaceutical solids, *Adv. Drug Delivery Rev.*, 56, 275–300 (2004).
6. S. Byrn, *Solid-State Chemistry of Drugs*, Academic Press, New York (1982).
7. M. Pudipeddi and A. Serajuddin, Trends in solubility of polymorphs, *J. Pharm. Sci.*, 94, 929–939 (2005).
8. D. Haynes, W. Jones, and W. Motherwell, Occurrence of pharmaceutically acceptable anions and cations in the Cambridge structural database, *J. Pharm. Sci.*, 94, 2111–2120 (2005).
9. Ö. Almarsson, M. Hickey, M. Peterson, S. Morissette, S. Soukasene, C. McNulty, M. Tawa, J. MacPhee, and J. Remenar, High-throughput surveys of crystal form diversity of highly polymorphic pharmaceutical compounds, *Cryst. Growth Des.*, 3, 927–933 (2003).
10. J. Remenar, J. MacPhee, B. Larson, V. Tyagi, J. Ho, D. McIlroy, M. Hickey, P. Shaw, and Ö. Almarsson, Salt selection and simultaneous polymorphism assessment via high-throughput crystallization: the case of sertraline, *Org. Process Res. Dev.*, 7, 990–996 (2003).
11. J. Bernstein, Crystal growth, polymorphism and structure-property relationships in organic crystals, *J. Phys. D: Appl. Phys.*, 26, B66–B76 (1993).
12. F. Allen, The Cambridge Structural Database: a quarter of a million crystal structures and rising, *Acta Crystallogr., Sect. B*, B58, 380–388 (2002).
13. G. Lewis, G. Steele, L. McBride, A. Florence, A. Kennedy, N. Shankland, W. David, K. Shankland, and S. Teat, Hydrophobic vs. hydrophilic: ionic competition in remacemide salt structures, *Cryst. Growth Des.*, 5, 427–438 (2005).
14. E. Collier, R. Davey, S. Black, and R. Roberts, 17 Salts of ephedrine: crystal structures and packing analysis, *Acta Crystallogr., Sect. B*, B62, 498–505 (2006).
15. M. Caira, L. Nassimbeni, and A. Wildervanck, Selective formation of hydrogen bonded co-crystals between a sulfonamide and aromatic carboxylic acids in the solid state, *J. Chem. Soc., Perkin Trans. 2*, 2213–2216 (1995).

16. N. Shan, F. Toda, and W. Jones, Mechanochemistry and co-crystal formation: effect of solvent on reaction kinetics, *Chem. Commun.*, 2372–2373 (2002).
17. A. Trask, W. Motherwell, and W. Jones, Solvent-drop grinding: green polymorph control of co-crystallisation, *Chem. Commun.*, 890–891 (2004).
18. M. Cheney, G. McManus, J. Perman, Z. Wang, and M. Zaworotko, The Role of co-crystals in solid-state synthesis: co-crystal-controlled solid-state synthesis of imides, *Cryst. Growth Des.*, 7(4), 616–617 (2007).
19. S. Childs, P. Mougin, and B. Stahly, Screening for solid pharmaceutical forms by ultrasound crystallization and co-crystallization using ultrasound, WO 2005089375 A2 (2005).
20. R. Storey, R. Docherty, P. Higginson, C. Dallman, C. Gilmore, G. Barr and W. Dong, Automation of solid form screening procedures in the pharmaceutical industry – How to avoid the bottlenecks, *Crystallogr. Rev.*, 10, 45–56 (2004).
21. N. Blagden, R. Davey, H. Lieberman, L. Williams, R. Payne, R. Roberts, R. Rowe, and R. Docherty, Crystal chemistry and solvent effects in polymorphic systems sulfathiazole. [Erratum to document cited in CA129:10804], *J. Chem. Soc., Faraday Trans.*, 94, 1919 (1998).
22. N. Blagden, R. Davey, H. Lieberman, L. Williams, R. Payne, R. Roberts, R. Rowe, and R. Docherty, Crystal chemistry and solvent effects in polymorphic systems Sulfathiazole, *J. Chem. Soc., Faraday Trans.*, 94, 1035–1044 (1998).
23. M. Lahav and L. Leiserowitz, The effect of solvent on crystal growth and morphology, *Chem. Eng. Sci.*, 56, 2245–2253 (2001).
24. T. Threlfall, Crystallisation of polymorphs: thermodynamic insight into the role of solvent, *OPRD*, 4, 384–390 (2000).
25. J. Lin, D. Ostovic, and J. Vacca, The integration of medicinal chemistry, drug metabolism, and pharmaceutical research and development in drug discovery and development: the story of Crixivan, an HIV protease inhibitor, *Pharm. Biotechnol.*, 11, 233–255 (1998).
26. B. Johnson, A. Howard, R. Varsolona, J. McCauley, and D. Ellison, Indinavar sulfate, in *Analytical Profiles of Drug Substances and Excipients*, H. Brittain (Ed.), Academic Press, New York, Vol. 26 (1999).
27. K. Yeh, P. Deutsch, H. Haddix, M. Hesney, V. Hoagland, W. Ju, S. Justice, B. Osborne, A. Sterrett, J. Stone, E. Woolf, and S. Waldman, Single-dose pharmacokinetics of indinavir and the effect of food, *Antimicrob. Agents Chemother.*, 42(5), 1308 (1998).
28. S. Wallis, B. Charles, L. Gahan, L. Filippich, M. Bredhauer, and P. Duckworth, Interaction of norfloxacin with divalent and trivalent pharmaceutical cations. In vitro complexation and in vivo pharmacokinetic studies in the dog, *J. Pharm. Sci.*, 85, 803–809 (1996).
29. I. Dos Santos, F. Fawaz, A. Lagueny, and F. Bonini, Improvement of norfloxacin oral bioavailability by EDTA and sodium caprate, *Int. J. Pharm.*, 260, 1–4 (2003).
30. R. Barbas, F. Marti, R. Prohens, and C. Puigjaner, Polymorphism of norfloxacin: evidence of the enantiotropic relationship between polymorphs A and B, *Cryst. Growth Des.*, 6, 1463–1467 (2006).

31. B. Sustar, N. Bukovec, and P. Bukovec, Polymorphism and stability of norfloxacin, (1-ethyl-6-fluoro-1,4-dihydro-4-oxo-7-(1-piperazinyl)-3-quinolinocarboxylic acid, *J. Therm. Anal.*, 40, 475–481 (1993).
32. A. Florence, A. Kennedy, N. Shankland, E. Wright, and A. Al Rubayi, Norfloxacin dihydrate, *Acta Crystallogr., Sect. C*, C56, 1372–1373 (2000).
33. Y. Wang, L. Sun, W. Wang, and L. Yan, Crystal structure of norfloxacin methanol solvate:  $C_{16}H_{18}FN_3O_3 \cdot CH_3OH \cdot H_2O$ , *Jiegou Huaxue*, 24, 1359–1362 (2005).
34. S. Basavoju, D. Bostroem, and S. Velaga, Pharmaceutical co-crystal and salts of norfloxacin, *Crys. Growth Des.*, 6, 2699–2708 (2006).
35. M. Martin, P. Perxes, Norfloxacin ascorbate, a water-soluble adduct of norfloxacin, ES2127036 A1 (1999).
36. X. Zhang, S. Wei, Q. Liao, J. Weng, D. Ma, and W. Huang, Preparation of norfloxacin lactate, *Guangxi Huagong*, 26, 15–17 (1997).
37. S. Velaga, S. Basavoju, and D. Bostroem, Norfloxacin saccharinate-saccharin dihydrate co-crystal - A new pharmaceutical co-crystal with an organic counter ion, *J. Mol. Struct.*, 889, 150–153 (2008).
38. M. Cordoba-Diaz, M. Cordoba-Borrego, and D. Cordoba-Diaz, Influence of pharmacotechnical design on the interaction and availability of norfloxacin in directly compressed tablets with certain antacids, *Drug Dev. Ind. Pharm.*, 26, 159–166 (2000).
39. Z. Chen, R. Xiong, J. Zuo, Z. Guo, X. You, and H. Fun, X-Ray crystal structures of  $Mg^{2+}$  and  $Ca^{2+}$  dimers of the antibacterial drug norfloxacin, *Inorg. Chem.*, 4013–4014 (2000).
40. Z. Chen, R. Xiong, J. Zhang, X. Chen, Z. Xue, and Z. You, 2D molecular square grid with strong blue fluorescent emission: a complex of norfloxacin with zinc(II), *Inorg. Chem.*, 40, 4075–4077 (2001).
41. Z. Qu, H. Zhao, L. Xing, X. Wang, Z. Chen, Z. Yu, R. Xiong, and X. You, Two polymeric complexes of norfloxacin with iron(II) and their magnetic properties, *Eur. J. Inorg. Chem.*, 2920–2923 (2003).
42. F. Fawaz, F. Bonini, M. Guyot, J. Bildet, M. Maury, and A. Lagueny, Bioavailability of norfloxacin from PEG 6000 solid dispersion and cyclodextrin inclusion complexes in rabbits, *Int. J. Pharm.*, 132, 271–275 (1996).
43. M. Guyot, F. Fawaz, J. Bildet, F. Bonini, and A. Lagueny, Physicochemical characterization and dissolution of norfloxacin/cyclodextrin inclusion compounds and PEG solid dispersions, *Int. J. Pharm.*, 123, 53–63 (1995).
44. T. Higuchi and K. Connors, Phase-solubility techniques, *Adv. Anal. Chem. Instr.*, 4, 117–212 (1965).
45. A. Bak, A. Gore, E. Yanez, M. Stanton, S. Tufekcic, R. Syed, A. Akrami, M. Rose, S. Surapaneni, T. Bostick, A. King, S. Neervannan, D. Ostovic, and A. Koparkar, The co-crystal approach to improve the exposure of a water-insoluble compound: AMG 517 sorbic acid co-crystal characterization and pharmacokinetics, *J. Pharm. Sci.*, 97, 3942–3956 (2008).
46. M. Stanton and A. Bak, Physicochemical properties of pharmaceutical co-crystals: a case study of ten AMG 517 co-crystals, *Cryst. Growth Des.*, 8, 3856–3862 (2008).

47. M. Stanton, S. Tufekcic, C. Morgan, and A. Bak, Drug substance and former structure property relationships in 15 diverse pharmaceutical co-crystals, *Cryst. Growth Des.*, 9, 1344–1352 (2009).
48. S. Paulson, M. Vaughn, S. Jessen, Y. Lawal, C. Gresk, B. Yan, T. Maziasz, C. Cook, and A. Karim, Pharmacokinetics of celecoxib after oral administration in dogs and humans: effect of food and site of absorption, *J. Pharmacol. Exp. Ther.*, 297, 638–645 (2001).
49. H. Guzman, M. Tawa, Z. Zhang, P. Ratanabanangkoon, P. Shaw, C. Gardner, H. Chen, J. Moreau, Ö. Almarsson, and J. Remenar, Combined use of crystalline salt forms and precipitation inhibitors to improve oral absorption of celecoxib from solid oral formulations, *J. Pharm. Sci.*, 96, 2686–2702 (2007).
50. J. Dressman, G. Amidon, C. Reppas, and V. Shah, Dissolution testing as a prognostic tool for oral drug absorption: immediate release dosage forms, *Pharm. Res.*, 15, 11–22 (1998).
51. J. Remenar, S. Morissette, M. Peterson, B. Moulton, J. MacPhee, H. Guzman, and Ö. Almarsson, Crystal engineering of novel co-crystals of a triazole drug with 1,4-dicarboxylic acids, *J. Am. Chem. Soc.*, 125, 8456–8457 (2003).
52. J. Remenar, M. Peterson, P. Stephens, Z. Zhang, Y. Zimenkov, and M. Hickey, Celecoxib:nicotinamide dissociation: using excipients to capture the co-crystal's potential, *Mol. Pharmaceut.*, 4, 386–400 (2007).
53. W. Jones, W. Motherwell, and A. Trask, Pharmaceutical co-crystals: an emerging approach to physical property enhancement, *MRS Bull.*, 31, 875–879 (2006).
54. M. Hickey, M. Peterson, E. Manas, J. Alvarez, F. Haefner, and Ö. Almarsson, Hydrates and solid-state reactivity: a survey of  $\beta$ -lactam antibiotics, *J. Pharm. Sci.*, 96, 1090–1099 (2007).
55. D. Dorman, L. Lorenz, J. Occolowitz, L. Spangle, M. Collins, F. Bashore, and S. Baertschi, Isolation and structure elucidation of the major degradation products of cefaclor in the solid state, *J. Pharm.*, 86, 540–549 (1997).
56. P. Proctor, N. Gensmantel, and M. Page, The chemical reactivity of penicillins and other  $\beta$ -lactam antibiotics, *J. Chem. Soc., Perkin Trans. 2*, 1185–1192 (1982).
57. Ö. Almarsson, R. Seburg, D. Godshall, E. Tsai, and M. Kaufman, Solid-state chemistry of a novel carbapenem with a releasable sidechain, *Tetrahedron*, 56, 6877–6885 (2000).
58. M. Pikal, A. Lukes, J. Lang, and K. Gaines, Quantitative crystallinity determinations for  $\beta$ -lactam antibiotics by solution calorimetry: correlations with stability, *J. Pharm. Sci.*, 67, 767–773 (1978).
59. Y. Takeuchi, Y. Takebayashi, M. Sunagawa, Y. Isobe, Y. Hamazume, A. Uemura, and T. Noguchi, The stability of a novel carbapenem antibiotic, meropenem (SM-7338), in a solid state formulation for injection, *Chem Pharm Bull (Tokyo)*, 41, 1998–2002 (1993).
60. Z. Dong, J. Salisbury, D. Zhou, E. Munson, S. Schroeder, I. Prakash, S. Vyazovkin, C. Wight, and D. Grant, Dehydration kinetics of neotame monohydrate, *J. Pharm. Sci.*, 91, 1423–1431 (2002).
61. S. Leung, B. Padden, E. Munson, and D. Grant, Hydration and dehydration behavior of aspartame hemihydrate, *J. Pharm. Sci.*, 87, 508–513 (1998).



62. S. Fleischman, S. Kuduva, J. McMahon, B. Moulton, R. Bailey Walsh, N. Rodriguez-Hornedo, and M. Zaworotko, Crystal engineering of the composition of pharmaceutical phases: multiple-component crystalline solids involving carbamazepine, *Cryst. Growth Des.*, 3, 909–919 (2003).
63. A. Grzesiak, M. Lang, K. Kim, and A. Matzger, Comparison of the four anhydrous polymorphs of carbamazepine and the crystal structure of form I, *J. Pharm. Sci.*, 92, 2260–2271 (2003).
64. R. Harris, P. Ghi, H. Puschmann, D. Apperley, U. Griesser, R. Hammond, C. Ma, K. Roberts, G. Pearce, J. Yates, and C. Pickard, Structural studies of the polymorphs of carbamazepine, its dihydrate, and two solvates, *OPRD*, 9, 902–910 (2005).
65. R. Harris, P. Ghi, H. Puschmann, D. Apperley, U. Griesser, R. Hammond, C. Ma, K. Roberts, G. Pearce, J. Yates, and C. Pickard, Structural studies of the polymorphs of carbamazepine, its dihydrate, and two solvates. [Erratum to document cited in CA143:396664], *OPRD*, 10, 165 (2006).
66. Y. Kobayashi, S. Ito, S. Itai, and K. Yamamoto, Physicochemical properties and bioavailability of carbamazepine polymorphs and dihydrate, *Int. J. Pharm.*, 193, 137–146 (2000).
67. M. Lovrecich, O. Orzincolo, and F. Rubessa, Physico-chemical aspects upon the dissolution behavior of different forms of carbamazepine, *Acta Technol. Legis Medicamenti*, 5, 29–48 (1994).
68. M. Etter, J. MacDonald, and J. Bernstein, Graph-set analysis of hydrogen-bond patterns in organic crystals, *Acta Crystallogr., Sect. B*, B46, 256–262 (1990).
69. M. Etter, Encoding and decoding hydrogen-bond patterns of organic compounds, *Acc. Chem. Res.*, 23, 120–126 (1990).
70. W. Porter, S. Elie, and A. Matzger, Polymorphism in carbamazepine co-crystals, *Cryst. Growth Des.*, 8, 14–16 (2008).
71. C. Betlach, M. Gonzalez, B. McKiernan, C. Neff-Davis, and N. Bodor, Oral pharmacokinetics of carbamazepine in dogs from commercial tablets and a cyclodextrin complex, *J. Pharm. Sci.*, 82, 1058–1060 (1993).
72. M. Hickey, M. Peterson, L. Scoppettuolo, S. Morrisette, A. Vetter, H. Guzman, J. Remenar, Z. Zhang, M. Tawa, S. Haley, M. Zaworotko, and Ö. Almarsson, Performance comparison of a co-crystal of carbamazepine with marketed product, *Eur. J. Pharm. Biopharm.*, 67, 112–119 (2007).
73. T. Gelbrich and M. Hursthouse, Systematic investigation of the relationships between 25 crystal structures containing the carbamazepine molecule or a close analog: a case study of the XPac method, *CrystEngComm*, 8, 448–460 (2006).
74. F. Krahn and J. Mielck, Relations between several polymorphic forms and the dihydrate of carbamazepine, *Pharm. Acta Helvet.*, 62, 247–254 (1987).
75. J. Wang, Z. Berkovitch-Yellin, and L. Leiserowitz, Location of ‘tailor-made’ additives in the crystal and their effect on crystal habit. A study on the host-additive system L-asparagine-L-aspartic acid monohydrate, *Acta Crystallogr., Sect. B*, B41, 341–348 (1985).
76. I. Weissbuch, M. Lahav, and L. Leiserowitz, Crystal morphology control with tailor-made additives; A stereochemical approach, *Adv. Cryst. Growth Res.*, 381–400 (2001).

77. C. Santesson, Addition combinations, *Arch. Exp. Pathol. Pharmacol.*, 118, 313–324 (1926).
78. M. O'Neil, A. Smith, P. Heckelman, and S. Budavari (Eds), *The Merck Index: An Encyclopedia of Chemicals, Drugs, and Biologicals*, 13th Edition, John Wiley & Sons, Ltd, New York (2001).
79. I. Zimmermann, H. Schulz, and W. Ulmer, Blood level courses after i.v., i.m. and p.o. administration of theophylline ethylenediamine and pH-neutral solution of theophylline (author's transl.), *Arzneim.-Forsch.*, 28, 1652–1654 (1978).
80. G. Jenkins and W. Hartung, *The Chemistry of Organic Medicinal Products*, 2nd Edition, John Wiley & Sons, Ltd, New York (1944).
81. J. Bettis, J. Lach, and J. Hood, Effect of complexation with phenobarbital on the biologic availability of theophylline from three tablet formulations, *Am. J. Hosp. Pharm.*, 30, 240–243 (1973).
82. W. Higgins and M. Dunker, Reaction of theophylline with barbiturates, *J. Am. Pharm. Assoc.*, 33, 310–314 (1944).
83. S. Nakao, S. Fujii, T. Sakaki, and K. Tomita, The crystal and molecular structure of the 2:1 molecular complex of theophylline with phenobarbital, *Acta Crystallogr., Sect. B*, B33, 1373–1378 (1977).
84. M. Bender and J. Nelson, Interaction of theophylline and phenobarbital in solution, *J. Chem. Soc., Faraday Trans. 1*, 69, 2074–2079 (1973).
85. Y. Kawashima, S. Lin, M. Ueda, and H. Takenaka, Spray-drying agglomeration. XVI. Preparation of directly compressible powders of a physical mixture and a complex of theophylline-phenobarbital using spray-drying, *Int. J. Pharm.*, 18, 335–343 (1984).
86. A. Leal, M. Leite Inacio, A. Tavares, and M. Baptista, The incompatibility of theophylline and phenobarbital in preparations of theophylline compound suppositories, *Rev. Portuguesa Farmac.*, 29, 75–83 (1979).
87. H. Wiedenfeld and F. Knoch, The crystal structure of the theophylline-urea complex, *Archiv. Pharm. (Weinheim, Ger.)*, 319, 654–659 (1986).
88. J. Madarasz, P. Bombicz, K. Jarmi, M. Ban, G. Pokol, and S. Gal, Thermal, FTIR and XRD study on some 1:1 molecular compounds of theophylline, *J. Therm. Anal. Calor.*, 69, 281–290 (2002).
89. E. Shefter, Structural studies on complexes. IV. Crystal structure of a 1:1 5-chlorosalicylic acid and theophylline complex, *J. Pharm. Sci.*, 58, 710–714 (1969).
90. S. Zaitu, Y. Miwa, and T. Taga, A 2:1 molecular complex of theophylline and 5-fluorouracil as the monohydrate, *Acta Crystallogr., Sect. C*, C51, 1857–1859 (1995).
91. S. Zaitu, Y. Miwa, and T. Taga, 1:1 Molecular complex of theophylline and p-nitroaniline, *Acta Crystallogr., Sect. C*, C51, 2390–2392 (1995).
92. A. Trask, W. Motherwell, and W. Jones, Physical stability enhancement of theophylline via co-crystallization, *Int. J. Pharm.*, 320, 114–123 (2006).
93. A. Antipas and M. Landis, Solid-state excipient compatibility testing, *Drugs Pharm. Sci.*, 153, 419–458 (2005).
94. G. Damien, Development of a solid dosage form compatibility studies on the active substance-excipients, *STP Pharma Prat.*, 14, 303–310 (2004).



95. A. Serajuddin, A. Thakur, R. Ghoshal, M. Fakes, S. Ranadive, K. Morris, and S. Varia, Selection of solid dosage form composition through drug-excipient compatibility testing, *J. Pharm. Sci.*, 88, 696–704 (1999).
96. J. Sims, J. Carreira, D. Carrier, S. Crabtree, L. Easton, S. Hancock, and C. Simcox, A new approach to accelerated drug-excipient compatibility testing, *Pharm. Dev. Technol.*, 8, 119–126 (2003).
97. M. Vrbinc and V. Kmetec Physical stability and compatibility of solid-state substances in preformulation studies of solid dosage forms, *Farm. Vestn. (Ljubljana, Slov.)*, 53, 367–376 (2002).
98. N. Wyttenbach, C. Birringer, J. Alsenz, and M. Kuentz, Drug-excipient compatibility testing using a high-throughput approach and statistical design, *Pharm. Dev. Technol.*, 10, 499–505 (2005).
99. M. Bergren, R. Chao, P. Meulman, R. Sarver, M. Lyster, J. Havens, and M. Hawley, Solid phases of delavirdine mesylate, *J. Pharm. Sci.*, 85, 834–841 (1996).
100. B. Rohrs, T. Thamann, P. Gao, D. Stelzer, M. Bergren, and R. Chao, Tablet dissolution affected by a moisture mediated solid-state interaction between drug and disintegrant, *Pharm. Res.*, 16, 1850–1856 (1999).
101. J. Schmidt and W. Snipes, Free radical formation in a gamma-irradiated pyrimidine-purine co-crystal complex, *Int. J. Rad. Bio. Related St. Phys., Chem. Med.*, 13, 101–109 (1967).
102. M. Etter, Hydrogen bonds as design elements in organic chemistry, *J. Phys. Chem.*, 95, 4601–4610 (1991).
103. G. Desiraju, Crystal engineering: solid state supramolecular synthesis, *Curr. Opin. Solid State Mater. Sci.*, 2, 451–454 (1997).
104. C. Aakeroy, J. Desper, and M. Fasulo, Improving success rate of hydrogen-bond driven synthesis of co-crystals, *CrystEngComm*, 8, 586–588 (2006).
105. C. Aakeroy, A. Beatty, and B. Helfrich, ‘Total synthesis’ supramolecular style: design and hydrogen-bond-directed assembly of ternary supermolecules, *Angew. Chem., Int. Ed.*, 40, 3240–3242 (2001).
106. B. Saha, A. Nangia, and M. Jaskolski, Crystal engineering with hydrogen bonds and halogen bonds, *CrystEngComm*, 7, 355–358 (2005).
107. W. Cross, N. Blagden, R. Davey, R. Pritchard, M. Neumann, R. Roberts, and R. Rowe, A whole output strategy for polymorph screening: combining crystal structure prediction, graph set analysis, and targeted crystallization experiments in the case of diflunisal, *Cryst. Growth Des.*, 3, 151–158 (2003).
108. M. Rafilovich and J. Bernstein, Serendipity and four polymorphic structures of benzidine, C<sub>12</sub>H<sub>12</sub>N<sub>2</sub>, *J. Am. Chem. Soc.*, 128, 12185–12191 (2006).



# 4

## Complex Formation of Surfactants with Aromatic Compounds and their Pharmaceutical Applications

*Yuji Ohashi, Keiju Sawada and Nahoko Iimura*

### 4.1 Introduction

Surface-active agents (surfactants) are a series of compounds consisting of hydrophobic and hydrophilic moieties. An aqueous solution containing a surfactant forms micelles above a critical concentration, which is called the critical micelle concentration (cmc). The value of the cmc is approximately  $10^{-4}$ – $10^{-2}$  M ( $1\text{ M} = 1\text{ mol dm}^{-3}$ ) in the case of cationic or anionic surfactants, and below  $10^{-4}$  in the case of nonionic surfactants [1]. When a third nonreactive component, such as a hydrocarbon or polar organic molecule, is added to an aqueous solution of surfactant below its cmc, it is found that the solubility of the additive in the solution is almost the same as that in pure water [2]. When the surfactant solution is above its cmc, the solubility of the additive is markedly greater than that in pure water and the thermodynamic activity of the additive is appreciably smaller than the corresponding thermodynamic activity in pure water. These results were firmly established by McBain and Hutchinson [3]. McBain proposed the term ‘solubilization’ to describe these phenomena.

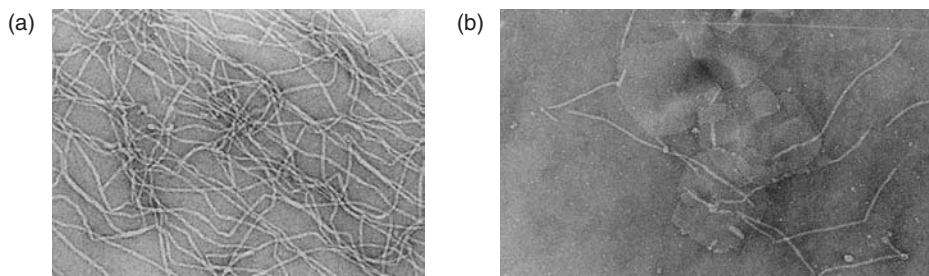
Although a number of methods have been proposed to determine the cmc of surfactants, such as electrical conductance [4, 5], transference number [6, 7], dye absorption [8–10], solubilization [11–15], surface tension [16, 17], partial molar volume [18, 19],

refractive index [20], light scattering [21–23], freezing point [24], osmotic pressure [25], vapour pressure [26], solubility [27], electromotive force [28], sound velocity [29], polarographic determination [30, 31], diffusion [32], viscosity [33, 34], dialysis [35], electrophoretic mobility [36], ultraviolet and infrared absorption [37, 38], and fluorescence depolarization [39], there are very few methods to determine the structure of the micelle. From  $^1\text{H-NMR}$  chemical shifts it was proposed that in aqueous solution the benzene molecule should be placed near the surface of the micelle of hexadecyltrimethylammonium bromide [40]. The same model was proposed from pulse radiolysis [41]. However, the opposite model, that is, with the benzene molecule situated at the centre of the micelle, was proposed from the analysis of UV absorption spectra [42]. The current model of micellar aggregates in an aqueous solution affords insufficient explanation for the available data and often gives rise to contradictions [43, 44].

Micelles with platelet or rod-like shapes were obtained in supersaturated aqueous solution containing a surfactant and an aromatic compound. In particular, chiral surfactants are known to produce enantiomeric precipitation, while most of the others give symmetric micelles [45–47].

Electron microscopy showed that gigantic rod-like micelles were formed in viscoelastic solutions in the presence of several aromatic compounds as shown in Figure 4.1(a) [48–50]. It was very surprising that single crystals, mixed with the gigantic micelles, were found to co-exist in these viscoelastic solutions, as shown in Figure 4.1(b) [51]. From the aqueous solution composed of hexadecyltrimethylammonium bromide and several phenolic derivatives, similar crystalline materials were obtained. These were assumed to be complexes composed of the surfactant and the respective phenolic derivative based on elementary and thermal analyses [52]. The crystal structure was analysed by X-rays and a variety of complexes with similar structures were obtained [53].

In the earlier part of this chapter, the various structures of the complexes between the surfactants and aromatic compounds are compared and the main factors behind complex formation are discussed, based on the X-ray crystallographic data [54]. Several examples of complex formation involving pharmaceutical substances are given in the latter part of the chapter [55].



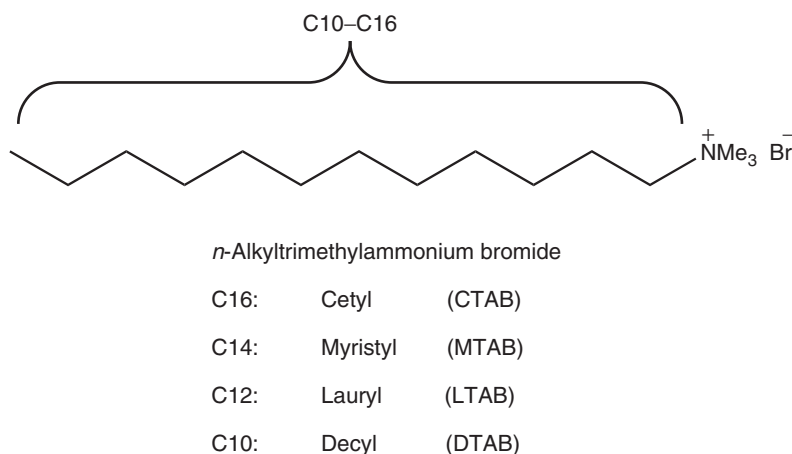
**Figure 4.1** Electron micrographs showing (a) rod-like micelles and (b) single crystals in addition to rod-like micelles. The photographs were kindly supplied by Professor H. Hirata

## 4.2 Structures of the Complexes Formed Between Surfactants and Aromatic Compounds [53]

A variety of complex crystals were obtained from aqueous methanol solutions of cationic surfactants and aromatic compounds in the molar ratio of 1:1, 2:1, 2:3 or 3:2. The formation of complex crystals between four surfactant molecules such as hexadecyl-, tetradecyl-, dodecyl- and decyltrimethylammonium bromide (CTAB, MTAB, LTAB and DTAB, respectively), which are shown in Figure 4.2, and various aromatic compounds are listed in Table 4.1. Nine kinds of complex crystals suitable for X-crystallography were characterized. All crystals are colourless and plate-like in appearance.

### 4.2.1 Crystal structure of I (CTAB-*p*-hydroxybenzoic acid)

The crystal structure of I viewed along the *b*- and *c*-axes is shown in Figure 4.3. The asymmetric unit comprises three hexadecyltrimethylammonium (CTA) cations, three bromide anions, two *p*-hydroxybenzoic acid (*p*-HBA) molecules, and one solvent water molecule. The CTA cations, each with an all trans conformation, run anti-parallel alternately along the *a*-axis. The *p*-HBA molecule is sandwiched by the alkyl chains of CTA and bromide anions. The carboxyl groups of the two independent *p*-HBA molecules self-associate to a dimer via the familiar carboxylic acid dimer synthon. The hydroxyl groups of A and B *p*-HBA also make hydrogen bonds. The hydroxyl group of A forms a hydrogen bond with the solvent water molecule which is hydrogen bonded with the bromide anions. The bromide anions and trimethylammonium moieties form ribbons along the *c*-axis. The hydrophobic ribbon comprising the alkyl group of CTA and *p*-HBA also extends along the *c*-axis. The ribbons are packed alternately along the *a*-axis, thereby forming a 'planar sheet' parallel to the *ac*-plane.

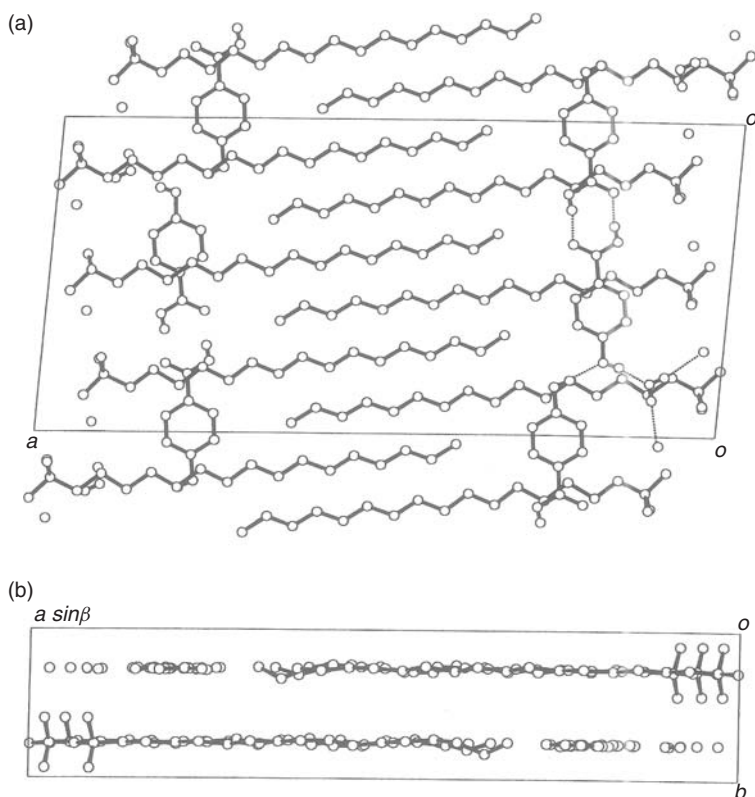


**Figure 4.2** Chemical formulations of four cationic surfactants: CTAB, MTAB, LTAB, and DTAB

**Table 4.1** Complex crystal formation between aromatic compounds and cationic surfactants, CTAB, MTAB, LTAB and DTAB<sup>a</sup>

Aromatic compound	CTAB	MTAB	LTAB	DTAB
<i>o</i> -Iodophenol	A	A	A	A
<i>m</i> -Iodophenol	B	N	B	B
<i>p</i> -Iodophenol	B	N	B	B
<i>p</i> -Cresol	A	N	N	N
<i>m</i> -Cyanophenol	A	N	B	B
<i>o</i> -Hydroxybenzene	B	B	B	B
<i>m</i> -Hydroxybenzene	B	A	B	B
<i>p</i> -Hydroxybenzene	A	N	C	A
<i>o</i> -Toluic acid	C	N	B	B
<i>m</i> -Toluic acid	C	N	B	B
<i>p</i> -Toluic acid	C	N	N	B
<i>o</i> -Phthalic acid	C	N	N	N
<i>m</i> -Phthalic acid	C	N	N	N
<i>p</i> -Phthalic acid	C	N	N	N
Hydroquinone	A	N	B	B
1,4-Dimethoxybenzene	C	N	C	N
<i>p</i> -Benzoquinone	N	N	C	N
1,4-Cyclohexanediol	C	N	C	C
Naphthalene	C	N	N	B
1-Naphthol	B	N	B	B
2-Naphthol	B	N	B	B
Indole	A	N	B	B
7-Hydroxycoumarine	C	C	C	B
Coumarine	C	N	C	B
2-Aminopyridine	B	B	B	B
3-Aminopyridine	B	N	B	B
4-Aminopyridine	C	N	B	B
2-Hydroxypyridine	C	N	B	B
3-Hydroxypyridine	C	N	N	B
4-Hydroxypyridine	C	N	N	B
3-Cyanopyridine	C	C	C	C
4-Cyanopyridine	C	N	C	C
4-Dimethylaminopyridine	C	N	N	B
2-Phenylpyridine	A	B	B	B
3-Phenylpyridine	B	B	B	B
4-Phenylpyridine	A	B	B	B
2,2'-Bipyridine	B	B	B	B
2,4-Bipyridine	B	B	B	B
4,4'-Bipyridine	B	B	B	B
Anthracene	C	N	B	B
Phenanthrene	B	B	B	B
Benzo[ <i>h</i> ]quinoline	C	N	B	B
Benzo[ <i>f</i> ]quinoline	C	N	B	B
1,1'-Biphenyl-4-ol	A	A	A	B
Bibenzyl	C	N	C	C
Biphenyl	A	A	A	A
Carbazole	A	B	B	B
Dibenzofuran	A	B	B	B
Diphenylamine	A	B	B	B

<sup>a</sup>The letters A, B, C, and N indicate single crystal formed, crystalline powder formed, crystals not formed, and not tried, respectively.



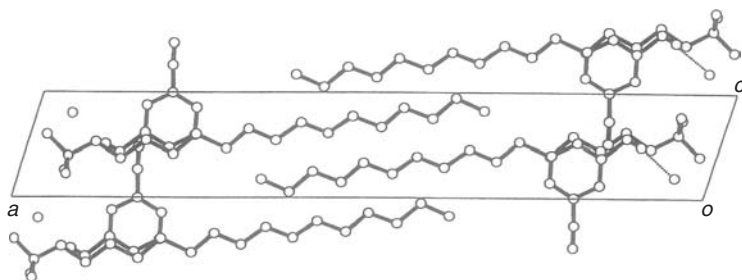
**Figure 4.3** Crystal structure of I viewed along the (a) *b*-axis and (b) *c*-axis. Reprinted from [53] with permission from The Chemical Society of Japan. Copyright 1998

#### 4.2.2 Crystal structure of II (CTAB-*m*-cyanophenol)

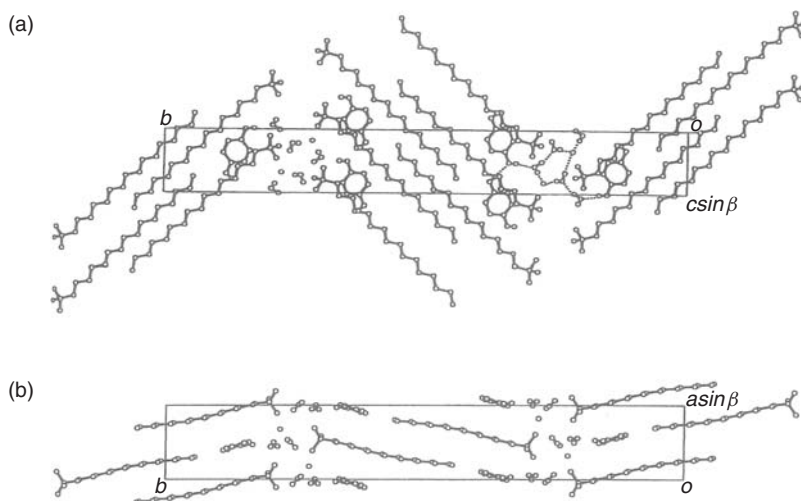
The crystal structure of II viewed along the *b*-axis is shown in Figure 4.4. Since the occupancy factor of the *m*-cyanophenol (*m*-CP) molecule is 0.5, there are no unusually short contacts between the *m*-CP molecules. The CTA cations, each with an all trans conformation, run anti-parallel alternately along the *a*-axis. The *m*-CP molecule is sandwiched by the alkyl chains of CTA along the *b*-axis and is situated between the CTA cation and the bromide anion along the *a*-axis, resembling the structure of I. The OH group of *m*-CP is hydrogen bonded to the bromide anion. The hydrophilic ribbon composed of the bromide anion and trimethylammonium cation and the hydrophobic ribbon composed of the alkyl group of CTA and *m*-CP are packed alternately, making a 'planar sheet' parallel to the *ac*-plane. Again, these features are substantially the same as those of crystal I.

#### 4.2.3 Crystal structure of III (CTAB-*p*-cresol)

The crystal structure of III is isomorphous to the crystal structure of II. The *p*-cresol molecule is disordered over two positions with occupancy factors of 0.65 and 0.35,



**Figure 4.4** Crystal structure of II viewed along the *b*-axis. Hydrogen bonds are shown with dashed lines. Reprinted from [53] with permission from The Chemical Society of Japan. Copyright 1998



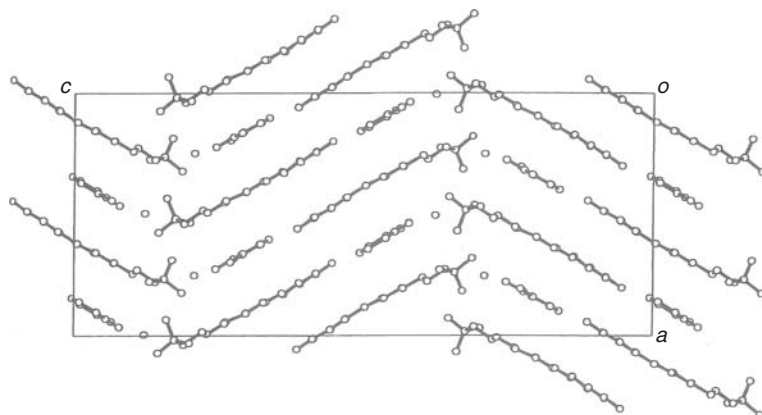
**Figure 4.5** Crystal structure of IV viewed along the (a) *a*-axis and (b) *c*-axis. Hydrogen bonds are shown with dashed lines. Reprinted from [53] with permission from The Chemical Society of Japan. Copyright 1998

respectively. To avoid short contacts between the *p*-cresol molecules, the conformations of the neighbouring *p*-cresol molecules along the *c*-axis are different from each other. The O-H group of *p*-cresol is hydrogen bonded to the bromide anion.

#### 4.2.4 Crystal structure of IV (CTAB-hydroquinone)

The crystal structure of IV viewed along the *a*- and *c*-axes is shown in Figure 4.5. There is one CTA cation, one bromide anion, one hydroquinone, and two water molecules in the asymmetric unit. One OH group of hydroquinone is hydrogen bonded with another OH group of the neighbouring hydroquinone molecule along the *c*-axis. The latter OH group is hydrogen bonded to the water molecule, which is also hydrogen bonded to the bromide anion and another water molecule. These hydrogen bonds connect the hydrophilic residues into a layer parallel to the *ac*-plane. As shown in the crystal





**Figure 4.6** Crystal structure of V viewed along the *b*-axis. Reprinted from [53] with permission from The Chemical Society of Japan. Copyright 1998

structures I–III, the CTA cation, the bromide anion and hydroquinone are packed as a sheet along the *a*-axis. However, in IV, the sheet is bent at the hydrophilic layer. This makes a ‘pleated sheet’ parallel to the *bc*-plane.

#### 4.2.5 Crystal structure of V (CTAB–*o*-iodophenol)

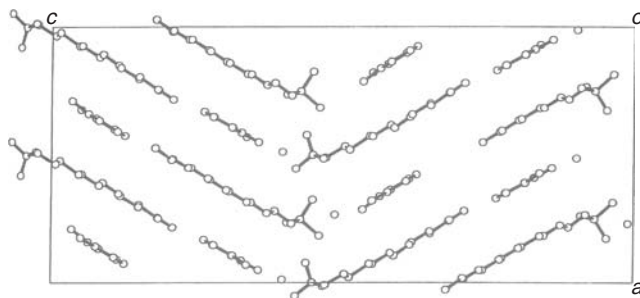
The crystal structure of V viewed along the *b*-axis is shown in Figure 4.6. There are two CTA cations and bromide anions and two *o*-iodophenol molecules in the asymmetric unit. The OH groups of the *o*-iodophenol molecules make hydrogen bonds with the bromide anions. The bromide anions and trimethylammonium moieties form ribbons along the *b*-axis. The hydrophobic ribbon comprising the alkyl group of CTA and *o*-iodophenol extends along the *a*-axis. The ribbons are packed alternately, making a ‘pleated sheet’ parallel to the *ac*-plane. The pleated sheets are piled up along the *b*-axis to form the crystal. These features are substantially the same as those of crystal IV.

#### 4.2.6 Crystal structure of VI (MTAB–*o*-iodophenol)

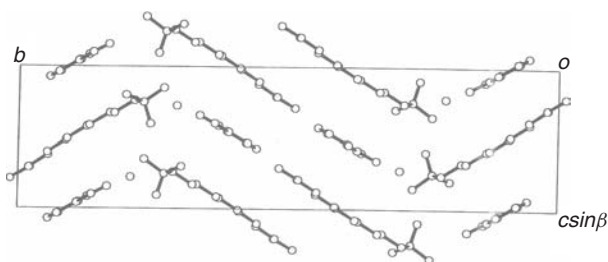
The crystal structure of VI viewed along the *b*-axis is shown in Figure 4.7. There are two MTA cations and bromide anions and two *o*-iodophenol molecules in the asymmetric unit. The OH group of the *o*-iodophenol is hydrogen bonded with the bromide anions. The packing mode of the MTA cations and the bromide anions is nearly the same as that of the CTAB complex, crystal V. The ‘pleated sheet’ is parallel to the *ac*-plane and the sheets are stacked along the *b*-axis.

#### 4.2.7 Crystal structures of VII (LTAB–*o*-iodophenol) and VIII (DTAB–*o*-iodophenol)

The crystal structure of VII viewed along the *a*-axis is shown in Figure 4.8. The asymmetric unit comprises one LTA cation, one bromide anion and one *o*-iodophenol molecule. The OH group of iodophenol is hydrogen bonded with the bromide. The



**Figure 4.7** Crystal structure of VI viewed along the *b*-axis. Reprinted from [53] with permission from The Chemical Society of Japan. Copyright 1998

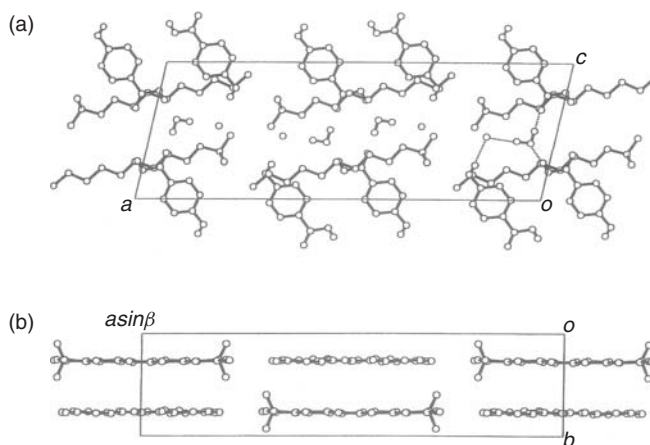


**Figure 4.8** Crystal structure of VII viewed along the *a*-axis. Reprinted from [53] with permission from The Chemical Society of Japan. Copyright 1998

packing mode of the LTA cations and the bromide anions are nearly the same as those of V and VI. The ‘pleated sheet’ is parallel to the *bc*-plane and the sheets are stacked along the *a*-axis. The crystal structure of VIII is isostructural to that of the LTAB complex, VII.

#### 4.2.8 Crystal structure of IX (DTAB-*p*-hydroxybenzoic acid)

The crystal structure of IX viewed along the *b*- and *c*-axes is shown in Figure 4.9. There is one DTA cation, one bromide anion, one *p*-HBA and one solvent water molecule in the asymmetric unit. The anti-parallel DTA cations with an all trans conformation run alternately along the  $[1\ 0\ \bar{1}]$  direction. The carboxyl groups of two *p*-HBA molecules make a dimer through the mediation of a water molecule (see Figure 4.9a). The hydroxyl group of *p*-HBA makes a hydrogen bond with the solvent water molecule. The *p*-HBA molecule is sandwiched by the alkyl groups of DTA cations. The bromide anions and trimethylammonium moieties form ribbons along the *c*-axis, and the hydrophobic ribbon, composed of the alkyl group of DTA, extends along the *c*-axis. Since the direction of the dimer structure of *p*-HBA is not parallel to the *c*-axis, anti-parallel contacts of the alkyl groups of DTA cations, see previously, are not formed. This causes the layer structure to be parallel to the *bc*-plane. The inner part of the layer is hydrophobic while both sides of the layer are hydrophilic. A ‘planar sheet’ is formed parallel to the *ac*-plane.



**Figure 4.9** Crystal structures of IX viewed along the (a) *b*-axis and (b) *c*-axis. Hydrogen bonds are shown with dashed lines. Reprinted from [53] with permission from The Chemical Society of Japan. Copyright 1998

### 4.3 Complex Formation of Aromatic Compounds Containing an Hetero Ring [54]

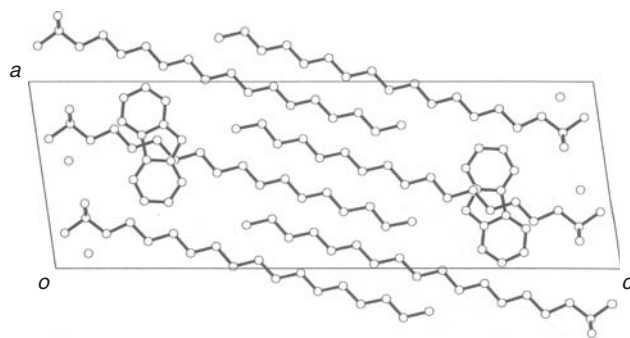
In the previous section, only the complexes comprising aromatic compounds without hetero rings were analysed. A variety of aromatic compounds with a hetero ring also form complexes with cationic surfactants. In this section, five kinds of complex crystal structures are discussed.

#### 4.3.1 Crystal structures of X (CTAB–acridine), XI (CTAB–phenoxazine) and XII (CTAB–phenothiazine)

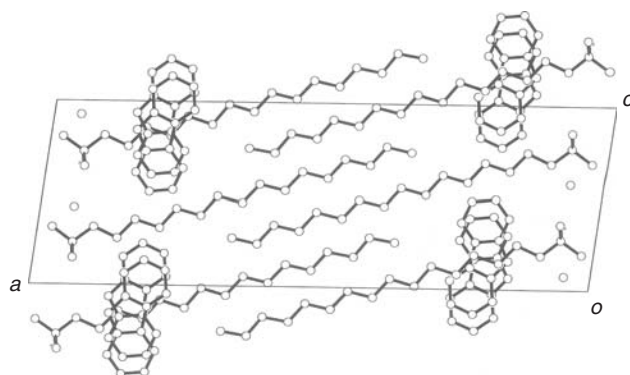
The crystal structure of X is isomorphous to crystals II and III. Since the occupancy factor of the acridine molecule is 0.5, there is one acridine molecule, on average, in the two unit cells along the *c*-axis. Although there is no hydrogen bond between acridine and the bromide anion, the intermolecular distances suggest the presence of a weak C–H...Br contact. The hydrophilic ribbon comprises the bromide anion and trimethylammonium cation, and the hydrophobic ribbon comprises the alkyl group and acridine. The ribbons are packed alternately to define a ‘planar sheet’ parallel to the *ac*-plane. The crystal structures of XI and XII are isostructural to X, with the same pattern of occupancy of the heteroaromatic rings.

#### 4.3.2 Crystal structures of XIII (CTAB–dibenzofuran) and XIV (CTAB–carbazole)

The *a*-axes of crystals XIII and XIV are twice the length of those of X (and XI and XII), as shown in Figures 4.10 and 4.11. This means that there are two CTABs and one dibenzofuran or carbazole molecule in the asymmetric unit, respectively. Although the two crystal structures are isostructural, the dibenzofuran molecule is ordered whereas



**Figure 4.10** *Crystal structure of XIII viewed along the b-axis*

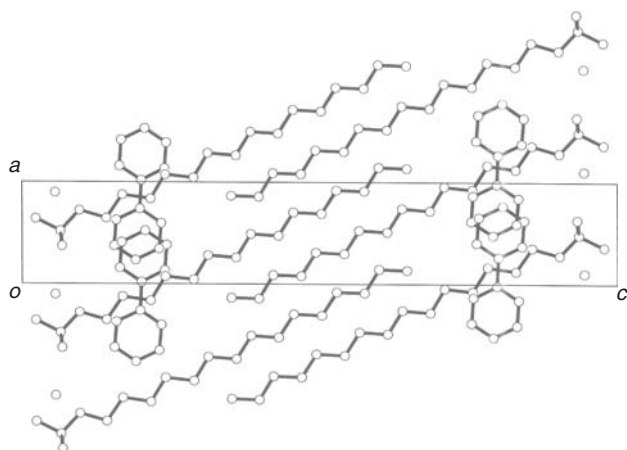


**Figure 4.11** *Crystal structure of XIV viewed along the b-axis*

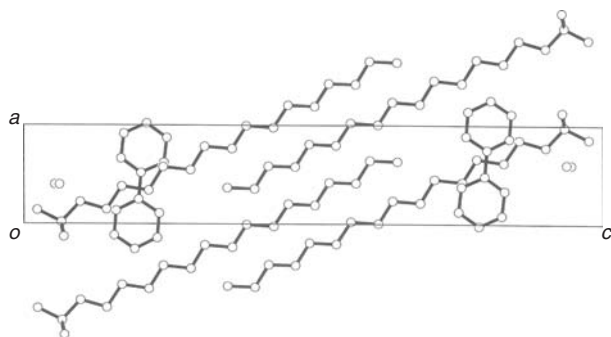
carbazole occupies two disordered positions in the ratio of 1:1. This may indicate that the interaction between dibenzofuran and the alkyl chain is stronger than that of the carbazole complex.

#### 4.4 Complex Formation of Biphenyl with Cationic Surfactants [54]

In the two previous sections, aromatic compounds carry functional groups such as hydroxyl, carboxyl, or halide or contain hetero atoms within the ring. As shall be now demonstrated, it is also possible to form complexes with cationic surfactants even if the aromatic compounds have no functional groups or hetero atoms, such as biphenyl, anthracene or phenanthrene. Single crystals suitable for X-ray crystallography were obtained for a series with biphenyl when the surfactant was CTAB (XV), MTAB (XVI) and DTAB (XVII) but not with LTAB. However, from the X-ray powder pattern, the LTAB complex was shown to be isostructural to the other cationic surfactant complexes. There is one surfactant and a half biphenyl molecule in the asymmetric unit, as shown in Figure 4.12 for the CTAB–biphenyl (XV) structure, being representative for this series.



**Figure 4.12** Crystal structure of XV viewed along the *b*-axis



**Figure 4.13** Crystal structure of XVIII viewed along the *b*-axis

Since the occupancy factor of the biphenyl molecule in each site is 0.5, the biphenyl molecule occupies alternate positions along the *a*-axis. Therefore, there are no short contacts between them along the *a*-axis. This observation is distinct from the anthracene complex crystal, in which the anthracene molecules adopt a disordered structure along the *a*-axis, see above. This may suggest that there are fairly strong interatomic interactions between the atoms of the biphenyl and surfactant molecules compared with those of the anthracene complex.

When the biphenyl molecule in the CTAB complex was replaced with 2-phenylpyridine or 4-phenylpyridine, in which one carbon atom of a ring is replaced with a nitrogen atom, isostructural crystals were obtained: CTAB–2-phenylpyridine (XVIII) (Figure 4.13) and CTAB–4-phenylpyridine (XIX). Since the occupancy factors of the 2- and 4-phenylpyridine molecules are 0.5, those molecules occupy alternate sites along the *a*-axis. There are no short contacts between the 2-phenylpyridine molecules or between the 4-phenylpyridine molecules in their respective crystal structures. These structures are essentially the same as that of the CTAB–biphenyl (XV) as shown in

Figure 4.12. This result suggests that the interatomic interaction of the alkyl group with the pyridine ring is nearly the same as that with the phenyl ring.

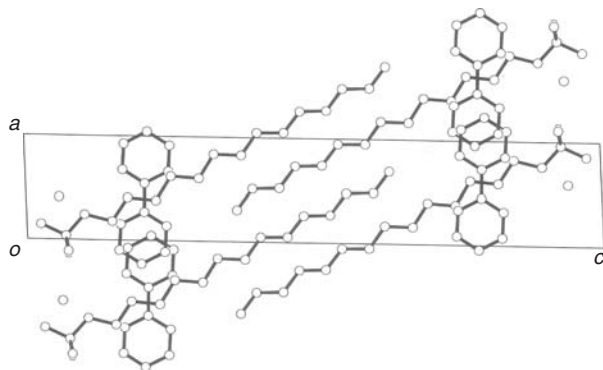
#### 4.5 Complex Formation of Odd-Number Surfactants with Biphenyl [54]

In the previous sections, all described complexes were formed between the aromatic compounds and surfactant molecules which have alkyl groups with even-number carbon atoms, such as hexadecyl, tetradecyl, dodecyl and decyl groups. These even-carbon-chain surfactants are utilized widely *in vitro* and are commercially available. However, there are several odd-carbon-chain fatty acids *in vivo*, although most fatty acids have even-carbon chains. The former acids are digested in the rumen of ruminants. The odd-carbon-chain surfactants were synthesized from the corresponding odd-carbon-chain fatty acids [56–58]. Two kinds of surfactant molecules with alkyl groups of odd-number carbon atoms, pentadecyl- and tridecyltrimethylammonium bromides (PTAB and TTAB, respectively), were prepared and their complexes with biphenyl were formed. The crystals were obtained from aqueous solutions.

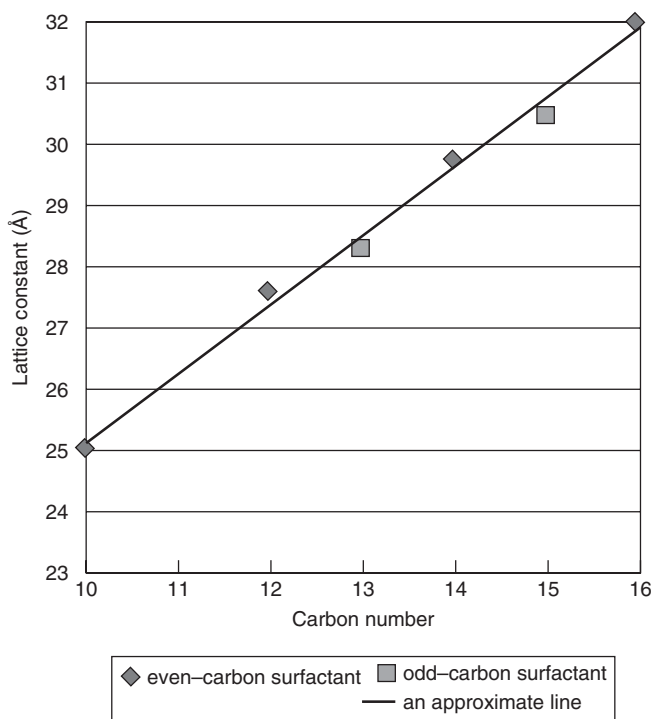
The two crystals, i.e. PTAB–biphenyl (XX) and TTAB–biphenyl (XXI), are isostructural. The crystal structure of XX viewed along the *b*-axis is shown in Figure 4.14. These structures are also isostructural to those of the even-carbon-chain surfactant complexes.

Since the crystal structures of the odd-carbon-chain surfactant complexes with biphenyl have essentially the same packing pattern as those of the even-carbon-chain surfactant complexes, it may be assumed that the unit-cell length is proportional to the carbon number of the surfactant whether the carbon number is even or odd. The relationship between the number of carbon atoms in the alkyl chain of the surfactant and the *c*-axis length of the complex crystal with biphenyl is shown in Figure 4.15. The *c*-axis length has a linear correlation with the carbon number of the alkyl chain of the surfactant. However, the *c*-axis lengths with odd numbers, 13 and 15, are shorter than the corresponding ones estimated from the average line, whereas those with even number, 12, 14 and 16, are longer than the averaged values.

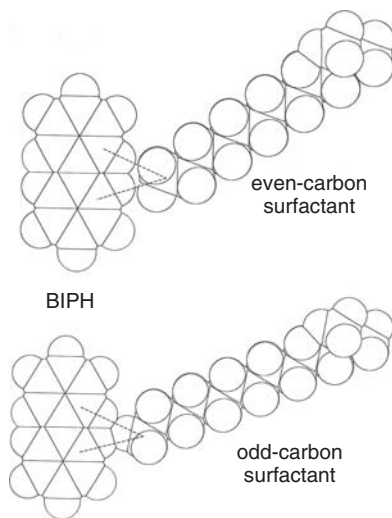
As shown in Figure 4.16, the alkyl group of the surfactant molecule contacts with one side of the biphenyl molecule. The contact modes of even-carbon-chain surfactant



**Figure 4.14** Crystal structure of XX viewed along the *b*-axis



**Figure 4.15** Relationship between the number of carbon atoms in the alkyl chain of the surfactant to the length of the c-axis in the crystal of the complex. The rhombuses and squares indicate surfactants with alkyl groups containing an even and an odd number of carbon atoms, respectively



**Figure 4.16** Contact modes of the even- and odd-numbered alkyl groups of the surfactants with the biphenyl group



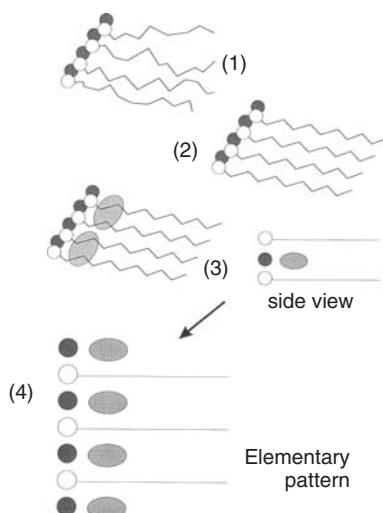
complex and odd-carbon-chain surfactant complex are drawn using space-filling models. The distances between the terminal carbon atom of the alkyl group and the ring carbon atoms of biphenyl are 3.98 and 3.89 Å for the even-carbon-chain surfactant complex and the comparable distances are 4.17 and 3.81 Å for the odd-carbon-chain surfactant complex. Since the orientations of the terminal methyl groups of the even- and odd-carbon-chain surfactant molecules are different from each other, the contact of the terminal methyl group of the odd carbon chain is closer than the corresponding one of the even carbon chain. This is a reason why the *c*-axis lengths of the odd- and even-carbon-chain surfactant complexes are shorter and longer than the averaged ones, respectively.

#### 4.6 Common Packing Mode in the Complexes [53]

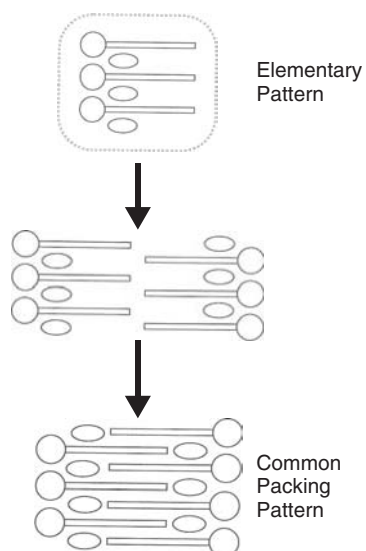
The complexes formed between the cationic surfactants and the aromatic compounds, as described above, have the following common features:

- (1) The trimethylammonium moieties of the surfactant molecules and bromide anions form a hydrophilic layer.
- (2) The alkyl chains of the surfactant molecules with an all trans conformation grow from the hydrophilic layer described in (1).
- (3) The aromatic molecule is sandwiched by the alkyl groups of the two surfactant molecules and makes hydrogen bond(s) with the bromide anion directly or through the solvent water molecule if the aromatic molecule has the hydrogen donor group.
- (4) The above aggregation forms a layer structure, which is schematically shown in Figure 4.17. One side of the layer is hydrophilic composed of the trimethylammonium cation and the bromide anion. In some cases the solvent water molecules are included. The other side of the layer is hydrophobic, being composed of the alkyl groups of the surfactant molecules and the aromatic molecules if the alkyl chain is short.
- (5) The alkyl chains of the two layers are used to make a common packing pattern of the layer structure in which anti-parallel stacking of the alkyl chains is formed, as shown in Figure 4.18.
- (6) If the alkyl chain of the surfactant is long enough, the anti-parallel alkyl chains have close contacts between them as shown in the CTAB complexes. As the alkyl chain becomes shorter, the interaction between the alkyl group and the aromatic compound prevents such a close contact between the alkyl groups.
- (7) When the above common packing pattern is stacked up along an axis of the crystal, there are two modes of stacking, the straight and inclined, as shown in Figure 4.19. In the straight mode a flat sheet is formed, whereas a pleated sheet is formed in the bent mode.

Figure 4.20 shows the crystal structures of CTAB itself viewed along the *b*- and *a*-axes [59]. The structure also has the characteristics of the complexes described above except that it contains no aromatic compound. A flat sheet composed of CTAs and the bromide anions is formed. When CTAB makes a complex with an aromatic compound, the anti-parallel alkyl groups of the neighbouring surfactant molecules slide in opposite

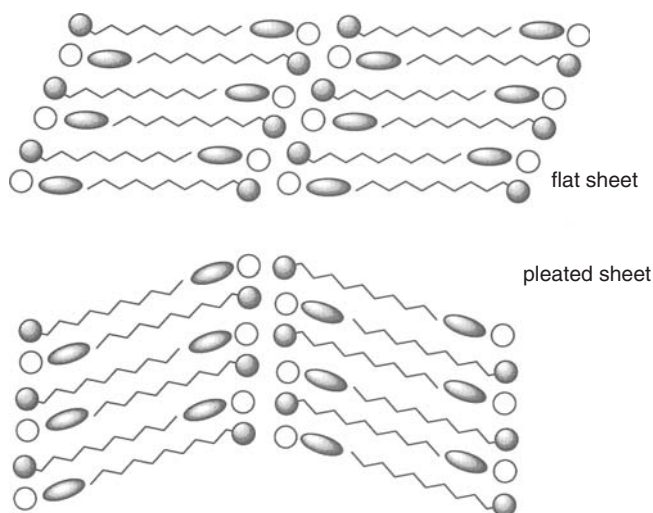


**Figure 4.17** Schematic drawing of the elementary interactions between the cationic surfactants and aromatic compounds

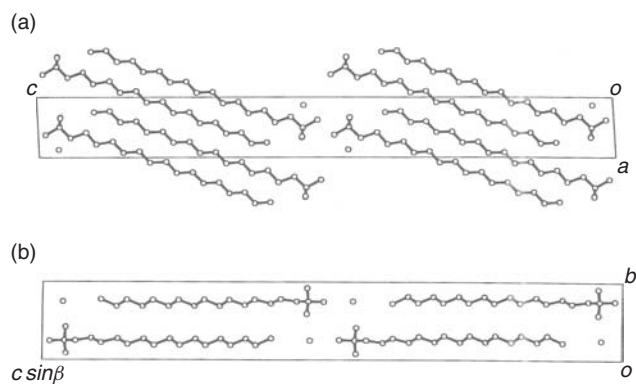


**Figure 4.18** Schematic drawing of the common packing pattern formed from the elementary pattern

directions to each other, as indicated in Figure 4.21. A gap is formed in the region between the bromide anion and the end of the hexadecyl group. The aromatic molecules occupy these gaps. The aromatic molecule in the gap is therefore sandwiched by the alkyl chains of the upper and lower CTA molecules. The sandwiched aromatic molecule makes a hydrogen bond with the bromide anion or a solvent water molecule if it has



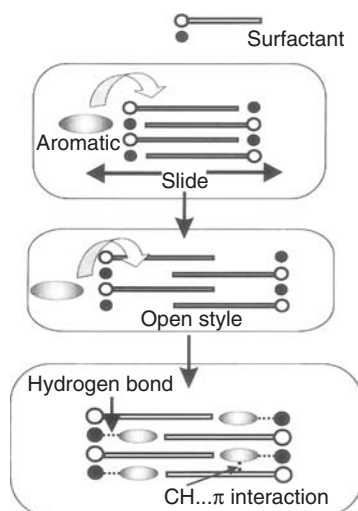
**Figure 4.19** Two modes of common packing pattern: flat and pleated sheets



**Figure 4.20** Crystal structure of CTAB viewed along the (a) *b*-axis and (b) *a*-axis

a hydrogen donor. However, since the aromatic molecule has no hydrogen donor, it also occupies the same position as shown in the biphenyl–CTAB complex. These facts suggest that the layer structure, as shown in Figure 4.18, is a common pattern of the complexes formed between the surfactant and the aromatic compound when they are dissolved in an aqueous solution.

The common pattern of the complex formation led to the idea that a similar structure should be formed when the aromatic compound is dissolved in an aqueous solution containing surfactant. Sakaiguchi *et al.* reported that elongated rod-like micelles were observed by electron microscopy when CTAB and several aromatic compounds, such as *o*-iodophenol and salicylic acid, were dissolved in aqueous solutions [48–50]. These

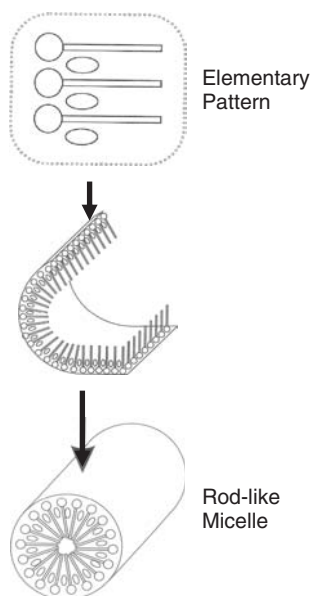


**Figure 4.21** Schematic drawing of the formation of the complex with aromatic compounds starting from the pure CTAB structure

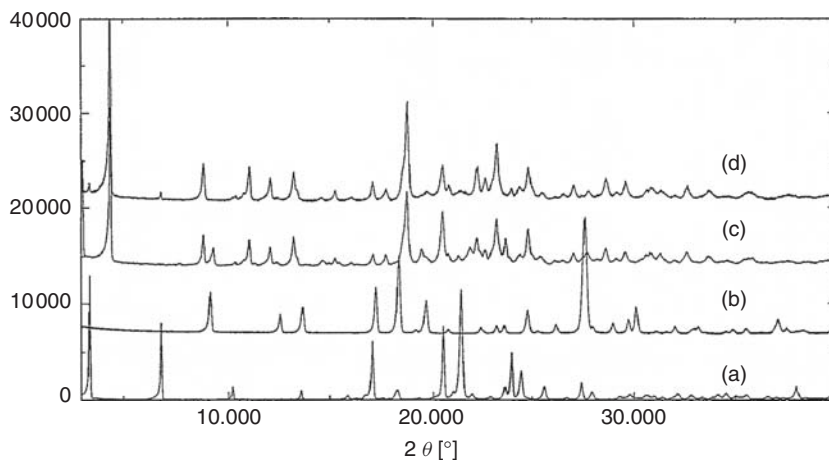
solutions often had high viscoelasticity, which is caused by the cobweb-like entanglement of the elongated rod-like micelles. These results suggest that the common pattern of the complex formation is responsible for the formation of the rod-like micelles, since the rod-like micelles and the single crystals co-exist as seen in a transmission electron micrograph of the solution. The layer compound of the CTA cation and the bromide anion and aromatic compound, as shown in Figure 4.18, is rolled to form a tube, which is the rod-like micelle. Figure 4.22 shows a schematic model of a section of the rod-like micelles. The aromatic molecules connect the surfactant molecules near the surface of the rod. Complex formation between the surfactant and aromatic molecules probably strengthens the rod formation in the aqueous solution. The diameter of the rod may be more than 70 Å, since the CTA cation is about 30 Å long and about 5–10 Å may be necessary in the central part of the rod to accommodate the entangled terminal parts of the alkyl chains. Moreover, the hydrophilic exterior should be covered with water molecules. This suggests the diameter of the rod-like micelles may be estimated to be 70–80 Å. This value is shorter than the diameter of 100–120 Å suggested by electron microscopy [50], but is longer than the diameter suggested by the cryo-TEM method [60].

#### 4.7 Complex Formation by Grinding in a Mortar [54]

In the crystallization of the complex crystals, we found that the complex crystals were formed when the powdered component crystals were ground in a mortar [61]. For example, the diffraction patterns of the CTAB and *o*-iodophenol powdered crystals are shown in Figure 4.23(a) and 4.23(b), respectively. After the powdered crystals were mixed and ground in a mortar, the diffraction pattern, Figure 4.23(d), is completely



**Figure 4.22** Schematic drawing of the rod-like micelle formation from the elementary pattern

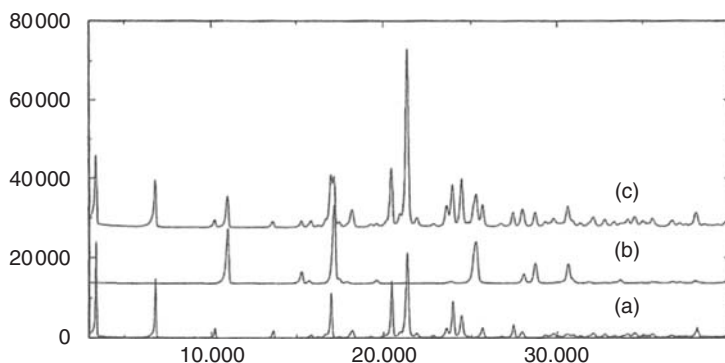


**Figure 4.23** Powder diffraction patterns of (a) CTAB only, (b) o-iodophenol only, (c) powdered crystals made from single crystals of the CTAB-o-iodophenol complex and (d) powdered crystals of CTAB and o-iodophenol mixed and ground in a mortar

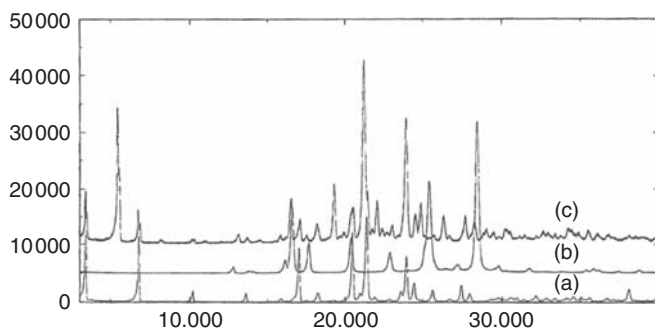
different from the sum of the component powder patterns. This indicates that the new crystals were produced only by grinding the powdered mixture in a mortar. The produced new pattern is essentially the same as the pattern in Figure 4.23(c) of the powdered complex crystal, which was produced by grinding the single crystals grown from solution in a mortar.

Complex formation upon grinding the component crystals in a mortar was not observed in all the combinations of surfactants and aromatic compounds. For example, the X-ray powder patterns of CTAB and *o*-hydroxybenzoic acid are shown in Figure 4.24(a) and 4.24(b), respectively. After the mixed samples were ground in a mortar, the diffraction pattern was measured. The obtained pattern, Figure 4.24(c), is the sum of the component patterns, (a) and (b). Of course, the pattern is completely different from the pattern of the powdered sample obtained by grinding the single crystal of the complex in a mortar.

For the complex between CTAB and *p*-cresol, the complex was not completely formed by grinding the component samples. The powder patterns of CTAB and *p*-cresol are shown in Figure 4.25(a) and 4.25(b), respectively. After the mixture was ground in a mortar, the X-ray powder pattern was measured, Figure 4.25(c). Although some peaks are consistent with those of the component crystals, several different peaks appear. The new peaks are consistent with those of the powdered sample produced by grinding the single crystal of the complex. This indicates that the complex can be formed only partly by grinding.



**Figure 4.24** Powder diffraction patterns of (a) CTAB only, (b) *o*-hydroxybenzoic acid and (c) powdered crystals of CTAB and *o*-hydroxybenzoic acid mixed and ground in a mortar



**Figure 4.25** Powder diffraction patterns of (a) CTAB only, (b) *p*-cresol and (c) powdered crystals of CTAB and *p*-cresol mixed and ground in a mortar

**Table 4.2** Ease of complex formation only by mixing and grinding the component crystals in a mortar<sup>a</sup>

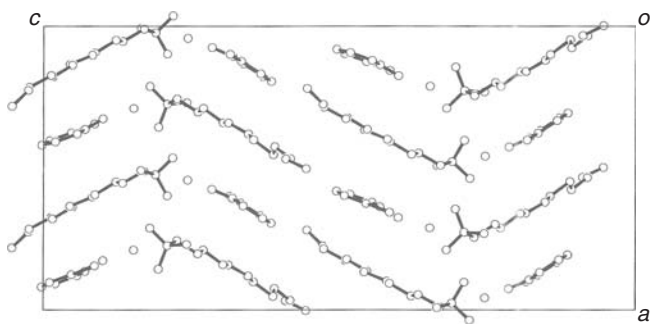
	m.p.	CTAB	MTAB	LTAB	DTAB
<i>Hydrophilic aromatic compounds</i>					
<i>(i) Hydroxyl group</i>					
<i>o</i> -Iodophenol	37–40	A*	A*	A*	A*
<i>m</i> -Iodophenol	42–44	A	A	A	A
<i>p</i> -Iodophenol	93–95	A	A	A	A
<i>p</i> -Cresol	32–34	A*			
<i>m</i> -Cyanophenol	79–81	A*	A	A	A
<i>(ii) Carboxyl group</i>					
<i>o</i> -Hydroxybenzoic acid	158–161	C*	C*	B	B
<i>m</i> -Hydroxybenzoic acid	201–203	B	A*	A	A
<i>p</i> -Hydroxybenzoic acid	215–217	C	C	C	B*
<i>o</i> -Toluic acid	103–105	C	C	B	B
<i>m</i> -Toluic acid	108–110	C	C	B	A
<i>p</i> -Toluic acid	180–182	C	C	C	B
<i>(iii) Imino group</i>					
Indole	52–54	A*	A	A	A
Carbazole	245	B*	B	B	B
Diphenylamine	54	B*	B	B	B
<i>Hydrophobic aromatic compounds</i>					
Acridine	107–110	C*	C	B	B
Anthracene	216–218	C	C	B	B
Biphenyl	69–72	B	A	A	A
Phenanthrene	99–101	B	B	B	A

<sup>a</sup>The asterisk indicates that the single crystal has been analysed by X-rays.

Complex formation only by grinding a mixture in a mortar was examined for a variety of combinations of the surfactant and aromatic compounds. The CTAB, MTAB, LTAB and DTAB surfactants were selected along with hydrophilic aromatic compounds containing hydroxyl, carboxyl and imino groups, and hydrophobic aromatic compounds such as acridine, anthracene, biphenyl and phenanthrene. Table 4.2 shows the results of the complex formation, in which the letters A, B and C mean that the complex was formed completely, partly and hardly, respectively. (The asterisk indicates that the single crystals have been analysed by X-rays). It is clear that:

- (1) aromatic compounds with a hydroxyl group can form the complex completely;
- (2) aromatic compounds with a carboxyl group form the complex incompletely; and
- (3) with hydrophobic aromatic compounds the complex only partly forms.

In order to clarify the above different characteristics, the structures of the crystals obtained from an aqueous solution were compared. The crystal structures of the complexes composed of aromatic compounds with a hydroxyl group have already been shown in the previous section; CTAB–*m*-cyanophenol (II), CTAB–*p*-cresol (III), CTAB–*o*-iodophenol (V), MTAB–*o*-iodophenol (VI), LTAB–*o*-iodophenol (VII) and DTAB–*o*-iodophenol (VIII). For the *o*-iodophenol complexes there are two structural groups;



**Figure 4.26** Crystal structure of VII' viewed along the *b*-axis

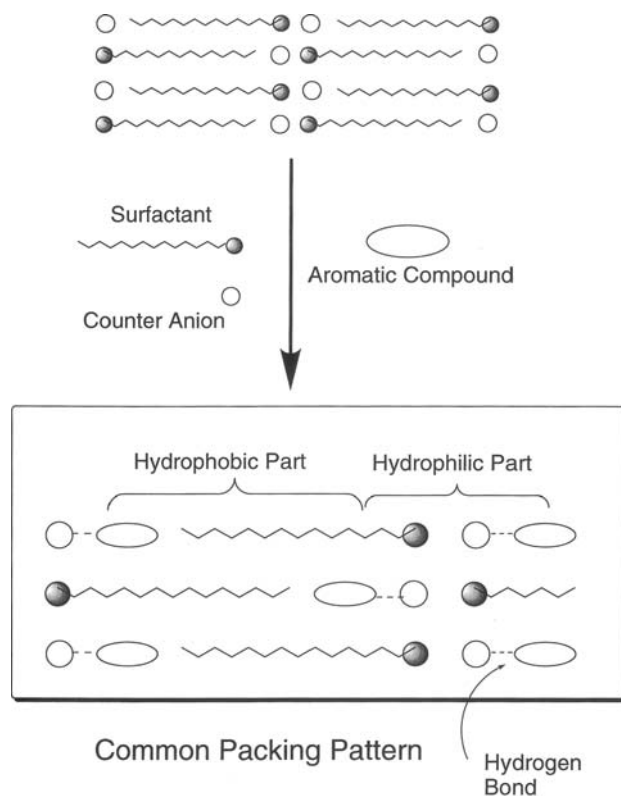
orthorhombic crystals (V and VI) and monoclinic crystals (VII and VIII). Within the two groups the two crystal structures are isostructural. Another crystal form was obtained for the complex LTAB-*o*-iodophenol (VII'). The crystal structure is shown in Figure 4.26. It is isostructural to V and VI as shown in Figures 4.6 and 4.7, but is different from the structure of VII shown in Figure 4.8. This suggests that although the four *o*-iodophenol complex crystals take the common packing mode, the slight change of the alkyl chains modifies the crystal structures. It must be emphasized that in all of the crystal structures V, VI, VII, VII' and VIII, the hydroxyl group of the *o*-iodophenol forms a hydrogen bond with the bromide anion. In the other complex crystals with a hydroxyl group, i.e. I–IV and IX, the hydroxyl group makes a hydrogen bond with the bromide anion directly or through the solvent water molecule.

This may suggest a process for complex formation when the component crystals were mixed and ground in a mortar. As shown in Figure 4.27, the aromatic molecule should be inserted between the alkyl chains when the neighbouring alkyl chains slide in an anti-parallel direction in a mortar. If the hydroxyl group makes a hydrogen bond with the bromide anion directly or through the water molecule, the complex formation should be enhanced. This is a reason why the aromatic compounds with a hydroxyl group can form the complex in a mortar very easily.

For the complex crystals with the carboxyl group, I and IX, the carboxyl groups make a dimer structure directly or through solvent water molecules. When the aromatic crystals with the dimer structure were ground in a mortar, the dimer structure is probably conserved. Since the dimer structure is too large, it may be difficult to be inserted between the anti-parallel alkyl chains. When the aromatic compounds with a carboxyl group are dissolved in aqueous solution, the dimer structure may be loosened allowing easy insertion between the alkyl chains.

Complexes composed of *m*-hydroxybenzoic acid can be easily formed by grinding the component crystals in a mortar. Since the single crystal for the combination of MTAB and *m*-hydroxybenzoic acid, XXI, was obtained from an aqueous solution, the structure was analysed by X-rays. The crystal structure viewed along the *c*-axis shown in Figure 4.28. The dimer structure of *m*-hydroxybenzoic acid as shown in the *m*-hydroxybenzoic acid crystal itself [59] is inserted between the two anti-parallel MTAB molecules. The hydroxyl group makes a hydrogen bond with the solvent water





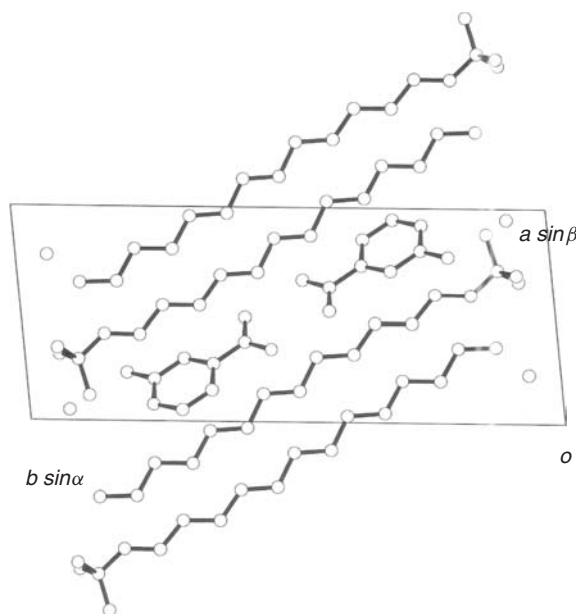
**Figure 4.27** Schematic drawings of the complex formation when the component crystals were mixed and ground in a mortar

molecule, which has hydrogen bonds with the two bromide anions. Since such conditions are satisfied, *m*-hydroxybenzoic acid can make complexes easily with the surfactant molecules.

For the dibenzofuran, carbazole, phenoxazine and phenothiazine complexes, formation ratios were estimated comparing the low-angle powder diffractions of the complex produced in a mortar with the corresponding ones of the powdered sample of the single crystal. The values are listed in Table 4.3. Although there is a significant difference in the complex formations of dibenzofuran and carbazole, there seems to be no significant difference in those of phenoxazine and phenothiazine. This may indicate that there is no significant difference between the aromatic compounds containing oxygen and sulfur atoms in the complex formation ground in a mortar.

#### 4.8 C-H... $\pi$ interactions [53]

In the previous section, it has been made clear that the hydrogen bond formation of the aromatic compound with the bromide anion reinforces the complex formation. However,



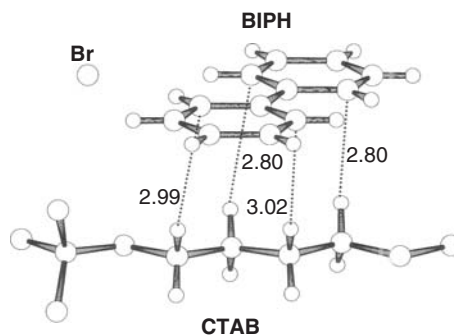
**Figure 4.28** Crystal structure of XXI viewed along the *c*-axis

**Table 4.3** Ratio of complex formation obtained by mixing and grinding the component crystals of aromatic compounds with heteroatoms and the cationic surfactants. The ratios were calculated using the peak heights of the low-angle diffractions of their X-ray powder patterns

Aromatic compound	CTAB	MTAB	LTAB	DTAB
Dibenzofuran	36	29	64	49
Carbazole	5	4	20	25
Phenoxazine	25	45	37	37
Phenothiazine	29	27	36	60

biphenyl, anthracene, phenanthrene and acridine form complex crystals with cationic surfactants not only by dissolving the component compounds in aqueous methanol solutions but also by grinding them in a mortar without any solvent, although such aromatic compounds have no donor groups for hydrogen bond formation. The crystal structures of crystals formed by the two different methods are essentially identical, showing a common packing pattern. This indicates that the most important factor to form the complex should be the intermolecular interaction between the aromatic ring and the alkyl group of the surfactant.

Figure 4.29 indicates the short interatomic distances between biphenyl and CTAB in crystal XV. There are short contacts of 2.80–3.02 Å between the hydrogen atoms of the C-H groups of CTAB and the aromatic carbon atoms of biphenyl. Similar interatomic interactions are also found in the MTAB and DTAB complexes, with short distances



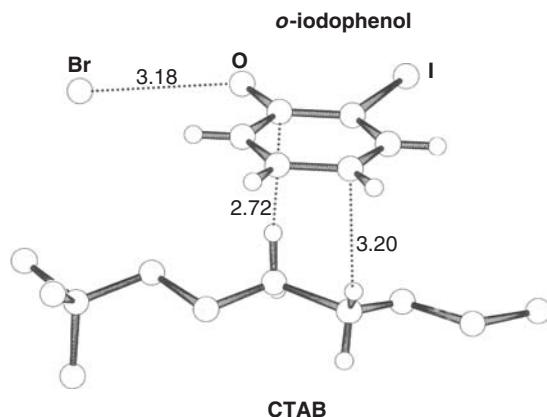
**Figure 4.29** Short interatomic C-H...C distances between biphenyl and CTAB in the complex crystal (XV)

of 2.74–3.20 Å and 2.52–3.24 Å, respectively. In all three complexes, the four C-H groups participating in these interactions begin at the  $\beta$ -carbon atom from the terminal ammonium nitrogen atom through three more carbon atoms.

For the odd-carbon-chain surfactant complexes, PTAB–biphenyl (XX) and TTAB–biphenyl (XXI), similar interactions are also observed. The short interatomic C-H...C(aromatic) distances are 2.71–3.32 Å and 2.97–3.18 Å for the PTAB and TTAB complexes, respectively. For the complexes of CTAB–2-phenylpyridine (XVIII) and CTAB–4-phenylpyridine (XIX), the C-H...C or C-H...N distances are 2.71–3.21 Å and 2.65–3.58 Å, respectively.

The weak intermolecular interaction between the C-H group and the aromatic ring has been well studied. A typical example of this type of interaction was found in the benzene-induced up-field shift of the  $^1\text{H}$ -NMR signal of chloroform in benzene solution [62–65]. From NMR spectroscopy, the chloroform molecule is supposed to form a 1:1 complex with benzene in which the chloroform molecule is located just above the plane of the benzene ring directing its C-H bond toward the  $\pi$ -electron cloud of the ring. There are several early references in which the aromatic rings act as acceptors for C-H hydrogen atoms in crystals [66]. Green includes benzene, methyl-substituted benzenes and halo benzenes in a list of hydrogen bond acceptors [67]. The C-H... $\pi$  contacts are reported in crystal structures by Hardy and MacNicol [68], Atwood *et al.* [69, 70] and Ferguson *et al.*, [71] in which H... $\pi$  distances range from 2.1 to 2.7 Å. Nishio and Hirota also reported a variety of C-H... $\pi$  interactions in organic compounds [72]. They suggested that the energy of stabilization of such an interaction is about  $3 \text{ kcal mol}^{-1}$ . Nishio *et al.* collected data on C-H... $\pi$  interactions from the Cambridge Structural Database and proposed that the C-H... $\pi$  interaction is not simply due to dispersion forces but involves other types of interactions which are orientation dependent [73]. Recently, Desiraju and Steiner have discussed C-H... $\pi$  contacts in their book [74].

Such C-H... $\pi$  interactions play an important role in the complex formation between the aromatic compounds with hydrogen donor groups and surfactant molecules. Figure 4.30 shows the molecular structures in the CTAB–o-iodophenol complex, V, where the strong hydrogen bond formed between the hydroxyl group and the



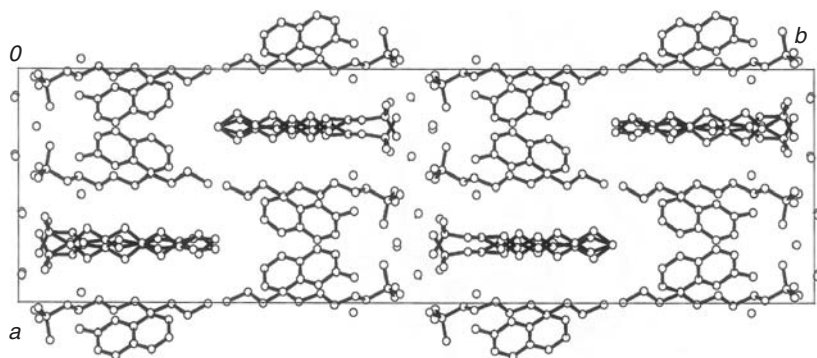
**Figure 4.30** Short interatomic C-H...C distances between *o*-iodophenol and CTAB in the complex crystal

bromide anion is highlighted. There are also short contacts between the phenyl ring of *o*-iodophenol and C-H groups of the CTA cation. In this case, there are only two short C-H...C(aromatic) distances, i.e. 2.72 and 3.20 Å, which are comparable with those of the biphenyl complex crystals, each of which forms four such contacts. The C-H... $\pi$  interactions probably determine the relative orientation of the two molecules in the complex formation. Since the complexes without strong hydrogen donor groups were formed only by grinding the component crystals in a mortar without solvent, the C-H... $\pi$  interactions should play a most important role in complex formation. The hydrogen bonds between the hydrogen donor such as a hydroxyl or amino group and the bromide anion should accelerate complex formation and stabilize the resultant complex.

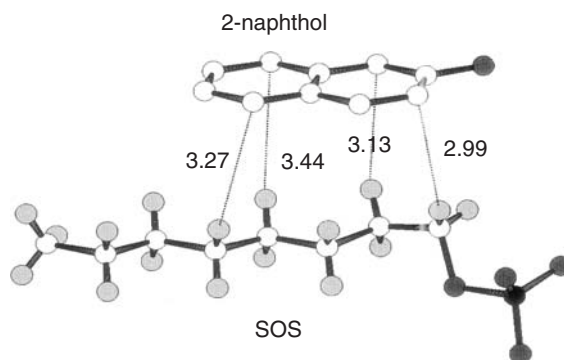
#### 4.9 Complex Formation of Anionic Surfactants with Aromatic Compounds [54]

In the above sections, the surfactant molecules are cationic. However, there are many anionic and nonionic surfactants. The anionic surfactants, such as sodium dodecyl sulfate (SDS), were reported to form a micelle in an aqueous solution with aromatic compounds [75]. Although much spectroscopic work has been conducted, the micelle structure is still unclear [76]. We tried to obtain crystals of the representative complexes suitable for X-ray work. After many trials, the single crystals were obtained from an aqueous methanol solution for the complex formed between the anionic surfactant, sodium octyl sulfate (SOS), and the aromatic compound, 2-naphthol.

The crystal structure of SOS–2-naphthol (XXII) viewed along the *c*-axis is shown in Figure 4.31. In the asymmetric unit, there are four octyl sulfate anions, four sodium cations, two 2-naphthol molecules and six water molecules. There are two types of octyl sulfate anions, each with an all trans conformation: one has a disordered alkyl



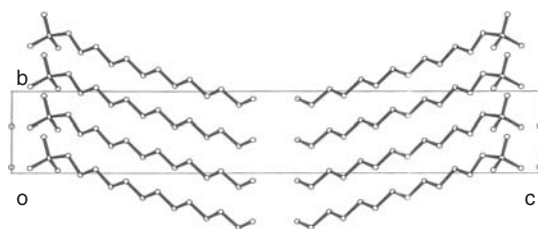
**Figure 4.31** Crystal structure of XXII viewed along the *c*-axis



**Figure 4.32** Short interatomic C-H...C distances between SOS and 2-naphthol in the complex crystal

chain and the other has an ordered alkyl chain. The latter octyl anion forms a complex with 2-naphthol. The two crystallographically independent complexes are related by a pseudo-translational symmetry along the *b*-axis. The interaction between the octyl sulfate anion and 2-naphthol is shown in Figure 4.32. The relatively short contacts between the C-H groups and the carbon atoms of 2-naphthol are given (2.99–3.44 Å). The shorter contacts suggest C-H... $\pi$  interactions between the molecules, as have been observed in the complex crystals between the cationic surfactants and aromatic compounds.

The crystal structure of SDS viewed along the *a*-axis is shown in Figure 4.33 [77], since SOS has not yet been analysed by X-ray crystallography. The all *trans* alkyl chains stack in a parallel way along the *b*-axis. It seems adequate to assume that the 2-naphthol molecule is inserted between the alkyl chains as shown in Figure 4.31. Since the SOS molecule has a shorter alkyl chain and the sodium cation has many water molecules around it, it may be difficult to envisage a similar packing pattern to that of the cationic surfactant complexes. However, the C-H... $\pi$  interaction between SOS and 2-naphthol is approximately the same as those of the cationic complexes.

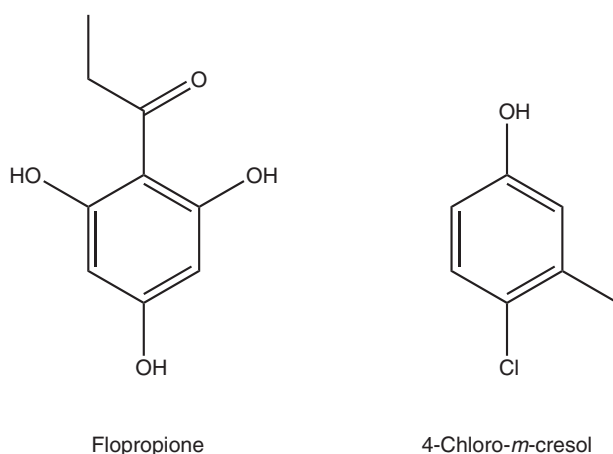


**Figure 4.33** Crystal structure of SDS viewed along the *a*-axis

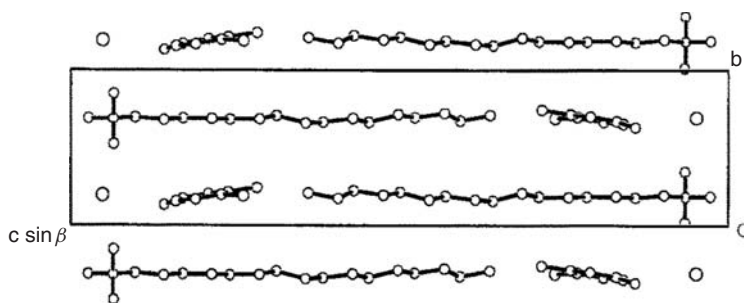
#### 4.10 Increased Solubility of Insoluble Drugs [55, 78]

Normally water-insoluble drugs with a phenyl ring can form molecular complexes with surfactants. The chemical structures of 1-(2,4,6-trihydroxyphenyl)-1-propanone (flopropione), an anti-spasmodic, and 4-chloro-*m*-cresol (a germicide) are shown in Figure 4.34. They became soluble in aqueous solutions in the presence of surfactants, probably because of complex formation. In order to understand the nature of complex formation, the cationic surfactants CTAB and MTAB were used. The complex crystals between CTAB and 4-chloro-*m*-cresol (XXIII), MTAB and flopropione (XXIV), and CTAB and flopropione (XXV) were obtained from the solutions of methanol, methanol and aqueous solutions, respectively.

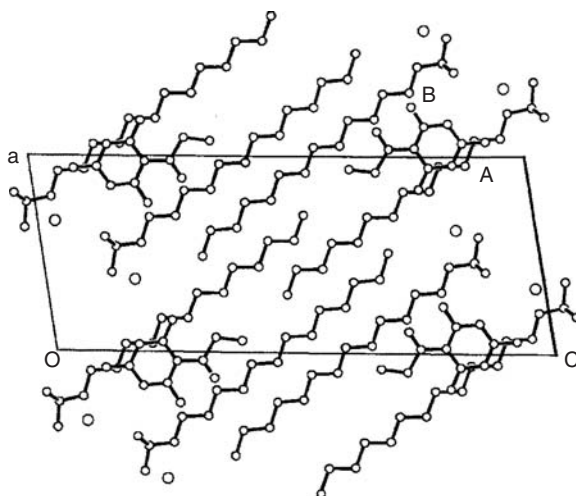
The crystal structure of the complex XXIII viewed along the *a*-axis is shown in Figure 4.35. The alkyl groups of the CTA cations with an all trans conformation are arranged in an anti-parallel way along the *b*-axis. The interaction between the cation and 4-chloro-*m*-cresol is essentially the same as that of the common packing pattern observed above. The 4-chloro-*m*-cresol molecule is sandwiched by the alkyl groups of the CTA cations.



**Figure 4.34** Structural formulae of flopropione and 4-chloro-*m*-cresol



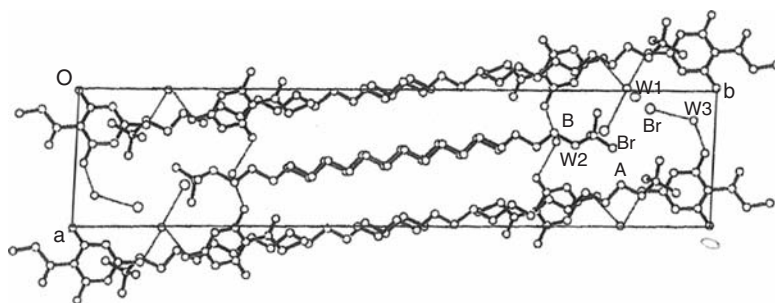
**Figure 4.35** Crystal structure of XXIII viewed along the *a*-axis. Reprinted from [78] with permission from The Chemical Society of Japan. Copyright 1999



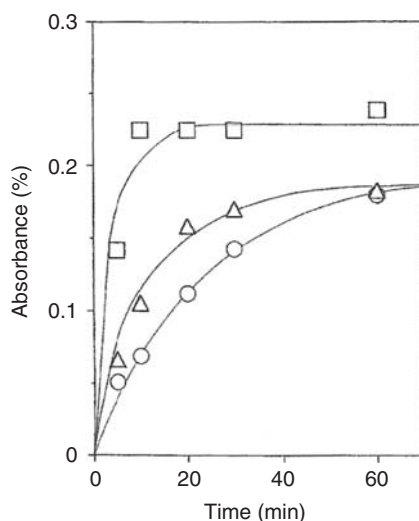
**Figure 4.36** Crystal structure of XXIV viewed along the *b*-axis. Reprinted from [78] with permission from The Chemical Society of Japan. Copyright 1999

The crystal structure of the complex XXIV viewed along the *b*-axis is shown in Figure 4.36. All the molecules lie on the crystallographic mirror planes. There are one flopropione and two MTABs, A and B, in the asymmetric unit. The benzene ring of flopropione is sandwiched by the alkyl groups of the A-MTA cations. Since the flopropione molecule is so large, another MTA cation and bromide anion of B is necessary to maintain the common packing pattern.

The crystal structure of the complex XXV viewed along the *c*-axis is shown in Figure 4.37. There are two crystallographically independent CTABs and flopropiones (A and B) and three solvent water molecules in the asymmetric unit. Both of the CTA cations take an all trans conformation. The interaction between the A-CTA cation and flopropione A is essentially the same as that of the common packing pattern, whereas the B-CTA and flopropione B molecules occupy the void space produced by the 3-D



**Figure 4.37** Crystal structure of XXV viewed along the *c*-axis

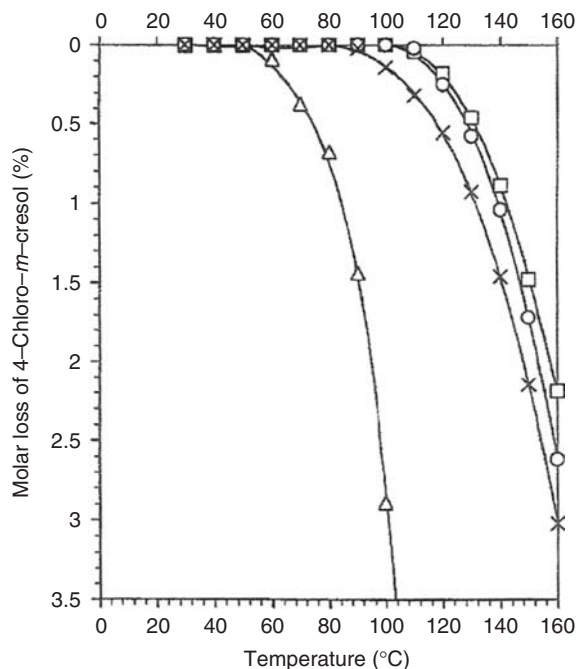


**Figure 4.38** Dissolution rates of flopropione only (○), the simple mechanical mixture of the powdered flopropione and CTAB (Δ), and the powdered complex crystals between flopropione and CTAB (□). Reprinted from [78] with permission from The Chemical Society of Japan. Copyright 1999

hydrogen bond network composed of flopropiones, solvent water molecules and bromide anions.

Figure 4.38 shows the dissolution rates of the three samples, flopropione only, the simple mechanical mixture of the powdered flopropione and CTAB, and the powdered complex crystals between flopropione and CTAB. The dissolution rate of the complex is significantly greater than those of the other two. This suggests that the complex species may be tightly kept as an entity in solution, including such a condition that the surfactant molecules may partly prevent the solvent water molecules from direct contact with the flopropione molecules. In fact, we already established that the surfactant complexes have different values of cmc or Kraft point from the corresponding ones of their component surfactant molecules [79]. The rate of the simple mechanical mixture is slightly greater





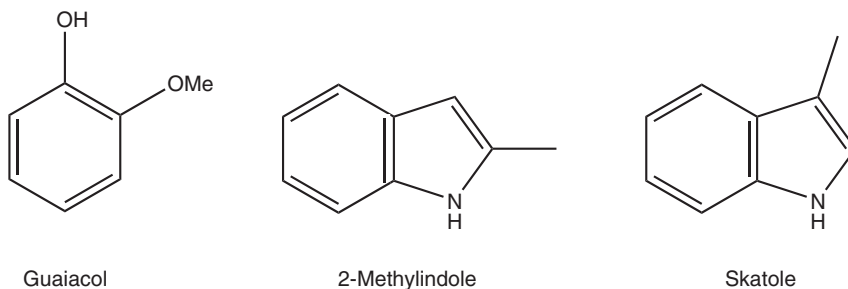
**Figure 4.39** Molar loss of 4-chloro-*m*-cresol with increasing temperature in the crystals of 4-chloro-*m*-cresol only ( $\Delta$ ) and its complexes with CTAB ( $\square$ ), MTAB ( $\circ$ ) and LTAB ( $\times$ ). Reprinted from [78] with permission from The Chemical Society of Japan. Copyright 1999

than that of flopropione only. This may be explained by two factors; one is that the same structure as that of the complex may be formed when the powdered samples of both components are mechanically ground [54]; the other is that the dissolved CTAB may form a complex with dissolved flopropione in solution.

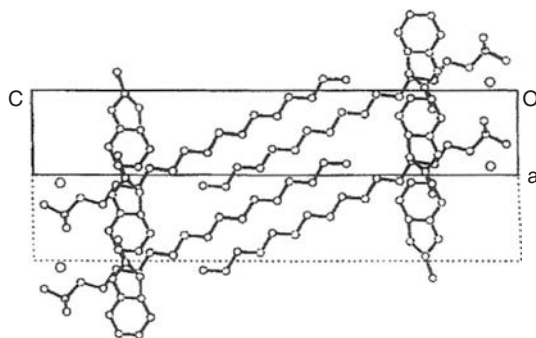
Figure 4.39 shows the molar loss of 4-chloro-*m*-cresol with increasing temperature in the crystals of 4-chloro-*m*-cresol only and its complexes with CTAB, MTAB and LTAB. The vaporization of 4-chloro-*m*-cresol is clearly prevented when it forms complexes with surfactants. The thermal stability depends on the length of the alkyl chains of the surfactants. The above results clearly indicate that complex formation of insoluble drugs with surfactants increases the solubility and thermal stability. This may be a useful application to improve the characteristics of drugs.

#### 4.11 Enhanced Thermal Stability of Perfumes [55, 80]

The thermal stability of the molecular complexes just described inspired the idea that the volatilization of perfumes may be delayed if the perfumes were incorporated into molecular complexes with surfactants. After many trials, it was found that perfumes such as guaiacol, 2-methylindole and skatole, as shown in Figure 4.40, formed crystalline molecular complexes with cationic surfactants such as CTAB, MTAB, LTAB



**Figure 4.40** Structural formulae of guaiacol, 2-Methylindole and skatole

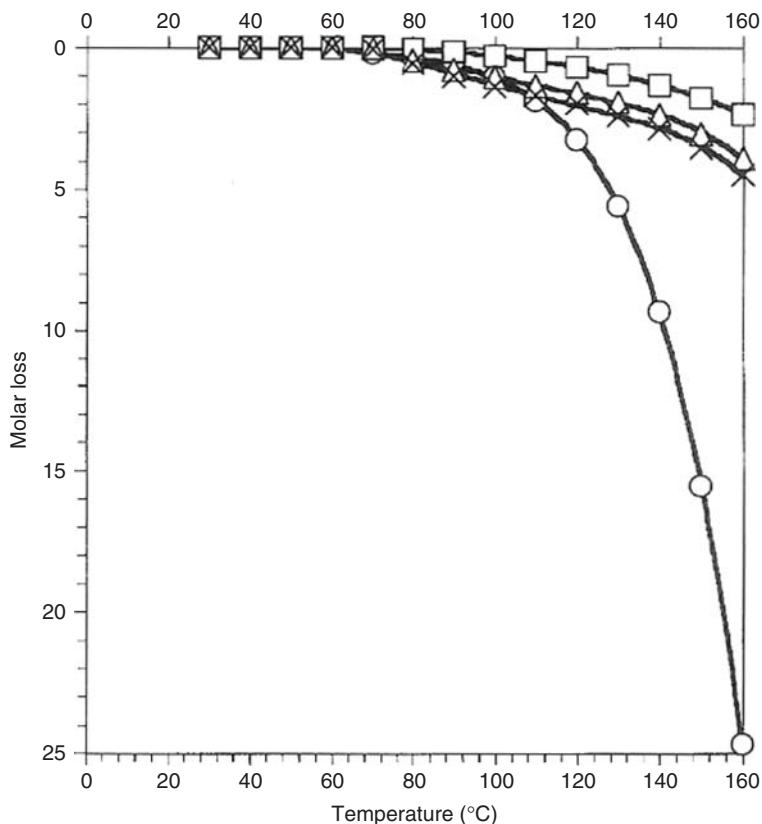


**Figure 4.41** Crystal structure of XXVI viewed along the *b*-axis. Reprinted from [80] with permission from The Chemical Society of Japan. Copyright 2000

and DTAB. Crystals suitable for X-ray analysis were obtained for the seven complexes between 2-methylindole and CTAB, MTAB and LTAB (XXVI, XXVII and XXVIII, respectively), and between skatole and CTAB, MTAB and LATB (XXIX, XXX and XXXI, respectively), and between guaiacol and CTAB (XXXII).

The crystal structure of XXVI viewed along the *b*-axis is shown in Figure 4.41. Since there are rather loose contacts between the 2-methylindole molecules along the *a*-axis, this molecule takes four disordered conformations (A, B, C and D), the occupancy factors of which are 0.37, 0.23, 0.12 and 0.28 for A, B, C and D, respectively; only the A conformation is shown in Figure 4.41. The alkyl group of the surfactant molecule has an all trans conformation. The N-H group of 2-methylindole forms a hydrogen bond with the bromide anion. The aromatic ring of 2-methylindole is sandwiched by the alkyl chains of CTAB, which is similar to those in the common packing pattern described above. Slightly short contacts are observed between the alkyl group of CTAB and the ring carbons of 2-methylindole, which may be C-H... $\pi$  interactions as suggested earlier.

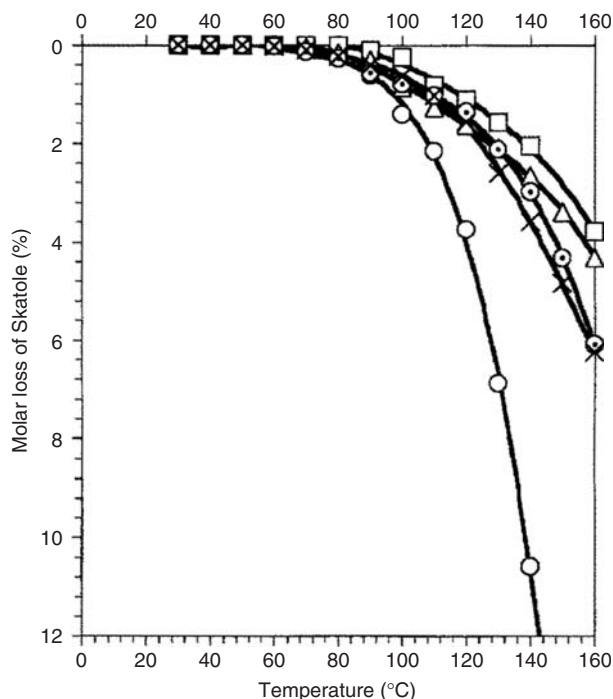
Similar packing patterns were observed in the other six crystals (XXVI–XXXII), since all the complex crystals are isostructural. The structures are substantially the same as that of XXVI except for the differences in the lengths of the alkyl chains of the surfactant molecules.



**Figure 4.42** Molar loss of 2-methylindole in the crystal of 2-methylindole only (○) and in its three molecular complex crystals with CTAB (□), MTAB (△) and LTAB (×). Reprinted from [80] with permission from The Chemical Society of Japan. Copyright 2000

The rate of molar losses of 2-methylindole and its three molecular complex crystals with CTAB, MTAB and LTAB, respectively, are shown in Figure 4.42. Although the volatilization of 2-methylindole begins at about 70 °C not only in pure crystal but also in the complex crystals, the volatilization rate of the pure crystal is significantly greater than those of the complex crystals. Moreover, as the length of the alkyl chain increases, the rate of volatilization decreases.

Figure 4.43 shows the rate of molar loss of skatole with increasing temperature for the pure skatole crystal and the four molecular complex crystals with CTAB, MTAB, LTAB and DTAB. The volatilization of skatole begins at about 70 °C but the rate of volatilization of skatole in its pure crystal form is significantly greater than those in the complex crystals. Moreover, the rates of volatilization of skatole are significantly greater than the corresponding ones of 2-methylindole, although the volatilization of skatole begins at nearly the same temperatures. It is clear that the perfumes are more stable in the molecular complexes than in pure crystal and that the stability depends on the alkyl chain length of the surfactants.



**Figure 4.43** Molar loss of skatole in the crystal of skatole only (○) and in its four molecular complex crystals with CTAB (□), MTAB (△), LTAB (⊙) and DTAB (×) with increasing temperature. Reprinted from [80] with permission from The Chemical Society of Japan. Copyright 2000

In order to explain the stability on the basis of the crystal structure, the van der Waals energy for each molecular complex was calculated using the Lennard–Jones potential. Only C...C interactions less than 5 Å between the neighbouring alkyl groups were taken into account in the first approximation, since the positions of the hydrogen atoms of the alkyl group were obtained geometrically from the positions of the carbon atoms. Moreover, the electrostatic energy should be approximately the same, since the ammonium cation and bromide anion occupy nearly the same position in each crystal. Table 4.4 shows the van der Waals energies for the complex crystals between 2-methylindole or skatole and CTAB, MTAB or LTAB. As the length of the alkyl chain decreases, the number of contacts less than 5 Å decreases. This causes an increase in van der Waals energy as the alkyl group is shorter, which is clearly shown in Table 4.4. The van der Waals energy of each 2-methylindole complex is lower than the corresponding one of the skatole complex. This correlates with the greater rates of volatilization of skatole compared with those of 2-methylindole, although the volatilization of skatole begins at nearly the same temperature as that of 2-methylindole.

In summary, complex formation of the chosen perfumes with surfactants decreases the volatilization rate of the perfume and the rate can be controlled by the length of the alkyl group of the surfactant.

**Table 4.4** Calculated van der Waals energies for the complexes between aromatic compounds and cationic surfactants

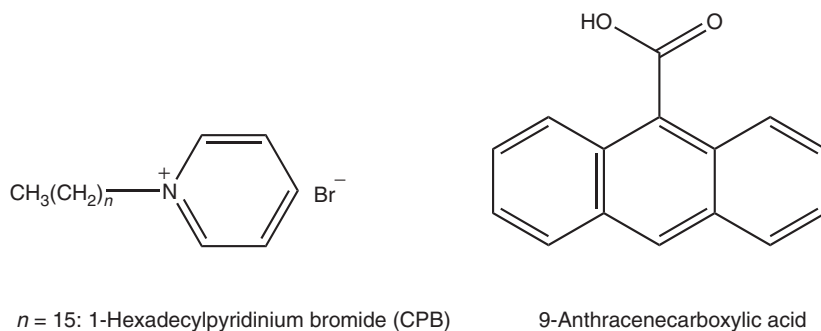
Complex	$\text{kJ mol}^{-1}$
CTAB–2-methylindole	–4.155
MTAB–2-methylindole	–3.265
LTAB–2-methylindole	–2.429
CTAB–skatole	–3.858
MTAB–skatole	–3.041
LTAB–skatole	–2.215

Complex formation between surfactants and aromatic compounds revealed two important characteristics: one is an enhancement of the solubility of insoluble drugs in water; the other is an improvement in thermal stability of unstable aromatic compounds. Such characteristics will surely be useful for pharmaceutical and industrial chemistries. The improved characteristics of aromatic compounds obtained by the complex formation with the surfactants were registered as a patent at the Japanese Patent Office (Patent No. 3900237).

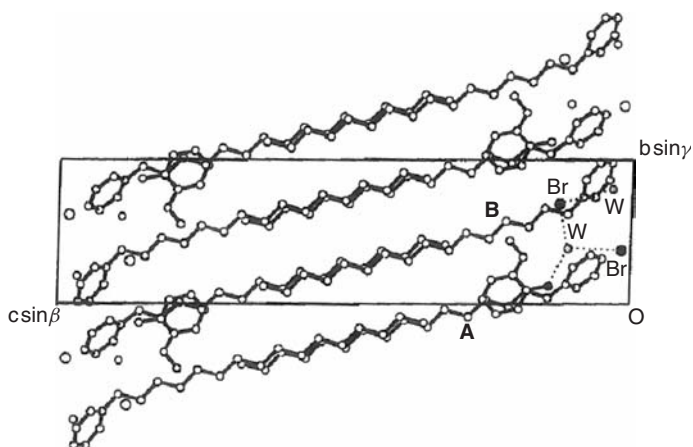
#### 4.12 Complex Formation with Surfactants other than Quaternary Alkylammonium Salts [81]

There are a variety of cationic surfactants other than quaternary alkylammonium salts. Accordingly, complex formation of 1-hexadecylpyridinium bromide (CPB) was attempted (Figure 4.44). After many trials, ten complex crystals were obtained, two of which were suitable for X-ray analysis, i.e. with guaiacol (XXXIII) and 9-anthracenecarboxylic acid (XXXIV) (Figure 4.44). The two crystals are isostructural.

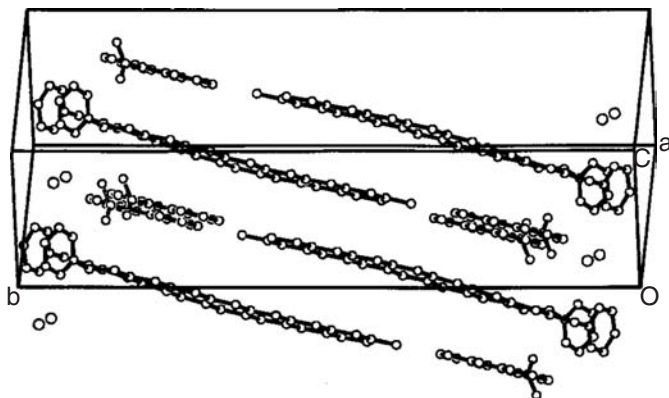
The crystal structure of XXXIII viewed along the *a*-axis is shown in Figure 4.45. There are two CP cations (A and B) and two Br (A and B) anions, one guaiacol and two solvent water molecules (A and B) in the asymmetric unit. However, there is no



**Figure 4.44** Structural formulae of hexadecylpyridinium bromide (CPB) and 9-anthracenecarboxylic acid



**Figure 4.45** Crystal structure of XXXIII viewed along the *a*-axis



**Figure 4.46** A view of the crystal structure of XXXIV. Reprinted from [81] with permission from The Chemical Society of Japan. Copyright 2002

solvent water molecule in the crystal of XXXIV. The A and B molecules of CP, both with an all trans conformation, are packed in a head-to-tail fashion along the *a*-axis and in a parallel mode along the *b*-axis. A guaiacol molecule is sandwiched by the alkyl groups of the CP molecules. A view of the crystal structure of XXXIV is shown in Figure 4.46. It is clear that in both XXXIII and XXXIV the molecules are packed in a common manner. The OH group of the guaiacol molecule makes a hydrogen bond with a solvent water molecule, which also makes hydrogen bonds with the two bromide anions, whereas the OH group of the 9-anthracenecarboxylic acid in XXXIII forms a hydrogen bond with one of the bromide anions.

The crystal structure of CPB itself was already reported by Ballirano *et al* [82]. If the CPB only and the complex structures are compared, a mechanism of formation of the latter may be envisaged. Thus, the complex structures are formed as the alkyl chains

**Table 4.5** Void space of the complex crystals of CTAB and CPB<sup>a</sup>

Complex	A (Å)	B (Å)	C (Å)
CTAB- <i>p</i> -cresol	9.80	7.47	5.57
CTAB- <i>m</i> -cyanophenol	10.27	7.47	5.58
CTAB-guaiacol	9.91	7.43	5.51
CTAB-vanilline	9.93	7.35	5.60
CTAB-skatole	9.90	7.41	5.49
CTAB-2-methylindole	9.39	7.43	5.52
CTAB-acridine	9.64	7.40	5.59
CTAB-phenanthrene	10.29	7.39	5.52
CPB-guaiacol	10.93	8.50	9.73
CPB-9-anthracenecarboxylic acid	12.26	9.59	9.06

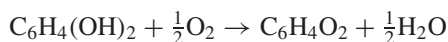
<sup>a</sup>The parameters A, B and C are defined in the text.

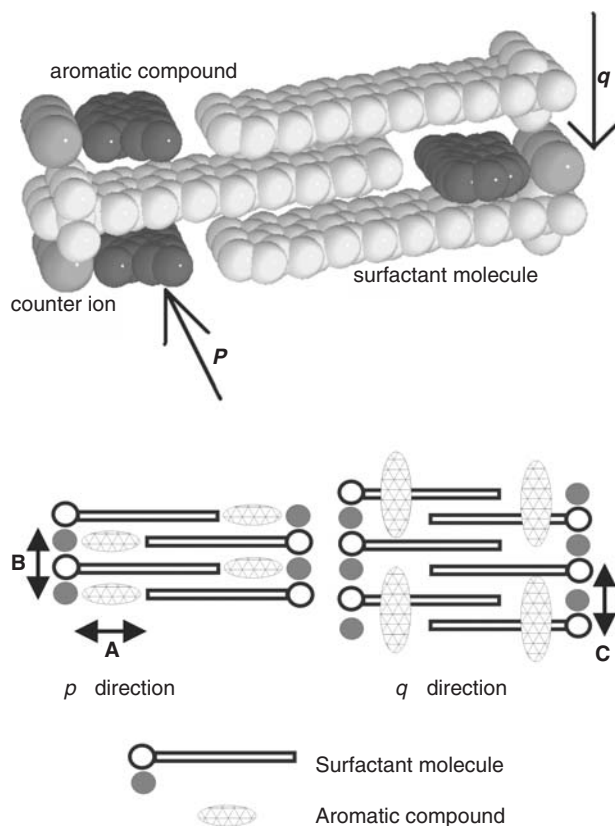
of CPB slide along the chain direction to form void space between the alkyl chain and the bromide anion (A) and between the neighbouring alkyl chains (B). The aromatic molecules slip into the void space. Such a structure is very similar to the common packing pattern shown in Figure 4.18. The size of the void space in the CTAB and CPB complexes are compared in Table 4.5, in which the parameters A, B and C are the distances between the terminal methyl carbon atom of the alkyl chain and the bromide anion, between the neighbouring aromatic molecules or neighbouring alkyl chains along the *p* direction (parallel to the arrangement of the aromatic molecules) and between the alkyl chains along the *q* direction (parallel to the stacking of the surfactant molecules), respectively, as shown in Figure 4.47.

For the CTAB complexes, the A, B and C values are less than 10.29, 7.48 and 5.60 Å, respectively. Along the *q* direction the void space is connected as a column. This means that the value of C corresponds to the molar ratio between the surfactant and the aromatic compound. However, the A and B values indicate the size of the molecule included in the void space. The A and B values in the CPB complexes, especially in XXXIV, are larger than the threshold values of the CTAB complexes. This indicates that CPB can accommodate bulkier aromatic compounds than CTAB. All efforts to make a complex between CTAB and 9-anthracenecarboxylic acid were in vain, probably because 9-anthracenecarboxylic acid is too bulky for the void space available in the CTAB complex, although a guaiacol complex with CTAB was made. These results suggest that it is possible to discriminate aromatic compounds by making complexes with different surfactants.

### 4.13 Hydroquinone Complexes [83]

The aromatic compound hydroquinone is well known to dermatologists as a drug to treat skin pigmentation, and has been used clinically [84–86]. However, hydroquinone is very easily discoloured to black [87]. This is because hydroquinone [C<sub>6</sub>H<sub>4</sub>(OH)<sub>2</sub>] is easily oxidized to quinone (C<sub>6</sub>H<sub>4</sub>O<sub>2</sub>) as follows:



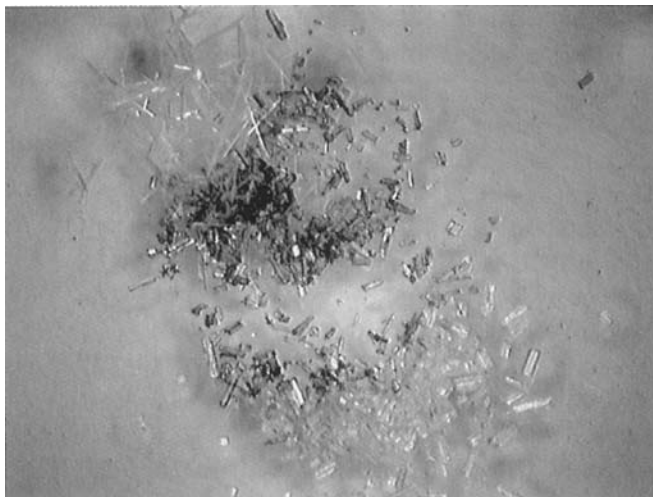


**Figure 4.47** The void space parameters  $A$ ,  $B$  and  $C$  viewed along the  $p$  and  $q$  directions. Reprinted from [81] with permission from The Chemical Society of Japan. Copyright 2002

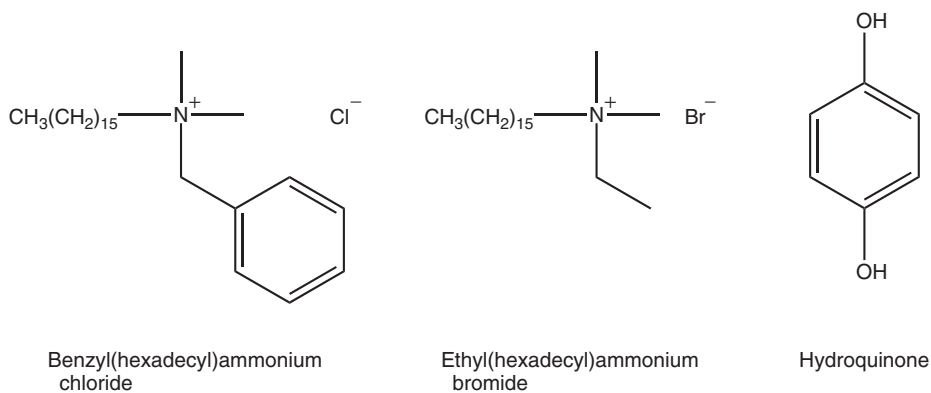
The produced quinone makes black complexes within the original hydroquinone very easily. As shown in Figure 4.48, when the crystals of hydroquinone and quinone are placed separately, the dark-violet crystals of the complex, which are called quinhydron, appear between the two crystalline samples. We tried to stabilize hydroquinone by making complexes with various surfactant molecules containing quaternary ammonium salts, such as alkyltrimethylammonium halides, benzyl(alkyl)dimethylammonium chloride, and ethyl(hexadecyl)dimethylammonium bromide (ECDAB). After many trials, complex crystals suitable for X-ray analysis were formed with benzyl(hexadecyl)dimethylammonium chloride (BCDAC) and ECDAB, as shown in Figure 4.49.

The crystal structure of the complex of hydroquinone and BCDAC (XXXV) viewed along the  $b$ -axis is shown in Figure 4.50. The complex is composed of one BCDAC and one and a half hydroquinone molecules (A- and B-HQ, respectively) in the asymmetric unit. The B-HQ is situated on the twofold axis between the cationic and anionic moieties of the surfactant molecules. The A- and B-HQ molecules have two OH groups with a *cis* conformation. The conformation of the hexadecyl group of BCDA cation is all *trans*.

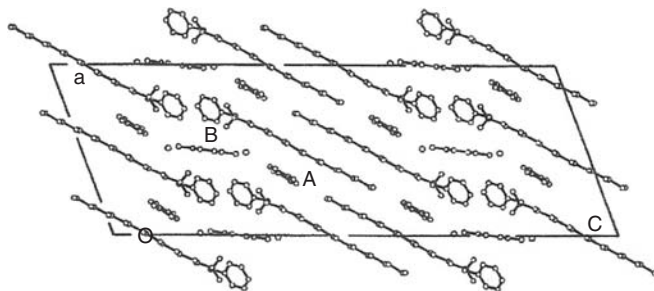




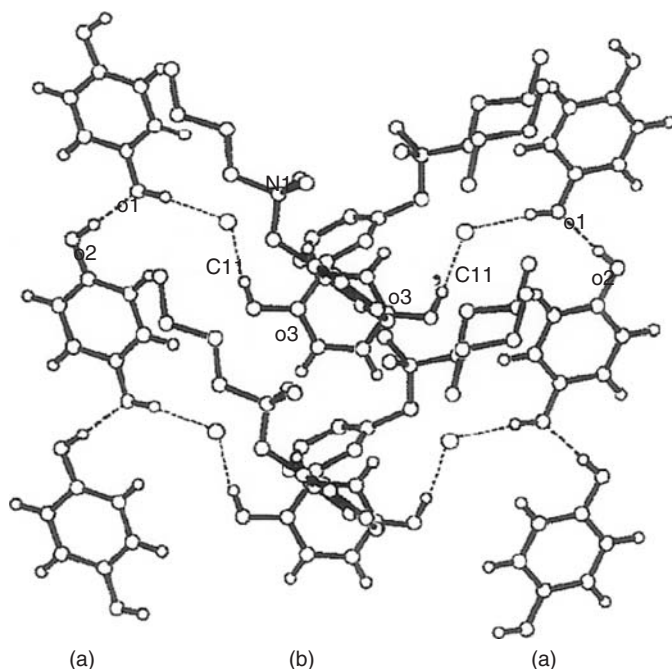
**Figure 4.48** Dark-violet crystals of quinhydrone formed from the reaction between the crystals of hydroquinone (colourless) and quinone (light yellow) Reprinted from [88] with permission from the Japanese Cosmetic Science Society. Copyright 2005 (see colour plate section)



**Figure 4.49** Structural formulae of BCDAC, ECDAC and hydroquinone



**Figure 4.50** Crystal structure of XXXV viewed along the *b*-axis. Reprinted from [83] with permission from The Chemical Society of Japan. Copyright 2005

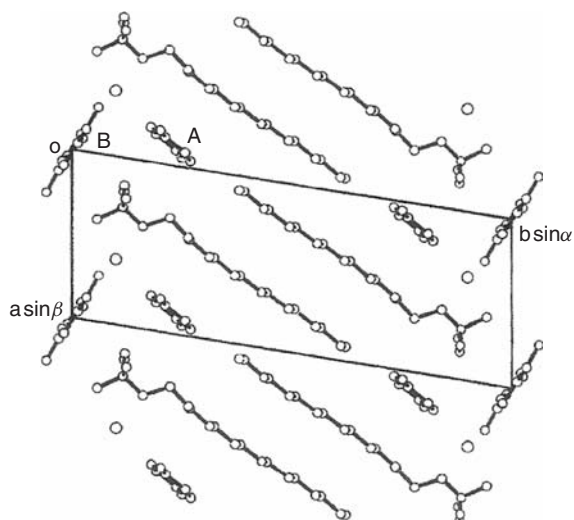


**Figure 4.51** Hydrogen bonding pattern in XXXV. Reprinted from [83] with permission from The Chemical Society of Japan. Copyright 2005

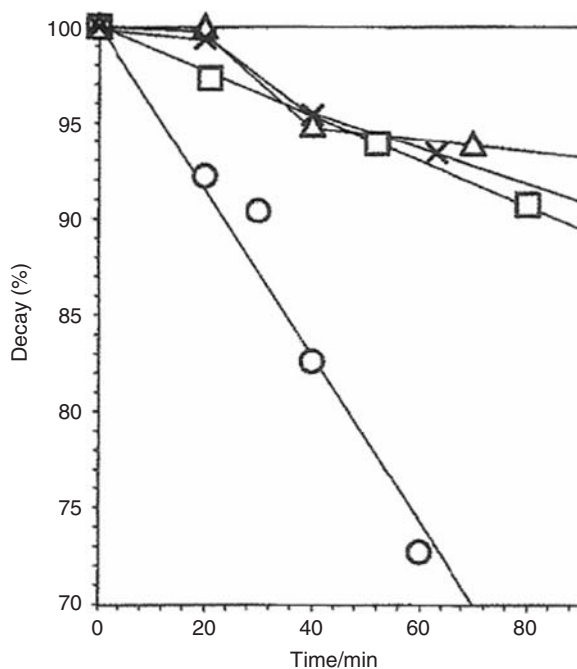
The A-HQ molecule is sandwiched by the two BCDA cations in the crystal structure. One of the two OH groups of A-HQ makes a hydrogen bond with the oxygen atom of the OH group of the neighbouring A-HQ molecule to form a linear chain along the *b*-axis. The acceptor OH group of A-HQ makes a hydrogen bond with the chloride anion. The hydrogen bonding pattern is shown in Figure 4.51 as (a). The OH group of the B-HQ molecule makes a hydrogen bond with the chloride anion to form a linear chain along the *c*-axis, shown as (b) in Figure 4.51.

The crystal structure of the complex of hydroquinone and ECDAB (XXXVI) viewed along the *c*-axis is shown in Figure 4.52. The complex is also composed of one ECDAB and one and a half hydroquinone molecules (A- and B-HQ, respectively) in the asymmetric unit. The A-HQ molecule is situated in the common packing pattern. The two OH groups of the A-HQ molecule form hydrogen bonds with the OH groups of the neighbouring A-HQ molecules and the bromide anions. The hexadecyl group of the ECDA cation has an all trans conformation. The B-HQ molecule is disposed about an inversion centre and takes on a disordered structure. The two OH groups of the disordered hydroquinone molecule make hydrogen bonds with the OH groups of the neighbouring B-HQ molecules and the bromide anions to form a linear chain along the *c*-axis.

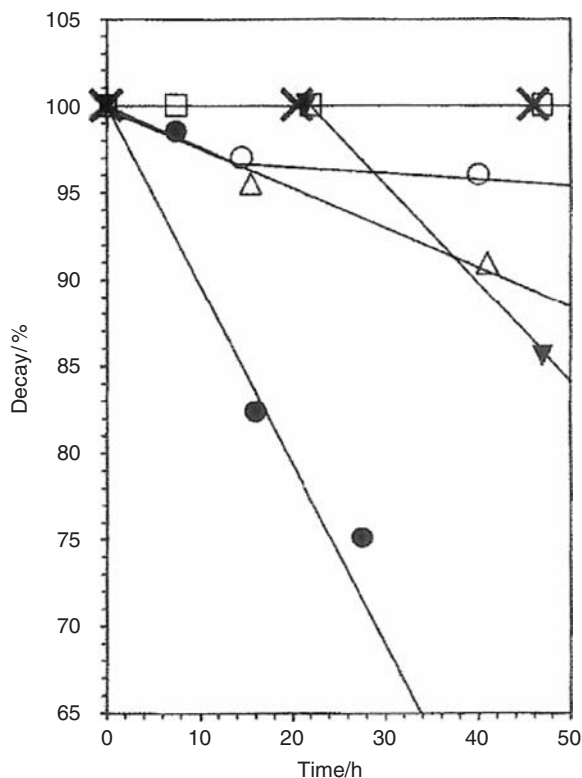
Although the two complexes have different crystal structures, the molecular arrangements are very similar to each other. The results of stabilization experiments on exposure to light or air are shown in Figures 4.53 and 4.54, respectively, in which only



**Figure 4.52** Crystal structure of XXXVI viewed along the  $c$ -axis. Reprinted from [83] with permission from The Chemical Society of Japan. Copyright 2005



**Figure 4.53** Decay curves of hydroquinone on exposure to a xenon lamp at 25°C. The curves show the pure hydroquinone crystal (○) and the three molecular complex crystals with BCDAC (□), BMDAC (△), and CTAC (×). Reprinted from [83] with permission from The Chemical Society of Japan. Copyright 2005



**Figure 4.54** Decay curves of hydroquinone on exposure to oxygen gas at 37°C. The curves show the pure hydroquinone crystal (●) and the five molecular complex crystals with STAC (□), BCDAC (×), MTAC (○), CTAC (△) and BMDAC (▼). Reprinted from [83] with permission from The Chemical Society of Japan. Copyright 2005

the surfactant complexes with chloride anions are given. Although the crystal structures of hydroquinone with benzyl(tetradecyl)dimethylammonium chloride (BMDAC), octadecyltrimethylammonium chloride (STAC), hexadecyltrimethylammonium chloride (CTAC) and tetradecyltrimethylammonium chloride (MTAC) have not yet been determined, complex formation was estimated from the UV spectra and powder X-ray diffraction. The hydroquinone molecules in the surfactant complexes with chloride anions are clearly more stable than hydroquinone only. Similar results for surfactant complexes with bromide anions have already been reported in Section 4.11.

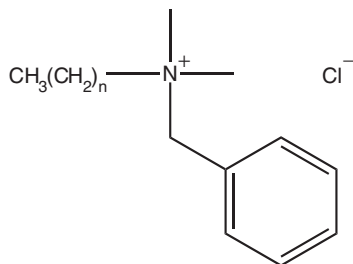
The hydroquinone molecule in its crystalline complexes with the BCDAC and ECDAB surfactants forms fairly strong hydrogen bonds with themselves and with the anions. However, if the intermolecular interaction is only such hydrogen bonds, the complex would be expected to be easily soluble in aqueous solution and may be moisture-sensitive. As shown in Figures 4.50 and 4.52, one of the hydroquinone molecules is sandwiched with the alkyl groups from both sides of the benzene ring in the common packing pattern. The sandwiched structure, composed of fairly strong C-H... $\pi$  contacts, stabilizes the A-HQ molecules in the complex formation. The B-HQ molecule is

also sandwiched with the benzyl groups or ethyl groups of the neighbouring surfactant molecules from both sides of the benzene ring, although the interatomic distances are somewhat longer than those of A-HQ. Not only the O-H...O and O-H...halogen hydrogen bonds but also the C-H... $\pi$  contacts ensure the hydroquinone molecule is fully surrounded by surfactant molecules, and the crystalline complex becomes very stable to external reagents and to light. The above results suggest that if more suitable surfactants with longer alkyl chains were selected, it may be possible to design more thermally stable complexes.

#### 4.14 Application to Whitening Agents [88]

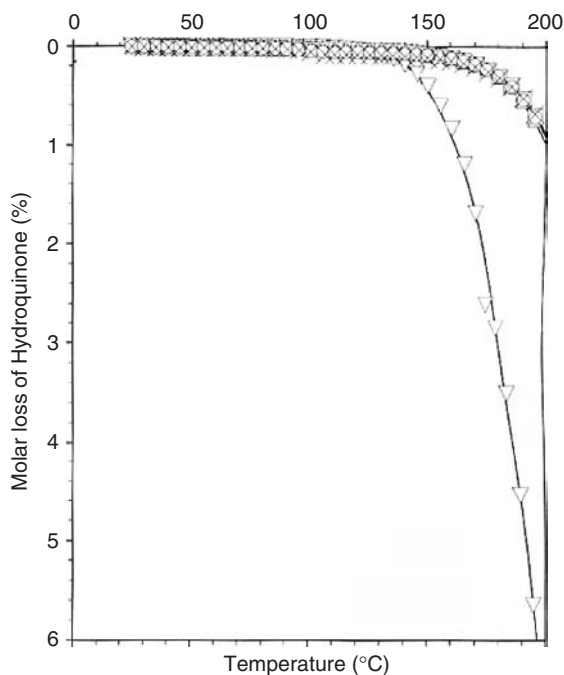
A variety of whitening agents have been developed to mask the manifestations of facial markings such as freckles [89–91]. Amongst whitening agents, hydroquinone has been well known to dermatologists to cure skin pigmentation because of its melanogenesis inhibitory effect [92–96]. Although the potential efficacy is very high, hydroquinone is very unstable on exposure to oxygen and light or upon heating as shown in the previous section. Since the complex of quinhydrone is different from hydroquinone, the coloured species obviously has no whitening effect. In order to inhibit the formation of quinhydrone, whitening agents, including hydroquinone, should be stored in a refrigerator and not be exposed to sun or even ambient light. However, it is very difficult to reduce the deterioration [97–100].

Consequently, complex formation of hydroquinone with surfactant molecules was investigated. In this section, four surfactant molecules with different alkyl chains are described, because they are considered to be safe for human health. As shown in Figure 4.55, they are benzyl(octadecyl)dimethylammonium chloride (BSDAC), benzyl(hexadecyl)dimethylammonium chloride (BCDAC), benzyl(tetradecyl)dimethylammonium chloride (BMDAC) and benzyl(dodecyl)dimethylammonium chloride



$n = 17$	Benzyl(octadecyl)dimethylammonium chloride	(BSDAC)
$n = 15$	Benzyl(hexadecyl)dimethylammonium chloride	(BCDAC)
$n = 13$	Benzyl(tetradecyl)dimethylammonium chloride	(BMDAC)
$n = 11$	Benzyl(dodecyl)dimethylammonium chloride	(BLDAC)

**Figure 4.55** Structural formulae of BSDAC, BCDAC, BMDAC and BLDAC



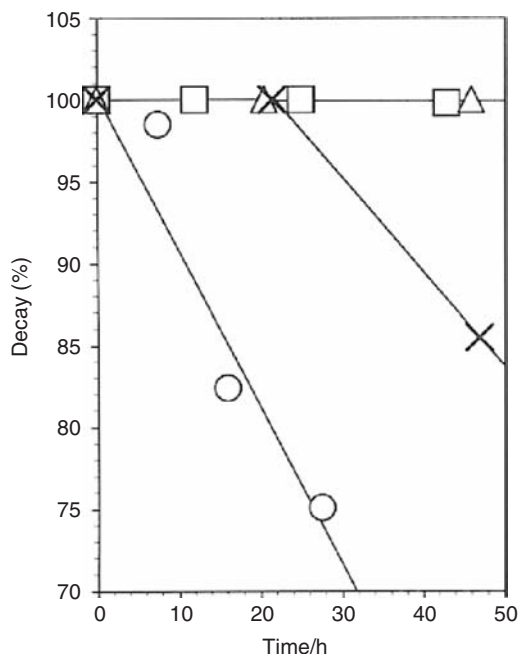
**Figure 4.56** Molar loss of hydroquinone due to the volatilization upon heating. The curves show the pure hydroquinone crystal (▽) and the four molecular complex crystals with BSDAC (□), BCDAC (Δ), BMDAC (○) and BLDAC (×). Reprinted from [88] with permission from the Japanese Cosmetic Science Society. Copyright 2005

(BLDAC). The molar loss of hydroquinone due to volatilization upon heating for the four complex crystals and hydroquinone only is shown in Figure 4.56. It is clear that the stability upon heating increases significantly.

The changes in the characteristic spectra of hydroquinone on exposure to oxygen gas at 37°C for the powdered samples of three complex crystals and hydroquinone only are shown in Figure 4.57. Although the sample of hydroquinone only is easily deteriorated, the samples of the three complexes revealed significant resistivity to the oxygen exposure.

The powdered samples of four complex crystals and hydroquinone only were exposed to a xenon lamp in a vacuum at 25°C and the characteristic spectra of hydroquinone were measured. For the sample of hydroquinone only, it was found that about 30% of hydroquinone had deteriorated, whereas the hydroquinone in the complex crystals showed insignificant change. The resistivity of the complex crystals to the light exposure was clearly confirmed.

After this, the complex of BCDAC and hydroquinone was selected, because the surfactant BCDAC has been used previously in ophthalmic solutions. First, the stability of the complex was checked when it was mixed with several ointments. The complex was mixed with hydrophilic ointment (HO), hydrophilic petrolatum (HP), white petrolatum (WP), plastibase (PB) and zinc oxide ointment (ZO). The colour value,



**Figure 4.57** Decay curves of hydroquinone on exposure to oxygen gas at 37°C. The curves show the pure hydroquinone crystal (○) and the three molecular complex crystals with BMDAC (×), BSDAC (□), and BCDAC (△). Reprinted from [88] with permission from the Japanese Cosmetic Science Society. Copyright 2005

luminosity, of each sample ( $L^*$ ) was measured with a colorimeter (COLOR-READER CR-13, MINOLTA). The highest or lowest brightness is 100 or 0, respectively. The values of the fresh sample and the corresponding sample kept at room temperature for 2 months were compared. The results are shown in Table 4.6. For the second sample of HO with hydroquinone only, a small portion of ascorbic acid was added to prevent oxidation. However, the deterioration was not stopped although it is noted that the complexes showed only a small change.

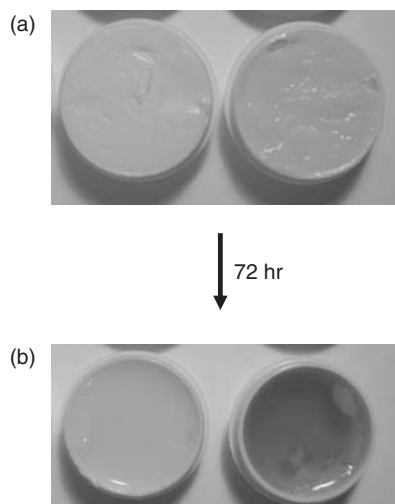
As there was no significant difference in stability of the complex mixed with the ointment, emollient cream was used instead. The colour change of the complex and hydroquinone only, each mixed with emollient cream, were compared as a fresh sample and samples kept at 40 °C for 72 h, as shown in Figure 4.58. The stability of the complex sample is conserved when the complex is mixed with the cream. The change in the  $L^*$  value is shown in Figure 4.59.

In order to use the above sample as a whitening agent, it is necessary to perform appropriate safety checks for human skin. Since this requires extensive knowledge of cosmetology, we can only state that the results with the complex showed no significant defects in terms of primary skin irritation, acute toxicity, skin sensitization, chronic toxicity, and human patch testing. All the above results indicated that the BCDAC–hydroquinone complex might be safely used as a whitening agent.

**Table 4.6** Preparation decay under ambient conditions determined the using  $L^*$  value measurement<sup>a</sup>

Samples	$L^*$
	Start → kept for 2 months
HO only	84.8 → 80.6
HO with HQ only	88.3 → 50.1
HO with BCDAC–HQ complex	86.5 → 77.8
HP only	73.8 → 72.8
HP with BCDAC–HQ complex	70.4 → 68.7
WP only	75.8 → 74.7
WP with BCDAC–HQ complex	69.2 → 66.5
PB only	78.5 → 76.9
PB with BCDAC–HQ complex	72.3 → 71.0
ZO only	86.8 → 81.4
ZO with BCDAC–HQ complex	89.3 → 80.5

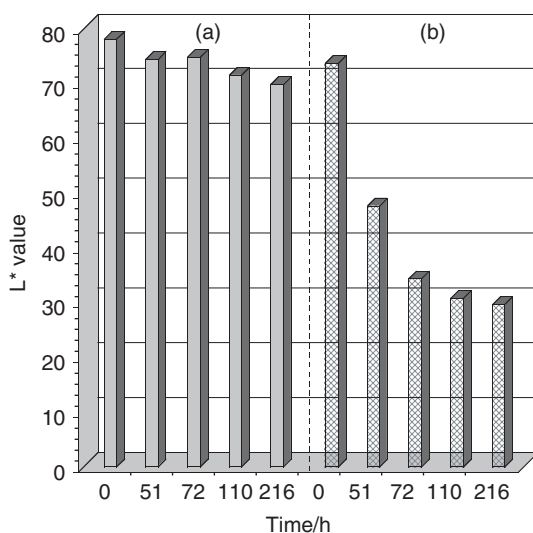
<sup>a</sup>HO, hydrophilic ointment; HQ, hydroquinone; HP, hydrophilic petrolatum; WP, white petrolatum; PB, plastibase; ZO, zinc oxide ointment.



**Figure 4.58** Colour changes of the complex (a) and hydroquinone only (b) mixed with emollient cream are compared between the fresh sample (upper) and samples kept at 40°C for 72 h (lower). Reprinted from [88] with permission from the Japanese Cosmetic Science Society. Copyright 2005 (see colour plate section)

Although it has been made clear that the BCDAC–hydroquinone complex is more stable than hydroquinone only and might be safely used as a whitening agent for human skin, the most important point is whether or not the efficiency of whitening is satisfactory for human skin. In order to check this efficiency, two tests were performed: one was the monitoring of the reduction of melanogenesis using a 3-D tissue culture model of human epidermis; and the other was medical examination of the patient. It was





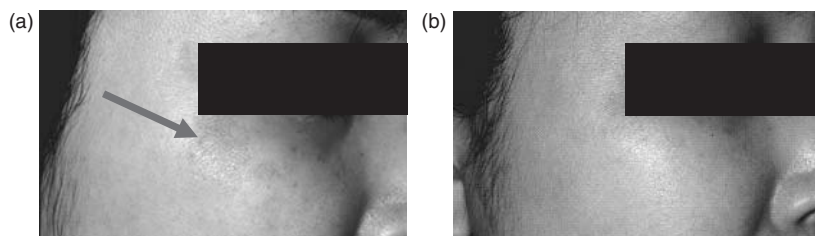
**Figure 4.59** The  $L^*$  value measurements on emollient creams with BCDAC–hydroquinone complex (a) and hydroquinone only (b). Reprinted from [88] with permission from the Japanese Cosmetic Science Society. Copyright 2005

**Table 4.7** Evaluation for hydroquinone only and for the BCDAC–hydroquinone complex

Type of dyschromatosis	Hydroquinone only	BCDAC–hydroquinone complex
Chloasma	36% (9/25)	50% (5/10)
<i>Lentigo senilis</i>	54.2% (13/24)	46.2% (6/13)
Postinflammatory hyperpigmentation	63% (17/24)	81.3% (13/16)
Freckle	75% (3/4)	80% (4/5)
<i>Nevus spilus</i>	0% (0/3)	0% (0/2)
Lichenoid keratosis	0% (0/3)	33.3% (1/3)
Total	48.8%	59.2%

found that the melanogenesis was not reduced by adding BCDAC only whereas the BCDAC–hydroquinone complex reduced the melanogenesis to nearly the same degree as that by hydroquinone only. Table 4.7 shows the evaluation for different disorders comparing hydroquinone only and the BCDAC–hydroquinone complex. For any type of dyschromatosis except *Lentigo senilis*, the BCDAC–hydroquinone complex gave a higher value than hydroquinone only. This indicates that the surfactant BCDAC may increase the permeability of hydroquinone to human skin.

Finally, the BCDAC–hydroquinone complex mixed with cream was applied to a patient who suffered from chloasma. The complex with cream was applied to the skin once a day for 1 month. Photographs of the patient's face before and after the treatment are compared in Figure 4.60, from which it is clear that the chloasma disappeared. The above results suggest the BCDAC–hydroquinone complex should be a very effective whitening agent.



**Figure 4.60** Photographs before (a) and after (b) the application of the BCDAC–hydroquinone complex with cream to a patient's skin once a day for 1 month. Reprinted from [88] with permission from the Japanese Cosmetic Science Society. Copyright 2005 (see colour plate section)

These findings were registered at the Japanese Patent Office (Patent No. 3712066). The whitening creams, including the BCDAC-hydroquinone complex, have been commercialized by several companies.

Although successful applications are currently limited, complex formation with surfactants holds great promise for further applications and is likely to be used in a variety of fields in the future.

## Acknowledgements

The authors express their thanks to Professor Hirotaka Hirata of Niigata University of Pharmacy and Applied Life Sciences and Dr Tomohiro Maruyama of Saiseikai Niigata Second Hospital and Mr Tatsuo Kitamura, Dr Hidehiro Uekusa and Dr Akiko Sekine of Tokyo Institute of Technology for their kind help and valuable discussion.

## References

1. J. H. Fendler and E. J. Fendler, *Catalysis in Micellar and Macromolecular Systems*, Academic Press, New York (1975), p 20.
2. R. S. Stearns, H. Oppenheimer, E. Simon and W. D. Harkins, *J. Chem. Phys.*, **15**, 496 (1947).
3. M. E. L. McBain and E. Hutchinson, *Solubilization and Related Phenomena*, Academic Press, New York (1955).
4. R. Williams, J. N. Phillips and K. J. Mysels, *Trans. Faraday Soc.*, **51**, 728 (1955).
5. W. Lin, *Bull. Chem. Soc. Jpn.*, **28**, 227 (1955).
6. C. W. Hoerr and A. W. Ralston, *J. Am. Chem. Soc.*, **65**, 976 (1943).
7. M. Shirai and B. Tamamushi, *Bull. Chem. Soc. Jpn.*, **29**, 733 (1956).
8. P. Mukerjee and K. J. Mysels, *J. Am. Chem. Soc.*, **77**, 2937 (1955).
9. P. Becker, *J. Phys. Chem.*, **63**, 1675 (1959).
10. K. Shinoda, *J. Phys. Chem.*, **58**, 541 (1954).
11. J. W. McBain, R. C. Merrill, Jr and J. R. Vinograd, *J. Am. Chem. Soc.*, **62**, 2880 (1940).
12. C. R. Singleterry and L. A. Weinberger, *J. Am. Chem. Soc.*, **73**, 4574 (1951).

13. S. Ross, *J. Colloid Sci.*, **6**, 497 (1951).
14. A. Kitahara, *Bull. Chem. Soc. Jpn.*, **28**, 234 (1955).
15. A. Kitahara, *Bull. Chem. Soc. Jpn.*, **29**, 15 (1956).
16. L. M. Kushner and W. D. Hubbard, *J. Phys. Chem.*, **57**, 898 (1953).
17. R. Matsuura, H. Kimizuka, S. Miyamoto and R. Shimazawa, *Bull. Chem. Soc. Jpn.*, **31**, 532 (1958).
18. E. Hutchinson and C. S. Mosher, *J. Colloid Sci.*, **11**, 352 (1956).
19. K. Shinoda and J. H. Hildebrand, *J. Phys. Chem.*, **61**, 789 (1957).
20. H. B. Klevens, *J. Phys. Colloid Chem.*, **52**, 130 (1948).
21. J. N. Phillips and K. J. Mysels, *J. Phys. Chem.*, **59**, 325 (1955).
22. W. Prins and J. J. Hermans, *J. Phys. Chem.*, **59**, 576 (1955).
23. K. J. Mysels and L. H. Princen, *J. Phys. Chem.*, **63**, 1696 (1959).
24. E. Gonick and J. W. McBain, *J. Am. Chem. Soc.*, **69**, 334 (1947).
25. L. Hess and L. A. Suranyi, *Z. Physik. Chem. (Leipzig)*, **A184**, 321 (1939).
26. A. Kitahara, *Bull. Chem. Soc. Jpn.*, **31**, 288 (1958).
27. S. Miyamoto, *Bull. Chem. Soc. Jpn.*, **33**, 371 (1960).
28. L. L. Neff, O. L. Wheeler, H. V. Tartar and E. C. Lingafelter, *J. Am. Chem. Soc.*, **70**, 1989 (1948).
29. A. Weissler, J. W. Fitzgerald and I. Resnick, *J. Appl. Phys.*, **18**, 434 (1947).
30. E. L. Colichman, *J. Am. Chem. Soc.*, **72**, 4036 (1950).
31. R. Tamamushi and T. Yamanaka, *Bull. Chem. Soc. Jpn.*, **28**, 673 (1955).
32. H. Okuyama and K. Tyuzyo, *Bull. Chem. Soc. Jpn.*, **27**, 259 (1954).
33. L. M. Kushner, B. C. Duncan and J. I. Hoffman, *J. Res. Natl. Bur. Stand.*, **49**, 85 (1952).
34. N. Sata and K. Tyuzyo, *Bull. Chem. Soc. Jpn.*, **26**, 177 (1953).
35. B. S. Harap and I. J. O'Donnell, *J. Phys. Chem.*, **58**, 1097 (1954).
36. D. Stigter and K. J. Mysels, *J. Phys. Chem.*, **59**, 45 (1955).
37. A. Kitahara, *Bull. Chem. Soc. Jpn.*, **28**, 586 (1955).
38. A. Kitahara, *Bull. Chem. Soc. Jpn.*, **28**, 653 (1955).
39. J. E. Lovelock, *J. Chem. Soc.*, **24**, 115 (1951).
40. J. C. Ericksson and G. Gilberg, *Acta Chem. Scand.*, **20**, 2019 (1966).
41. J. H. Fendler and L. K. Patterson, *J. Phys. Chem.*, **75**, 3907 (1971).
42. S. J. Rehfeld, *J. Phys. Chem.*, **75**, 3905 (1971).
43. K. A. Dill, D. E. Koppel, R. S. Cantor, J. D. Dill, D. Bendedouch and S. Chen, *Nature*, **309**, 42 (1984).
44. F. M. Menger and D. W. Doll, *J. Am. Chem. Soc.*, **106**, 1109 (1984).
45. T. Tachibana, T. Yoshimizu and K. Hori, *J. Am. Chem. Soc.*, **107**, 509 (1985).
46. J.-H. Fuhrhop, P. Schneider, J. Rosenberg and E. Boekema, *J. Am. Chem. Soc.*, **109**, 3387 (1987).
47. J.-H. Fuhrhop, P. Schneider, E. Boekema and W. Helfrich, *J. Am. Chem. Soc.*, **110**, 2861 (1988).
48. Y. Sakaiguchi, T. Shikata, H. Urakami, A. Tamura and H. Hirata, *J. Electron Microsc.*, **36**, 168 (1987).
49. Y. Sakaiguchi, T. Shikata, H. Urakami, A. Tamura and H. Hirata, *Colloid Polym. Sci.*, **265**, 750 (1987).

50. T. Shikata, Y. Sakaiguchi, H. Urakami, A. Tamura and H. Hirata, *J. Colloid Interface Sci.*, **119**, 291 (1987).
51. (a) H. Hirata and Y. Sakaiguchi, *Bull. Chem. Soc. Jpn.*, **62**, 581 (1989);  
(b) H. Hirata and Y. Sakaiguchi, *J. Colloid Interface Sci.*, **127**, 589 (1989).
52. H. Hirata, Y. Kanda and S. Ohashi, *Colloid Polym. Sci.*, **270**, 781 (1992).
53. K. Sawada, T. Kitamura, Y. Ohashi, N. Iimura and H. Hirata, *Bull. Chem. Soc. Jpn.*, **71**, 2109 (1998).
54. K. Sawada, Doctor Thesis, Tokyo Institute of Technology (1999).
55. N. Iimura, Doctor Thesis, Tokyo Institute of Technology (2000).
56. T. Shimizu and M. Masuda, *J. Am. Chem. Soc.*, **119**, 2812 (1997).
57. N. Yamada and M. Kawasaki, *J. Chem. Soc. Chem. Commun.*, 568 (1990).
58. N. Yamada, K. Okuyama, T. Serizawa and M. Kawasaki, *J. Chem. Soc., Perkin Trans. 2*, 2707 (1996).
59. A. R. Campanelli and L. Scaramuzza, *Acta Crystallogr.* **C42**, 1380 (1986).
60. P. K. Vinson and Y. Talmon, *J. Colloid Interface Sci.*, **133**, 288 (1989).
61. T. Kitamura, Master Thesis, Tokyo Institute of Technology (1993).
62. R. F. Zurcher, *Helv. Chim. Acta*, **44**, 1380 (1961).
63. R. F. Zurcher, *Helv. Chim. Acta*, **46**, 2054 (1963).
64. N. S. Bhacca and D. H. Williams, *Tetrahedron Lett.*, 3127 (1964).
65. D. H. Williams and N. S. Bhacca, *Tetrahedron*, **21**, 2021 (1965).
66. G. C. Pimentel and A. L. McClellan, *The Hydrogen Bond*, Freeman, San Francisco (1960).
67. R. D. Green, *Hydrogen-Bonding by C-H Groups*, Wiley Interscience, New York (1974).
68. A. D. U. Hardy and D. D. MacNicol, *J. Chem. Soc. Perkin Trans. 2*, 1140 (1976).
69. J. L. Atwood, F. Hamada, D. K. Robinson, G. W. Orr and R. L. Vincent, *Nature*, **349**, 603 (1991).
70. J. L. Atwood, S. Bott, C. Jones and C. L. Raston, *J. Chem. Soc. Chem. Commun.*, 1349 (1992).
71. G. Ferguson, J. F. Gallacher, C. Gildewell and C. Zakaria, *Acta Cryst.*, **C50**, 70 (1994).
72. M. Nishio and M. Hirota, *Tetrahedron*, **45**, 7201 (1989).
73. Y. Umezawa, S. Tsuboyama, K. Honda, J. Uzawa and M. Nishio, *Bull. Chem. Soc. Jpn.*, **71**, 1207 (1998).
74. G. R. Desiraju and T. Steiner, *The Weak Hydrogen Bond*, Oxford Science Publications, Oxford (1999).
75. P. Mukerjee and K. J. Mysels, *Critical Micelle Concentration of Aqueous Surfactant Systems*, US Department of Documents, Washington, DC (1971).
76. K. A. Dill, D. E. Koppel, R. S. Cantor, J. D. Dill, D. Bendedouch and S. Chen, *Nature*, **309**, 42 (1984).
77. V. M. Coiro, F. Mazza and G. Pochetti, *Acta Crystallogr.*, **C42**, 991 (1986).
78. N. Iimura, K. Sawada, Y. Ohashi and H. Hirata, *Bull. Chem. Soc. Jpn.*, **72**, 2417 (1999).
79. H. Hirata, A. Ohira and N. Iimura, *Langmuir*, **12**, 6044 (1996).
80. N. Iimura, Y. Ohashi and H. Hirata, *Bull. Chem. Soc. Jpn.*, **73**, 1097 (2000).

81. N. Iimura, H. Hirata and Y. Ohashi, *Bull. Chem. Soc. Jpn.*, **75**, 71 (2002).
82. P. Ballirano, R. Caminiti, C. Sadun, V. M. Coiro, G. Mancini, and A. Maras, *Z. Kristallogr.*, **213**, 123 (1998).
83. N. Iimura, Y. Fujimura, A. Sekine, S. Kitagawa and Y. Ohashi, *Bull. Chem. Soc. Jpn.*, **78**, 418 (2005).
84. P. G. Engasser and H. I. Maibach, *J. Am. Acad. Dermatol.*, **5**, 143 (1981).
85. A. Garcia and J. E. Fulton, Jr, *Dermatol. Surg.*, **22**, 443 (1996).
86. K. Yoshimura, A. Momosawa, A. Watanabe, K. Sato, D. Matsumoto, E. Aiba, K. Hattori, T. Yamamoto, T. Aoyama and T. Iga, *Dermatol. Surg.*, **28**, 1153 (2002).
87. N. Tatsumoto, S. Fujii and K. Yamaoka, *J. Acoust. Soc. Jpn.*, **11**, 63 (1990).
88. N. Iimura, T. Maruyama, S. Kitagawa and Y. Ohashi, *J. Jpn. Cosmet. Sci. Soc.*, **29**, 301 (2005).
89. S. Uehara, Y. Mizutani, M. Nakayama, O. Sakata, I. Sasaki, K. Arakane, T. Suzuki, R. Uchida and M. Tokutake, *J. Jpn. Cosmet. Sci. Soc.*, **27**, 247 (2003).
90. K. Koide and J. Miyazaki, *Fragrance J.*, **18**, 97 (2003).
91. T. Kitahara and T. Hachitani, *Fragrance J.*, **18**, 76 (2003).
92. T. Ohshima, K. Ohshima, F. Morikawa and K. Takahashi, *Nishinohon J. Dermatol.*, **42**, 115 (1980).
93. G. P. Engasser and I. H. Maibach, *J. Am. Acad., Dermatol.*, **5**, 143 (1981).
94. N. Hayashi, H. Fukinaka, M. Yukawa, K. Nakamura and K. Tamaoki, *Nishinohon J. Dermatol.*, **62**, 668 (2000).
95. Y. Tanaka, A. Okumura, H. Shimada, M. Tamura, S. Sasada and K. Takagi, *Iyaku J.*, **37**, 155 (2001).
96. K. Ueda and H. Yoshida, *Iyaku J.*, **20**, 99 (1984).
97. N. Karashima, S. Anjo, Y. Nagata, Y. Suzuki, Y. Kondo and M. Maruyama, *JJSHP*, **24**, 801 (1988).
98. T. Matsubayashi, T. Sakaeda, T. Kita, M. Nara, Y. Funasaka, M. Ichihashi, T. Fujita, F. Kamiyama, A. Yamamoto, J. J. Nordlund, M. Kaneko, A. Iida, M. Hirai and K. Okumura, *Biol. Pharm. Bull.*, **25**, 92 (2002).
99. T. Mastibayashi, T. Sakaeda, T. Kita, T. Nakamura, M. Kakumoto, Y. Funasaka, M. Ichihashi, T. Fujita, F. Kamiyama, A. Yamamoto, J. J. Nordlund, M. Kaneko, A. Iida and K. Okumura, *Biol. Pharm. Bull.*, **26**, 120 (2003).
100. Y. Kawasaki, M. Moriyama, H. Araki and H. Gomida, *Nihonbyouinyakuza-isikaizassi*, **40**, 41 (2004).

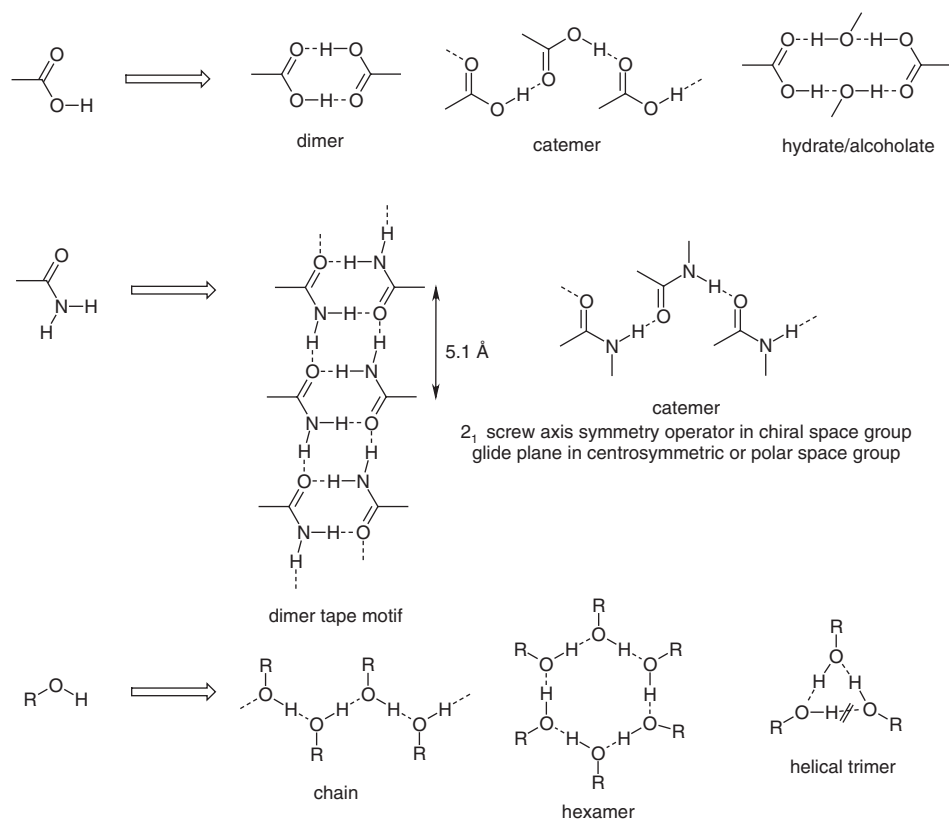
# 5

## Hydrogen Bonding and Molecular Packing in Multi-functional Crystal Structures

*Ashwini Nangia*

### 5.1 Introduction

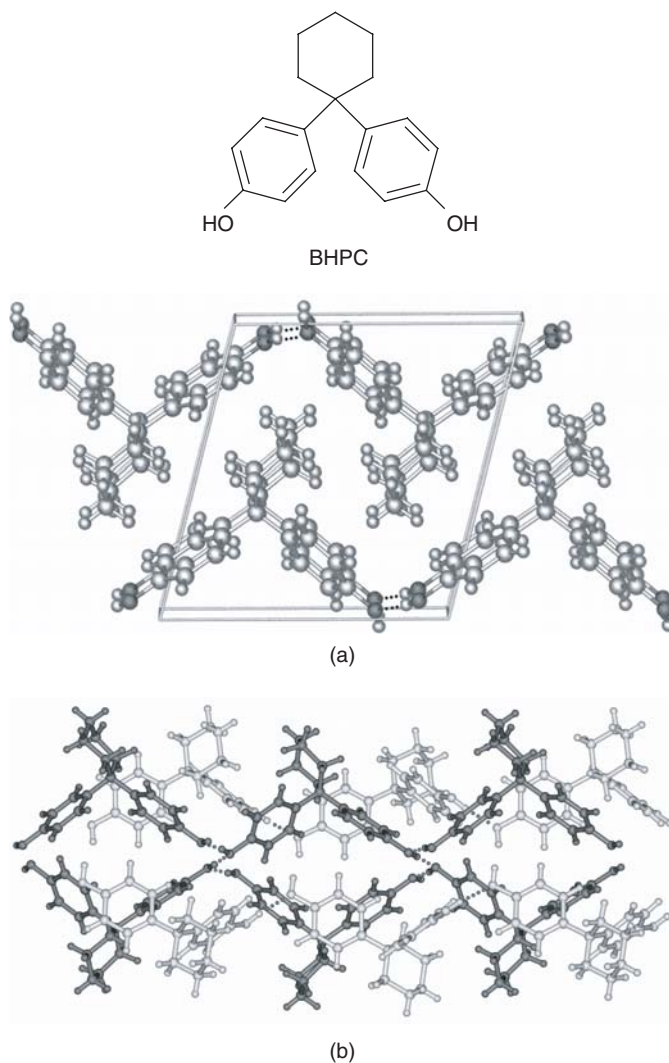
The functional group is such a fundamental concept in organic chemistry that textbooks often cover this topic chapter-wise. The reactivity of a molecule is related to the functional group(s) present, such as carboxylic acid, carboxamide, alcohol, phenol, amine, aniline, aromatic and halogen. Thus, knowledge of the functional group chemistry provides information about the reactions and products of compounds in a given category [1]. For example, carboxylic acids react with alcohols and amines to give esters and amides, aromatic compounds react with electrophilic reagents to give substituted aromatics, alkyl halides undergo substitution or elimination with nucleophilic and basic species, and so on. However, in supramolecular chemistry [2] the behaviour of a given functional group is not so easily derived. For example, carboxylic acids not only form the expected dimer but also give hydrogen-bonded catemer and hydrated motifs (Figure 5.1). Different aggregation patterns of functional groups in the solid state have been termed supramolecular synthons [3]. Thus, the COOH group is generally associated with the finite dimer synthon but occasionally infinite catemers are also formed. The distinction between supramolecular synthons originating from the same functional group can be subtle. Alcohols assemble via  $\text{O-H} \cdots \text{O}$  hydrogen bond chains of molecules related by screw-axis or glide plane symmetry operators [4] (Figure 5.1). A casual glance will not



**Figure 5.1** One functional group can give more than one supramolecular synthon in crystal structures. Three common functional groups and their common synthons are shown

reveal differences between these similar packing arrangements, but they are very distinct because the molecules comprising them have the same or the opposite chirality.

Thanks to rapid growth in crystal engineering [5], notably in the last two decades, it is now possible to carry out the premeditated construction of ordered crystalline assemblies from functionalized molecules. The first volume of *Frontiers in Crystal Engineering* [6] provides an up-to-date status of the subject. We now know that crystal structures are determined by a convolution of several types of intermolecular interactions that have different strengths and directional preferences [7]. For example, the crystal structure of 1,1-bis(4-hydroxyphenyl)cyclohexane (BHPC) has linear chains of  $\text{O}-\text{H}\cdots\text{O}$  hydrogen bonds that connect the bis-phenol molecules along the  $b$ -axis. The second OH group engages in  $\text{O}-\text{H}\cdots\pi$  interaction. Cyclohexyl and phenyl groups dovetail in the voids between the hydrogen-bonded chains such that bumps fit into hollows (Figure 5.2) [8]. There are three types of intermolecular interactions in the first example structure: strong and directional  $\text{O}-\text{H}\cdots\text{O}$  hydrogen bonds [9], weak  $\text{O}-\text{H}\cdots\pi$  interaction of moderate directionality [10], and finally the numerous isotropic van der Waals interactions. Hydrogen bonding is unquestionably the most important cohesive force in directing the supramolecular organization in the solid state. However in energy terms the hydrogen



**Figure 5.2** (a)  $O-H \cdots O$  chain along  $[010]$  in the triclinic polymorph ( $P\bar{1}$ ) of BHPC having syn-oriented bis-phenol OH groups. (b) The rectangular voids in the molecular ladder along  $[010]$  of anti OH molecules (black) in the orthorhombic polymorph ( $Pbca$ ) have a cross-section of  $3.6 \times 2.2 \text{ \AA}$ .  $O-H \cdots O$  chains down  $[100]$  are not shown for clarity. The grey syn molecules have similar hydrogen bonding to that shown in (a) for the triclinic polymorph

bond contribution is no more than 20% for typical organic molecules. Alkyl and aryl groups provide the bulk of the stability to the crystal structure by space filling and close packing. The matter would be relatively simple if this was the only crystal structure for BHPC. There is a second polymorph [11], or form, of the same compound in which there are two molecules (syn and anti OH conformers) compared to syn conformer only in the first polymorph. The syn OH conformer arrangement is similar to the first structure



whereas the anti OH conformer engages in cooperative O–H···O chains along the *a*-axis using both OH groups. Cooperative hydrogen bonds are stronger than isolated and finite hydrogen bond motifs [9]. Thus, differences in molecular conformation, in this case syn and anti orientations of phenol OH groups, result in different hydrogen bond synthons and molecular packing. The two crystal structures of BHPC make it clear that it is very difficult to establish connections between the functional group and hydrogen bonding from the molecular diagram of an organic molecule. Crystal Structure Prediction blind tests [12] are a testimony to the challenges in the computational approach.

This chapter begins by showing organized crystalline assemblies in model compounds. New and at times unexpected hydrogen bond synthons begin to appear as molecular complexity is incrementally added. Starting from hydrogen bond rules developed for noncompetitive or minimal interference systems, hydrogen bond donor–acceptor recognition is described for conformationally flexible diaryl urea crystal structures. A better understanding of hydrogen bond donor and acceptor strengths and their selective pairing in multi-functional organic molecules emerge from these studies.

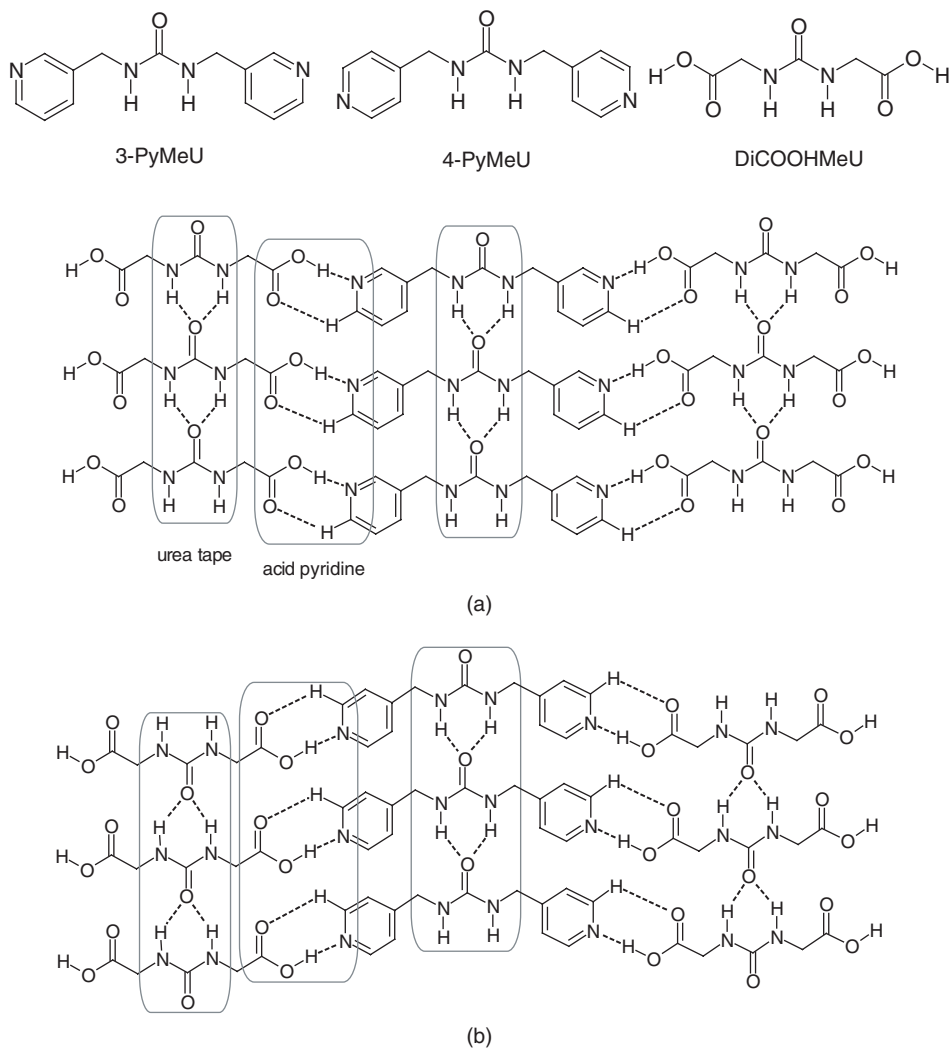
## 5.2 Hydrogen Bonding in Ureas and Amides

Etter [13] proposed hydrogen bonding rules in molecular crystals containing N–H donors and C=O acceptors:

- (1) All good proton donors and acceptors are used in hydrogen bonding.
- (2) Six-membered-ring intramolecular hydrogen bonds form in preference to intermolecular hydrogen bonds.
- (3) The best proton donors and acceptors remaining after intramolecular hydrogen bond formation form intermolecular hydrogen bonds with one another.

The third rule deals with ranking hydrogen bond donors and acceptors based on their strength: the strongest donor interacts with the strongest acceptor, the second best donor bonds to the second best acceptor, and so on. Lauher and co-workers [14] reported a series of crystal structures that follow the hierarchic hydrogen bond model containing urea, pyridyl and acid functional groups. The strongest COOH donor bonds to the best pyridine acceptor and the left over urea group assembles as 1-D tape (Figure 5.3). Aakeröy and co-workers [15] extended Etter's hydrogen bond hierarchy rule by showcasing termolecular supramolecular construction via acid–pyridine and amide–amide hydrogen bond synthons (Figure 5.4). However, preliminary results suggested to us that modular hydrogen bond assembly calls for a second look in multi-functional organic crystals. For example, the *m*-hydroxybenzoic acid and isonicotinamide co-crystal has acid–pyridine and amide dimer hierarchic hydrogen bonds [15] but the isomeric *p*-hydroxybenzoic acid and isonicotinamide co-crystal is sustained by completely different phenol–pyridine and acid–amide synthons [16] (Figure 5.5). In the  $\alpha,\omega$ -dicarboxylic acid and isonicotinamide homologue series [17], glutaric acid and adipic acid (Figure 5.6) afforded co-crystals with acid–pyridine and amide–amide synthons (1:2 complex) as well as acid–pyridine and acid–amide synthon combinations (1:1 complex).

Not just multi-component crystal structures but even simple molecular crystals can be quite difficult to classify using a simple predictive ranking of hydrogen bonds. Phenyl

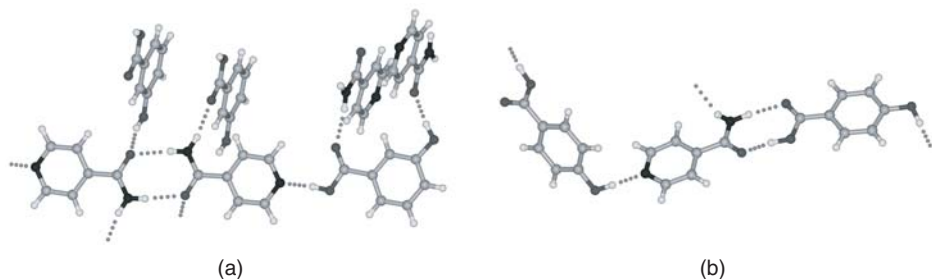


**Figure 5.3** Acid··pyridine heterosynthon and urea tape synthon in (a) co-crystal of *m*-pyridylmethylurea (3-PyMeU) and dicarboxylic acid methylurea (DiCOOHMeU) and (b) co-crystal of 4-PyMeU and DiCOOHMeU. Note the modular assembly via structurally insulated urea tape and acid-pyridine synthons in distinct domains of the crystal structure

urea crystallizes with the characteristic  $\alpha$ -network of bifurcated  $\text{N-H}\cdots\text{O}$  hydrogen bonds [18]. However, replacing the phenyl group by pyridyl disrupts this robust tape synthon of aryl ureas. The crystal structure of 3-pyridyl urea [19] is more like that of an aromatic amide. The  $\text{pyr-NH}$  forms the  $\text{N-H}\cdots\text{N}$  hydrogen bond and the left-over  $\text{CONH}_2$  group assembles as the characteristic amide tape synthon (Figure 5.7). Similar interference to  $\text{N-H}\cdots\text{O}$  synthons is seen in polymorphs of isonicotinamide [20] and pyrazinamide [21]. In the two polymorphs of isonicotinamide, one structure contains the  $\text{N-H}\cdots\text{O}$  dimer but the other has the  $\text{N-H}\cdots\text{N}$  hydrogen bond (Figure 5.8).



**Figure 5.4** Hierarchic hydrogen bonding between acid, pyridine and amide groups in ternary co-crystals: strongest acid $\cdots$ pyridine and next best acid $\cdots$ amide heterosynthons. (a) 4-Dimethylaminobenzoic acid•isonicotinamide•3,5-dinitrobenzoic acid and (b) 3,4-dimethoxycinnamic acid•isonicotinamide•3,5-dinitrobenzoic acid

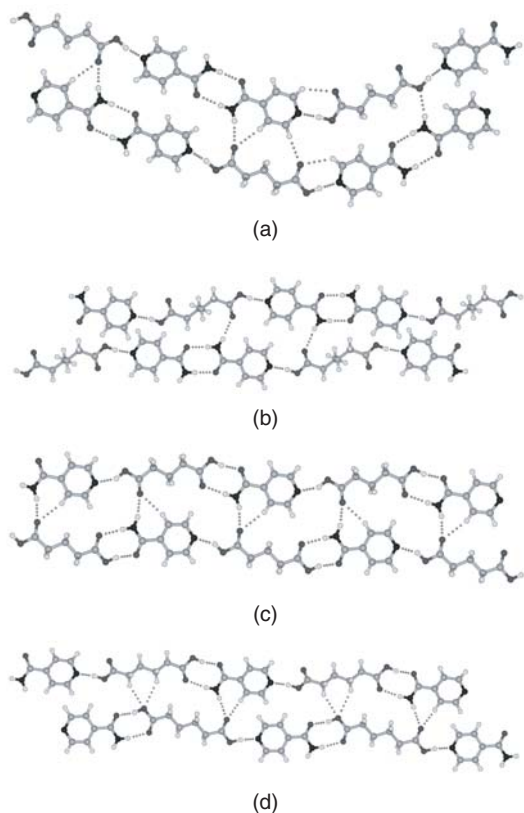


**Figure 5.5** Very different hydrogen bond synthons in isomeric co-crystals containing COOH, pyridine, CONH<sub>2</sub>, and OH functional groups. (a) Hierarchic hydrogen bonding in *m*-hydroxybenzoic acid•isonicotinamide and (b) nonhierarchic hydrogen bonding in *p*-hydroxybenzoic acid•isonicotinamide

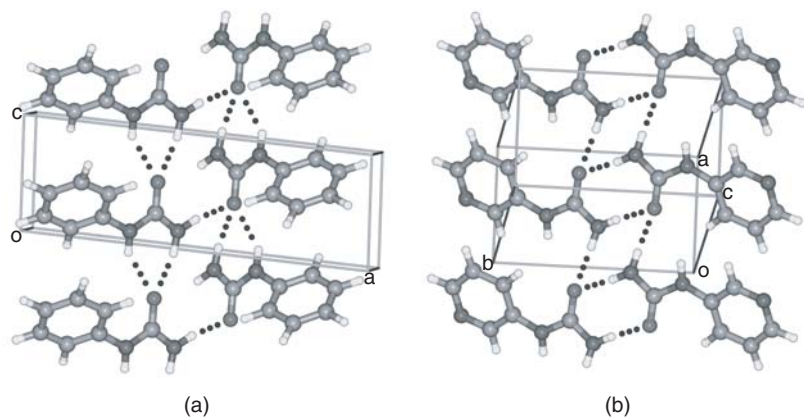
Pyrazinamide contains the amide dimer in its polymorphic crystal structures but there are differences in N–H $\cdots$ N interactions with pyrazine N acceptors. All this suggested that phenyl to pyridyl or pyrazinyl isosteric replacement can lead to very different hydrogen bonding and crystal packing motifs, even though the N–H $\cdots$ N interaction is relatively weak compared with the N–H $\cdots$ O hydrogen bond. The ability of a functional group to disturb a standard hydrogen bond synthon is referred to as structural interference, a term coined by Desiraju in the context of *gem*-alkynols [22]. Structural interference to the N–H $\cdots$ O urea tape and amide dimer synthons from the pyridyl functional group is discussed in the first half of this chapter. The second half deals with competition from the NO<sub>2</sub> group and a possible crystal engineering solution to restore the urea tape N–H $\cdots$ O synthon in otherwise divergent nitrophenyl urea crystal structures by exploiting the hard and soft acid–base (HSAB) concept.

### 5.3 Pyridyl Ureas and Amides

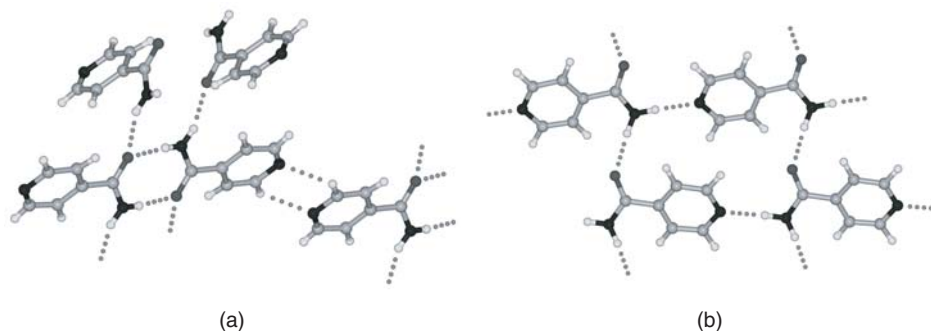
Starting from 3-pyridylurea whose crystal structure lacks the N–H $\cdots$ O tape synthon we studied the structural chemistry of *N,N'*-bis(3-pyridyl)urea, abbreviated as MDPU or *m*-dipyridyl urea [23]. The molecule can adopt at least three distinct planar molecular conformations (Figure 5.9) and all three were observed in the solid state. That the



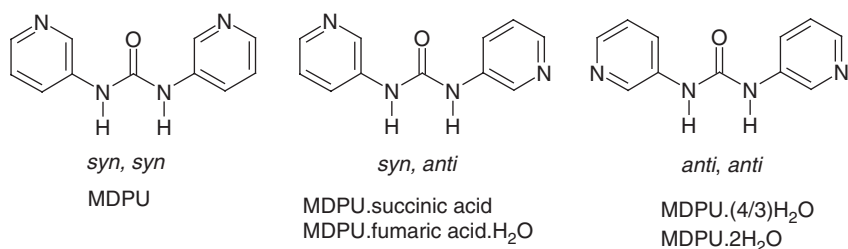
**Figure 5.6** Acid–pyridine and amide dimer synthons in 1:2 co-crystals of glutaric acid and adipic acid with isonicotinamide (a and b), and acid–pyridine and acid–amide synthons in 1:1 co-crystals (c and d)



**Figure 5.7** (a)  $N-H \cdots O$   $\alpha$ -network in the crystal structure of phenylurea and (b)  $N-H \cdots O$  dimer tape in crystal structure of 3-pyridylurea resembles a typical amide tape. The pyridyl-NH group interferes by engaging in  $N-H \cdots N$  interaction (not shown)



**Figure 5.8** (a) Amide dimer in isonicotinamide form 1 and  $N-H \cdots N$  hydrogen bond in form 2 (b). The pyridyl group participates in weak  $C-H \cdots N$  interactions in form 1 but plays a decisive role in form 2. The crystal structure of form 1 contains the expected synthon but form 2 is different

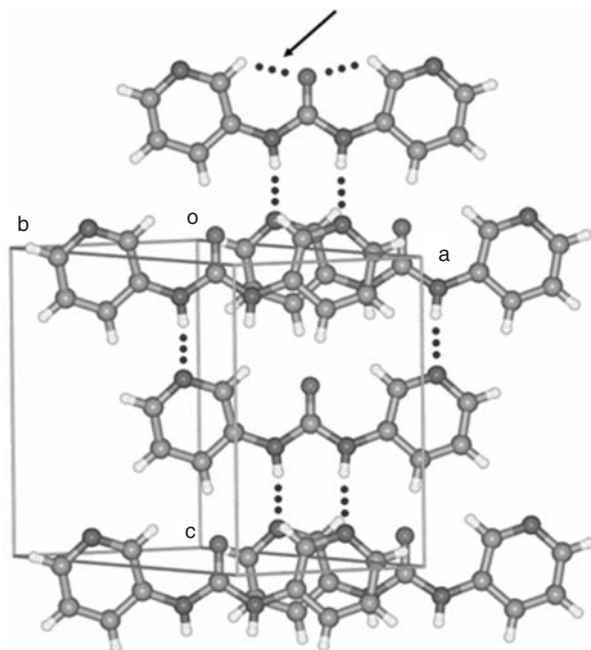


**Figure 5.9** Different conformations of pyridyl groups in MDPU. The syn conformer has carbonyl-O and pyridyl-N on the same side

pyridyl group is able to disrupt the urea  $\alpha$ -network was convincingly demonstrated in the crystal structure of MDPU. The compound crystallized in the uncommon space group *Aba*2 with the molecule lying on a twofold axis in the syn, syn planar conformation. NH donors hydrogen bond to different pyridine-N acceptors through short  $N-H \cdots N_{\text{pyridyl}}$  interactions (2.08 Å, 164°). Intramolecular  $C-H \cdots O_{\text{urea}}$  interactions (2.18 Å, 124°) from activated donor groups ortho to pyridyl-N (see arrow in Figure 5.10) provide auxiliary support to the planar conformation. Glide related molecules produce layers mediated via  $N-H \cdots N$  and  $C-H \cdots O$  interactions, and such layers are connected through  $N-H \cdots N$  hydrogen bonds.

Crystallization of MDPU from alcoholic solvents afforded hydrate forms,  $MDPU \cdot (4/3)H_2O$  from MeOH and  $MDPU \cdot 2H_2O$  from EtOH in space group *C*2/*c* and *P*2<sub>1</sub>/*c*, respectively. The pyridyl groups now adopt the anti, anti conformation. The hydrates, too, adopt nonurea type structures in that the characteristic  $N-H \cdots O$  tape is absent but the intramolecular  $C-H \cdots O_{\text{urea}}$  synthon persists (see arrows in Figure 5.11).

A crystal engineering solution to favour formation of the  $\alpha$ -network in dipyrindyl urea would be to engage the interfering pyridyl group in acid  $\cdots$  pyridine hydrogen bonding, such as shown in Figure 5.3. Crystal structures of MDPU with succinic acid (SA) and fumaric acid (FA) are shown in Figure 5.12. Now MDPU adopts the syn, anti conformation. The only structural feature common to the MDPU series is the recurring

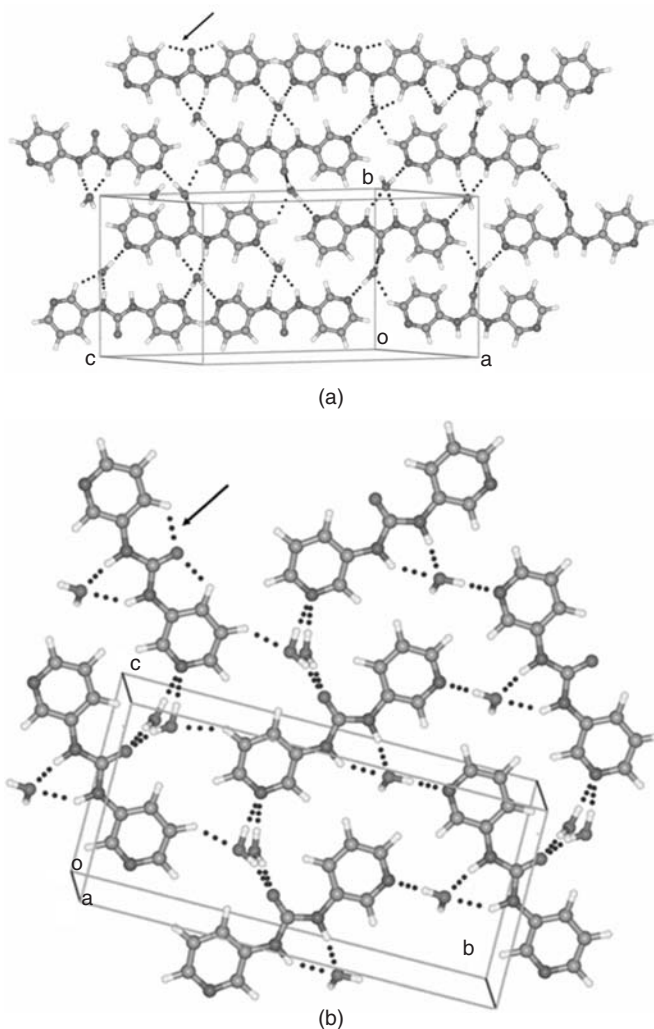


**Figure 5.10**  $N\text{-H}\cdots N_{\text{pyridyl}}$  hydrogen bonds and  $\text{C-H}\cdots\text{O}_{\text{urea}}$  interactions in MDPU. The urea tape synthon is absent.  $\text{C-H}\cdots\text{O}$  interactions are highlighted with an arrow in this and subsequent figures

$\text{C-H}\cdots\text{O}_{\text{urea}}$  synthon. The molecular conformation and hydrogen bonding are different from one structure to another but the urea tape is conspicuously absent. The urea tape is absent in *N*-phenyl-*N*-3-pyridylurea (a phenyl-pyridyl urea example) but the  $\text{C-H}\cdots\text{O}$  synthon is present.

Interference to the robust urea tape by  $\text{N-H}\cdots\text{N}$  hydrogen bonds is not limited to heteroaryl ureas. The  $\beta$ -sheet structure is sustained exclusively by  $\text{N-H}\cdots\text{O}$  hydrogen bonds in aromatic amides such as PhAm (Figure 5.13). However, introduction of the pyridyl moiety, e.g. as in the 3-PyAm and 3-PyRevAm homologue series [24], gave structures sustained by  $\text{N-H}\cdots\text{O}$  hydrogen bonds as well as by  $\text{C-H}\cdots\text{N}$  interactions and  $\text{N-H}\cdots\text{O}$  and  $\text{N-H}\cdots\text{N}$  hydrogen bonds (Figure 5.14,  $n = 6$ ). Thus, while the pyridyl group plays an auxiliary role in 3-PyAm structures it is definitely interfering via  $\text{N-H}\cdots\text{N}$  hydrogen bonds in 3-PyRevAm. 4-PyAm compounds are similar to 3-PyAm but 4-PyRevAm structures crystallize in hydrate form.

That there is structural interference in 3-PyRevAm but lack of it in 3-PyAm means that the same combination of functional groups may behave very differently in terms of supramolecular interactions. The same was noted in MDPU structures (Figures 5.10–12) versus PyMeU•DiCOOHMeU co-crystals (Figure 5.3) as well as for the isomeric co-crystal structures in Figure 5.4 that have the same functional groups but different hydrogen bond motifs. Thus, unpredictable structural interference from weak hydrogen bonding groups is the real challenge in establishing a molecular functional group to supramolecular synthon relationship in multi-functional crystal structures.

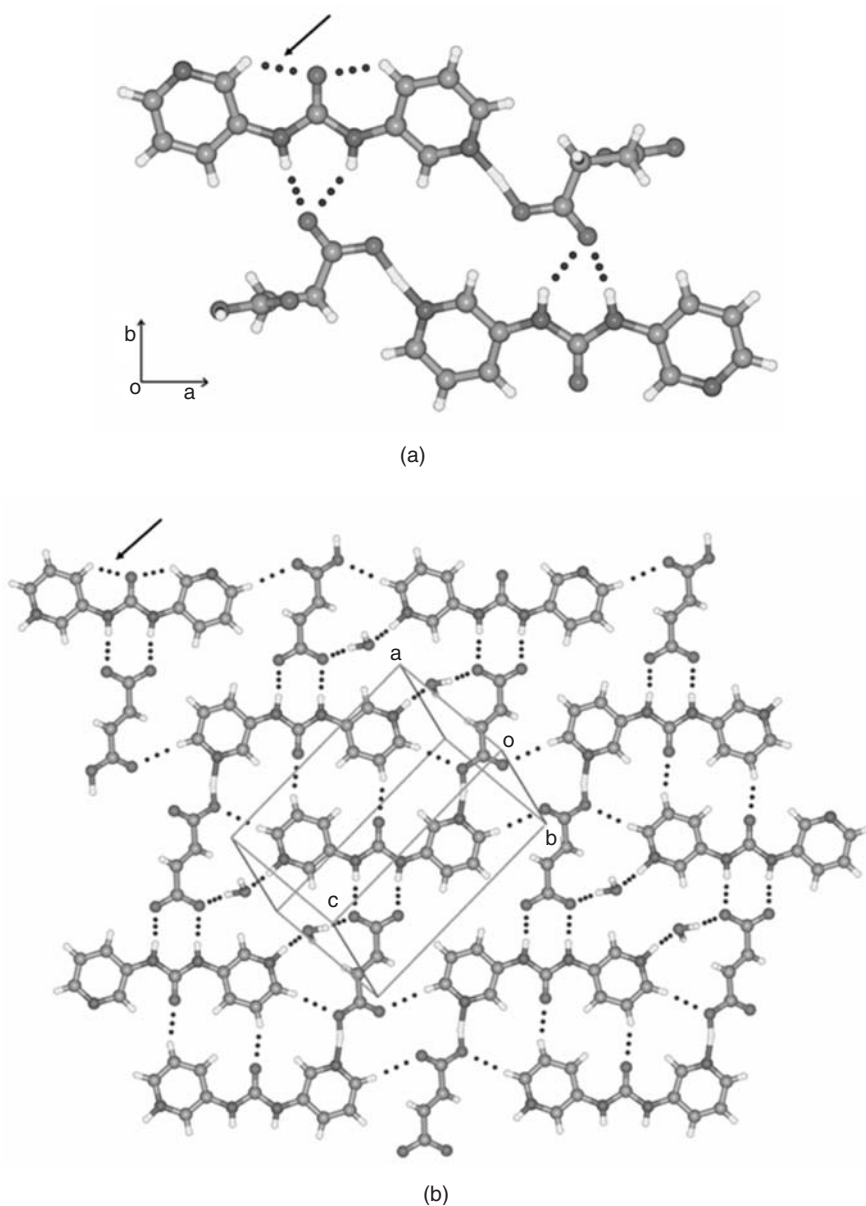


**Figure 5.11**  $N\text{-H}\cdots O_{\text{water}}$ ,  $O\text{-H}\cdots O_{\text{urea}}$  and  $O\text{-H}\cdots N_{\text{pyridyl}}$  hydrogen bonding in MDPU hydrates. (a)  $\text{MDPU}\bullet(4/3)\text{H}_2\text{O}$ . (b)  $\text{MDPU}\bullet 2\text{H}_2\text{O}$ . Note the absence of urea tape and the persistent  $\text{C-H}\cdots O_{\text{urea}}$  synthon (marked with arrow) in both structures

## 5.4 Nitrophenyl Ureas and Amides

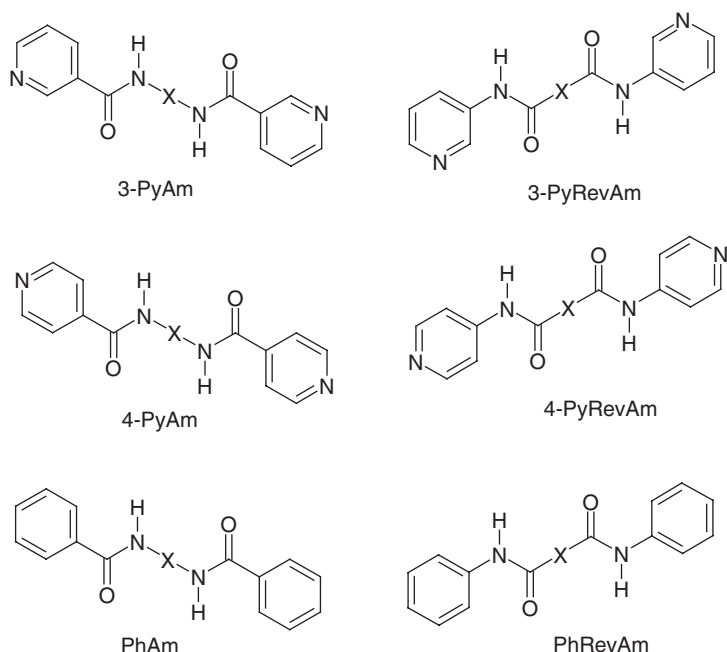
*N,N'*-di(*m*-nitrophenyl)urea or DNPU (dinitrophenyl urea) exists in three concomitant polymorphic forms ( $\alpha$ ,  $\beta$  and  $\delta$ ) of different colours [25]. It readily forms solvates and co-crystals with compounds having strong to moderate hydrogen bond acceptors [26]. The characteristic urea network is absent except in the noncentrosymmetric  $\beta$ -polymorph of DNPU whereas  $\alpha$  and  $\gamma$  polymorphs are sustained by  $\text{N-H}\cdots O_{\text{nitro}}$  H bonds. We examined hydrogen bond competition in crystal structures of *N*-X-phenyl-*N'*-*p*-nitrophenyl urea, abbreviated as PNPU-X [27] (Figure 5.15). A variety of X groups were considered,





**Figure 5.12** (a) Acid...pyridine and urea...carboxylate hydrogen bonding in MDPU•SA and (b) urea...carboxylate and acid...pyridine hydrogen bonding in MDPU•FA hydrate. Note the divergent nature of hydrogen bonding in this co-crystal structure pair compared with the modular assembly shown in Figure 5.3. The functional groups are the same in both cases although the molecular species are slightly different. The C-H...O<sub>urea</sub> synthon (marked with arrow) persists



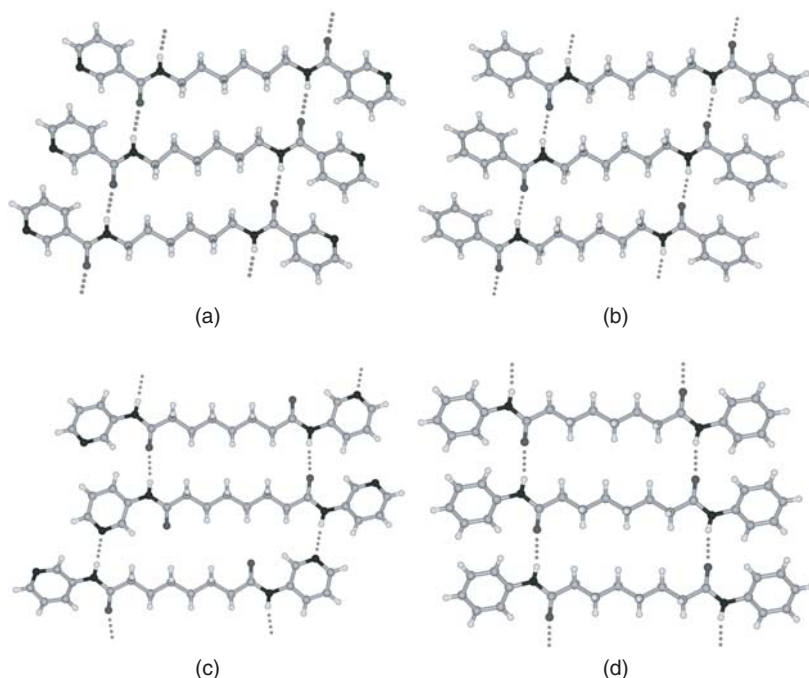


**Figure 5.13** 3-PyAm and 3-PyRevAm compounds, their 4-pyridyl analogues and the parent phenyl compounds.  $X = (\text{CH}_2)_n$ ,  $n = 2, 4, 6, 8$  in the homologue series

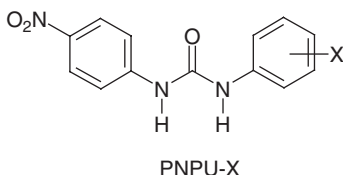
mostly at the para position (PNPUs;  $X = \text{H}, \text{F}, \text{Cl}, \text{Br}, \text{I}, \text{CN}, \text{C}\equiv\text{CH}, \text{CONH}_2, \text{COCH}_3, \text{OH}, \text{Me}$ ) but a few meta-substituted phenyl ureas as well to address the following questions: (1) Does the X group interfere with the urea functionality or does it play a bystander role? (2) What is the nature of the competition from the  $\text{NO}_2$  group and the status of  $\text{NO}_2 \cdots \text{X}$  interactions in urea structures? (3) Is it possible to crystallize the urea  $\text{N-H} \cdots \text{O}$  tape in a multi-functional milieu?

Crystal structures of halogen-substituted PNPUs ( $X = \text{F}, \text{Cl}, \text{Br}$  and  $\text{CN}$ ) are isomorphous and isostructural, as illustrated by the fluoro structure PNPu-F (Figure 5.16). The molecule adopts a planar conformation stabilized by intramolecular  $\text{C-H} \cdots \text{O}$  hydrogen bonds (2.24 Å,  $118^\circ$ ; 2.25 Å,  $118^\circ$ ), further assisted by resonance in the extended aromatic scaffold. The characteristic urea  $\alpha$ -network of  $\text{N-H} \cdots \text{O}$  hydrogen bonds is absent. Glide related urea molecules interact via  $\text{N-H} \cdots \text{O}_{\text{nitro}}$  H bonds (1.90 Å,  $160^\circ$ ; 2.22 Å,  $149^\circ$ ) such that one of the oxygen atoms of the nitro group acts as a bifurcated acceptor to form 1-D zigzag tapes, and these tapes are further connected by  $\text{C-H} \cdots \text{O}$  (2.45 Å,  $170^\circ$ ) interaction in a layered structure. Adjacent  $\pi$ -stacked layers are held together by van der Waals interactions. The nitro-O competes successfully to replace the stronger urea carbonyl-O by hydrogen bonding with the strongest N-H donors. The urea-C=O does not accept strong hydrogen bonds but only makes weak  $\text{C-H} \cdots \text{O}$  interactions. There are no significant inter-fluoride contacts.

A second category of structures are solvates of dimethylformamide (DMF) and dimethyl sulfoxide (DMSO) in which the dipolar solvent interferes with the urea tape synthon, e.g. PNPu-H, -CN, -Me, -CONH<sub>2</sub>, -COCH<sub>3</sub>. The solvent-O successfully



**Figure 5.14**  $\beta$ -Sheet structure in (a) 3-PyAm sustained via  $N-H \cdots O$  hydrogen bonded chain and  $C-H \cdots N$  catemer (not shown) and (c) 3-PyRevAm has alternating  $N-H \cdots O$  and  $N-H \cdots N$  hydrogen bonds. The corresponding phenyl structures are shown on the right (b and d) for reference. Note that (a) and (b) are similar but (c) and (d) are different, once again emphasizing the importance of regiochemistry in hydrogen bond synthons for isomeric compounds



**Category I** Urea  $N-H \cdots O_{\text{nitro}}$  H bonding

X = 4-F, 4-Cl, 4-Br, 4-CN, 3-I, 3-Br, 3-CH<sub>3</sub>

**Category II** Urea  $N-H \cdots O_{\text{solvent}}$  H bonding

X = H (DMF), 4-CN (DMF), 4-CONH<sub>2</sub> (DMF)

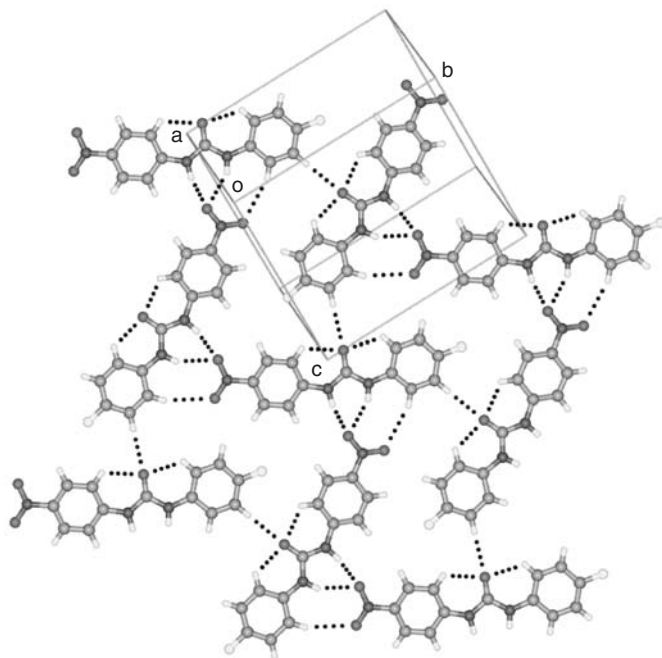
X = 4-I (DMSO), 4-CONH<sub>2</sub> (DMSO), 4-CH<sub>3</sub> (DMSO), 4-COCH<sub>3</sub> (DMSO)

X = 3-OH (H<sub>2</sub>O)

**Category III** Urea  $N-H \cdots O$  tape synthon

X = 4-I, 4-ethynyl, 4-NMe<sub>2</sub>, 4,4'-iodophenyl, 4,4'-ethynylphenyl, 4-CH<sub>3</sub>

**Figure 5.15** Crystal structures of *p*-nitrophenyl ureas (PNPU-X) are classified in three categories

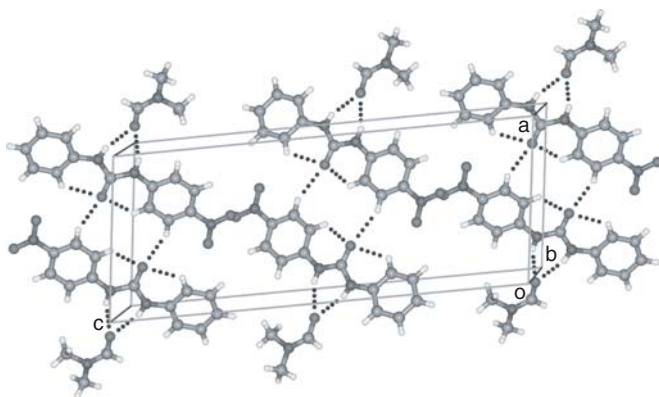


**Figure 5.16** Crystal structure of PNPU-F to show the  $N-H \cdots O_{\text{nitro}}$  and intramolecular  $C-H \cdots O_{\text{carbonyl}}$  interactions. The urea tape synthon is absent in the PNPU-Cl, -Br and -CN structures which have similar hydrogen bonding and crystal packing characterized by the nonurea tape category

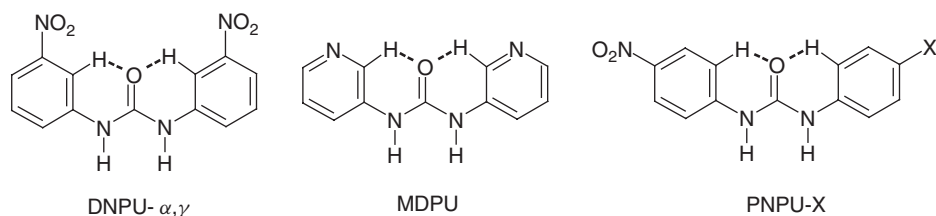
replaces the urea carbonyl or nitro acceptor groups for the NH donors. Surprisingly, the hydrogen bond ‘active’ nitro and urea-O acceptors are not involved in strong hydrogen bonding but make  $C-H \cdots O$  interactions only. The DMF solvent disrupts the urea  $\alpha$ -network through  $N-H \cdots O_{\text{DMF}}$  hydrogen bonds (1.88 Å, 159°; 1.90 Å, 157°) in PNPU-H (Figure 5.17).

A persistent theme in MDPU, DNPU and PNPU crystal structures is the absence of the urea tape network but recurrence of the intramolecular  $C-H \cdots O_{\text{urea}}$  synthon (Figure 5.18). A slight activation of the proximal C-H donor by an electron-withdrawing group is necessary. The  $C-H \cdots O$  interaction occurs on the pyridyl side but not the phenyl side in unsymmetrical urea [23]. Similarly, the  $\text{NO}_2$ -Ph moiety is co-planar with the urea group but the X-Ph side adopts different conformations in PNPU structures [27]. The recurrence of the intramolecular  $C-H \cdots O_{\text{urea}}$  interaction was also noted by Steed and co-workers [28] in closely related pyridyl urea systems (Figure 5.19).

A third category of crystal structures offered an elegant strategy for crystal design in multi-functional systems. The *p*-iodo, *p*-ethynyl, *p*-*N,N*-dimethylamino, *p*-iodophenyl, *p*-ethynylphenyl and *p*-methyl compounds (Figure 5.15) belong to category III. The iodo compound PNPU-I crystallized in the polar space group *Cc* with one molecule in the asymmetric unit. Now, self-assembly occurs via the urea  $N-H \cdots O$  tape synthon to produce a 1-D network along the *b*-axis (Figure 5.20). The iodine atom is in close contact with the nitro group via bifurcated  $I \cdots O$  interactions (3.29 Å, 169°; 3.62 Å,



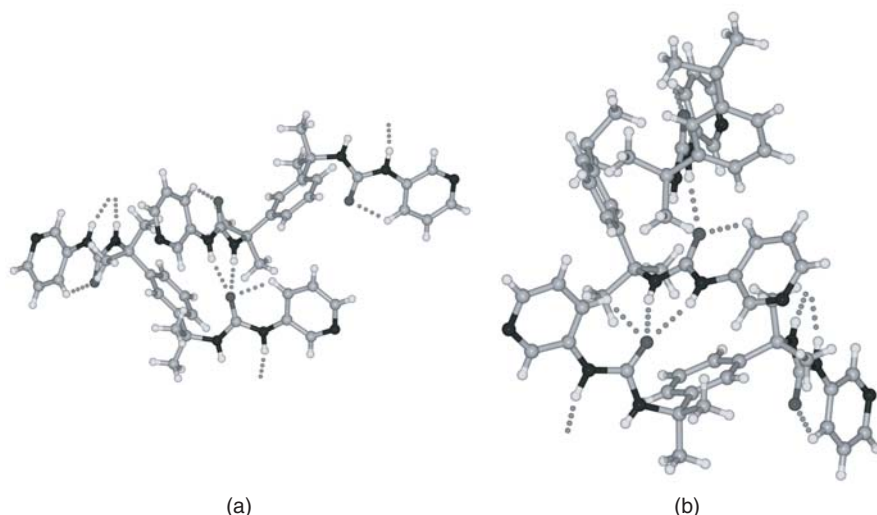
**Figure 5.17** Urea  $N-H \cdots O_{DMF}$  and intramolecular  $C-H \cdots O_{carbonyl}$  interactions in the DMF solvate of PNPu-H. The interrupted urea tape due to hydrogen bonding of urea  $N-H$  donors to solvent- $O$  is a common pattern in solvates and hydrates



**Figure 5.18** The intramolecular  $C-H \cdots O$  synthon is observed in the  $\alpha$ - and  $\gamma$ - polymorphs of *m*-dinitrophenyl urea (DNPU), *m*-dipyridyl urea (MDPU) and *p*-nitrophenyl urea (PNPU-X)

$154^\circ$ ) [29]. Urea molecules align along the short  $b$ -axis ( $4.67 \text{ \AA}$ ) and the aryl groups are twisted out of the urea plane ( $45.4, 49.2^\circ$ ) to accommodate the hydrogen bonded tape. The *p*-ethynyl derivative is isomorphous and isostructural to the iodo structure, consistent with iodo-ethynyl exchange in crystal structures [30].

The nature of the iodo  $\cdots$  nitro interaction has been discussed in recent papers [29]. Such interactions, referred to as halogen bonds [31], involve the close approach of soft donor halogens (I, Br) with electronegative heteroatoms (O, N) in a Lewis acid-Lewis base type interaction with  $n \rightarrow \sigma^*$  charge-transfer character. The electrostatic origin of the iodo  $\cdots$  nitro interaction is analogous to classical hydrogen bonds. The positive iodine lobe in the electrostatic surface potential (ESP) map points towards the oxygen lone pairs in a bifurcated motif (Figure 5.21) [32]. However, the chloro  $\cdots$  nitro interaction is not as specific and strong. This explains why the PNPU-I urea tape structure is different from the chloro and bromo derivatives that have urea  $\cdots$  nitro hydrogen bonding (compare Figures 5.16 and 5.20). The ability of the nitro group in steering crystal structures towards the urea tape synthon was successfully reproduced in *p*-NMe (methyl  $C-H \cdots O$ ), *p*-iodophenyl ( $I \cdots O_2N$ ) and tolyl derivatives (phenyl  $C-H \cdots O$ ). These crystal structures are shown in Figure 5.22.



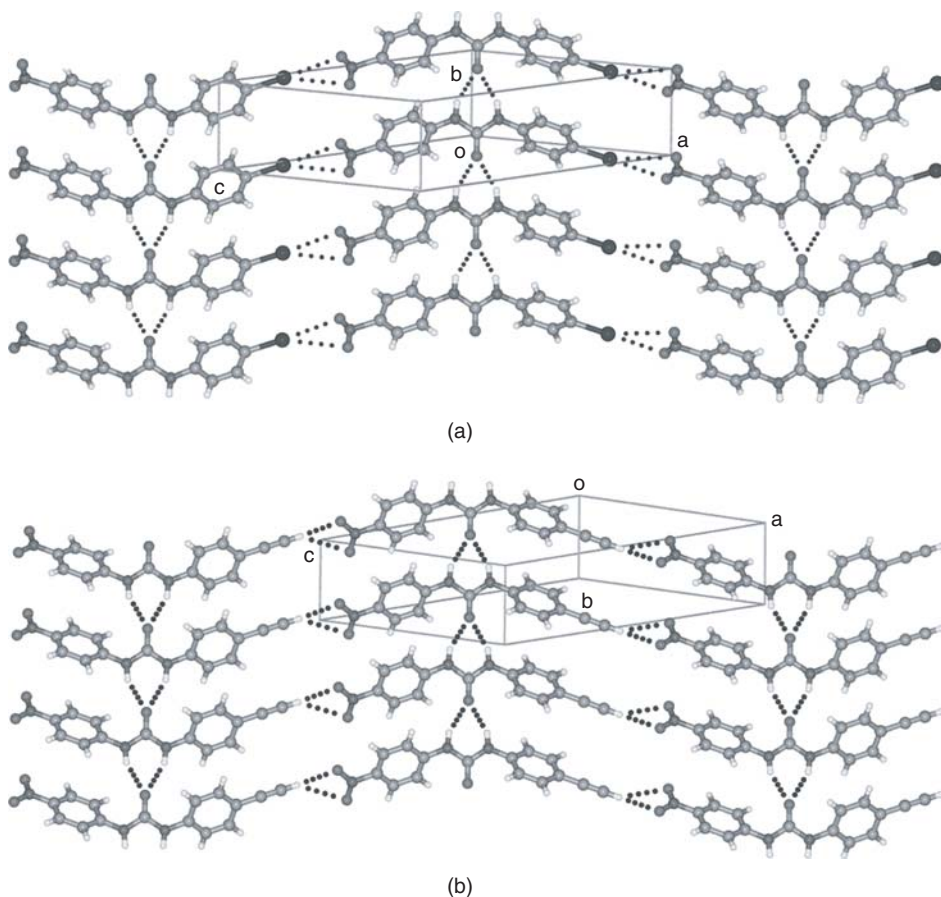
**Figure 5.19** *N,N'*-(Benzene-1,3-diyl)dipropene-2,2-diylbis(3-pyridin-3-ylurea) hydrate (a) and methanol clathrate (b). Note the intramolecular C–H···O interaction with urea–O

Structural interference by the nitro group was analysed in aromatic amides. A plethora of interactions were observed in *N*-iodophenyl nitrophenyl benzamide crystal structures [33] depending on the regioisomer (Figure 5.23): N–H···O<sub>amide</sub> and C–I···O<sub>amide</sub> in 2-I-Ph-NHCO-3-NO<sub>2</sub>-Ph and 3-I-2-NO<sub>2</sub>, N–H···O<sub>amide</sub> and C–H···π in 2-I-4-NO<sub>2</sub>, N–H···O<sub>amide</sub> and C–I···O<sub>nitro</sub> in 3-I-3-NO<sub>2</sub> and 4-I-4-NO<sub>2</sub>, and N–H···O<sub>nitro</sub> and C–I···O<sub>nitro</sub> in 3-I-4-NO<sub>2</sub>. The amide–amide and iodo–nitro interactions are observed only in the same iodophenyl nitrophenyl positional isomers but not in mixed isomer structures. This generalization is similar to PNPU in which 3-I-4-NO<sub>2</sub> belongs to the nonurea category I but 4-I-4-NO<sub>2</sub> belongs to the urea tape category III [27].

## 5.5 Molecular Conformation and Hydrogen Bonding

The molecular conformation plays a subtle yet definitive role during crystallization. A dramatic case of how different molecular conformations can lead to differences in hydrogen bonding and crystal packing was noted in polymorphs of bis(*p*-tolyl)ketone *p*-tosylhydrazone [34]. The crystal structure of form 1 has the N–H···O=S dimer synthon (N–N–S–C torsion angle  $\tau = 65.9^\circ$ ). Forms 2 and 3 adopt slightly different molecular conformers ( $\tau = 70.3^\circ$ ,  $62.4^\circ$ ) and are completely devoid of strong hydrogen bonds (Figure 5.24). That there is a one-to-one correspondence between molecular conformation and crystal packing was verified by crystal structure simulation starting from the experimental conformer.

We noted in over 100 diaryl urea structures (Table 5.1) extracted from the February 2009 update of the Cambridge Structural Database (version 5.30) [35] that the molecular conformation is planar in nonurea crystal structures, i.e. those having N–H···N<sub>pyridyl</sub>, N–H···O<sub>nitro</sub>, N–H···O<sub>solvent</sub> hydrogen bond as the primary interaction together with

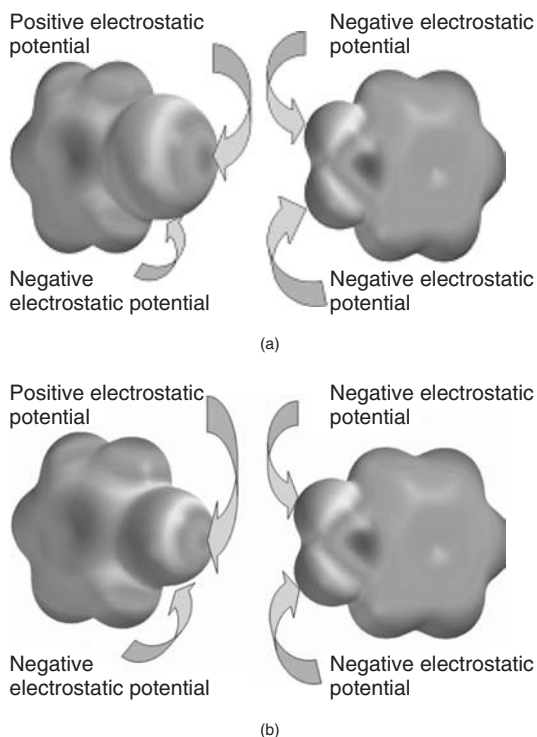


**Figure 5.20** (a)  $N-H \cdots O$  urea tape synthon along the  $b$ -axis and  $I \cdots O_2N$  halogen bonding in the crystal structure of the  $p$ -iodo compound PNPU-I. (b) The  $p$ -ethynyl crystal structure is isostructural, being mediated by urea tape synthons and  $C \equiv C-H \cdots O_2N$  interactions

intramolecular  $C-H \cdots O_{urea}$  interaction. Urea tape structures, however, have a twisted conformation (Figure 5.25) to accommodate the aryl rings in the 4.7 Å periodic linear tape motif. Similarly, amide  $\cdots$  amide hydrogen bonding is associated with an interplanar angle between the amide group and pyridine ring of  $>20^\circ$  whereas the angle can be small due to lack of steric repulsion in reverse amides which results in interference from  $N-H \cdots N_{pyridyl}$  hydrogen bonds [24].

Energy factors were computed to understand a correlation between molecular conformation and hydrogen bonding. The planar *trans*, *trans*-diphenyl urea conformer ( $\tau = 0, 180^\circ$ ) of the model compound is more stable than the twisted *trans*,*trans*-rotamer ( $\tau = 90^\circ$ ) by  $9.0 \text{ kcal mol}^{-1}$  (Figure 5.26) [36]. This energy difference may be partitioned into two components, intramolecular  $C-H \cdots O_{urea}$  interactions worth  $\sim 5 \text{ kcal mol}^{-1}$  and extended conjugation between the aromatic ring and the urea functional group contributing  $\sim 5 \text{ kcal mol}^{-1}$ . What is interesting is that the urea-O has very different acceptor strength in different conformers based on computed ESP charges. The value is

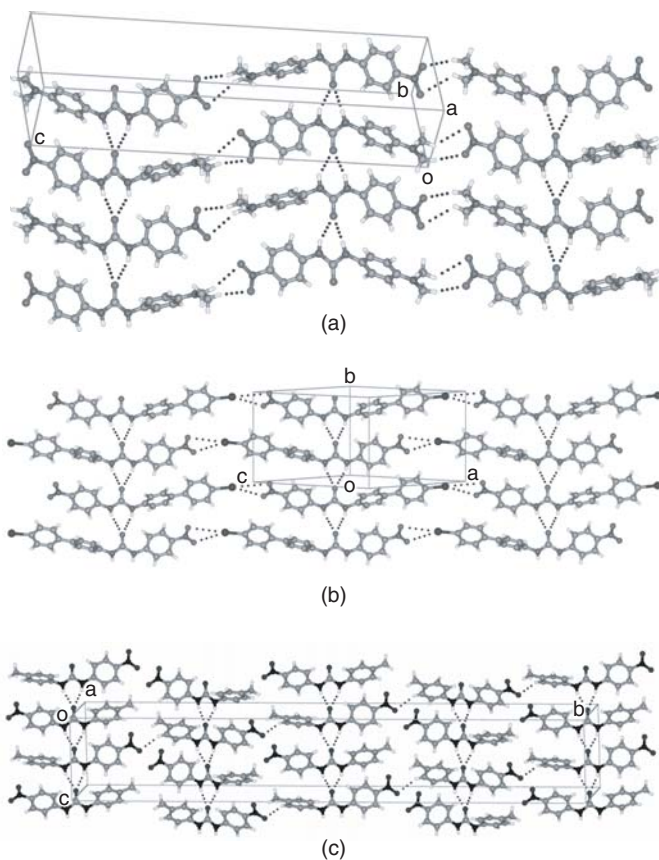




**Figure 5.21** (a) Electrostatic surface potential calculated in Spartan (AM1) showing the electropositive blue lobe pointing towards the electronegative oxygen atoms,  $I^{\delta+} \cdots O^{\delta-}$ . (b) The smaller surface of the chloride atom (see blue dot) interacts to form a weaker  $Cl \cdots O_2N$  contact. The depth of red and blue colour is calibrated to about the same negative and positive charge in (a) and (b). (see colour plate section)

$-45.7 \text{ kcal mol}^{-1}$  in the metastable twisted conformation ( $\tau \approx 45^\circ$ ) indicating a more electronegative urea-O compared with the less electronegative O ( $-38.6 \text{ kcal mol}^{-1}$ ) in the stable planar conformation ( $\tau = 0^\circ$ ). The urea  $C=O$  is a better hydrogen bond acceptor in the twisted conformer compared with the planar one. The change in hydrogen bond acceptor strength as a function of conformation (Figure 5.26) explains why the twisted molecule makes  $N-H \cdots O_{\text{urea}}$  and the planar conformer  $N-H \cdots N_{\text{pyridyl}}$  or  $N-H \cdots O_{\text{nitro}}$  H bonds in MDPU and PNPU structures. All this suggests that the nature of the strongest and the next best acceptor in a conformationally flexible molecule may change from one crystal structure to another, especially for closely ranked functional groups. Energies of hydrogen bond synthons in nitro-phenyl urea structures showed that urea  $\cdots$  urea is marginally stronger than urea  $\cdots$  DMF followed by urea  $\cdots$  nitro  $N-H \cdots O$  hydrogen bonds (Figure 5.27).

Quantitative measures of hydrogen bond acceptor strength are  $pK_{\text{HB}}$  values [37]. The urea carbonyl (tetramethylurea 2.44, tertraphenylurea 1.74) is stronger than nitro-O (nitrobenzene = 0.30, *N,N*-dimethyl-*p*-nitroaniline = 0.83). DMF and DMSO are equally strong hydrogen bond acceptors ( $pK_{\text{HB}}$  2.10, 2.58). The  $pK_{\text{HB}}$  scale ( $1.364 pK_{\text{HB}} = -\Delta G_{\text{HB}}$  in  $\text{kcal mol}^{-1}$ ) measures the free energy of the 1:1 hydrogen-bonded

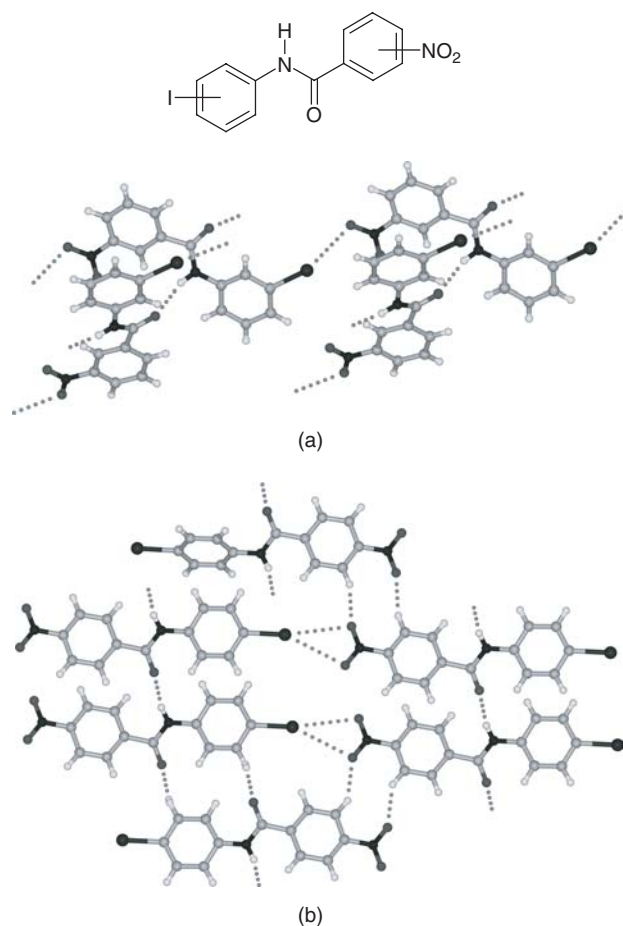


**Figure 5.22** Urea  $N-H \cdots O$  1-D tape and nitro-based interactions with (a)  $NMe_2$ , (b)  $I$  and (c)  $CH$  donors in the orthogonal direction

complex between 4-fluorophenol and approximately 1000 different acceptor groups. This thermodynamic scale naturally does not take into account changes in acceptor strength due to supramolecular effects such as hydrogen bonding, molecular conformation and steric effects. Electron density at the urea carbonyl-O is considerably reduced due to intramolecular  $C-H \cdots O$  interactions and hence, it becomes a weaker acceptor than solvent-O and even nitro-O in the planar conformation. This argument explains the interplay of molecular conformation and hydrogen bonding in diaryl urea with acceptor groups not only for the examples discussed herein but also for several structures reported in the literature (Table 5.1). Thus, twisted diaryl urea conformers ( $\tau = 25-45^\circ$ ) build up the urea  $N-H \cdots O$  tape whereas planar conformers ( $\tau = 2-10^\circ$ ) are present in nonurea structures.

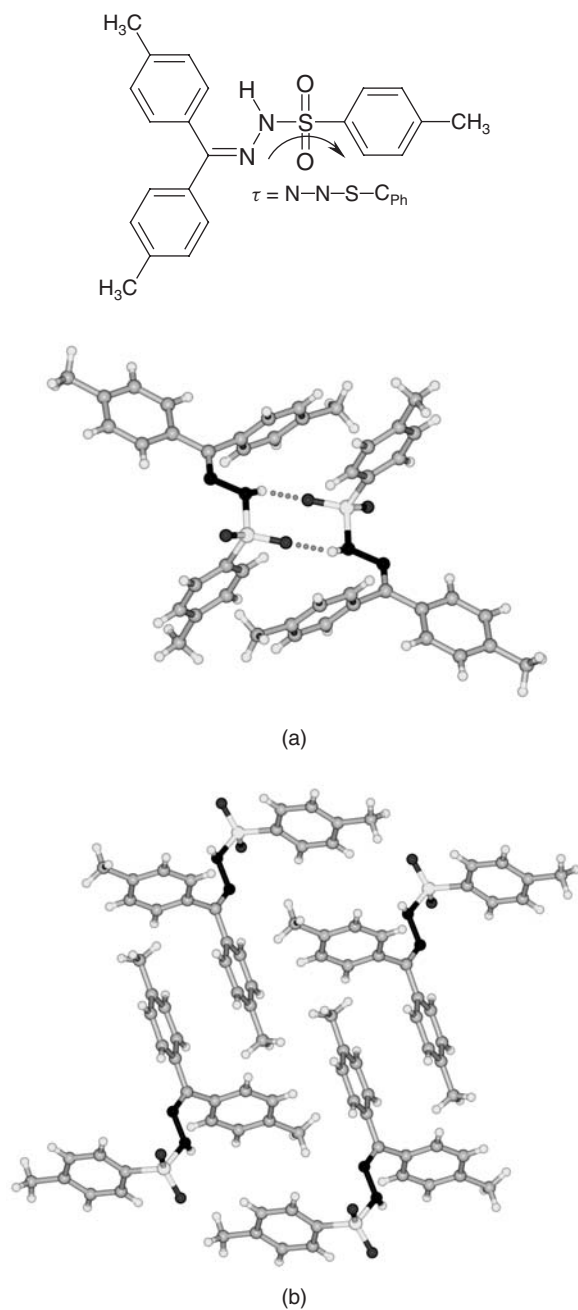
The most interesting interplay in flexible diaryl urea structures is that of molecular conformer and crystal lattice energy. The planar molecule resides in the stable conformation (lower  $E_{conf}$ ) but it does not make the strongest hydrogen bonds (higher  $E_{synthon}$ ) whereas the metastable, twisted conformer (higher  $E_{conf}$ ) has stronger supramolecular interactions (lower  $E_{synthon}$ ). This intramolecular and intermolecular energy balance was





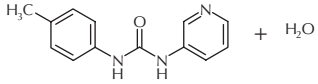
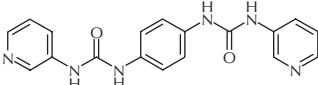
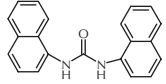
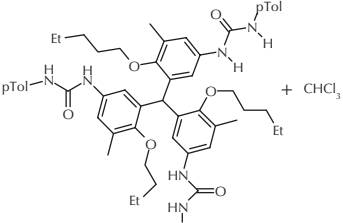
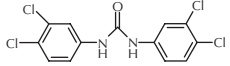
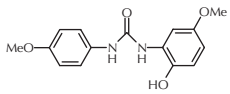
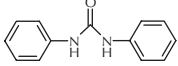
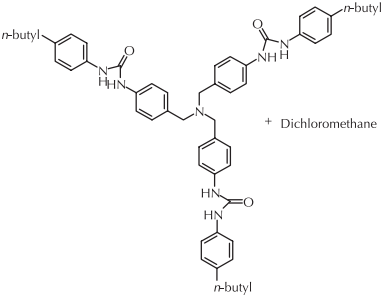
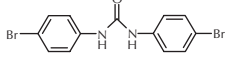
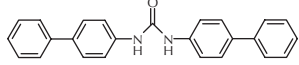
**Figure 5.23** Amide  $N-H \cdots O$  and iodo  $\cdots$  nitro interactions in 3-I-3-NO<sub>2</sub> (a) and 4-I-4-NO<sub>2</sub> (b) isomers of N-iodophenyl nitrobenzamide. Other isomers show cross-over of amide and nitro based interactions

analysed in polymorphs of DNPU and several other conformational polymorphs sets [38]. The  $\alpha$  and  $\gamma$  forms of DNPU have  $N-H \cdots O_{\text{nitro}}$  synthons whereas the  $\beta$  form has the urea tape. The conformer strain ( $E_{\text{conf}}$ ) was computed in Gaussian 03 [DFT, B3LYP/6-31G (d,p)] [39] and lattice energy ( $U_{\text{latt}}$ ) in Cerius<sup>2</sup> (COMPASS force field) [40] to give  $U_{\text{latt}} = -50.57 \text{ kcal mol}^{-1}$  and  $E_{\text{conf}} = 0.00$  (relative energy) for the  $\alpha$  form,  $-53.20$  and  $2.66 \text{ kcal mol}^{-1}$  for the  $\beta$  form and  $-51.04$  and  $1.34 \text{ kcal mol}^{-1}$  for the  $\gamma$  form. Even though lattice energies of the  $\alpha$  and  $\beta$  forms differ by  $2.63 \text{ kcal mol}^{-1}$ , the total energy ( $\Delta E_{\text{total}} = \Delta E_{\text{conf}} + \Delta U_{\text{latt}}$ ) between these polymorphs differs by a mere  $0.03 \text{ kcal mol}^{-1}$ , as shown by the intra- and intermolecular energy diagram of Figure 5.28. A strained conformer in the stable crystal lattice, and conversely the presence of the stable conformer in a not so strong crystal environment, i.e. intra- and intermolecular energy compensation, explains the concomitant occurrence of polymorphs due to marginal  $\Delta E_{\text{total}}$  difference.

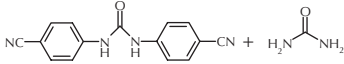
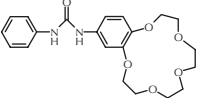
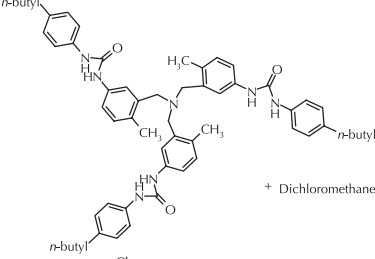
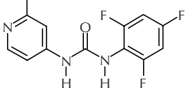
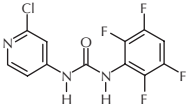
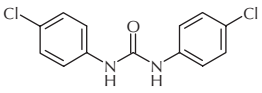
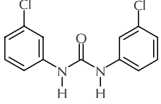
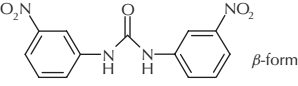
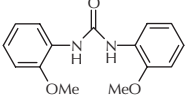
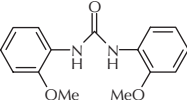
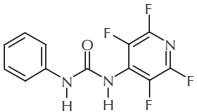


**Figure 5.24** (a) Sulfonamide  $\text{N-H}\cdots\text{O}$  dimer synthon in the crystal structure of bis(*p*-tolyl)ketone *p*-tosylhydrazone form 1 ( $\text{C2}/c$ ). (b) The sulfonamide NH donor does not participate in conventional hydrogen bonding or in an  $\text{N-H}\cdots\pi$  interaction in form 3 ( $\text{P2}_1/c$ ). The crystal structure of form 2 ( $\text{C2}/c$ ) is similar to polymorph 3. The conformation about the sulfonamide moiety ( $\tau$ ) is different in forms 1 and 3. Hirshfeld surface of polymorph 1 is shown on the cover page

**Table 5.1** CSD refcodes and structural summary of *N,N'*-diarylureas extracted from the Cambridge Structural Database (version 5.30, February 2009 update, ConQuest 1.11). In case of duplicate refcodes, the structure with lowest R-factor is selected. CSD refcodes are listed in alphabetical order for urea tape and nonurea tape categories

Entry	Structural formula	Refcode	Torsion angle <sup>a</sup> (C–C–N–C) (deg)	Hydrogen-bond synthon
1		AGOKUD	54.6, 56.5	urea tape
2		AGOMAL	43.4, 37.7 31.5, 22.5	urea tape
3		AMAFEZ	42.5, 49.6	urea tape
4		CEVWOQ	49.1, 33.3 55.0, 28.0 56.0, 35.0	urea tape
5		DCPHUR	20.2, 31.3 30.9, 46.6	urea tape
6		DEVVUV	59.0, 68.0	urea tape
7		DPUREA	39.9, 44.0	urea tape
8		EGUFER	23.9, 54.7, 33.5, 55.5, 18.1, 65.9	urea tape
9		ETOBiy	51.3, 51.3	urea tape
10		HAHCOJ	37.1, 57.1	urea tape

**Table 5.1** (continued)

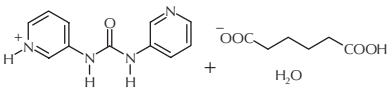
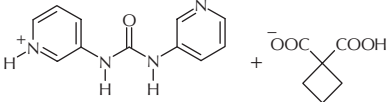
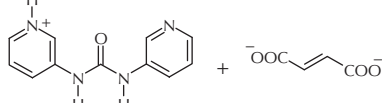
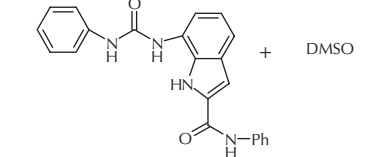
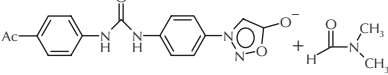
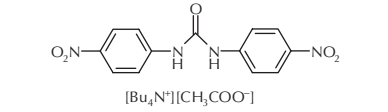
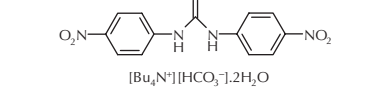
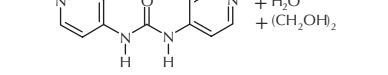
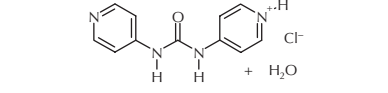
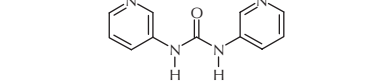
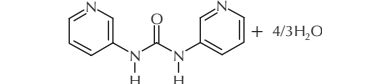
Entry	Structural formula	Refcode	Torsion angle <sup>a</sup> (C–C–N–C) (deg)	Hydrogen-bond synthon
11		HEZTEL	39.3, 39.3	urea tape
12		ILICAH	40.3, 62.6	urea tape
13		IXEMON	13.6, 60.0	urea tape
14		JEDSAN	39.4, 90.0	urea tape
15		JEDTIW	24.0, 61.9	urea tape
16		MIXZAV	52.6, 52.6	urea tape
17		QQQAGA03	47.2, 47.2	urea tape
18		SILTOW01	32.6, 32.6	urea tape
19		SILTUC	44.6, 44.6	urea tape
20		SILTUC01	24.4, 24.4	urea tape
21		TAPRAD	29.6, 51.7	urea tape

(continued overleaf)

**Table 5.1** (continued)

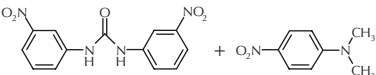
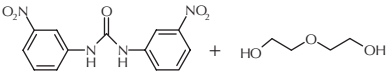
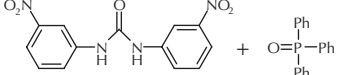
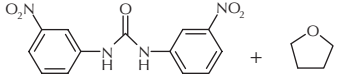
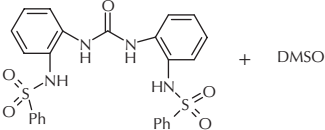
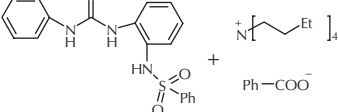
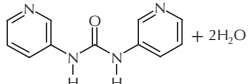
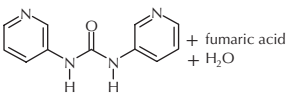
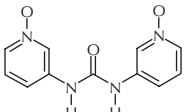
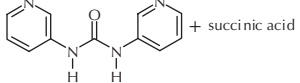
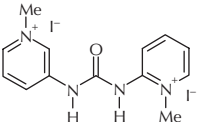
Entry	Structural formula	Refcode	Torsion angle <sup>a</sup> (C–C–N–C) (deg)	Hydrogen- bond synthon
22		TIVPUK	27.3, 51.4 87.9, 20.6	urea tape
23		TIVQAR	24.7, 63.0	urea tape
24		TIVQEV	38.7, 59.4	urea tape
25		TUDLUZ	50.0, 50.1	urea tape
26		VIMYEV	40.3, 62.6	urea tape
27		AGOKOX	2.9, 3.9	nonurea tape
28		AGOLAK	5.0, 6.9	nonurea tape
29		AGOMEP	14.0, 0.5 30.3, 18.6	nonurea tape
30		DIGBUR	2.1, 7.2 0.3, 9.1 0.2, 5.4 1.5, 8.2	nonurea tape
31		DIGCAY	10.3, 8.3	nonurea tape
32		DIGCEC	13.6, 15.0	nonurea tape

**Table 5.1** (continued)

Entry	Structural formula	Refcode	Torsion angle <sup>a</sup> (C–C–N–C) (deg)	Hydrogen-bond synthon
33		DIGCIG	1.5, 10.9 11.5, 7.0	nonurea tape
34		DIGCOM	4.8, 12.6	nonurea tape
35		DIGCUS	2.4, 7.1	nonurea tape
36		DISVAD	53.7, 11.3	nonurea tape
37		EWOZUL	0.5, 7.3	nonurea tape
38		FEMQEU	6.6, 7.3	nonurea tape
39		FEMQIY	1.5, 4.1	nonurea tape
40		GASMUI	4.2, 10.3	nonurea tape
41		GASNAQ	3.0, 9.99 1.3, 3.5	nonurea tape
42		GASNEU01	4.5, 4.5	nonurea tape
43		GASNIY01	9.0, 9.9, 10.8	nonurea tape

(continued overleaf)

**Table 5.1** (continued)

Entry	Structural formula	Refcode	Torsion angle <sup>a</sup> (C–C–N–C) (deg)	Hydrogen-bond synthon
44		GIMRUP10	2.1, 2.1	nonurea tape
45		GIMSAW10	5.2, 12.2	nonurea tape
46		GIMSEA10	2.1, 22.9	nonurea tape
47		GIRROJ10	3.0, 6.3	nonurea tape
48		HITFOG	13.7, 56.3	nonurea tape
49		HITGAT	36.4, 9.1	nonurea tape
50		JEDKIN	1.7, 5.1	nonurea tape
51		JEDKOT	0.7, 4.5	nonurea tape
52		JEDKUZ	7.7, 16.1	nonurea tape
53		JEDLAG	0.9, 0.9	nonurea tape
54		JEDLEK	2.1, 6.0	nonurea tape

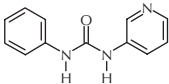
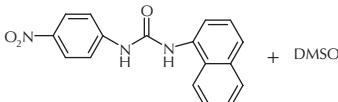
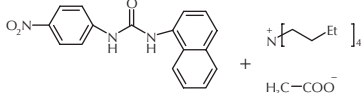
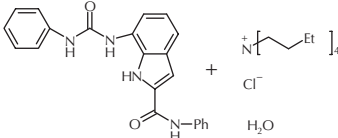
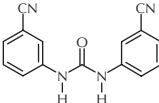
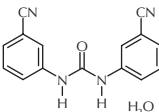
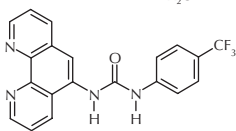
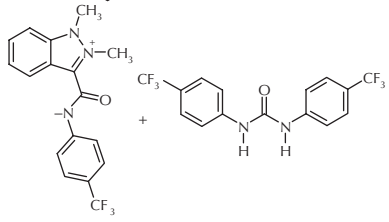
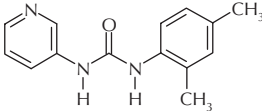
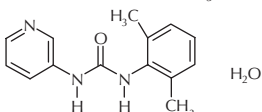
**Table 5.1** (continued)

Entry	Structural formula	Refcode	Torsion angle <sup>a</sup> (C–C–N–C) (deg)	Hydrogen-bond synthon
55		JEDRUG	16.3, 65.2	nonurea tape
56		JEDSER	3.3, 19.0	nonurea tape
57		JEDSIV	4.2, 6.6	nonurea tape
58		JEDSOB	5.5, 9.2	nonurea tape
59		JEDSUH	6.6, 11.5	nonurea tape
60		JEDTAO	6.8, 27.6	nonurea tape
61		JEDTES	16.5, 29.2	nonurea tape
62		JEHFIM	22.2, 18.7 55.5, 37.1 26.8, 27.7 37.7, 60.7	nonurea tape
63		MEPWUZ01	7.5, 17.3	nonurea tape
64		MEPXAG	1.9, 35.8	nonurea tape

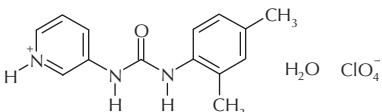
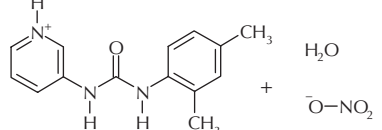
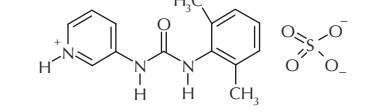
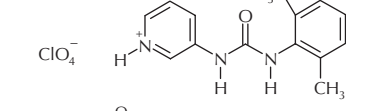
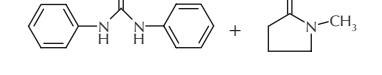
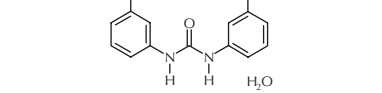
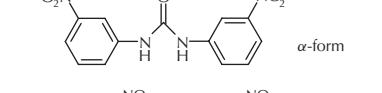
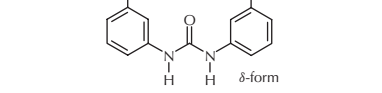
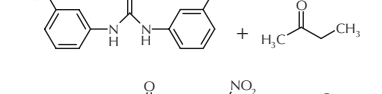
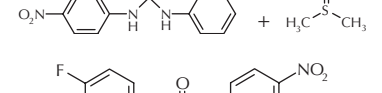
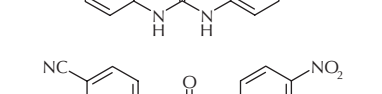
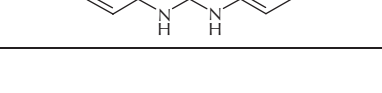
(continued overleaf)



**Table 5.1** (continued)

Entry	Structural formula	Refcode	Torsion angle <sup>a</sup> (C–C–N–C) (deg)	Hydrogen-bond synthon
65		MEPXK01	1.9, 15.5	nonurea tape
66		MIPLUT	22.9, 7.36	nonurea tape
67		MIPMAA	63.7, 3.88	nonurea tape
68		NIQFID	2.2, 5.7 4.0, 2.7	nonurea tape
69		PEFGIR	27.3, 30.8	nonurea tape
70		PEFGOX	5.6, 21.0	nonurea tape
71		PEZDEE	0.2, 34.5	nonurea tape
72		PEZJUA	18.61, 14.47	nonurea tape
73		PIHLAU	4.0, 1.3	nonurea tape
74		PIHLEY	3.28, 108.6	nonurea tape

**Table 5.1** (continued)

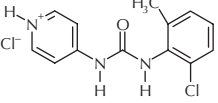
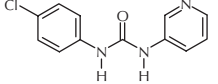
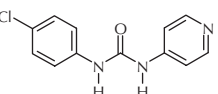
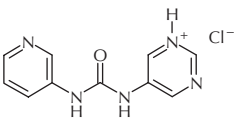
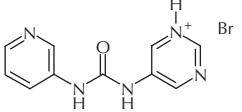
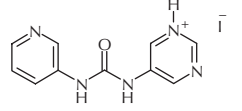
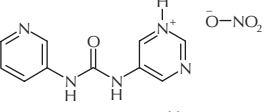
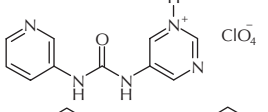
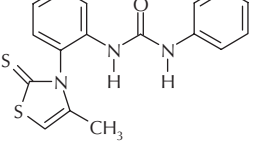
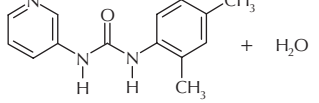
Entry	Structural formula	Refcode	Torsion angle <sup>a</sup> (C–C–N–C) (deg)	Hydrogen-bond synthon
75		PIHLIC	45.5, 7.2	nonurea tape
76		PIHLOI	20.4, 8.1	nonurea tape
77		PIHLUO	115.1, 5.4	nonurea tape
78		PIHMAV	4.1, 63.7	nonurea tape
79		QAZXUK	4.2, 10.1	nonurea tape
80		SAYQEP	22.1, 22.1	nonurea tape
81		SILTOW	1.6, 24.8	nonurea tape
82		SILTOW11	25.0, 31.7	nonurea tape
83		SILVAK	3.3, 4.6	nonurea tape
84		SILVOY	0.5, 16.7	nonurea tape
85		TIVMAN	5.1, 4.6	nonurea tape
86		TIVMER	6.5, 5.8	nonurea tape

(continued overleaf)

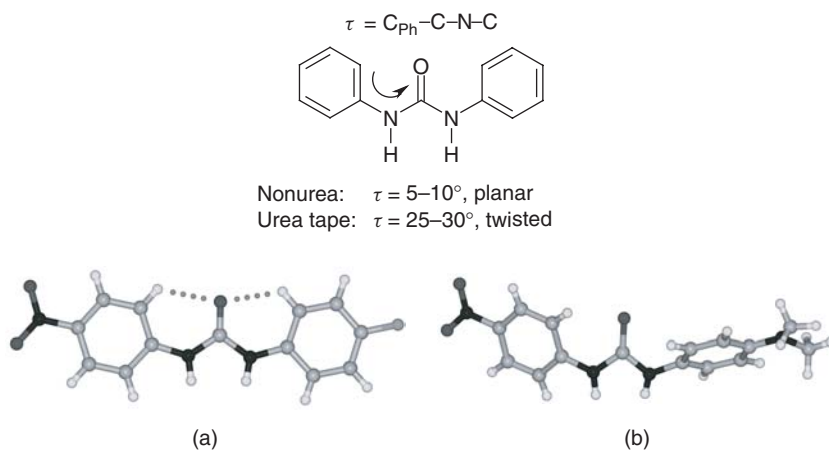
**Table 5.1** (continued)

Entry	Structural formula	Refcode	Torsion angle <sup>a</sup> (C–C–N–C) (deg)	Hydrogen- bond synthon
87		TIVMIV	12.9, 3.7	nonurea tape
88		TIVMOB	14.7, 3.8	nonurea tape
89		TIVMUH	4.2, 13.2 15.8, 13.4	nonurea tape
90		TIVNAO	12.7, 9.3	nonurea tape
91		TIVNES	6.7, 2.7	nonurea tape
92		TIVNIW	8.1, 4.2	nonurea tape
93		TIVNOC	14.5, 13.9 5.7, 4.6	nonurea tape
94		TIVNUI	0.1, 11.0	nonurea tape
95		TIVPEU	2.4, 5.4	nonurea tape
96		TIVPIY	9.3, 15.2	nonurea tape
97		TIVPOE	9.2, 0.7	nonurea tape

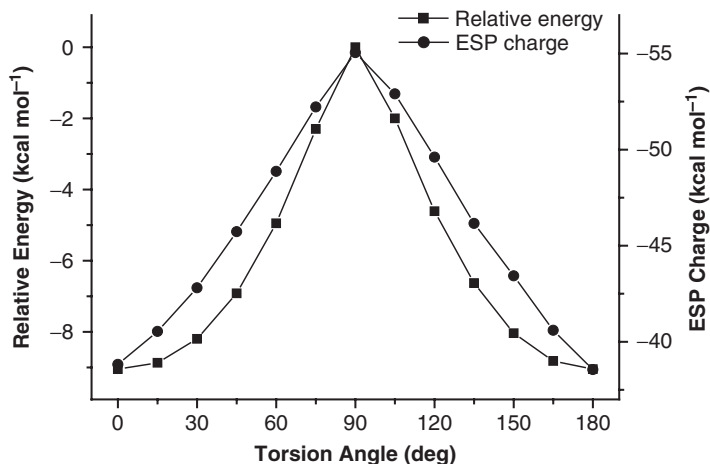
**Table 5.1** (continued)

Entry	Structural formula	Refcode	Torsion angle <sup>a</sup> (C–C–N–C) (deg)	Hydrogen-bond synthon
98		UCOHUQ	4.4, 74.3 7.7, 61.0	nonurea tape
99		WICBEP	6.6, 7.3	nonurea tape
100		WICBIT	1.6, 10.9	nonurea tape
101		WOCFAW	5.4, 8.0	nonurea tape
102		WOCFEA	6.8, 5.2	nonurea tape
103		WOCFIE	7.3, 0.4	nonurea tape
104		WOCFOK	5.7, 3.3	nonurea tape
105		WOCFUQ	4.1, 8.4	nonurea tape
106		XEWXON	4.1, 39.2	nonurea tape
107		XEYYOQ	12.2, 26.4	nonurea tape

<sup>a</sup>See Figure 5.25 for torsion angle.



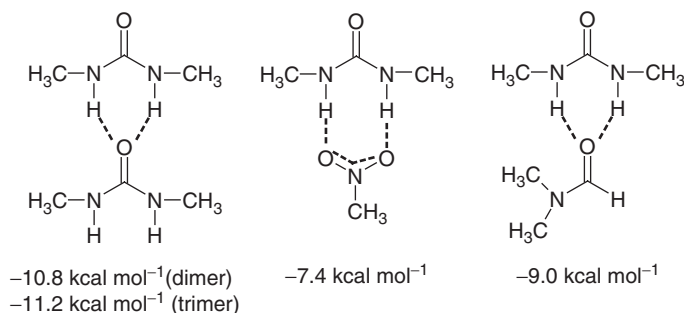
**Figure 5.25** Planar and twisted conformations of diaryl ureas occur in nonurea (a) and urea tape (b) structures, respectively



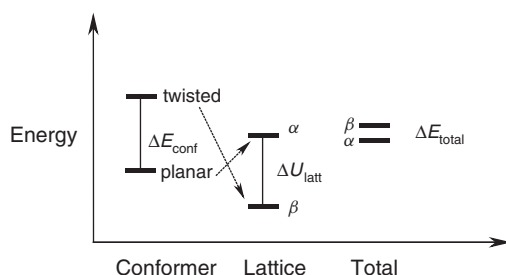
**Figure 5.26** Spartan computations (DFT-B3LYP/6-31G\*) on *trans,trans*-diphenyl urea as a function of the aryl C–C–N–C=O torsion angle. Relative energy is plotted such that the highest energy conformer is fixed at 0. The planar molecular conformation is the most stable rotamer. The ESP charge on the urea-O atom is higher (better hydrogen bond acceptor) in the twisted conformation and lower in the planar state. Rotation of the *m*-NO<sub>2</sub> group with respect to the phenyl ring imposes an insignificant energy penalty of <0.25 kcal mol<sup>-1</sup>

## 5.6 Supramolecular HSAB Interactions

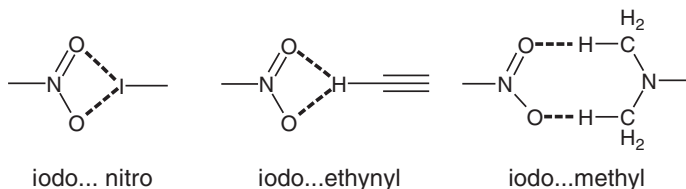
A crystal engineering strategy to restore the N–H···O tape in diaryl ureas is now discussed. First, there are clear differences between the diaryl ureas highlighted in this chapter (N–aromatic) [23, 25–28] and the dibenzyl ureas (N–CH<sub>2</sub>–aromatic) [14] reported over a decade ago. The same combination of functional groups, namely urea,



**Figure 5.27** Synthon energy calculations in Spartan (RHF/6-31G\*\*) on the urea...urea homodimer and urea...nitro and urea...DMF heterodimers in model compounds

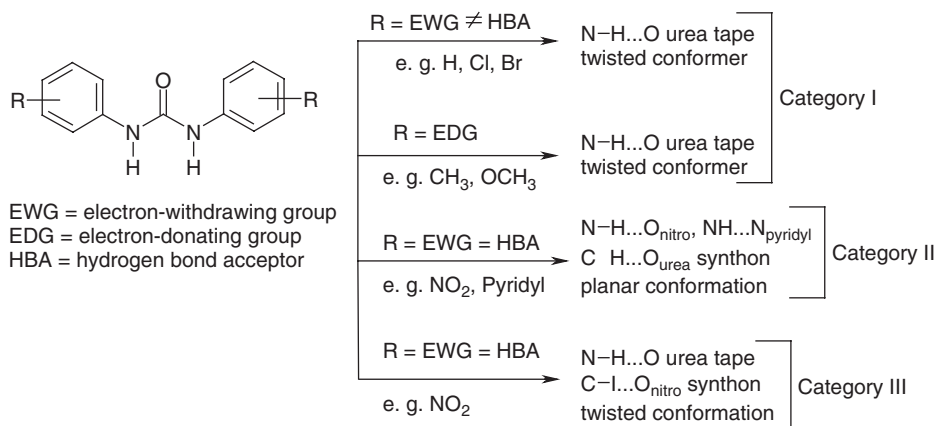


**Figure 5.28** Intramolecular ( $E_{\text{conf}}$ ) and intermolecular ( $U_{\text{latt}}$ ) energy compensation in  $\alpha$ - and  $\beta$ -conformational polymorphs of DNPU.  $E_{\text{total}} = E_{\text{conf}} + U_{\text{latt}}$



**Figure 5.29** Soft nitro-based synthons that restore the strong urea tape network

pyridine and acid, gave good structural control and modular assembly in one case [14] but structural diversity in the more recent systems. This dramatic difference in the two families of crystal structures is ascribed to the interfering nature of the weak  $\text{C-H} \cdots \text{O}_{\text{urea}}$  synthon that is possible in diaryl ureas but not in dialkylaryl ureas because phenyl-CH is a stronger donor than alkyl-CH. The weakening of urea-O due to intramolecular  $\text{C-H} \cdots \text{O}$  changes acceptor rankings such that nitro-O or pyridyl-N compete favourably for the strong NH donors, resulting in the dominant nonurea structures due to competing hydrogen bond acceptors. The presence of donor groups that can engage in selective interactions with the nitro-O, but do not interact with urea-O, is a potential solution to the divergent structural behaviour. This occurs with soft CH and iodo donors, as shown



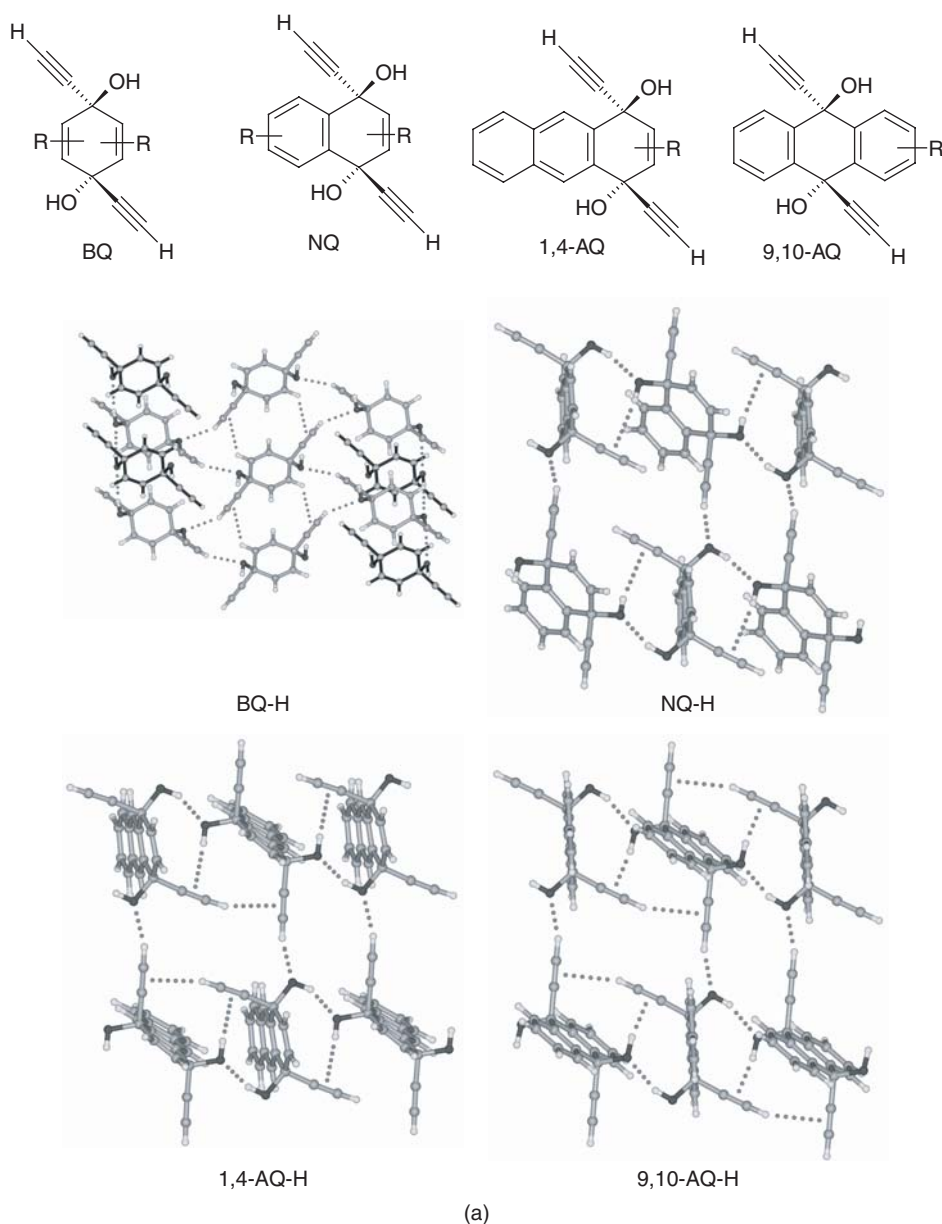
**Figure 5.30** Correlation between molecular functional groups and hydrogen bonding in diaryl urea structures. The nonurea category II structures can be steered towards the urea tape category III via iodo...nitro synthons in PNPU's

for PNPU-I, -NMe<sub>2</sub> and -Me compounds (Figure 5.22). Thus, the strong hydrogen bonds of the urea tape (hard) and weak interactions with the NO<sub>2</sub> group (soft) are mutually exclusive and operate in orthogonal domains in these crystal structures. The ability of nitro...iodo, nitro...ethynyl and nitro...*N,N*-dimethylamino synthons (Figure 5.29) in restoring the urea tape illustrates a supramolecular HSAB principle in crystal engineering [27]. Apart from functional groups, regiochemistry too plays a role, e.g. *p*-iodo PNPU is steered toward the urea tape network but the *m*-iodo structure has nonurea N-H...O<sub>nitro</sub> hydrogen bonds. Such behaviour is not new in multi-functional crystal structures, e.g. in isomeric *N*-iodophenyl nitrobenzamides [33].

A take home message is that the weak C-H...O<sub>urea</sub> synthon disturbs the strong N-H...O<sub>urea</sub> tape and, once again, it is the weak C-I...O<sub>nitro</sub> or C-H...O<sub>nitro</sub> interaction that restores the tape synthon. Weak interactions not only cause deviation from the expected strong hydrogen bond synthon but also provide a solution to enable the urea tape. There are three kinds of situations possible in diaryl ureas (Figure 5.30): (1) The strong H bonds dominate and the weak interactions play a bystander role. The expected urea tape synthon is easily predicted. (2) The presence of strong hydrogen bond acceptors promotes non-urea structures, e.g. in pyridyl and nitrophenyl ureas. (3) A third category is when weak intermolecular interactions are exploited to control the strong hydrogen bond network via HSAB partitioning. This is the new PNPU subset of crystal structures. These generalizations may be extended to other types of hard-soft combinations of functional groups.

## 5.7 *gem*-Alkynols

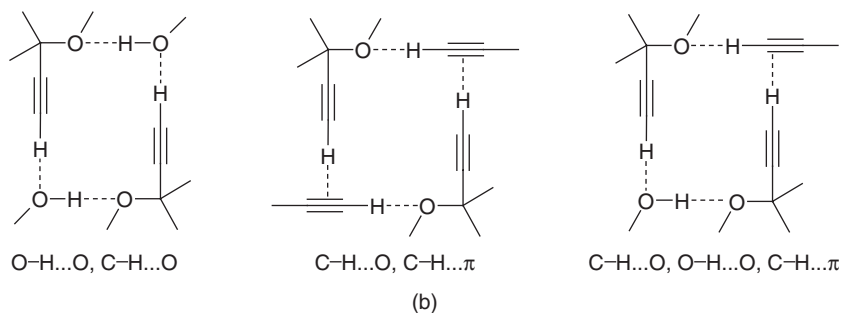
Structural diversity in *gem*-alkynol compounds, which have hydroxyl and ethynyl groups juxtaposed on the same carbon atom, is now discussed. Both functional groups can behave as hydrogen bond donors (O-H and C $\equiv$ C-H) and acceptors (O-H and



(a)

**Figure 5.31** (a) Hydrogen bonding in gem-alkynols: BQ-H, NQ-H, 1,4-AQ-H and 9,10-AQ-H. The structural synthons are similar in NQ-H and 1,4-AQ-H. (b) Synthon identification in the family of crystal structures ( $R = H, CH_3, Cl$ ). Note that these synthons are larger in size than the ones discussed earlier in this chapter



**Figure 5.31** (continued)

$\text{C}\equiv\text{C}-\text{H}$ ). Normally one would expect  $\text{O}-\text{H}\cdots\text{O}$  chains and  $\text{C}-\text{H}\cdots\pi$  interactions based on hydrogen bond hierarchy. However, in reality all four interactions  $\text{O}-\text{H}\cdots\text{O}$ ,  $\text{O}-\text{H}\cdots\pi_{\text{C}\equiv\text{C}}$ ,  $\text{C}\equiv\text{C}-\text{H}\cdots\text{O}$ , and  $\text{C}\equiv\text{C}-\text{H}\cdots\pi_{\text{C}\equiv\text{C}}$  were observed [22]. Changes in hydrogen bonding by the successive introduction of phenyl rings was studied in the compound series benzoquinone (BQ), naphthoquinone (NQ), 1,4-anthraquinone (1,4-AQ), and 9,10-anthraquinone (9,10-AQ), and their derivatives. For example, hydrogen bonding in the parent set ( $\text{R} = \text{H}$ ) shown in Figure 5.31 highlights the structural variation when a phenyl ring was added to the *gem*-alkynol group in benzoquinone. The only repetitive structure is the similarity of NQ to 1,4-AQ. Recurring synthons were identified in this otherwise structurally divergent family by the analysis of several compounds ( $\text{R} = \text{CH}_3, \text{Cl}$ ) [41]. A possible reason why OH and  $\text{C}\equiv\text{C}-\text{H}$  based interactions cross-over could be the proximal placement of these active functional groups. The trans series gave better diffracting single crystals than the cis isomers.

## 5.8 Conclusions

Acyclic diaryl ureas were classified into three major families (Figure 5.30). Surprisingly, only 26 out of 107 crystal structures analysed (Table 5.1) contain the urea tape synthon whereas three-quarters are nonurea type motifs. Those with electron-donating and electron-withdrawing groups adopt the classic urea tape prototype (category I). The presence of hydrogen bond acceptor groups gives the nonurea structures (category II). The structure-directing role of soft iodo $\cdots$ nitro and iodo $\cdots$ ethynyl synthons restores the urea tape in PNPu structures (category III). The last category underscores the significance of crystal engineering in multi-functional ureas. Intramolecular and intermolecular energy balance in flexible diaryl ureas provide a dynamic system to study the interplay of molecular conformation and hydrogen bonding in crystallization. Supramolecular HSAB recognition and interaction pairing should find application not only in the design of crystalline materials but also to help understand and control polymorphism of drug molecules. A correlation of molecular conformation with hydrogen bonding is very useful in crystal structure prediction.

## Acknowledgements

I thank the Department of Science and Technology, Council of Scientific and Industrial Research and University Grants Commission for continued support of my research. L. Sreenivas Reddy and N. Jagadeesh Babu are thanked for preparing figures and extracting CSD refcodes.

## References

1. (a) M. B. Smith and J. March, *March's Advanced Organic Chemistry: Reactions, Mechanisms, and Structure*, 6th Edn, John Wiley & Sons, Ltd, Hoboken, 2007; (b) F. A. Carey and R. J. Sundberg, *Advanced Organic Chemistry*, 4th Edn, Part A and B, Kluwer Academic, New York, 2002.
2. (a) J.-M. Lehn, *Supramolecular Chemistry*, VCH, Weinheim, 1995; (b) J. W. Steed and J. L. Atwood, *Supramolecular Chemistry*, John Wiley & Sons, Ltd, Chichester, 2000. (c) C. A. Schalley (Ed.), *Analytical Methods in Supramolecular Chemistry*, Wiley-VCH, Weinheim, 2007.
3. G. R. Desiraju, *Angew. Chem., Int. Ed. Engl.*, **34**, 2311–2327 (1995).
4. C. Giacovazzo (Ed.), *Fundamentals of Crystallography*, International Union of Crystallography, Oxford University Press, Oxford, 2000.
5. (a) G. R. Desiraju, *Crystal Engineering: The Design of Organic Solids*, Elsevier, Amsterdam, 1989; (b) G. R. Desiraju, *J. Mol. Struct.*, **656**, 5–15 (2003); (c) D. Braga, L. Brammer and N. R. Champness, *CrystEngComm*, **7**, 1–19 (2005); (d) M. J. Zaworotko, *Cryst. Growth Des.*, **7**, 4–9 (2007); (e) A. Nangia, *Cryst. Growth Des.*, **8**, 1079–1081 (2008).
6. E. R. T. Tiekink and J. J. Vittal (Eds), *Frontiers in Crystal Engineering*, John Wiley & Sons, Ltd, Chichester, 2006.
7. J. D. Dunitz and A. Gavezzotti, *Angew. Chem., Int. Ed.*, **44**, 1766–1787 (2005).
8. B. Sarma, S. Roy and A. Nangia, *Chem. Commun.*, 4918–4920 (2006).
9. G. A. Jeffrey, *An Introduction to Hydrogen Bonding*, Oxford University Press, New York, 1997.
10. G. R. Desiraju and T. Steiner, *The Weak Hydrogen Bond in Structural Chemistry and Biology*, Oxford University Press, Oxford, 1999.
11. J. Bernstein, *Polymorphism in Molecular Crystals*, Clarendon Press, Oxford, 2002.
12. (a) G. M. Day, W. D. S. Motherwell, H. Ammon, S. X. M. Boerrigter, R. G. Della Valle, E. Venuta, A. Dzyabchenko, J. D. Dunitz, B. Schweizer, B. P. van Eijck, P. Erk, J. C. Facelli, V. E. Bazterra, M. B. Ferraro, D. W. M. Hofmann, F. J. J. Leusen, C. Liang, C. C. Pantelides, P. G. Karamertzanis, S. L. Price, T. C. Lewis, H. Nowell, A. Torrisi, H. A. Scheraga, Y. A. Arnautova, M. U. Schmidt and P. Verwer, *Acta Crystallogr., Sect. B*, **61**, 511–527 (2005); (b) G. M. Day, T. G. Cooper, A. J. Cruz-Cabeza, K. E. Hejczyk, H. L. Ammon, S. X. M. Boerrigter, J. S. Tan, R. G. Della Valle, E. Venuti, J. Jose, S. R. Gadre, G. R. Desiraju, T. S. Thakur, B. P. van Eijck, J. C. Facelli, V. E. Bazterra, M. B. Ferraro, D. W. M. Hofmann, M. A. Neumann, F. J. J. Leusen, J. Kendrick, S. L. Price, A. J. Misquitta,

- P. G. Karamertzanis, G. W. A. Welch, H. A. Scheraga, Y. A. Arnautova, M. U. Schmidt, J. van de Streek, A. K. Wolf and B. Schweizer, *Acta Crystallogr., Sect. B*, **65**, 107–125 (2009).
13. (a) M. C. Etter, *J. Phys. Chem.*, **95**, 4601–4610 (1991); (b) M. C. Etter, *Acc. Chem. Res.*, **23**, 120–126 (1990).
14. (a) Y.-L. Chang, A.-N. West, F. W. Fowler and J. W. Lauher, *J. Am. Chem. Soc.*, **115**, 5991–6000 (1993); (b) J. J. Kane, R. F. Liao, J. W. Lauher and F. W. Fowler, *J. Am. Chem. Soc.*, **117**, 12 003–12 004 (1995).
15. (a) C. B. Aakeröy, A. M. Beatty and B. A. Helfrich, *Angew. Chem. Int. Ed.*, **40**, 3240–3242 (2001); (b) C. B. Aakeröy, A. M. Beatty and B. A. Helfrich, *J. Am. Chem. Soc.*, **124**, 14425–14432 (2002).
16. P. Vishweshwar, A. Nangia and V. M. Lynch, *CrystEngComm*, **5**, 164–168 (2003).
17. P. Vishweshwar, A. Nangia and V. M. Lynch, *Cryst. Growth. Des.*, **3**, 783–790 (2003).
18. R. Custelcean, *Chem. Commun.* 295–307 (2008).
19. S. George and A. Nangia, *Acta Crystallogr., Sect. E*, **57**, o719–o720 (2001).
20. C. B. Aakeröy, A. M. Beatty, B. A. Helfrich and M. Nieuwenhuyzen, *Cryst. Growth. Des.*, **3**, 159–165 (2003).
21. A. Nangia and G. R. Desiraju, *Top. Curr. Chem.*, **198**, 57–95 (1998).
22. N. N. L. Madhavi, C. Bilton, J. A. K. Howard, F. H. Allen, A. Nangia and G. R. Desiraju, *New J. Chem.*, **24**, 1–4 (2000).
23. L. S. Reddy, S. Basavoju, V. R. Vangala and A. Nangia, *Cryst. Growth Des.*, **6**, 161–173 (2006).
24. (a) M. Sarkar and K. Biradha, *Cryst. Growth Des.*, **6**, 202–208 (2006); (b) L. Rajput, S. Singha and K. Biradha, *Cryst. Growth Des.*, **7**, 2788–2795 (2007).
25. M. Rafilovich, J. Bernstein, R. K. Harris, D. C. Apperley, P. G. Karamertzanis and S. L. Price, *Cryst. Growth Des.* **5**, 2197–2209 (2005).
26. (a) M. C. Etter and T. W. Panunto, *J. Am. Chem. Soc.*, **110**, 5896–5897 (1988); (b) M. C. Etter, Z. Urbańczyk-Lipkowska, M. Zia-Ebrahimi and T. W. Panunto, *J. Am. Chem. Soc.*, **112**, 8415–8426 (1990).
27. L. S. Reddy, S. K. Chandran, S. George, N. J. Babu and A. Nangia, *Cryst. Growth Des.*, **7**, 2675–2690 (2007).
28. A. M. Todd, K. M. Anderson, P. Byrne, A. E. Goeta and J. W. Steed, *Cryst. Growth Des.*, **6**, 1750–1752 (2006).
29. (a) S. George, A. Nangia, C.-K. Lam, T. C. W. Mak and J.-F. Nicoud, *Chem. Commun.*, 1202–1203 (2004); (b) R. Thaimattam, C. V. K. Sharma, A. Clearfield and G. R. Desiraju, *Cryst. Growth Des.*, **1**, 103–106 (2001).
30. B. K. Saha and A. Nangia, *Cryst. Growth Des.*, **7**, 393–401 (2007).
31. (a) P. Metrangolo and G. Resnati, *Chem. Eur. J.*, **7**, 2511–2519 (2001); (b) P. Metrangolo, H. Neukirch, T. Pilati and G. Resnati, *Acc. Chem. Res.*, **38**, 386–395 (2005); (c) R. G. Gonnade, M. S. Shashidhar and M. M. Bhadbhade, *J. Ind. Inst. Sci.*, **87**, 149–165 (2007).
32. B. K. Saha, A. Nangia and M. Jaskólski, *CrystEngComm*, **7**, 355–358 (2005).
33. J. L. Wardell, J. N. Low, J. M. S. Skakle and C. Glidewell, *Acta Crystallogr., Sect. B*, **62**, 931–943 (2006).
34. S. Roy and A. Nangia, *Cryst. Growth Des.*, **7**, 2047–2058 (2007).

35. (a) F. H. Allen, *Acta Crystallogr., Sect. B*, **58**, 380–388 (2002); (b) Cambridge Crystallographic Data Centre, Cambridge, UK, [www.ccdc.cam.ac.uk](http://www.ccdc.cam.ac.uk).
36. Spartan 04, Wavefunction Inc., Irvine, USA, [www.wavefun.com](http://www.wavefun.com).
37. C. Laurence and H. Berthelot, *Perspect. Drug Discovery Des.*, **18**, 39–60 (2000).
38. A. Nangia, *Acc. Chem. Res.*, **40**, 595–604 (2008).
39. Gaussian 03, DFT, B3LYP/6-31G (d,p) level, Gaussian Inc., Walingford, USA, [www.gaussian.com](http://www.gaussian.com).
40. Cerius<sup>2</sup>, COMPASS force field, Accelrys Inc., San Diego, USA, [www.accelrys.com](http://www.accelrys.com).
41. R. Banerjee, R. Mondal, J. A. K. Howard and G. R. Desiraju, *Cryst. Growth Des.*, **6**, 999–1009 (2006).



# 6

## Persistence of N–H...S Hydrogen Bonding in Thiocarbamide Structures

*Edward R. T. Tiekink*

### 6.1 Introduction

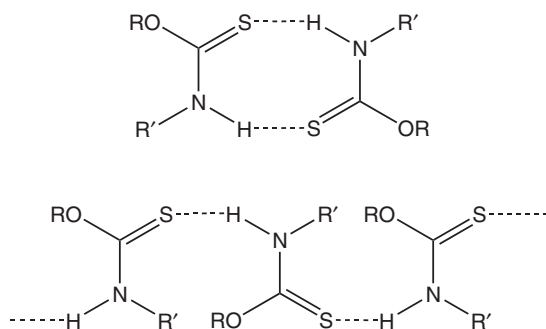
Successful exercises in crystal engineering are predicated upon the ability to control supramolecular organisation in the solid state. Practitioners of crystal engineering would like to assemble molecules at will via robust supramolecular synthons and this, succinctly, is the challenge of crystal engineering. The most popular form of associating molecules in organic crystal engineering involves the exploitation of hydrogen bonding interactions owing to their ease of formation, robustness, directionality, predictability, etc. But, is this really true? Arguably, one of the most recognisable supramolecular synthons is the eight-membered  $\{...HOC=O\}_2$ ,  $R_2^2(8)$ , carboxylic acid dimer homosynthon [1]. Building upon an earlier bibliographic survey [2], Zaworotko and co-workers recently confirmed from their examination of the crystallographic data included in the Cambridge Structural Database (CSD), that of all the organic carboxylic acid structures, only 33% aggregate via the  $\{...HOC=O\}_2$  homosynthon owing to the presence of competing supramolecular synthons in their crystal structures [3]. Further investigation by these authors showed that when competing supramolecular synthons were absent, the percentage of structures featuring the  $\{...HOC=O\}_2$  homosynthon increased to 93% [3]. These percentages clearly indicate that the  $\{...HOC=O\}_2$  homosynthon is readily subjugated when other modes of association are possible in the structures of carboxylic

acid derivatives. The related  $\{\dots \text{HN(R)C=O}\}_2$ ,  $R_2^2(8)$ , homosynthon, potentially formed by secondary amides, is found in less than 10% of their organic structures but this may be due to near universal adoption of an anti configuration in their compounds which would lead to supramolecular chains [2]; this consideration is discussed in greater detail below. Motivated by the imperative that a hierarchy of supramolecular synthons needs to be determined in order to ascertain reliable supramolecular synthons [3–5], the present work details an investigation of the supramolecular aggregation patterns in the crystal structures of thiocarbamide derivatives of  $\text{ROC(=S)N(H)R'}$ , whereby  $\text{N-H}\dots\text{S}$  hydrogen bonding interactions leading to dimeric aggregates or supramolecular chains might be envisaged, see Scheme 6.1.

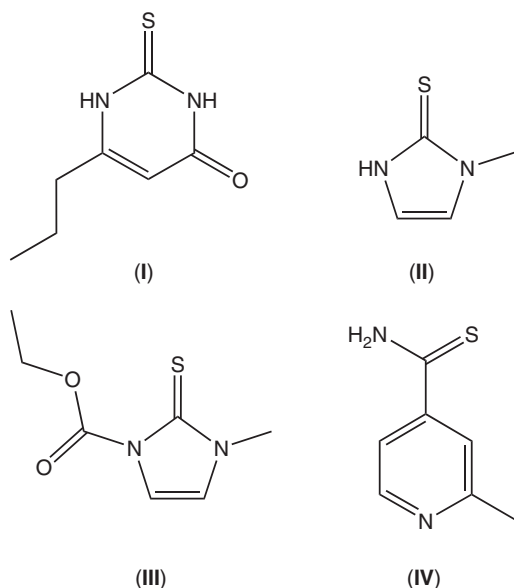
While hydrogen bonding interactions involving sulfur are less well studied than their oxygen congeners, they, e.g.  $\text{O-H}\dots\text{S}$  and  $\text{N-H}\dots\text{S}$ , can be exploited in crystal engineering studies [6–19]. Database analyses of these interactions have demonstrated their prevalence, their relative weakness and lack of directionality compared with more conventional hydrogen bonding interactions [6, 7]. Mak's pioneering work on stabilising unusual species using the principles of crystal engineering is particularly noteworthy in this context [14–18].

Our motivation for investigating thioamide derivatives, which potentially can form  $\text{N-H}\dots\text{S}$  interactions, arises from the presence of the  $\text{-N(H)-C(=S)-}$  residue, and derivatives thereof, in several drugs [20], see Scheme 6.2 for chemical structures of these species. Thus, the antithyroid drugs 6-*n*-propyl-thiouracil (**I**) and 1-methyl-3*H*-imidazole-2-thione (**II**, Methimazole<sup>®</sup>, also marketed as Tapazole<sup>®</sup>) feature the  $\text{-N(H)-C(=S)-}$  functional group [21]. Hyperthyroidism may be treated with 3-methyl-2-thioxo-4-imidazoline-1-carboxylate (**III**, Carbimazole<sup>®</sup>) which is metabolised *in vivo* to the active form, known as Methimazole<sup>®</sup> [22]. Anti-tubercular agents containing the thioamide residue and which inhibit mycolic acid synthesis include 2-ethyl-thioisonicotinamide (**IV**, Ethionamide<sup>®</sup>) and its *n*-propyl derivative (Prothionamide<sup>®</sup>) [23–25].

The focus of the present chapter is upon thiocarbamide derivatives and in particular upon the persistence of  $\text{N-H}\dots\text{S}$  hydrogen bonding operating in their crystal structures.



**Scheme 6.1** Possible  $\text{N-H}\dots\text{S}$  hydrogen bonding interactions in the crystal structures of  $\text{ROC(=S)N(H)R'}$

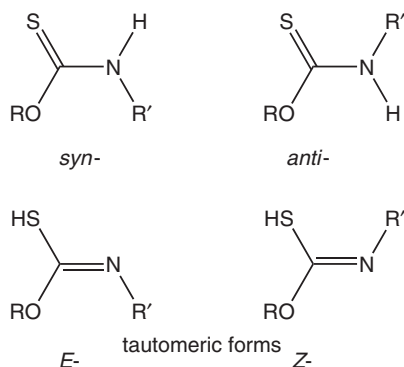


**Scheme 6.2** Chemical structures of drugs containing the  $-N(H)-C(=S)-$  residue and derivatives thereof: (I) 6-n-propyl-thiouracil, (II) 1-methyl-3H-imidazole-2-thione (Methimazole<sup>®</sup> or Tapazole<sup>®</sup>), (III) ethyl 3-methyl-2-sulfanylidene-imidazole-1-carboxylate (Carbimazole<sup>®</sup>) and (IV) 2-ethylpyridine-4-carbothioamide (Ethionamide<sup>®</sup>)

Several isomeric forms may be envisaged for these molecules, which are secondary amides, as shown in Scheme 6.3. For the N-protonated form of the molecule, the isomer whereby the N–H hydrogen lies to the same side of the N–C bond as the thione-sulfur atom is designated the syn isomer and the opposite configuration is designated as the anti isomer. Theoretical studies support crystallographic observations that suggest significant double-bond character in the N–C bond and hence, conventional E and Z terminology is usually employed [26]. This practice is also adopted herein. The tautomeric structures in which the acidic hydrogen atom has migrated to the sulfur atom to form a thiol enabling the formation of a formal N=C double bond are also possible, Scheme 6.3. However, there are no examples of crystal structures of thiocarbamide molecules existing in either the E- or Z- form of the thiol – all published crystal structure determinations feature the thione form in either the E- or Z- configuration (see below). This observation is substantiated by theoretical studies on representative thiocarbamide molecules that indicate the E- and Z-thiol tautomers are over  $70 \text{ kJ mol}^{-1}$  higher in energy, i.e. less stable, than the respective E- and Z-thione tautomers [26].

A survey of the crystallographic data deposited in the Cambridge Structural Database (CSD version 5.30 February 2009 update) [27] was conducted whereby CONQUEST (version 1.11) [28] was employed to search for the  $-O-C(=S)-N(H)-C-$  fragment. This resulted in 59 ‘hits’ for organic structures suitable for analysis, e.g. unambiguous location of hydrogen atoms. Structure analyses and interpretation were conducted with the aid of PLATON [29] and diagrams were drawn with the DIAMOND programme [30].





**Scheme 6.3** Isomeric forms of thiocarbamide molecules

## 6.2 Supramolecular Aggregation Patterns in the Thiocarbamides

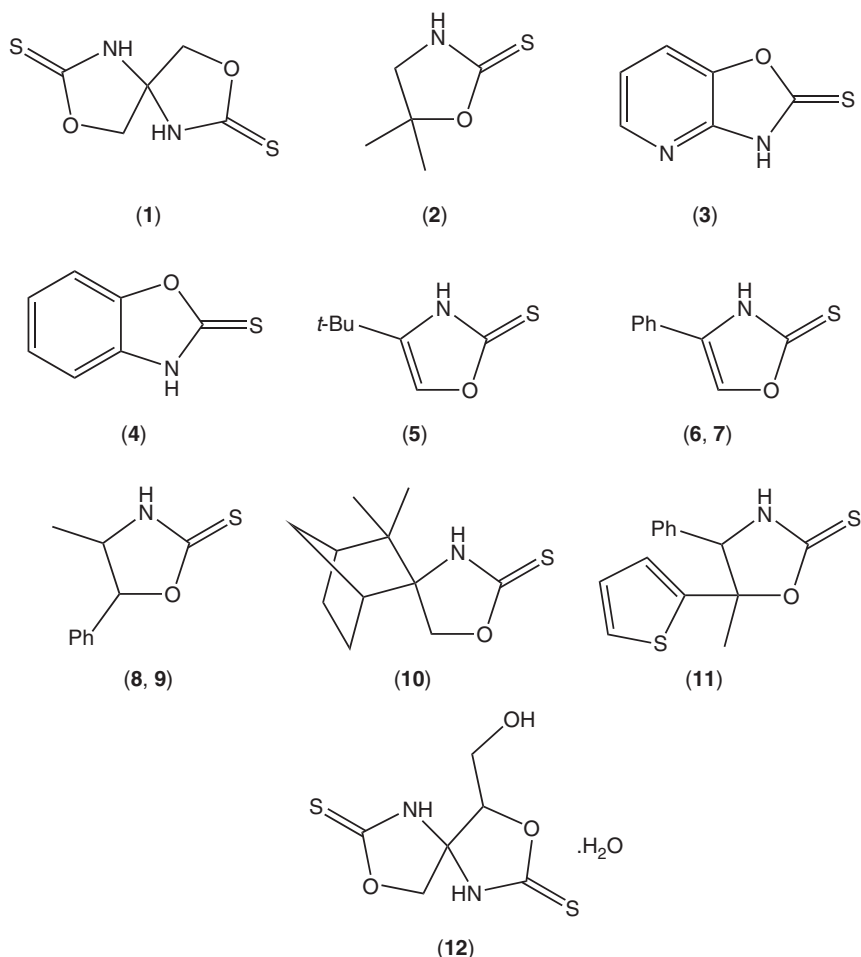
In the following discussion, structures containing thiocarbamide residues are classified into two groups depending on how the residue is integrated into the molecule, i.e. incorporated within a cyclic ring or not. In the former, the molecules are further divided into groups comprising five- and six-membered rings. For the acyclic species, molecules are described in accordance with their isomeric form, i.e. E or Z configurations. Finally, within each category, molecules are arranged in order of increasing molecular formulae with solvates listed after non-solvated structures.

### 6.2.1 Cyclic thiocarbamides – five-membered rings

Chemical structures for molecules in this category, i.e. (1)–(12) [31–40] are shown in Scheme 6.4 and geometric parameters are collected in Table 6.1. The key feature of these structures is that, by virtue of its incorporation within a five-membered ring, the thiocarbamide adopts a syn configuration which, presumably, will favour the adoption of the  $R_2^2(8)$  ring synthon.

Indeed, of the molecules collated in Table 6.1, N–H...S hydrogen bonding leading to eight-membered {...HNC(=S)}<sub>2</sub> motifs are found in seven structures, a representative example showing this mode of association, (4), is shown in Figure 6.1(a). In most examples, the eight-membered synthon is disposed about a crystallographic centre of inversion but in the cases of (9) [37] and (10) [38], the dimer is formed between two crystallographically independent molecules comprising the respective asymmetric unit. Two examples exist whereby N–H...S hydrogen bonding leads to supramolecular chains, C(4), instead, namely (1) [31] and (7) [36], illustrated in Figure 6.1(b).

The fickleness of the adoption of dimer or chain motifs is highlighted in the following two cases. Firstly, in (1), which contains two thioamide functionalities, see Scheme 6.4, the chains formed involve the participation of only one of the thioamide groups, the other does not engage in significant intermolecular interactions. Secondly, chirality can play a role in the adoption of the dimer or chain motifs as seen in the structures of (6) and (7) [36]. Thus, racemic-(6) forms a dimer whereas chiral-(7) adopts the chain motif [36]. By contrast, both racemic-(8) and chiral-(9) form dimer motifs [37]. The



**Scheme 6.4** Chemical structures of thiocarbamide residues incorporated within five-membered rings, (1)–(12)

three remaining structures in this category feature alternative supramolecular aggregation patterns owing to the intervention of different supramolecular synthons.

Dimeric aggregates are formed in the structure of (3), Figure 6.1(c), but mediated by N–H...N rather than N–H...S hydrogen bonding [33], clearly the former dominates over the later in accordance with Etter's rules [41]. The remaining two structures form supramolecular chains. In (4), Figure 6.1(d), association between molecules occurs via N–H...S interactions but connections are reinforced by weak N–H...O contacts, and so this motif is a variant of the C(4) chains described above [34]. A bifunctional thiocarbamide molecule features in the only solvate in this category, i.e. a hydrate, (12) [40]. The water molecules play a pivotal role in the supramolecular organisation, linking molecules via O–H...O and N–H...O hydrogen bonds; no N–H...S interactions are noted. However, O–H...S contacts link chains; the remaining N–H group does not participate in significant intermolecular interactions, Figure 6.1(e).

**Table 6.1** Details of supramolecular association of the thiocarbamide moiety when incorporated within a five-membered ring, (1)–(12)

Compound	Molecular formula	Motif	Distance N...S (Å)	Angle N–H...S (°)	Ref.
(1)	C <sub>5</sub> H <sub>6</sub> N <sub>2</sub> O <sub>2</sub> S <sub>2</sub>	C(4)	3.34	155	[31]
(2)	C <sub>5</sub> H <sub>9</sub> NOS	R <sub>2</sub> <sup>2</sup> (8)	3.36	172	[32]
(3)	C <sub>6</sub> H <sub>4</sub> N <sub>2</sub> OS	R <sub>2</sub> <sup>2</sup> (8)	–	–	[33]
(4)	C <sub>7</sub> H <sub>5</sub> NOS	C(4)	3.44	149	[34]
(5)	C <sub>7</sub> H <sub>11</sub> NOS	R <sub>2</sub> <sup>2</sup> (8)	3.39	168	[35]
(6)	C <sub>9</sub> H <sub>9</sub> NOS	R <sub>2</sub> <sup>2</sup> (8)	3.44	176	[36]
(7)	C <sub>9</sub> H <sub>9</sub> NOS	C(4)	3.35	160	[36]
(8)	C <sub>10</sub> H <sub>11</sub> NOS	R <sub>2</sub> <sup>2</sup> (8)	3.44	171	[37]
(9)	C <sub>10</sub> H <sub>11</sub> NOS	R <sub>2</sub> <sup>2</sup> (8)	3.41	172	[37]
			3.28	169	
(10)	C <sub>11</sub> H <sub>17</sub> NOS	R <sub>2</sub> <sup>2</sup> (8)	3.36	172	[38]
			3.41	171	
(11)	C <sub>14</sub> H <sub>13</sub> NOS <sub>2</sub>	R <sub>2</sub> <sup>2</sup> (8)	3.35	176	[39]
(12)	C <sub>8</sub> H <sub>9</sub> N <sub>2</sub> O <sub>2</sub> S <sub>2</sub> ·H <sub>2</sub> O	Chain	–	–	[40]

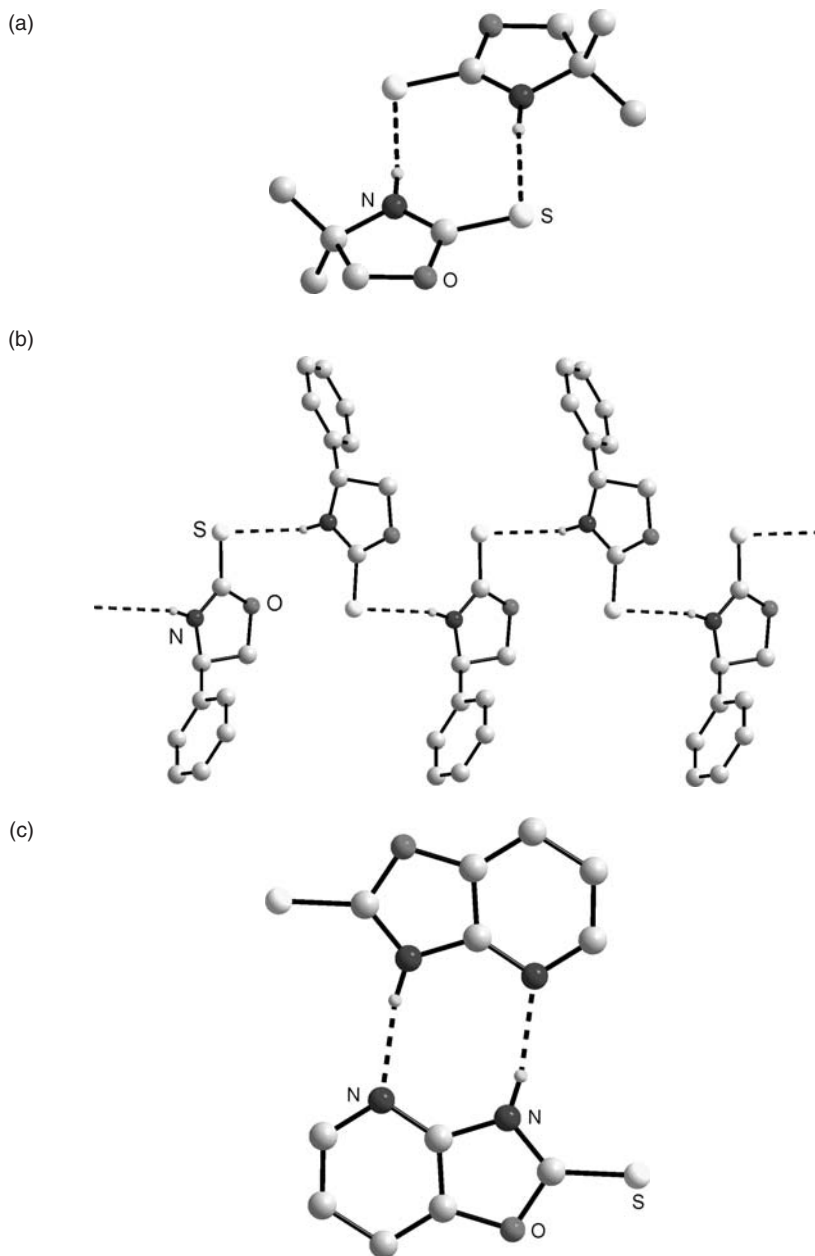
### 6.2.2 Cyclic thiocarbamides – six-membered rings

Chemical structures for all ten structures in this category, (13)–(22) [42–49], are illustrated in Scheme 6.5 and geometric data are collected in Table 6.2. By contrast to the structural variety exhibited by their five-membered ring analogues, (13)–(22) crystallise as centrosymmetric dimers via the {...HNC(=S)}<sub>2</sub> synthon, akin to that illustrated in Figure 6.1(a). Three pairs, i.e. (13) [42] and (14) [43], (17) and (18) [45], and (19) and (20) [46], crystallise as polymorphs but their supramolecular aggregation patterns involving N–H...S contacts in each pair are indistinguishable. While such consistency of supramolecular aggregation might be encouraging in terms of identifying a robust synthon, the close similarity of the structures coupled with the absence of additional functional groups capable of forming significant hydrogen bonding interactions needs to be noted.

### 6.2.3 Acyclic thiocarbamides – E configuration

The homogeneity in the supramolecular association patterns observed above is repeated for the monofunctional acyclic thiocarbamides having an E configuration, (23)–(44) [26, 49–64]; see Scheme 6.6 and Table 6.3. Thus, 17 of the 19 structures associate via the {...HNC(=S)}<sub>2</sub> synthon, see Figure 6.2(a) for a representative example, i.e. (25). Compound (23) crystallises with two independent molecules in the asymmetric unit and these associate to form the {...HNC(=S)}<sub>2</sub> synthon [26]. In (28), there are three independent molecules comprising the asymmetric unit [52]. Two associate as for (23) and the other forms the {...HNC(=S)}<sub>2</sub> synthon with a centrosymmetric mate. The remaining structures forming the {...HNC(=S)}<sub>2</sub> synthon do so across a centre of inversion. The two exceptional structures in this category each form supramolecular chains.

In (39), a supramolecular chain is formed via N–H...S hydrogen bonding, Figure 6.2(b) [52]. In (31) [55], there are no conventional intermolecular hydrogen bonding



**Figure 6.1** Supramolecular association (black-dashed lines) operating in thiocarbamide structures when incorporated within a five-membered ring. Hydrogen atoms not participating in intermolecular interactions are omitted for reasons of clarity. (a) Dimer formation in (**2**) via N–H...S hydrogen bonding, (b) chain formation in (**7**) via N–H...S hydrogen bonding, (c) dimer formation in (**3**) via N–H...N hydrogen bonding, (d) chain formation in (**4**) via N–H...O hydrogen bonding and (e) chain formation in (**12**) via O–H...O hydrogen bonding whereby the pivotal role of solvent water molecules is highlighted

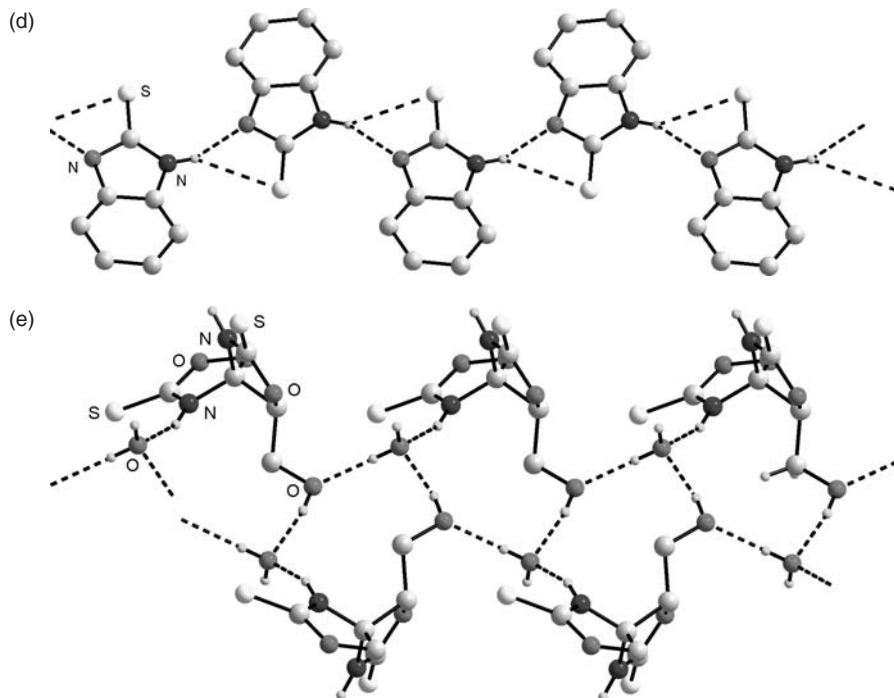
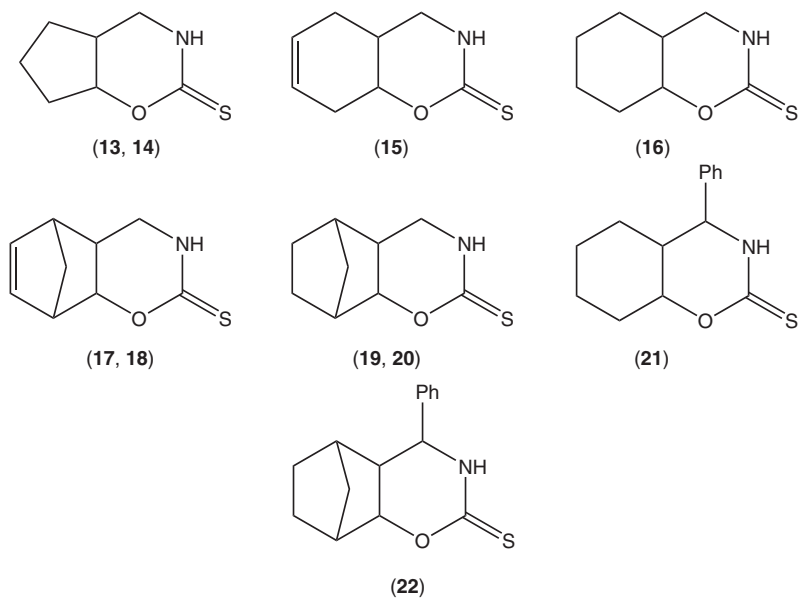


Figure 6.1 (continued)



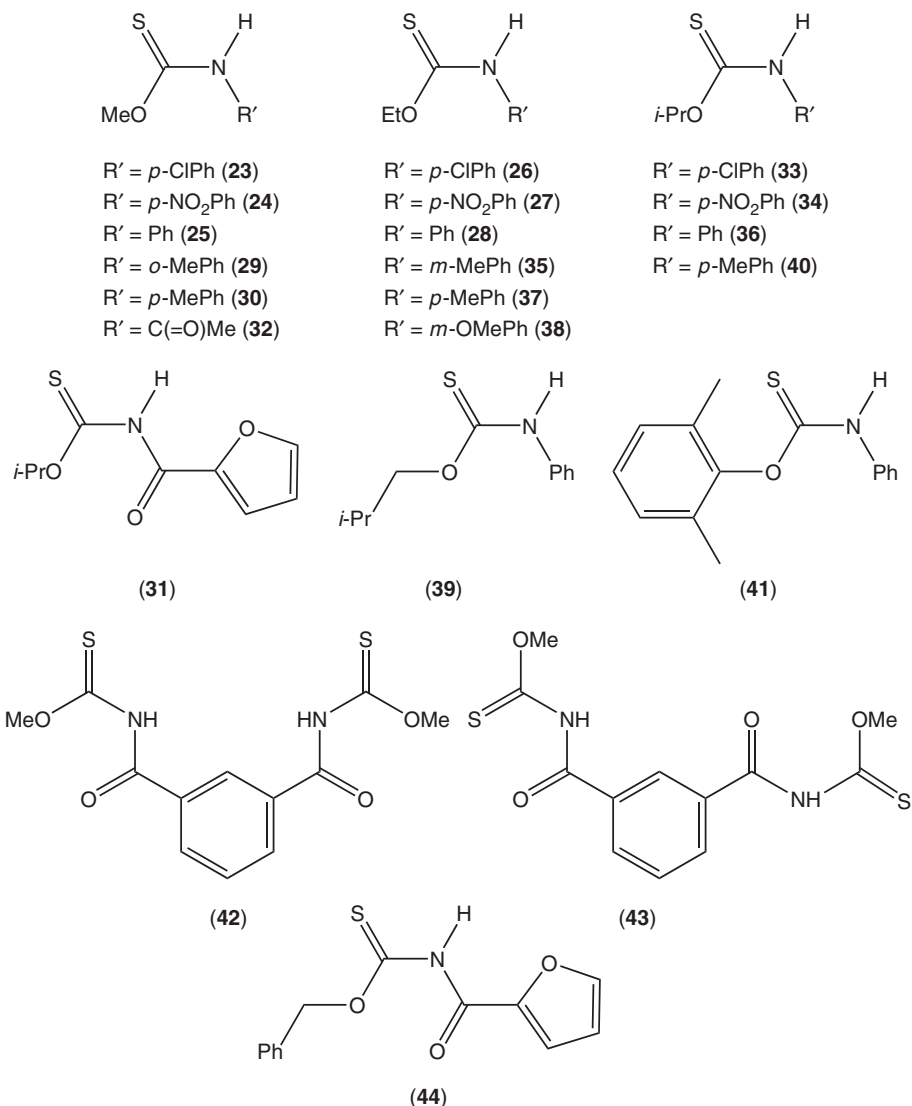
**Scheme 6.5** Chemical structures of thiocarbamide residues incorporated within six-membered rings, (13)–(22)

**Table 6.2** Details of supramolecular association of the thiocarbamide moiety when incorporated within a six-membered ring, (13)–(22)

Compound	Molecular formula	Motif	Distance N...S (Å)	Angle N–H...S (°)	Ref.
(13)	C <sub>7</sub> H <sub>11</sub> NOS	$R_2^2(8)$	3.33	165	[42]
(14)	C <sub>7</sub> H <sub>11</sub> NOS	$R_2^2(8)$	3.42	164	[43]
(15)	C <sub>8</sub> H <sub>11</sub> NOS	$R_2^2(8)$	3.37	167	[44]
(16)	C <sub>8</sub> H <sub>13</sub> NOS	$R_2^2(8)$	3.37	167	[44]
(17)	C <sub>9</sub> H <sub>11</sub> NOS	$R_2^2(8)$	3.39	145	[45]
(18)	C <sub>9</sub> H <sub>11</sub> NOS	$R_2^2(8)$	3.32	168	[45]
(19)	C <sub>9</sub> H <sub>13</sub> NOS	$R_2^2(8)$	3.38	141	[46]
(20)	C <sub>9</sub> H <sub>13</sub> NOS	$R_2^2(8)$	3.32	157	[46]
		$R_2^2(8)$	3.35	160	
(21)	C <sub>14</sub> H <sub>17</sub> NOS	$R_2^2(8)$	3.34	158	[47]
(22)	C <sub>15</sub> H <sub>17</sub> NOS	$R_2^2(8)$	3.40	165	[48]

interactions in the crystal structure as the acidic N–H hydrogen is intramolecularly hydrogen bonded to syn-disposed ether-O ( $H \cdots O = 2.19 \text{ Å}$ ) and thione-S ( $H \cdots S = 2.62 \text{ Å}$ ) atoms. Instead, a supramolecular chain mediated by C–H...O contacts ( $C \cdots O = 3.32 \text{ Å}$ ) is formed, Figure 6.2(c). While there is a clear reason to explain the deviation from the norm for (31), no obvious reason exists for the adoption of a C(4) supramolecular chain for (39) [61]. This structure is particularly exceptional especially when one considers that the related  $ROC(=S)N(H)Ph$  structures for  $R = Me$  (25) [49], Et (28) [52] and i-Pr (36) [56] each adopt the prevalent  $R_2^2(8)$  motif. In the absence of any obvious steric or electronic reasons, the adoption of the C(4) motif for (39) appears to be due to the adaptation of a different local energy-minimum structure in the crystal packing. Such considerations appear to contribute to the structural variations in the three remaining compounds in this category, namely (42)–(44); see Scheme 6.6 and Table 6.3.

The next two structures to be described, i.e. (42) and (43), are isomeric and each contain two thiocarbamide residues [63]. Compound (42) has crystallographic twofold symmetry and the thiocarbamide residue has an E configuration. By contrast, isomeric (43) has both E and Z configurations within the same molecule and therefore the structures of (42) and (43) might be described as being conformational polymorphs [65]. The proximity of carbonyl-O atoms to the amide groups indicates that these compounds are amides and therefore, not surprisingly, (42) crystallise as supramolecular C(4) chains via N–H...O hydrogen bonds, rather than N–H...S hydrogen bonds, as shown in Figure 6.3(a). However, in (43), both N–H...O and N–H...S hydrogen bonding is evident so that the N–H of the thiocarbamide in the E configuration hydrogen bonds to the sulfur atom of the thiocarbamide in the Z configuration; the N–H of the latter forms a regular amide N–H...O hydrogen bond with  $H \cdots O = 2.08 \text{ Å}$ , Figure 6.3(b). The carbonyl not involved in a formal hydrogen bond forms a weak N–H...O interaction ( $H \cdots O = 2.55 \text{ Å}$ ) to reinforce the N–H...S contact and hence, this N–H is bifurcated. The supramolecular chain in (42) comprises molecules arranged in a head-to-head fashion whereas in (43), the arrangement is head-to-tail. The different topologies of the resultant chains is highlighted in the end-on views shown in Figure 6.3(c) and (d), which emphasises the somewhat splayed disposition of the chain in (43).



**Scheme 6.6** Chemical structures of acyclic thiocarbamide structures having an *E* configuration, (**23**)–(**44**)

The final structure in this category, (**44**), in a sense resembles the situation in (**43**). Here, the asymmetric unit comprises two independent molecules, one with the thiocarbamide in an *E* configuration and the other in the *Z* configuration [64]. Also, the molecules may be properly described as amides. Molecules of (**44**) aggregate into a supramolecular chain comprising alternating *E* and *Z* molecules via N–H...O hydrogen bonding, as illustrated in Figure 6.3(e).

**Table 6.3** Details of supramolecular association of acyclic thiocarbamides having an *E* configuration, (23) – (44)

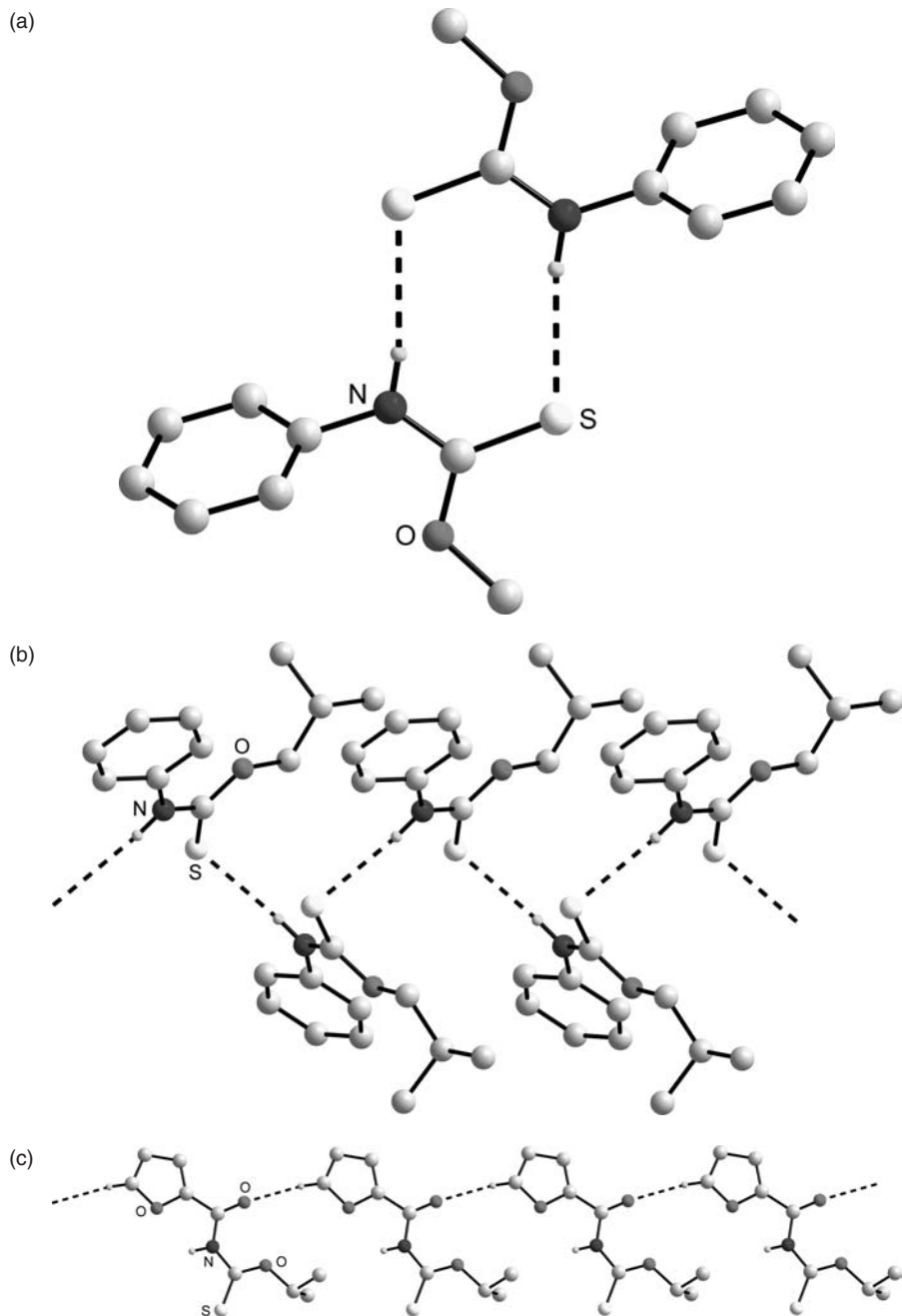
Compound	Molecular formula	Motif	Distance N...S (Å)	Angle N–H...S (°)	Ref.
(23)	C <sub>8</sub> H <sub>8</sub> ClNOS	$R_2^2(8)$	3.40	168	[26]
			3.47	169	
(24)	C <sub>8</sub> H <sub>8</sub> N <sub>2</sub> O <sub>3</sub> S	$R_2^2(8)$	3.41	165	[26]
(25)	C <sub>8</sub> H <sub>9</sub> NOS	$R_2^2(8)$	3.36	165	[49]
(26)	C <sub>9</sub> H <sub>10</sub> ClNOS	$R_2^2(8)$	3.42	159	[50]
(27)	C <sub>9</sub> H <sub>10</sub> N <sub>2</sub> O <sub>3</sub> S	$R_2^2(8)$	3.44	142	[51]
(28)	C <sub>9</sub> H <sub>11</sub> NOS	$R_2^2(8)$	3.39	167	[52]
			3.37	160	
			3.39	163	
(29)	C <sub>9</sub> H <sub>11</sub> NOS	$R_2^2(8)$	3.46	169	[53]
(30)	C <sub>9</sub> H <sub>11</sub> NOS	$R_2^2(8)$	3.29	169	[54]
(31)	C <sub>9</sub> H <sub>13</sub> NOS	C(6)	–	–	[55]
(32)	C <sub>10</sub> H <sub>11</sub> NO <sub>2</sub> S	$R_2^2(8)$	3.41	168	[56]
(33)	C <sub>10</sub> H <sub>12</sub> ClNOS	$R_2^2(8)$	3.36	165	[56]
(34)	C <sub>10</sub> H <sub>12</sub> N <sub>2</sub> O <sub>3</sub> S	$R_2^2(8)$	3.44	162	[57]
(35)	C <sub>10</sub> H <sub>13</sub> NOS	$R_2^2(8)$	3.41	160	[58]
(36)	C <sub>10</sub> H <sub>13</sub> NOS	$R_2^2(8)$	3.37	163	[56]
(37)	C <sub>10</sub> H <sub>13</sub> NOS	$R_2^2(8)$	3.42	168	[59]
(38)	C <sub>10</sub> H <sub>13</sub> NOS	$R_2^2(8)$	3.46	170	[60]
(39)	C <sub>11</sub> H <sub>15</sub> NOS	C(4)	3.40	173	[61]
(40)	C <sub>11</sub> H <sub>15</sub> NOS	$R_2^2(8)$	3.37	163	[56]
(41)	C <sub>15</sub> H <sub>15</sub> NOS	$R_2^2(8)$	3.34	170	[62]
(42)	C <sub>12</sub> H <sub>12</sub> N <sub>2</sub> O <sub>4</sub> S <sub>2</sub>	C(4)	–	–	[63]
(43)	C <sub>12</sub> H <sub>12</sub> N <sub>2</sub> O <sub>4</sub> S <sub>2</sub>	see text	3.42	162	[63]
(44)	C <sub>13</sub> H <sub>11</sub> NO <sub>3</sub> S	C(4)	–	–	[64]

#### 6.2.4 Acyclic thiocarbamides – *Z* configuration

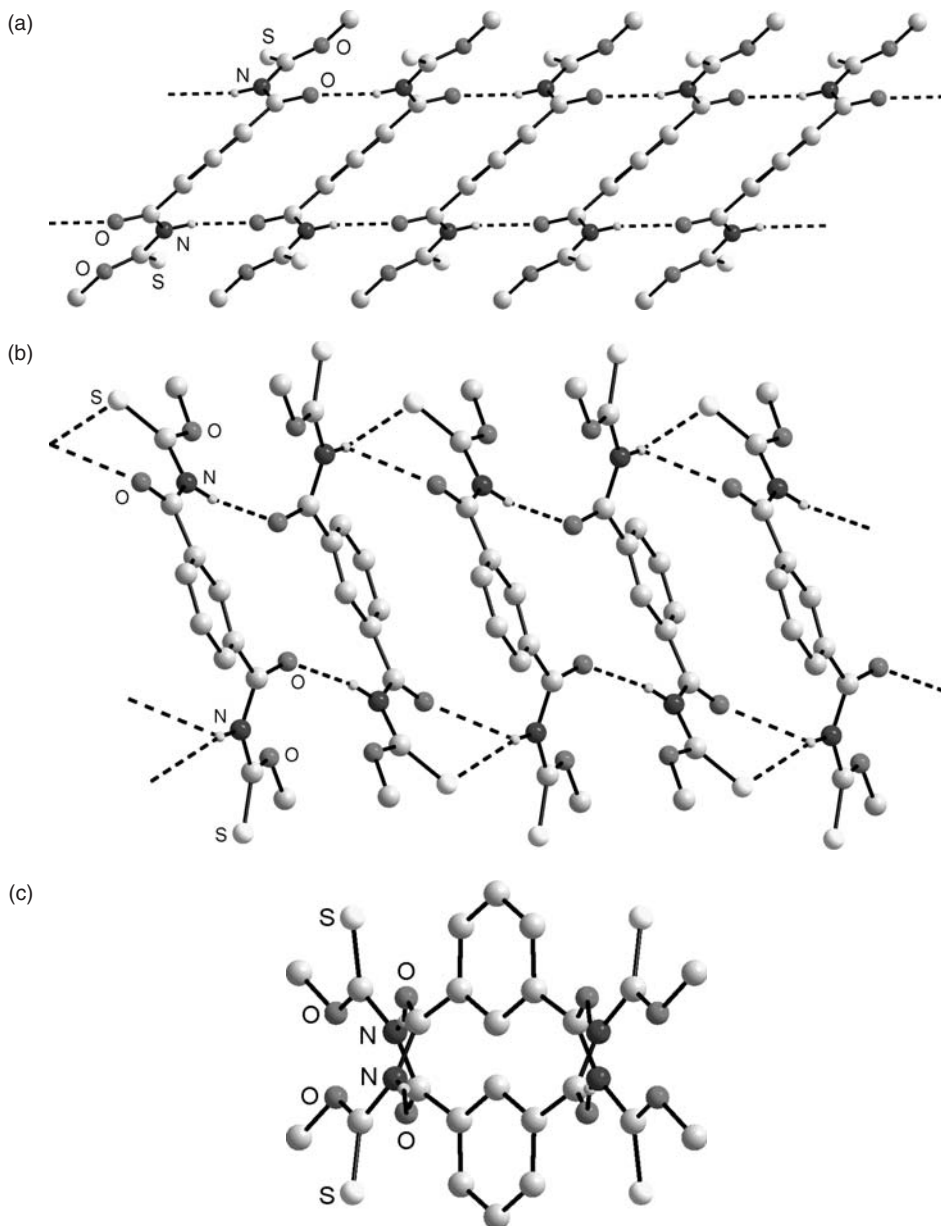
Fifteen structures, (45)–(59) [26, 66–77], fall into this category and of these, only one example adopts a supramolecular aggregation pattern based on N–H...S hydrogen bonding; see Scheme 6.7 and Table 6.4.

The structure of (55) is extraordinary in this category as being the sole example forming a supramolecular chain, C(4), based on N–H...S hydrogen bonding (N...S = 3.51 Å with angle at H = 154°) [74], Figure 6.4(a). All remaining structures have alternative hydrogen bonding synthons that overwhelm putative N–H...S hydrogen bonding interactions. The structures of (48) and (51) [69], see Figure 6.4(b), each associate via N–H...N hydrogen bonding to form a {...HNCN}<sub>2</sub> homosynthon; each eight-membered ring has crystallographic twofold symmetry. This mode of association clearly demonstrates that an N–H...N interaction, where the nitrogen atom is incorporated within a ring, is capable of overwhelming an N–H...S hydrogen bond. This principle is also in operation in the structure of (54) which features both a pyridine-N atom as well as carbonyl-O atom capable of functioning as hydrogen atom acceptors [73]. A dimer is formed in the crystal





**Figure 6.2** Supramolecular association (black-dashed lines) operating in acyclic thio-carbamide structures having an *E* configuration. Hydrogen atoms not participating in intermolecular interactions are omitted for reasons of clarity. (a) Dimer formation in (**23**) via  $N-H \cdots S$  hydrogen bonding, (b) chain formation in (**39**) via  $N-H \cdots S$  hydrogen bonding and (c) chain formation in (**31**) via  $C-H \cdots O$  contacts; the  $N-H$  atom is precluded from forming intermolecular contacts owing to  $N-H \cdots O$  and  $N-H \cdots S$  intramolecular contacts



**Figure 6.3** Supramolecular association (black-dashed lines) operating in acyclic bifunctional thiocarbamide structures having an *E* configuration. Hydrogen atoms not participating in intermolecular interactions are omitted for reasons of clarity. (a) Supramolecular chain in (**42**) mediated by N–H...O hydrogen bonding, (b) chain formation in (**43**) mediated by N–H...O and N–H...S hydrogen bonding, (c) end-on view of the chain in (**42**), (d) end-on view of the chain in (**43**) and (e) supramolecular chain in (**44**) mediated by N–H...O hydrogen bonding

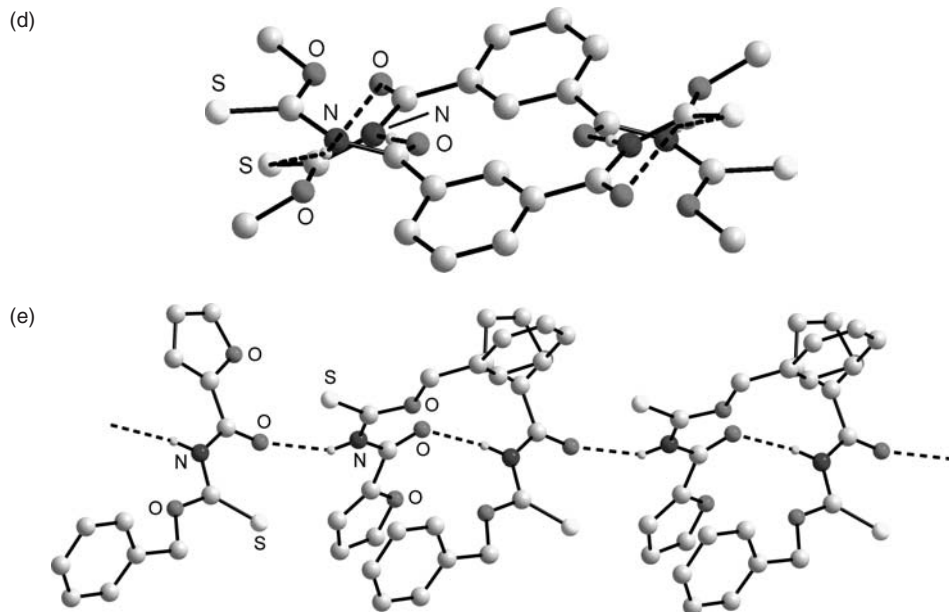
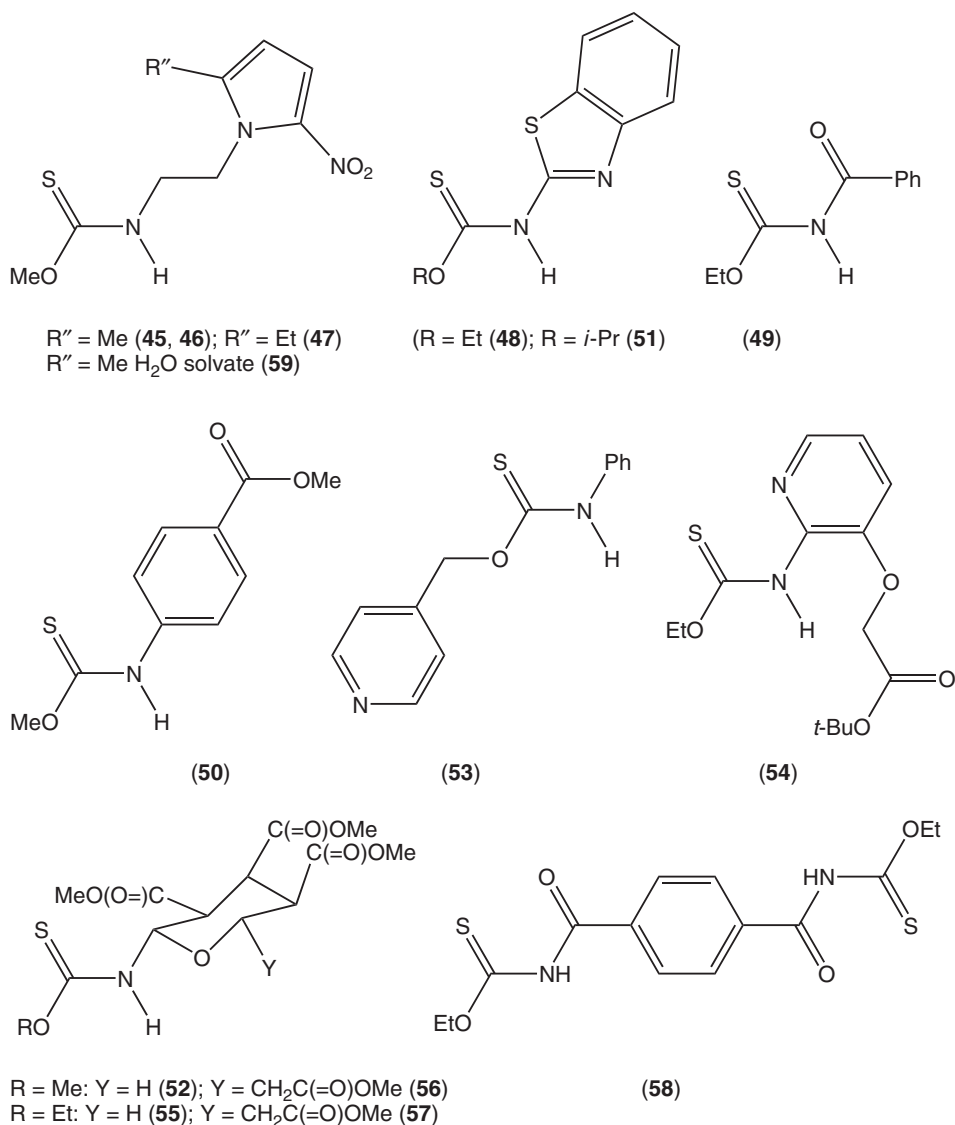


Figure 6.3 (continued)

structure of (**54**) mediated by a pair of cooperative  $\text{N-H} \cdots \text{O}_{\text{carbonyl}}$  and  $\text{N-H} \cdots \text{N}_{\text{pyridine}}$  hydrogen bonds leading to a  $\{\cdots \text{O}=\text{C}_2\text{OC}_2\text{NH} \cdots \text{NCNH}\}$ ,  $R_2^2(12)$ , heterosynthon, Figure 6.4(c). This association leaves one carbonyl-O and one pyridine-N derived from each loosely associate dimeric unit uninvolved in significant hydrogen bonding. This observation suggests that the acceptor ability of these atoms is approximately equivalent, at least when thioamide-nitrogen-H is the donor.

All but two of the remaining structures to be described feature supramolecular chains in their respective crystal structures. In (**49**) [70], the C(4) chain is stabilised by amide-type  $\text{N-H} \cdots \text{O}$  hydrogen bonding, as illustrated in Figure 6.5(a). There are four additional structures, i.e. those of (**50**) [26], (**52**) [71], (**56**) [75], and (**57**) [75], that feature similar supramolecular aggregation via  $\text{N-H} \cdots \text{O}$  hydrogen bonding but where the oxygen acceptor is derived from a carbonyl functionality placed elsewhere in the molecule. A representative example, i.e. (**50**), is shown in Figure 6.4(a). The observed aggregation patterns in the sugar derivatives (**52**), (**56**), and (**57**) might be as expected given that an  $\text{N-H} \cdots \text{O}$  hydrogen bond might be assumed to be stronger than the corresponding  $\text{N-H} \cdots \text{S}$  hydrogen bond. This, then, makes the observed supramolecular aggregation for (**55**) with  $\text{N-H} \cdots \text{S}$  hydrogen bonding, Figure 6.5(a), all the more curious [74]. The next four structures feature heterocyclic rings containing nitrogen atoms potentially available as hydrogen bond acceptors and in each of these structures, i.e. polymorphic (**45**) [66] and (**46**) [67], (**47**) [68] and (**53**) [72],  $\text{N-H} \cdots \text{N}$  hydrogen bonding leading to supramolecular chains is found; see Figure 6.5(c) for a representative example, (**45**). The topology of the resultant chains in (**45**) and (**46**) are quite distinct owing to variable orientations of successive molecules. Thus, in (**45**) [66], the supramolecular chain is approximately planar, whereas a zigzag topology is found in (**46**) [67]. In the sole



**Scheme 6.7** Chemical structures of acyclic thiocarbamide structures having a *Z* configuration, (45)–(59)

example of a bifunctional molecule, i.e. (58), amide-type N-H...O hydrogen bonding in three dimensions leads to a network architecture [76].

The final structure to be described is a hydrate, namely (59) [77], which is a pseudopolymorph [78] of (45) [66] and (46) [67]. The water molecule plays a pivotal role in the crystal packing by forming O-H...O and O-H...N hydrogen bonds with water-hydrogen as the donor, and accepts an N-H...O hydrogen bond. In this way, the water molecule links three different thioamide molecules with the result that a 2-D array is

**Table 6.4** Details of supramolecular association of acyclic thiocarbamides having a *Z* configuration, (45)–(59)

Compound	Molecular formula	Motif	Ref.
(45)	C <sub>8</sub> H <sub>12</sub> N <sub>4</sub> O <sub>3</sub> S	C(7)	[66]
(46)	C <sub>8</sub> H <sub>12</sub> N <sub>4</sub> O <sub>3</sub> S	C(7)	[67]
(47)	C <sub>9</sub> H <sub>14</sub> N <sub>4</sub> O <sub>3</sub> S	C(7)	[68]
(48)	C <sub>10</sub> H <sub>10</sub> N <sub>2</sub> OS <sub>2</sub>	<i>R</i> <sub>2</sub> <sup>2</sup> (8)	[69]
(49)	C <sub>10</sub> H <sub>11</sub> NO <sub>2</sub> S	C(4)	[70]
(50)	C <sub>10</sub> H <sub>11</sub> NO <sub>3</sub> S	C(8)	[26]
(51)	C <sub>11</sub> H <sub>12</sub> N <sub>2</sub> O <sub>2</sub> S	<i>R</i> <sub>2</sub> <sup>2</sup> (8)	[69]
(52)	C <sub>13</sub> H <sub>9</sub> NO <sub>8</sub> S <sub>2</sub>	C(7)	[71]
(53)	C <sub>13</sub> H <sub>12</sub> N <sub>2</sub> OS	C(9)	[72]
(54)	C <sub>14</sub> H <sub>20</sub> N <sub>2</sub> O <sub>4</sub> S	<i>R</i> <sub>2</sub> <sup>2</sup> (12)	[73]
(55)	C <sub>14</sub> H <sub>21</sub> N <sub>2</sub> O <sub>8</sub> S	C(4)	[74]
(56)	C <sub>16</sub> H <sub>23</sub> NO <sub>10</sub> S	C(8)	[75]
(57)	C <sub>17</sub> H <sub>25</sub> NO <sub>10</sub> S	C(8)	[75]
(58)	C <sub>14</sub> H <sub>16</sub> N <sub>2</sub> O <sub>4</sub> S <sub>2</sub>	3-D	[76]
(59)	C <sub>8</sub> H <sub>12</sub> N <sub>4</sub> O <sub>3</sub> S.H <sub>2</sub> O	2-D	[77]

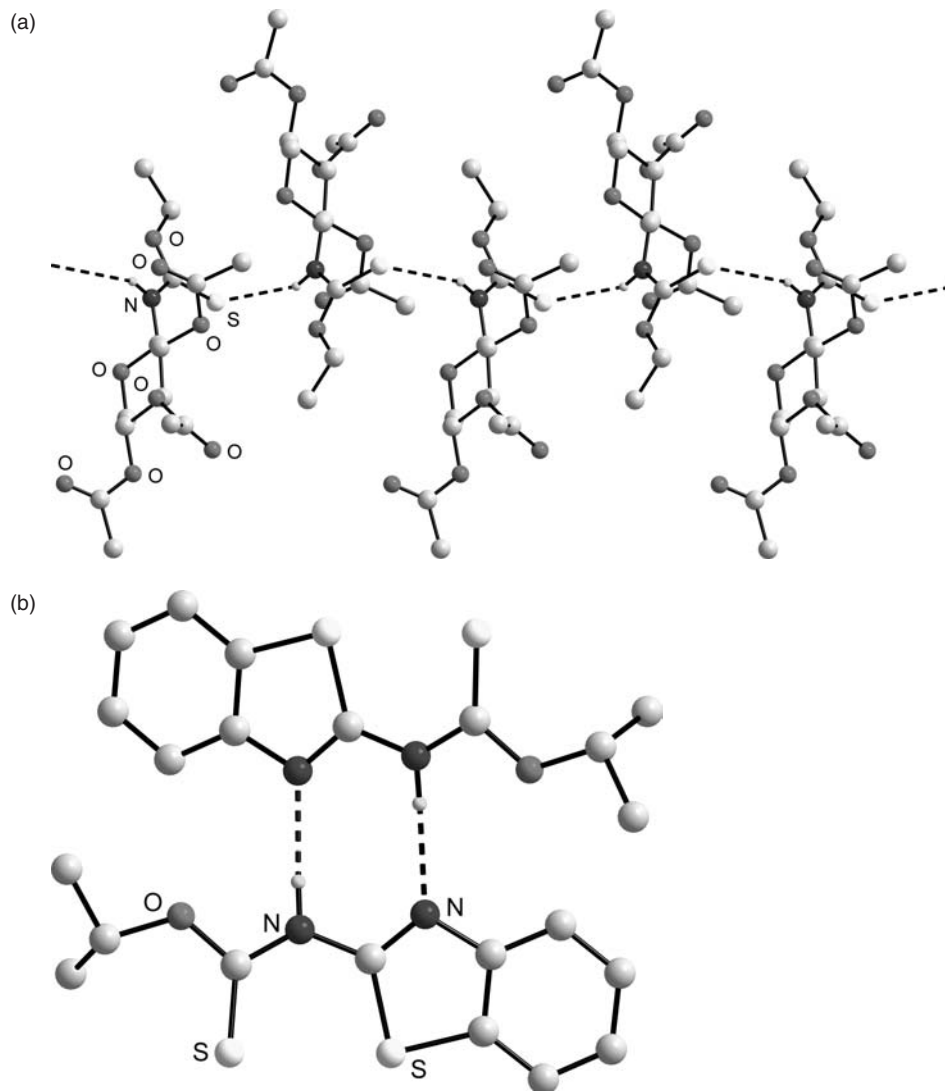
constructed, as illustrated in Figure 6.6. From the foregoing, it is clear that the water molecule induces additional supramolecular aggregation whereby the normally observed supramolecular chains observed in this category are associated into a 2-D array, (59).

### 6.2.5 Co-crystals containing thiocarbamides

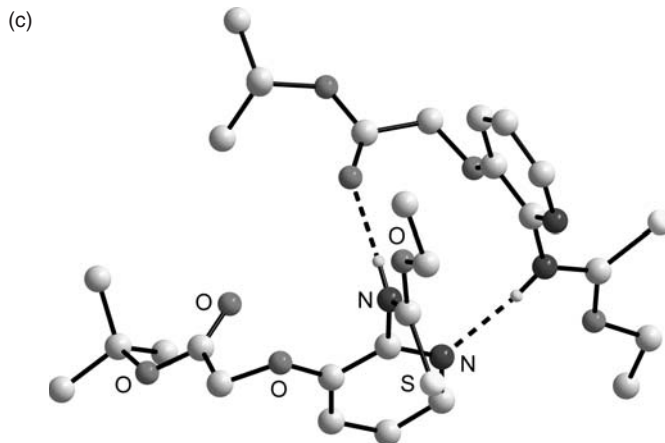
The exploration of co-crystal formation is largely motivated by pharmaceutical considerations as has been comprehensively reviewed in a companion volume [79]. In this respect, thiocarbamides are relatively unexplored despite the presence of the –N(H)–C(=S)– residue in some pharmaceuticals, as discussed above. A recent study describes crystallographic analyses of 2:1 co-crystal formation between (24), (25), (28) and (36) with *trans*-1,2-bis(4-pyridyl)ethane, and between (27) and 4,4'-bipyridine [19]. Without exception, the N–H...S hydrogen bonding operating in the structures of the pure thiocarbamides, leading to the eight-membered synthon, is disrupted in favour of N–H...N hydrogen bonding in keeping with Etter's rules [41]; see Figure 6.7.

## 6.3 Conclusions

The delineation of supramolecular synthons, in particular the prevalence of N–H...S hydrogen bonding in crystallographically characterised thiocarbamides has been the purpose of this bibliographic survey. Of the 59 structures available for examination, no less than 34 form the *R*<sub>2</sub><sup>2</sup>{8}, {...HNC(=S)}<sub>2</sub> homosynthon, with another five molecules forming C(4) supramolecular chains, again mediated by N–H...S hydrogen bonding. The remaining 20 structures feature alternate synthons, usually based on N–H...O and N–H...N hydrogen bonding. As a general principle, N–H...S hydrogen bonding forms in approximately two-thirds of the crystal structures and is only displaced when



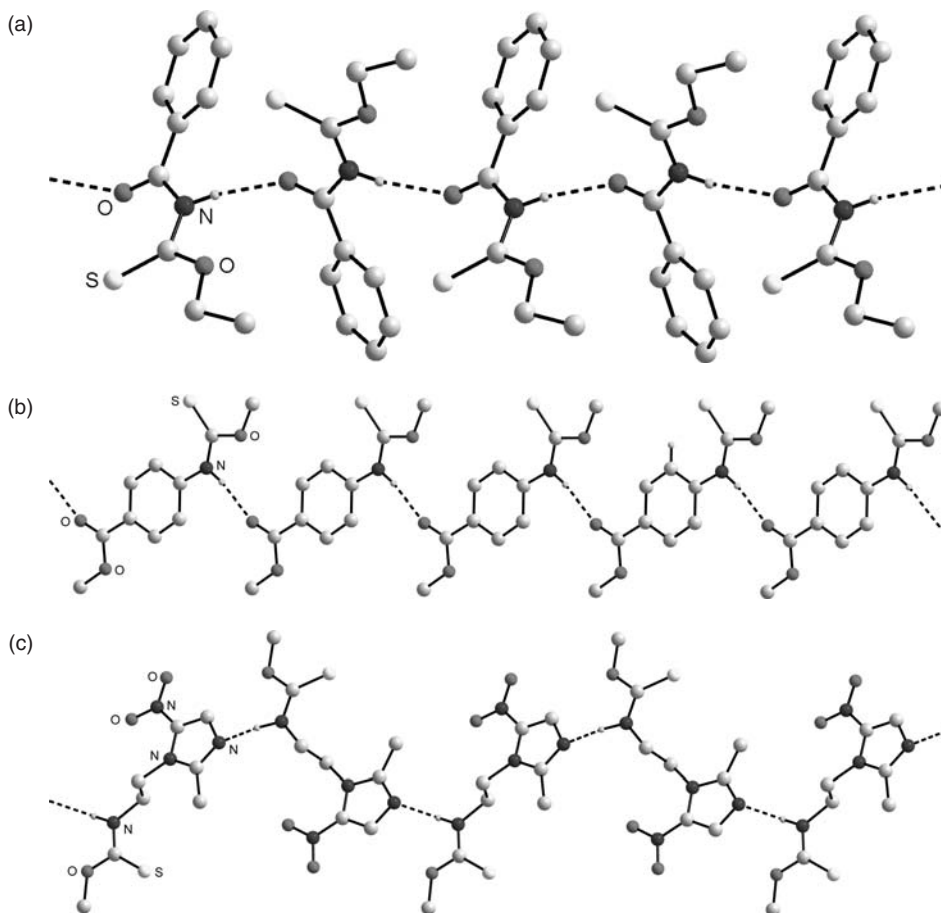
**Figure 6.4** Supramolecular association (black-dashed lines) operating in acyclic thiocarbamide structures having a *Z* configuration. Hydrogen atoms not participating in intermolecular interactions are omitted for reasons of clarity. (a) Sole example of N–H...S hydrogen bonding, leading to a supramolecular chain in the structure of (55), (b) dimer formation mediated by N–H...N hydrogen bonding in (51) and (c) dimer formation via N–H...O and N–H...N hydrogen bonding in (54)

**Figure 6.4** (continued)

carbonyl-O or pyridine-N atoms are also present in the molecules. Such subjugation of a supramolecular synthon has been described as ‘structural interference’ [80].

A correlation exists between the isomeric form of the thiocarbamide molecule, i.e. having an E or Z configuration, and the adoption of the  $\{ \dots \text{HNC}(=\text{S}) \}_2$  homosynthon. Molecules with an E configuration are predisposed to form the  $\{ \dots \text{HNC}(=\text{S}) \}_2$  homosynthon but the conformation can be altered depending on the dictates of crystal packing. A theoretical study on several of the compounds included in this survey shows that the energy differences between the E and Z configurations are of the order  $5 \text{ kJ mol}^{-1}$ , i.e. well within the normal energies of stabilisation of crystal structures [26]. Further, experimental evidence based on  $^1\text{H-NMR}$  measurements indicate that both configurations co-exist in solution. In summary, in the absence of steric or electronic reasons, potentially either the E or Z configuration may be adopted in the crystal structures of thiocarbamides. These theoretical and experimental studies are further vindicated by the observed crystallisation of (42) and (43) [63]. Thus, colourless-(42), with both thiocarbamides residues having an E configuration, and yellow-(43), with one thiocarbamide in an E configuration and the other in the Z configuration, crystallise concomitantly from the same solution [63]. Further, crystals of (44), with equal quantities of E and Z isomers, again indicate very small energy differences between the isomeric molecules [64]. Therefore, the observed solid-state configuration for thiocarbamides most probably depends on which local energy minimum structure is attained in the crystal packing. In the absence of competing synthons involving acceptor atoms such as carbonyl and pyridine-N, molecules will dimerise to form the eight-membered synthon or failing that, will form supramolecular chains, again mediated by  $\text{N-H} \dots \text{S}$  interactions.

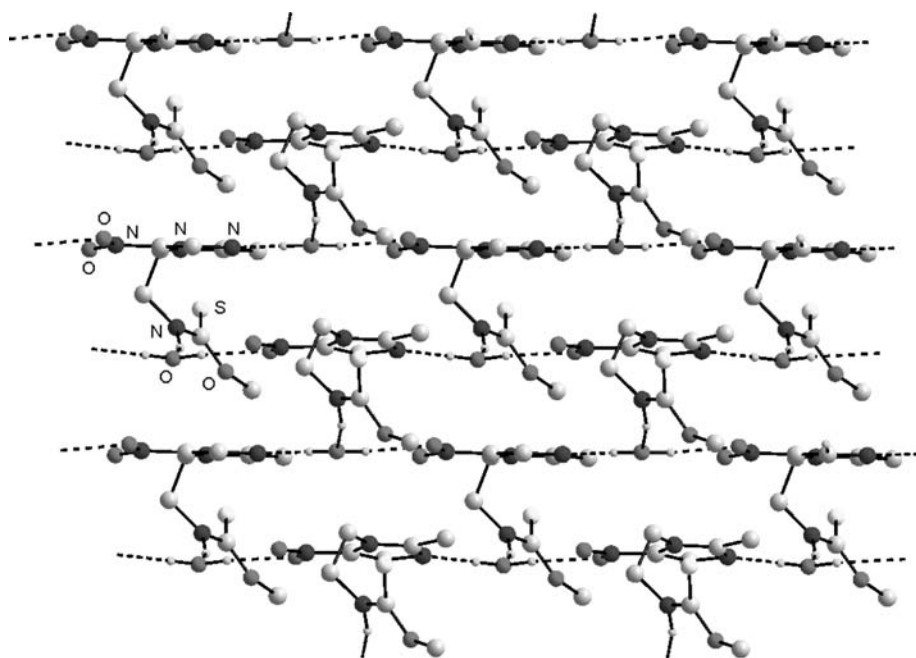
The observed behaviour for the thiocarbamides contrasts to that of secondary amides. First, it is noted for the molecules described herein that also contain the amide



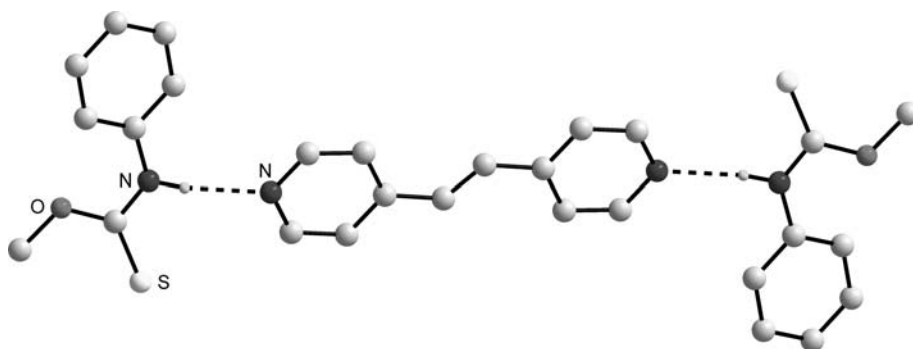
**Figure 6.5** Supramolecular association (black-dashed lines) found for acyclic thiocarbamide structures having a *Z* configuration. Hydrogen atoms not participating in intermolecular interactions are omitted for reasons of clarity. (a) Supramolecular chain mediated by N-H...O (amide) hydrogen bonding in (**49**), (b) supramolecular chains mediated by N-H...O hydrogen bonding in (**50**) and (c) supramolecular chains mediated by N-H...N hydrogen bonding in (**45**)

functionality, N-H...O hydrogen bonding leading to supramolecular chains is often found [81]. For all secondary amide structures where the amide is not constrained within a ring, the {...HNC(=O)}<sub>2</sub> homosynthon is found in less than 10% of their organic structures [2]. This rises to 50% in the thiocarbamides, the relatively small number of structures notwithstanding. Clearly, the steric and electronic effects leading to the adoption of the anti configuration in secondary amides does not apply in the thiocarbamides for which syn configurations are more often than not found in their solid-state structures.





**Figure 6.6** Supramolecular association (black-dashed lines) operating in acyclic thio-carbamide structures having a *Z* configuration. Hydrogen atoms not participating in intermolecular interactions are omitted for reasons of clarity. Supramolecular 2-D array mediated by O-H...O, O-H...N and N-H...O hydrogen bonding in (59), highlighting the pivotal role played by the solvent water molecules in the aggregation pattern



**Figure 6.7** The N-H...N hydrogen bonding (black-dashed lines) in the 2:1 co-crystal formed between (25) and trans-1,2-bis(4-pyridyl)ethene. Hydrogen atoms not participating in intermolecular interactions are omitted for reasons of clarity

## References

1. G. R. Desiraju, *Angew. Chem., Int. Ed.*, **46**, 8342 (2007).
2. F. H. Allen, W. D. S. Motherwell, P. R. Raithby, G. P. Shields and R. Taylor, *New J. Chem.*, **25** (1999).
3. T. R. Shattock, K. K. Arora, P. Vishweshwar and M. J. Zaworotko, *Cryst. Growth Des.*, **8**, 4533 (2008).
4. J. Bis, P. Vishweshwar, D. Weyna and M. J. Zaworotko, *Mol. Pharm.*, **4**, 401 (2007).
5. C. B. Aakeröy, J. Desper, M. Fasulo, I. Hussain, B. Levin and N. Schultheiss, *CrystEngComm*, **10**, 1816 (2008).
6. F. H. Allen, C. M. Bird, R. S. Rowland and P. R. Raithby, *Acta Crystallogr.*, **B53**, 680 (1997).
7. P. A. Wood, E. Pidcock and F. H. Allen, *Acta Crystallogr.*, **B64**, 491 (2008).
8. D. R. Smyth, B. R. Vincent and E. R. T. Tiekink, *Cryst. Growth Des.*, **1**, 113 (2001).
9. K. A. Wheeler, B. Harrington, M. Zapp and E. Casey, *CrystEngComm*, **5**, 337 (2003).
10. R. Alonso, E. Bermejo, R. Carballo, A. Castineiras and T. Perez, *J. Mol. Struct.*, **606**, 155 (2002).
11. J. Valdes-Martinez, S. Hernandez-Ortega, M. Rubio, D. T. Li, J. K. Swearingen, W. Kaminsky, D. R. Kelman, and D. X. West *J. Chem. Crystallogr.*, **34**, 533 (2004).
12. M. G. Fisher, P. A. Gale, M. E. Light and R. Quesada, *CrystEngComm*, **10**, 1180 (2008).
13. R. A. Howie, G. M. de Lima, D. C. Menezes, J. L. Wardell, S. M. S. V. Wardell, D. J. Young and E. R. T. Tiekink, *CrystEngComm*, **10**, 1626 (2008).
14. C.-K. Lam and T. C. W. Mak, *Tetrahedron*, **56**, 6657 (2000).
15. T. C. W. Mak and F. Xue, *J. Am. Chem. Soc.*, **122**, 9860 (2000).
16. C.-K. Lam and T. C. W. Mak, *Chem. Commun.*, 2660 (2003).
17. C.-K. Lam, T.-L. Chan and T. C. W. Mak, *CrystEngComm*, **6**, 290 (2004).
18. J. Han, L. Zhao, C. W. Yau and T. C. W. Mak, *Cryst. Growth Des.*, **9**, 308 (2009).
19. C. A. Ellis, M. A. Miller, J. Spencer, J. Zukerman-Schpector and E. R. T. Tiekink, *CrystEngComm*, **11**, 1352 (2009).
20. J. E. F. Reynolds (Ed.), Martidale. *The Extra Pharmacopoeia*, 28th Edn, The Pharmaceutical Press, London, 1982.
21. H. Nakamura, J. Y. Noh, K. Itoh, S. Fukata, A. Miyauchi and N. Hamada, *J. Clin. Endocrinol. Metabol.*, **92**, 2157 (2007).
22. L. G. Dover, A. Alahari, P. Gratraud, J. M. Gomes, V. Bhowruth, R. C. Reynolds, G. S. Besra and L. Kremer, *Antimicrob. Agents Chemother.*, **51**, 1055 (2007).
23. A. E. DeBarber, K. Mdluli, M. Bosman, L. G. Bekker and C. E. Barry, *Proc. Natl. Acad. Sci. USA*, **97**, 9677 (2000).
24. D. M. Yajko, P. S. Nassos and W. K. Hadley, *Antimicrob. Agents Chemother.*, **31**, 117 (1987).

25. F. Wang, R. Langley, G. Gulten, L. G. Dover, G. S. Besra, W. R. Jacobs and J. C. Sacchettini, *J. Exp. Med.*, **204**, 73 (2007).
26. S. Y. Ho, R. P. A. Bettens, D. Dakternieks, A. Duthie and E. R. T. Tiekink, *CrystEngComm*, **7**, 682 (2005).
27. F. H. Allen, *Acta Crystallogr., Sect. B*, **58**, 380 (2002).
28. I. J. Bruno, J. C. Cole, P. R. Edgington, M. Kessler, C. F. Macrae, P. McCabe, J. Pearson and R. Taylor, *Acta Crystallogr., Sect. B*, **B58**, 389 (2002).
29. A. L. Spek, *J. Appl. Cryst.* **36**, 7 (2003).
30. K. Brandenburg, *DIAMOND*. Visual Crystal Structure Information System, Version 3.1, CRYSTAL IMPACT, Bonn, Germany, 2006.
31. R. Saul, T. Kern, J. Kopf, I. Pinter and P. Koll, *Eur. J. Org. Chem.*, 205 (2000).
32. T. Gray, C. E. Laplaza and R. J. Staples, *Z. Kristallogr. - New Cryst. Struct.*, **214**, 230 (1999).
33. N. A. Aliev, B. Tashkhodzhaev, M. G. Levkovich, N. D. Abdullaev and V. G. Kartsev, *Khim. Get. Soedin., SSSR* (in Russian), 1545 (1997).
34. (a) P. Groth, *Acta Chem. Scand.*, **27**, 945 (1973); (b) M. N. Kamat, N. P. Rath and A. V. Demchenko, *J. Org. Chem.*, **72**, 6938 (2007).
35. S. Mohanraj, W. T. Ford, P. J. Wooldridge and E. M. Holt, *J. Org. Chem.*, **53**, 1113 (1988).
36. S. Kitoh, K.-K. Kunimoto, N. Funaki, H. Senda, A. Kuwae and K. Hanai, *J. Chem. Cryst.*, **32**, 547 (2002).
37. S. Kitoh, H. Senda and K.-K. Kunimoto, *Anal. Sci.: X-Ray Struct. Anal. Online*, **19**, x15 (2003).
38. S. Parsons, G. Howells, A. Hulme and P. Wood, Personal communication to CCDC (refcode: YAGQIH) (2004).
39. A. R. Katritzky, N. M. Khashab, D. N. Haase, M. Yoshioka, I. Ghiviriga and P. J. Steel, *J. Org. Chem.*, **72**, 6742 (2007).
40. G. Cioci, N. Leconte, A. Tatibouet, P. Rollin, S. Perez and A. Imberty, *Acta Crystallogr., Sect. E*, **60**, o2399 (2004).
41. M. C. Etter, *Acc. Chem. Res.*, **23**, 120 (1990).
42. G. Stajer, E. A. Szabo, F. Fulop, G. Bernath, A. Kalman, G. Argay and P. Sohar, *Tetrahedron*, **39**, 1829 (1983).
43. S. Stankovic, A. Kapor, B. Ribar, A. Kalman, G. Argay, L. Karanovic, G. Stajer and G. Bernath, *J. Mol. Struct.*, **131**, 45 (1985).
44. G. Argay, A. Kalman, A. Kapor, G. Stajer and G. Bernath, *J. Mol. Struct.*, **131**, 31 (1985).
45. A. Kapor, B. Ribar, V. Divjakovic, G. Stajer, G. Bernath and G. Argay, *Acta Crystallogr., Sect. C*, **43**, 1347 (1987).
46. A. Kapor, G. Argay, G. Stajer and G. Bernath, *Acta Crystallogr., Sect. C*, **46**, 1495 (1990).
47. A. Kalman, G. Argay, G. Stajer and G. Bernath, *J. Mol. Struct.*, **328**, 245 (1994).
48. A. Kalman, G. Argay, G. Stajer and G. Bernath, *J. Mol. Struct.*, **248**, 167 (1991).
49. S. Y. Ho, C. S. Lai and E. R. T. Tiekink, *Acta Crystallogr. Sect. E*, **59**, o1155 (2003).
50. P. P. Taduppa and E. R. T. Tiekink, *Acta Crystallogr. Sect. E*, **63**, o1779 (2007).

51. R. E. Benson, G. A. Broker, L. M. Daniels, E. R. T. Tiekink, J. L. Wardell and D. J. Young, *Acta Crystallogr. Sect. E*, **62**, o4106 (2006).
52. R. L. Taylor and E. R. T. Tiekink, *Z. Kristallogr.*, **209**, 64 (1994).
53. F. S. Kuan, P. Tadbuppa and E. R. T. Tiekink, *Z. Kristallogr. - New Cryst. Struct.*, **220**, 393 (2005).
54. S. Y. Ho, F. S. Kuan and E. R. T. Tiekink, *Acta Crystallogr. Sect. E*, **63**, o1723 (2007).
55. A. D. Morales, H. Novoa de Armas, N. M. Blaton, O. M. Peeters, C. J. De Ranter, H. Marquez and R. P. Hernandez, *Acta Crystallogr., Sect. C*, **56**, 1042 (2000).
56. F. S. Kuan, F. Mohr, P. P. Tadbuppa and E. R. T. Tiekink, *CrystEngComm*, **9**, 574 (2007).
57. C. A. Ellis, E. R. T. Tiekink and J. Zukerman-Schpector, *Acta Crystallogr., Sect. E*, **64**, o345 (2008).
58. P. Tadbuppa and E. R. T. Tiekink, *Z. Kristallogr. - New Cryst. Struct.*, **220**, 395 (2005).
59. P. P. Tadbuppa and E. R. T. Tiekink, *Acta Crystallogr., Sect. E*, **63**, o1885 (2007).
60. M. Hanif, G. Qadeer, L. Shi and N. H. Rama, *Acta Crystallogr., Sect. E*, **63**, o3062 (2007).
61. F.-F. Jian, H.-Q. Yu, Y.-B. Qiao and T.-L. Liang, *Acta Crystallogr., Sect. E*, **62**, o3416 (2006).
62. S. P. Abraham, A. G. Samuelson and M. Nethaji, *Proc. Indian Acad. Sci., Chem. Sci.*, **107**, 255 (1995).
63. G. Blewett, M. W. Bredenkamp and K. R. Koch, *Acta Crystallogr., Sect. C*, **61**, o469 (2005).
64. L. A. Montiel-Ortega, S. Rojas-Lima, E. Otazo-Sanchez and R. Villagomez-Ibarra, *J. Chem. Crystallogr.*, **34**, 89 (2004).
65. A. Nangia, *Acc. Chem. Res.*, **41**, 595 (2008).
66. G. Bernardinelli, T. Berclaz, M. Geoffroy and N. Rajalakshmi, *Acta Crystallogr., Sect. C*, **51**, 1390 (1995).
67. H. N. De Armas, O. M. Peeters, N. Blaton, G. van den Mooter, D. J. A. De Ridder and H. Schenk, *J. Pharm. Sci.*, **95**, 2123 (2006).
68. G. Germain, J. P. Declercq, M. van Meerssche and M. H. J. Koch, *Acta Crystallogr., Sect. B*, **33**, 2269 (1977).
69. F. Tellez, A. Cruz, H. Lopez-Sandoval, I. Ramos-Garcia, M. Gayosso, R. N. Castillo-Sierra, B. Paz-Michel, H. Noth, A. Flores-Parra and R. Contreras, *Eur. J. Org. Chem.*, 4203 (2004).
70. H. Arslan, U. Florke and N. Kulcu, *Spectrochim. Acta, Part A*, **67**, 936 (2007).
71. B. Yang, S.-S. Zhang, J.-Z. Li, X.-M. Li and K. Jiao, *J. Chem. Soc. Pak.*, **27**, 223 (2005).
72. H.-L. Xiao, K.-F. Wang and F.-F. Jian, *Acta Crystallogr., Sect. E*, **62**, o2852 (2006).
73. S. J. Coles, D. Douheret, M. B. Hursthouse, J. D. Kilburn, and S. Rossi, *Acta Crystallogr., Sect. C*, **56**, e224 (2000).
74. B. Yang, S.-S. Zhang, Y.-F. Wang, X.-M. Li and K. Jiao, *Chem. Res. Chin. Univ.*, **21**, 163 (2005).

75. S.-S. Zhang, Z.-W. Wang, M. Li, K. Jiao, I. A. Razak, S. S. S. Raj and H.-K. Fun, *Acta Crystallogr., Sect. C*, **57**, 566 (2001).
76. G. Blewett, C. Esterhuysen, M. W. Bredenkamp and K. R. Koch, *Acta Crystallogr., Sect. C*, **60**, o862 (2004).
77. N. M. Blaton, O. M. Peeters and C. J. De Ranter, *Acta Crystallogr., Sect. B*, **35**, 753 (1979).
78. A. Nangia, *Cryst. Growth Des.*, **6**, 2 (2006).
79. P. Vishweshwar, J. A. McMahon and M. J. Zaworotko, in *Frontiers in Crystal Engineering*, E. R. T. Tiekink and J. J. Vittal (Eds), Chapter 2, John Wiley & Sons, Ltd, Chichester, 2006.
80. N. N. L. Madhavi, C. Bilton, J. A. K. Howard, F. H. Allen, A. Nangia and G. R. Desiraju, *New J. Chem.*, **24**, 1 (2000).
81. C. B. Aakeröy, B. M. T. Scott and J. Desper, *New J. Chem.*, **31**, 2044 (2007).

# 7

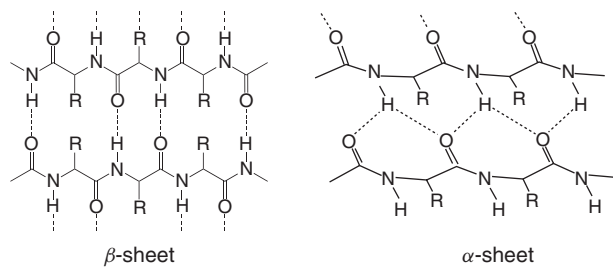
## Crystal Engineering with Molecules Containing Amide and Pyridine Functionalities

*Kumar Biradha and Lalit Rajput*

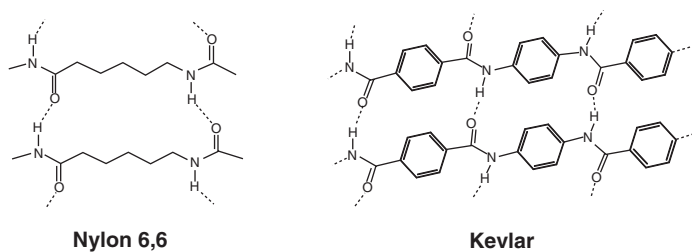
### 7.1 Introduction

Amide functional groups are ubiquitous in nature due to their abundant presence in small or complex synthetic, natural or biological molecules [1]. They also play a very important role in structural chemistry due to their ability to form complementary supramolecular synthons [2, 3]. The importance of N-H $\cdots$ O hydrogen bonds is well established in protein folding and conformations of nucleic acids. Some examples of hydrogen bonded networks which are formed in the secondary structure of proteins are depicted in Scheme 7.1 [4]. They also play an important role in synthetic polymers such as nylon 6,6, nylon 6 and Kevlar (Scheme 7.2).

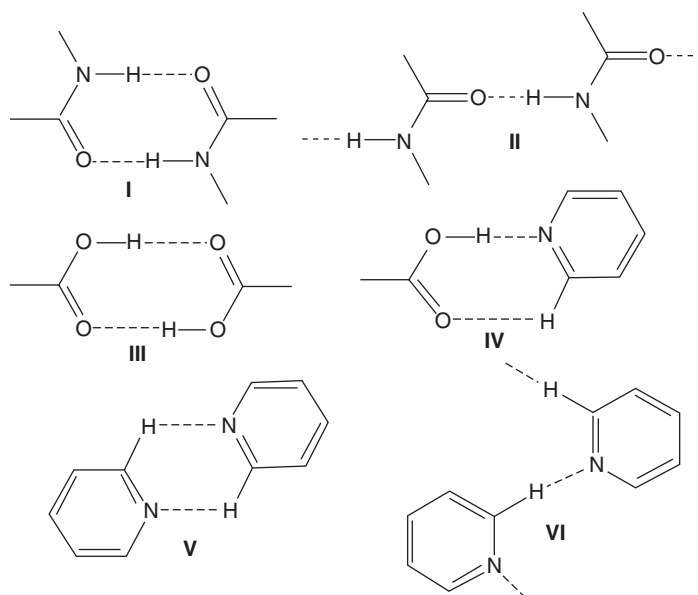
Successful strategies in crystal engineering rely on the building of crystalline solids from the functional group point of view. Functional groups such as carboxylic acid and amide are usually known to form robust synthons (**I**, **II** and **III**; Scheme 7.3) in the absence of other strong hydrogen bonding functional groups [5]. In drug molecules or in natural systems it is not possible to have exclusively –COOH or amide functional groups as one cannot avoid the presence of other functional groups. Therefore, for the study of the general systems it is important to have a knowledge of interactions between various functional groups. The change in the common pattern of a particular functional



**Scheme 7.1** Schematic diagram of a  $\beta$ -sheet and a  $\alpha$ -sheet



**Scheme 7.2** Schematic diagram of nylon 6,6 and Kevlar



**Scheme 7.3** Supramolecular synthons

group due to the presence of some other functional group in a molecule was referred by Desiraju as an interference of interactions [3a]. The understanding of interference effects is very important for the general applicability of crystal engineering. The factors that influence the interference are:

1. the matching of the strengths of donor and acceptor;
2. the cooperative effects of a synthon;
3. the balance between the number of donors and acceptors;
4. solvation effects; and
5. solvent inclusion or templation or interference.

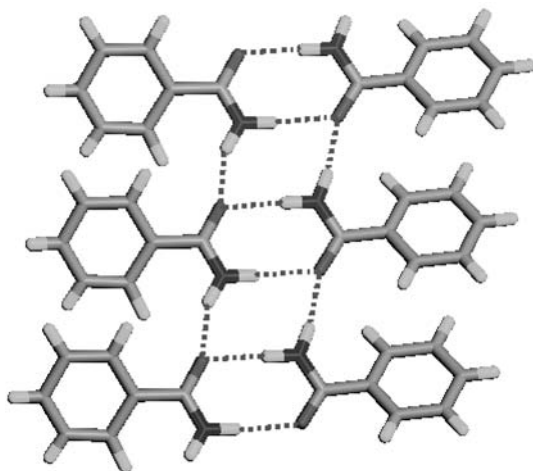
The matching of strengths of donors and acceptors was termed the hierarchy of interactions by Etter [6]. According to Etter's rules, the best donor pairs with the best acceptor and the next best donor pairs with the next best acceptor, and so on. This may be true in an ideal situation where the other four effects listed above are negligible. The cooperative effect deals with the coherence of the interactions which is not the strength of any one particular interaction but the combination of a series of interactions. The balance between the number of acceptors and donors is also very important. Solvation effects drive the crystallization process and have a major role in the final structural outcome. The inclusion or templation of solvents into the crystal lattice also significantly influences the structural outcome. Finally, the process of crystallization is the compromise of all the above effects. Therefore, a thorough understanding of each of these effects is necessary to predetermine the structural or functional properties of a material.

The amide group is known to form N-H...O synthons **I** and **II** in crystal structures [7]. However, pyridine only has N-acceptor, it does not have any strong donor moiety, and often forms weak C-H...N hydrogen bond synthons **V** and **VI** (Scheme 7.3). Therefore, pyridine competes with the other acceptors for strong proton donors. For example, the presence of a pyridine moiety can disrupt the conventional -COOH homosynthon (**III**) to form a heterosynthon (**IV**). Herein, we wish to present the interference of pyridine moieties in amide-to-amide hydrogen bonds by classifying the literature from a molecular structure point of view. Molecules having strong hydrogen bonding groups such as -OH, -NO<sub>2</sub> and -COOH in addition to amide and pyridine groups are not considered. The molecular entities covered in this chapter include mono-, bis- and tris-amidopyridines and their co-crystals. The Cambridge Structural Database (CSD) was used for fragment based searches and data retrieval [8].

## 7.2 Primary Amides Containing a Pyridine Moiety

In a primary amide the acceptors and donors are present in a ratio of 1:2. However, in the crystal lattice the amide carbonyl accepts two protons and forms an acceptor bifurcated interaction. Therefore, the introduction of pyridine into these systems balances the donor and acceptor ratio and the formation of both N-H...O and N-H...N hydrogen bonds in the resultant crystal structure is anticipated. Before discussing pyridine-containing amides, the crystal structure of benzamide is described. It has three polymorphs and in all three polymorphs the strong hydrogen bonding pattern is unique [9]. Each of the crystal structures contains 1-D aggregation which occurs through the cooperation of



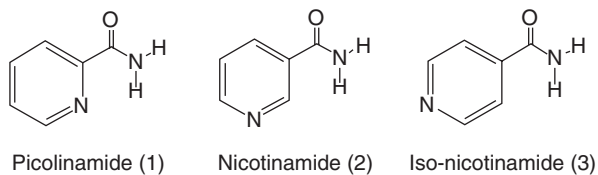


**Figure 7.1** Hydrogen bonding in the crystal structures of polymorphic benzamide

synthons **I** and **II** (Figure 7.1). The acceptor carbonyl is bifurcated by both the N-H groups.

### 7.2.1 Amides directly linked to a pyridine moiety

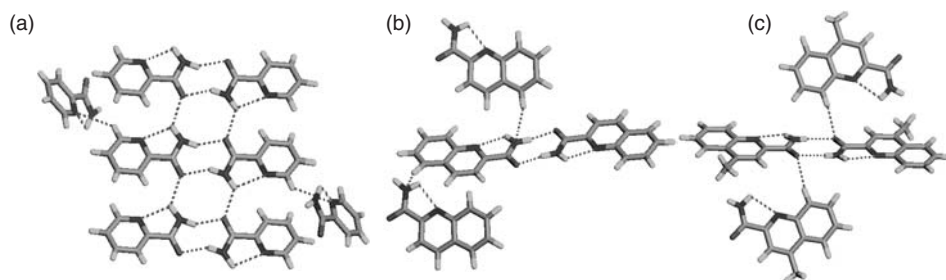
There are three prototype compounds that fall into this category, namely 2-picolinamide, nicotinamide and iso-nicotinamide. In the following sections each class of compounds will be presented separately.



#### 7.2.1.1 Picolinamide based systems

In the crystal structure of **1**, the amide N-H group, which is in the anti configuration to C=O, participates in an intramolecular N-H...N hydrogen bond and further links the molecules via synthon **II** (Figure 7.2a). The *syn*-N-H group of the amide participates via synthon **I** to form a molecular ribbon [10a]. Apart from the intramolecular hydrogen bond the structure bears a close resemblance to that of benzamide.

These interaction patterns do not persist in the crystal structures of the closely related quinoline derivatives [10b, c]. The larger groups such as quinoline (**4**) and methyl-substituted quinoline (**5**) disrupt the formation of synthon **II** (Figure 7.2b and c). Therefore, the *anti*-N-H group is involved in only an intramolecular N-H...N contact, and the *syn*-N-H group forms synthon **I** as usual. Although **4** and **5** are similar in terms of synthon **I**, the weak hydrogen bonding pattern differs due to the methyl substitution on the quinoline moiety in **5**. In the case of **4**, the C-H...N contact is responsible for

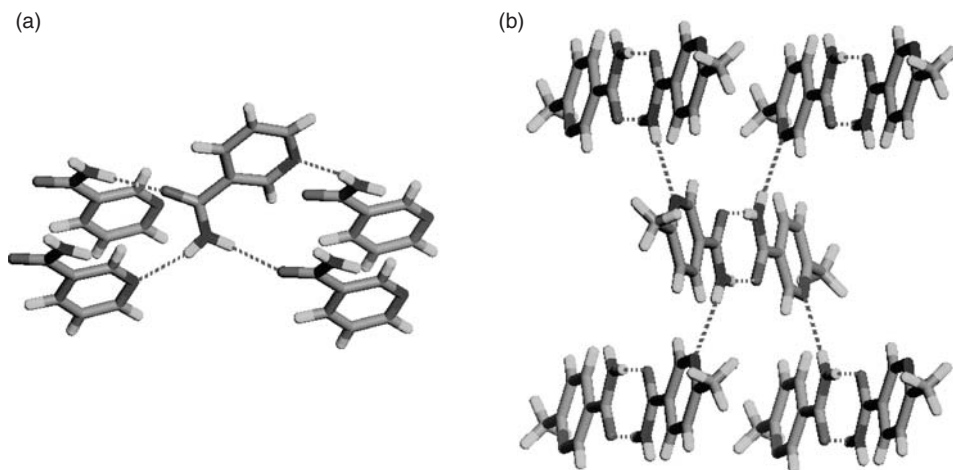


**Figure 7.2** Hydrogen bonding patterns in **1** (a), and related quinoline derivatives **4** (b) and **5** (c)

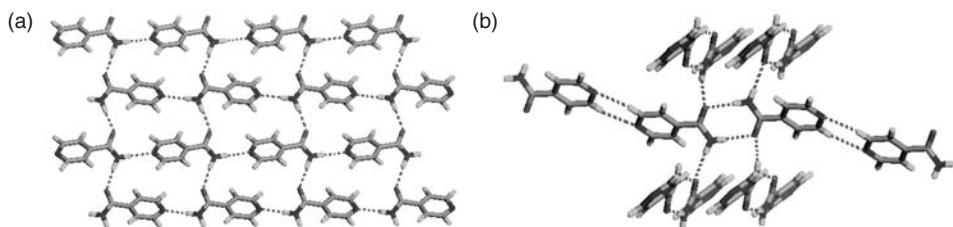
further molecular packing whereas in the case of **5**, C-H...O contacts serve to aggregate the amide dimers. In terms of molecular geometry, all three compounds have the amide-NH<sub>2</sub> close to the pyridine nitrogen due to the prevalent intramolecular N-H...N hydrogen bond. Therefore, these structures indicate that the interference from 2-pyridyl substitution in amide-to-amide hydrogen bonds is minimal.

#### 7.2.1.2 Nicotinamide based systems

Crystal structures for two compounds are known in this category, one is nicotinamide (**2**) and the other is methyl-substituted nicotinamide (**6**) [11]. Obviously, no intramolecular N-H...N hydrogen bond was found in these derivatives. As a result, the molecules deviate from planarity with an angle between the pyridyl and amide groups of 22.4 and 6.8° in **2** and **6**, respectively. The difference in molecular geometry is reflected in the supramolecular aggregation. Both structures contain an intermolecular N-H...N hydrogen bond due to the pyridine interference (Figure 7.3). However, they differ in terms of N-H...O hydrogen bonds: structure **2** contains synthon **II** whereas structure **6**



**Figure 7.3** Hydrogen bonding patterns in the crystal structures of (a) nicotinamide (**2**) and (b) 6-methyl-nicotinamide (**6**)



**Figure 7.4** Hydrogen bonding patterns in the polymorphic crystal structures of iso-nicotinamide **6**: (a) form **3A** and (b) form **3B**. Note that form **3A** contains synthon **II** and form **3B** contains synthon **I**

contains synthon **I**. Therefore, unlike 2-pyridyl substitution in isomeric **1**, the presence of 3-pyridine in **2** displays partial interference in amide-to-amide synthons.

#### 7.2.1.3 Iso-nicotinamide based systems

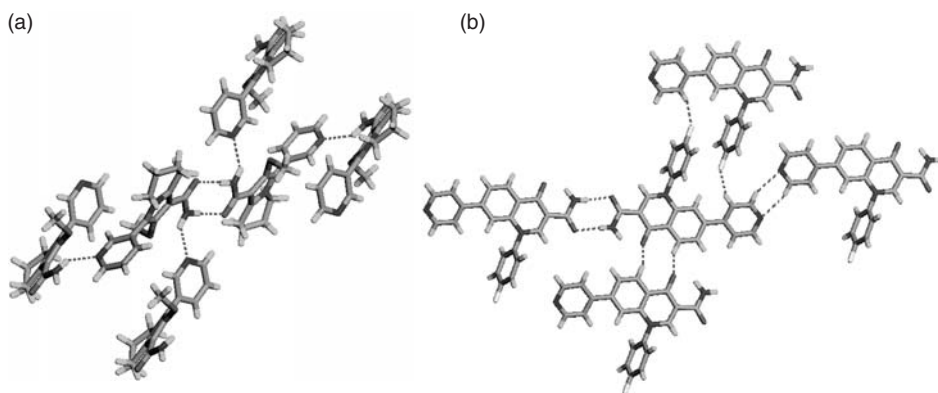
Iso-nicotinamide (**3**) exists in two polymorphic forms (**3A** and **3B**) [12]. Form **3A** was crystallized from a variety of solvents such as ethanol, water, THF, etc. as plate-like crystals whereas form **3B** was crystallized from nitrobenzene or nitromethane as rod-like crystals. In form **3A** the molecules are assembled into a 2-D sheet via synthon **II** and N-H...N hydrogen bonds (Figure 7.4). It is interesting to note that the hydrogen bonding components in form **3A** are similar to that of nicotinamide (**2**). However, the crystal structure of form **3A** contains a flat layer while that of **2** contains a highly corrugated layer. Interestingly, in form **3B** pyridine does not interfere in amide-to-amide hydrogen bonding and exhibits both synthons **I** and **II**. The pyridine moiety forms C-H...N synthon **V**. The crystal structure of form **3B** resembles that of picolinamide (**1**) as both structures contain synthons **I** and **II** as well as weak C-H...N contacts.

### 7.2.2 Primary amide and pyridine linked by spacer units

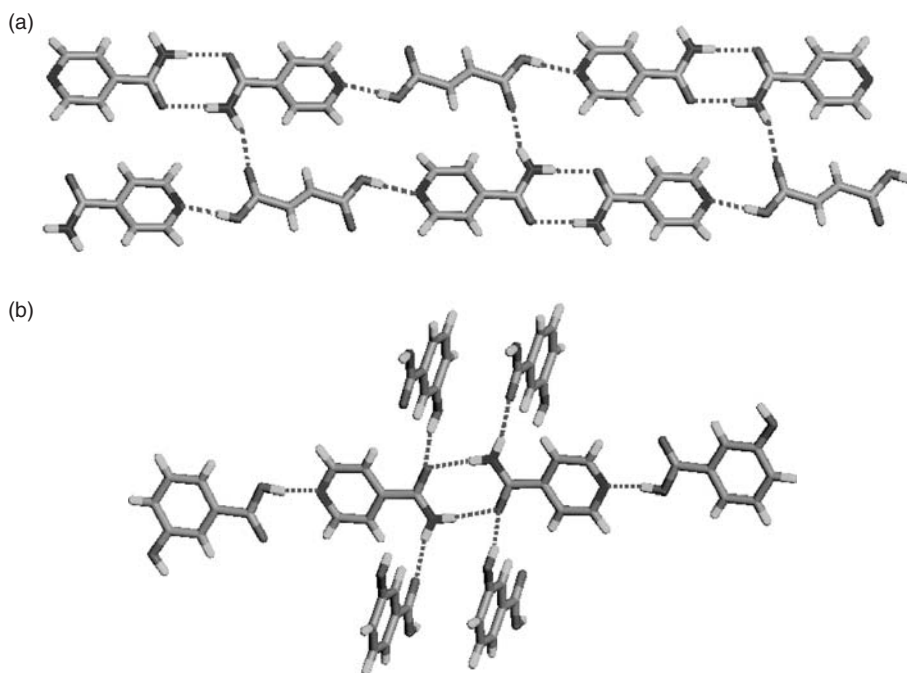
Only two crystal structures of such molecules are known, one with 3-pyridyl (**7**) and the other with 4-pyridyl (**8**) [13]. Both of these molecules, in addition to pyridine and primary amide functional groups, also contain two more functional groups such as methoxy and tertiary amine. The crystal structures exhibited amide-to-amide synthon **I** (Figure 7.5). The pyridine is involved in an N-H...N hydrogen bond only in **7** whereas in **8** it also participates in a C-H...N contact, synthon **V**. In **8**, the second N-H is involved in an intramolecular hydrogen bond with the adjacent keto group.

## 7.3 Co-crystals with Primary Amidopyridines

Aakeröy and co-workers and Nangia and co-workers were successful in the exploitation of Etter's hydrogen bonding hierarchy rule to design binary and ternary complexes of nicotinamide (**2**) and iso-nicotinamide (**3**) with carboxylic acids and phenols [14]. Synthon **I** was retained in iso-nicotinamide co-crystals with succinic acid and 3-hydroxybenzoic acid (Figure 7.6) [14d]. In these structures the carboxylic acid interacts, as a strong proton donor, with a strong proton acceptor such as pyridine.

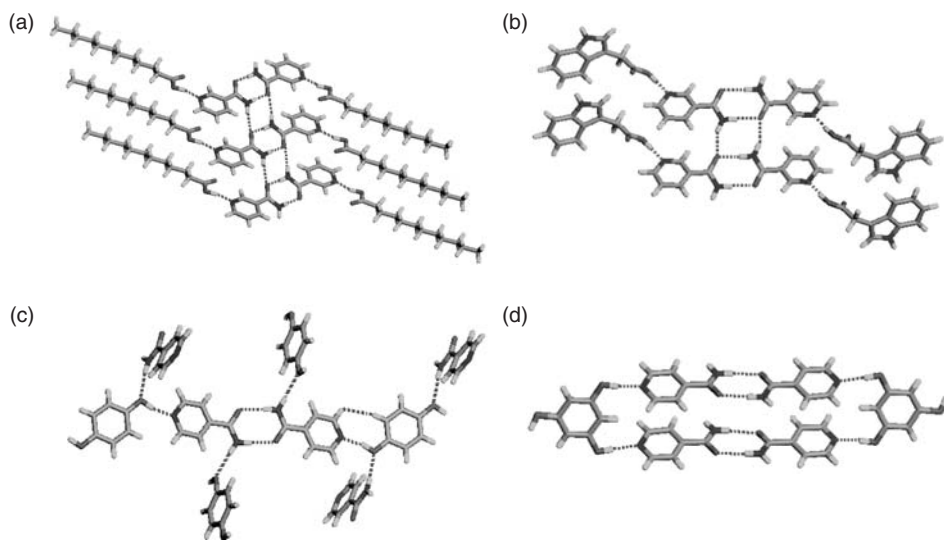


**Figure 7.5** Hydrogen bonding patterns in the crystal structures of (a) **7** and (b) **8**; both contain synthon **I**



**Figure 7.6** Hydrogen bonding patterns in the co-crystals of iso-nicotinamide (**3**) with (a) succinic acid and (b) 3-hydroxybenzoic acid

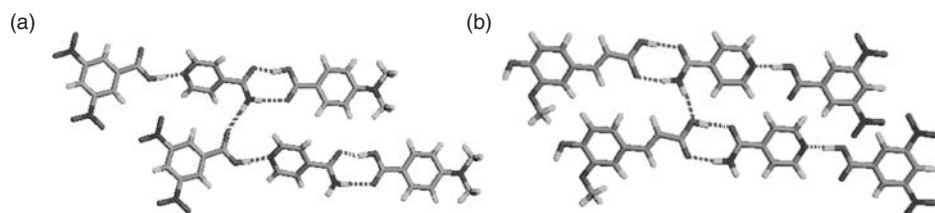
Similarly, nicotinamide was shown to form co-crystals with dodecanoic acid and indole acetic acid, and iso-nicotinamide with phloroglucinol and hydroquinone (Figure 7.7) [14b, c, e]. All these structures contain synthon **I**. The  $\text{-COOH}$  or  $\text{-OH}$  groups interact with the pyridine moiety and also with the second N-H group of amide which is not a part of synthon **I**. All these structures indicate that the acid-pyridine synthon is favoured over amide-acid according to the hierarchy of the interactions.



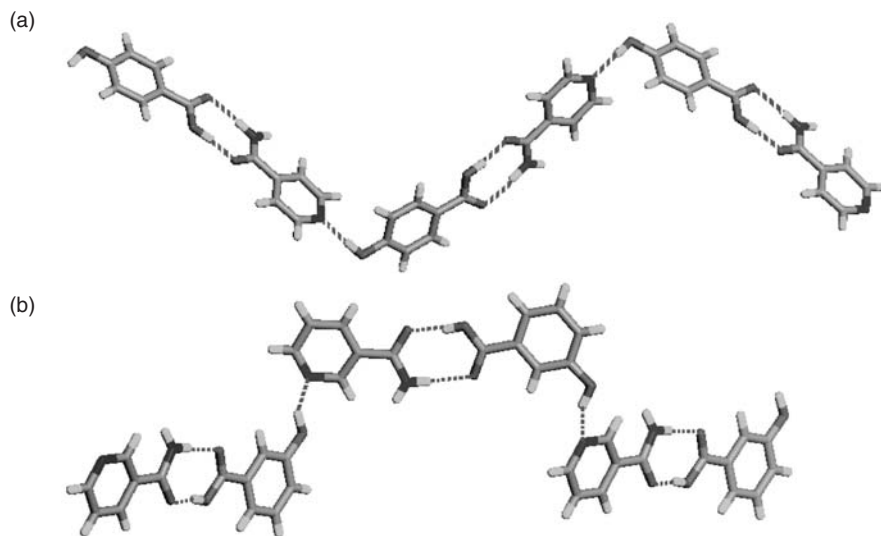
**Figure 7.7** Hydrogen bonding patterns in the co-crystals of iso-nicotinamide (**3**) with (a) dodecanoic acid, (b) indole acetic acid, (c) hydroquinone and (d) phloroglucinol

Thus, the  $\text{-COOH}$  group, despite giving a strong hydrogen donor, does not disrupt the amide recognition as it prefers to interact with the strong acceptor such as pyridine. This aspect has been proved even in ternary crystals of iso-nicotinamide (**3**). In these studies the authors considered two different carboxylic acids with varied donor strengths to co-crystallize with **3**. In this ternary crystal, as expected, the more acidic  $\text{-COOH}$  (e.g. 3,5-dinitrobenzoic acid) interacted with the more basic pyridine functionality while the less acidic  $\text{-COOH}$  (4-(dimethylamino)benzoic acid or 4-hydroxy-3-methoxybenzoic acid) interacted with the amide functionality (Figure 7.8) [14a]. Synthon **I** is disrupted as there are two carboxylic acids per iso-nicotinamide.

However, there exist several examples in the literature which do not follow the expected hierarchy of interactions. This is true even for the iso-nicotinamide and nicotinamide co-crystals with 4-hydroxybenzoic acid and 3-hydroxybenzoic acid, respectively. In these structures the amide recognition synthon **I** was disrupted by the amide-acid synthon (Figure 7.9) [14e, g]. The relatively weak donor ( $\text{-OH}$ ) interacts with the strong



**Figure 7.8** Hydrogen bonding patterns in the ternary co-crystals of iso-nicotinamide (**3**) with (a) 3,5-dinitrobenzoic acid and 4-(dimethylamino)benzoic acid and (b) 3,5-dinitrobenzoic acid and 4-hydroxy-3-methoxybenzoic acid

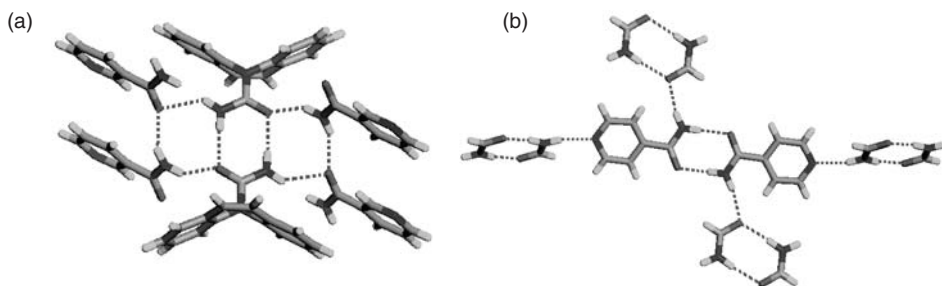


**Figure 7.9** Hydrogen bonding patterns in the crystal structures of co-crystals of iso-nicotinamide (**3**) with (a) 4-hydroxybenzoic acid and (b) 3-hydroxybenzoic acid

acceptor (pyridine) while the strong donor ( $-\text{COOH}$ ) forms a heterosynthon with the amide functional group. These results indicate that the crystallization experiments do not necessarily obey the hierarchy of interactions or the complementarities of the components. As discussed in Section 7.1, the final outcome of the crystallization depends on several other factors such as solvent, dilution, rate of evaporation and solvation of the components.

### 7.3.1 Combination of amides for co-crystallization

Co-crystals of nicotinamide and iso-nicotinamide have been prepared with carbamazepine and formamide, respectively [15]. In these co-crystals it is observed that the homodimeric synthon **I** of amides is always favoured despite there being several other possible hydrogen bonding synthons (Figure 7.10). The homodimers thus formed interact with each other via hydrogen bonding synthon **II**.



**Figure 7.10** Hydrogen bonding patterns in the co-crystals of two amide derivatives: (a) nicotinamide and carbamazepine; and (b) iso-nicotinamide and formamide



## 7.4 Secondary Amides Containing a Pyridine Moiety

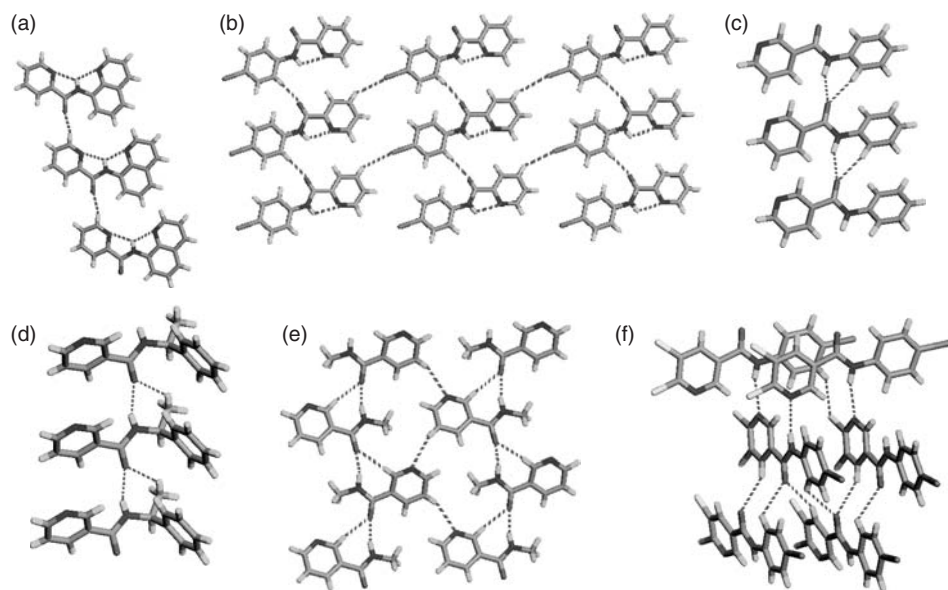
### 7.4.1 Monoamide and monopyridine derivatives

Secondary amides contain one acceptor and one donor each. Therefore, the presence of an additional acceptor site such as pyridine-nitrogen is expected to interfere with amide-to-amide hydrogen bonds. In the case of primary amidopyridines, the presence of pyridine balances the donor–acceptor ratio whereas here it destroys the balance. Having fewer acceptors and more donors (primary amide) is a better situation than vice versa as the acceptors can form a bifurcated interaction more frequently than can donor hydrogen atoms. In the case of secondary amidopyridines, one acceptor has to sacrifice its strong hydrogen bond for the sake of the other acceptor. In such instances it is interesting to see which acceptor ‘wins the race of strong hydrogen bonds’.

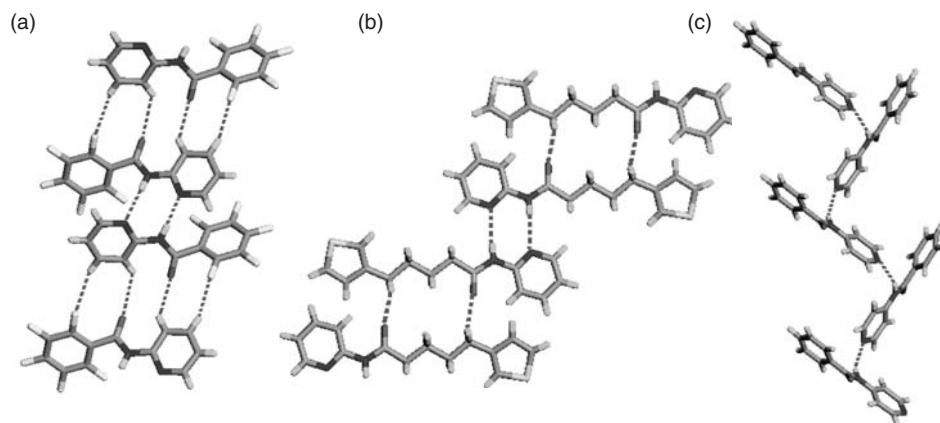
These amides can be categorized into two types based on the connection of the pyridyl moiety to the amide functionality: (1) C-terminal (amides), when pyridine is connected to the C-atom of the amide; and (2) N-terminal (reverse amides), when pyridine is connected to the N-atom of the amide. These two types are found to have different characteristics and therefore are discussed separately.

#### 7.4.1.1 C-Terminal amides

Two crystal structures are known with 2-pyridyl functionality [16]. In both structures, intermolecular N-H...N and N-H...O hydrogen bonds are absent due to geometrical constraints (Figure 7.11). The N-H group forms an intramolecular hydrogen bond with



**Figure 7.11** Hydrogen bonding patterns in C-terminal amides of 2-pyridyl (a and b) and 3-pyridyl derivatives (c–f)



**Figure 7.12** Hydrogen bonding patterns in the crystal structures of 2-pyridyl substituted *N*-terminal amides: (a and b) *N*-H...*N* dimers aggregated further by *C*-H...*O* hydrogen bonds; (c) 1-D *N*-H...*N* hydrogen bonded chain

the pyridine-nitrogen atom and C=O forms intermolecular *C*-H...*O* contacts. In the 3-pyridyl derivatives, *N*-H...*O* hydrogen bonds are favoured over *N*-H...*N* hydrogen bonds (3 out of 4) [17]. In the case of the 4-pyridyl derivatives, only one structure is known and it also prefers *N*-H...*O* aggregation over *N*-H...*N* aggregation [18]. The data presented here clearly indicate the preference for *N*-H...*O* over *N*-H...*N* hydrogen bonds.

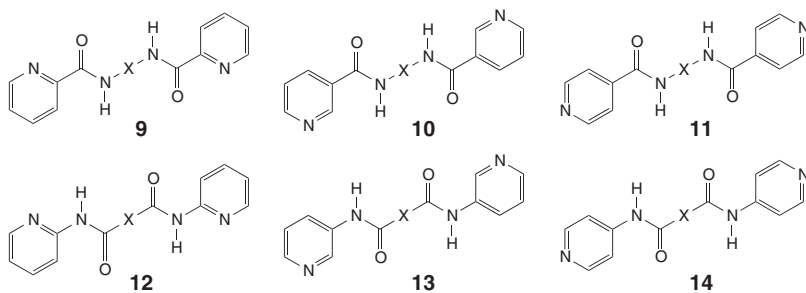
#### 7.4.1.2 *N*-Terminal amides

The crystal structures of 2-pyridyl substituted *N*-terminal amides, unlike the aforementioned *C*-terminal amides, exhibit a recurring *N*-H...*N* cyclic synthon which leads to the formation of dimers [19]. These dimers aggregate further via *C*-H...*O* contacts between the amide-carbonyl and aromatic *C*-H group (Figure 7.12). No 3-pyridyl structures of this category are known yet, and only one structure was reported for a 4-pyridyl derivative [20]. In this structure, *N*-H...*N* hydrogen bonds are preferred over *N*-H...*O* hydrogen bonds.

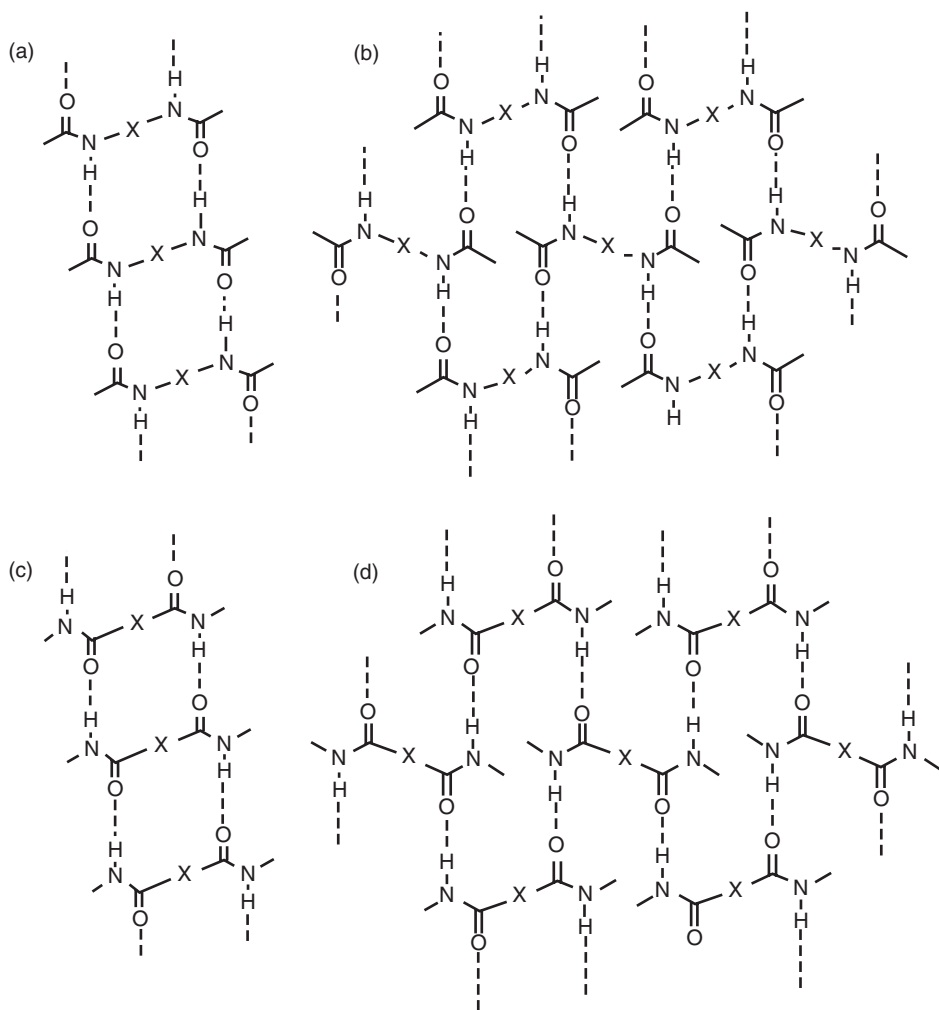
## 7.5 Bis-Amidopyridine Derivatives

The crystal structures of the homologous series of amides **9–11** and reverse amides **12–14** have been explored recently to address important factors such as isostructurality, interference of pyridine functionality in amide-to-amide hydrogen bonds, and the role of the spacer between the amide groups. The phenyl analogues of these derivatives have been identified to form two patterns such as a  $\beta$ -sheet and a 2-D layer via amide-to-amide hydrogen bonds (Scheme 7.4) [21]. The longer spacers such as  $-(\text{CH}_2)_6-$ ,  $-(\text{CH}_2)_8-$  and  $-\text{C}_6\text{H}_4-$  favoured  $\beta$ -sheet motifs and shorter spacers such as  $-(\text{CH}_2)_2-$  and  $-(\text{CH}_2)_4-$  favoured 2-D layers. The pyridine containing derivatives will be discussed in the following two sections by segregating them in to amides (**9–11**) and reverse amides (**12–14**).





X = a) no spacer; b)  $(\text{CH}_2)_2$ ; c)  $(\text{CH}_2)_4$ ; d)  $(\text{CH}_2)_6$ ; e)  $(\text{CH}_2)_8$ ; f)  $(\text{C}_6\text{H}_4)$ , g) adamantyl



**Scheme 7.4**  $\text{N-H} \cdots \text{O}$  hydrogen bonded  $\beta$ -sheet and 2-D layers in C-terminal amides (a and b) and N-terminal amides (c and d)

### 7.5.1 C-Terminal bis-amidopyridines

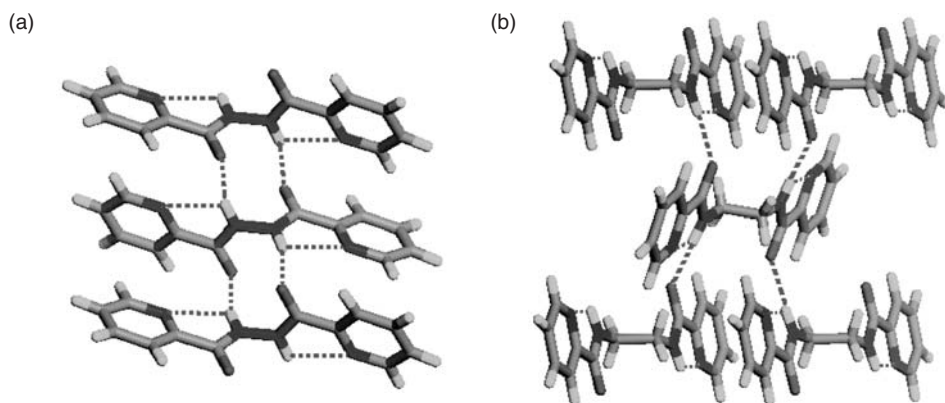
#### 7.5.1.1 2-Pyridyl based systems

Only two crystal structures are known in this category, one without a spacer (**9a**) and one with a spacer  $X = -(CH_2)_2-$  (**9b**) [22]. The molecules in the crystal structure of **9a** are assembled through amide-to-amide hydrogen bonds to form a  $\beta$ -sheet-like pattern (Figure 7.13). The second structure also contains amide-to-amide hydrogen bonds but results in a 2-D layer. In both structures the pyridine group is syn to the amide N-H group and therefore facilitates the intramolecular N-H...N hydrogen bond. Further, due to the donor bifurcated hydrogen bonds both the acceptors are involved in strong hydrogen bonds.

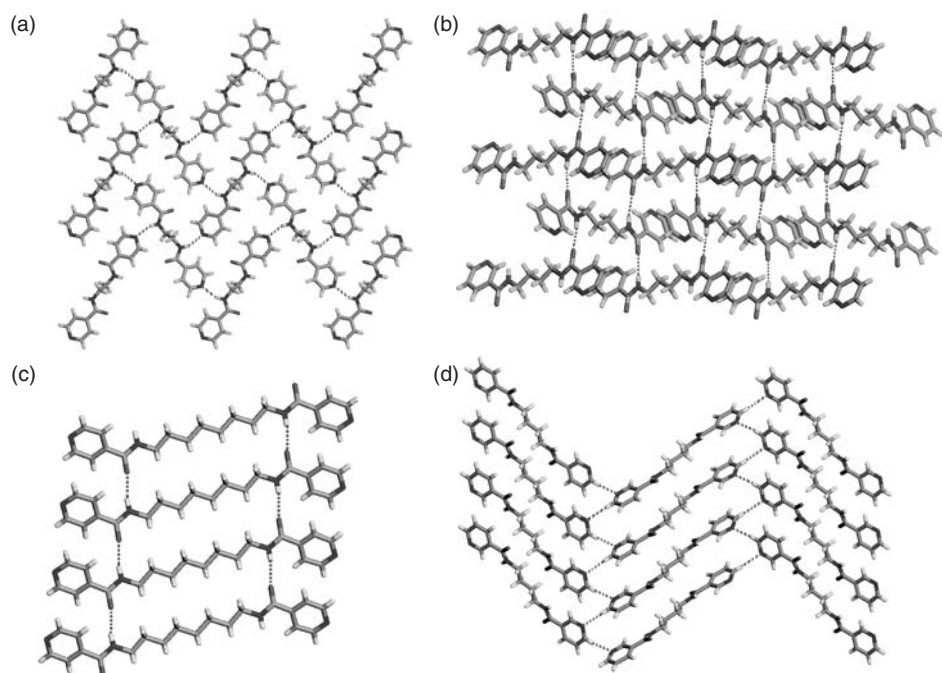
#### 7.5.1.2 3-Pyridyl and 4-pyridyl based systems

These derivatives were found to prefer two patterns, i.e. a  $\beta$ -sheet or a 2-D layer via amide-to-amide hydrogen bonds [23]. The interference of the pyridine group in amide-to-amide hydrogen bonds was found in only one structure, i.e. with 4-pyridyl substitution and  $X = -(CH_2)_2-$  (**11b**). In the crystal structure of **11b**, the pyridine group engages in N-H...N hydrogen bonds to form a herringbone layer and the amide-carbonyl forms weak C-H...O interactions (Figure 7.14). In the remaining derivatives, similar to the phenyl analogues, the small spacers such as  $-(CH_2)_2-$  and  $-(CH_2)_4-$  prefer 2-D layers whereas the longer spacers such as  $-(CH_2)_6-$ ,  $-(CH_2)_8-$  and  $-(C_6H_4)-$  prefer to form a  $\beta$ -sheet. The pyridine groups are involved in weak C-H...N contacts in all derivatives.

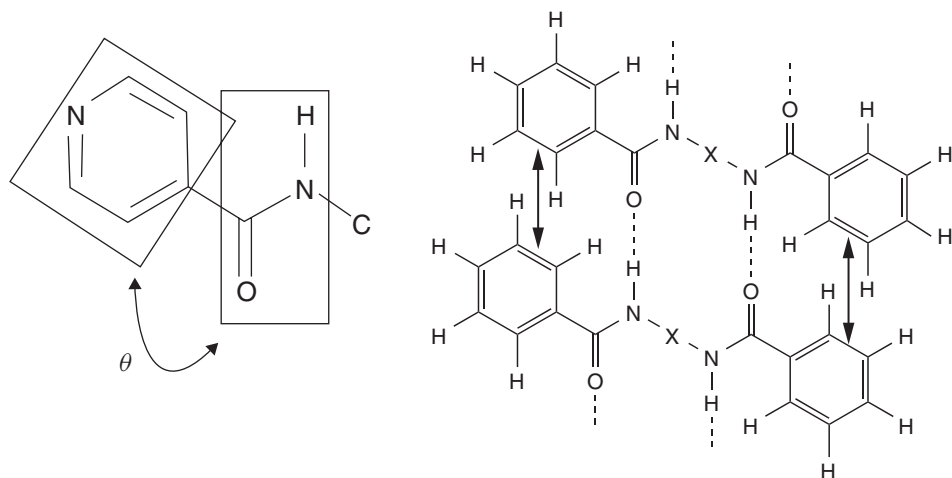
A close examination of the molecular geometries reveals that the interplanar angle ( $\theta$ ) between the amide and pyridine groups is important in these derivatives. When the  $\theta$  value is below  $20^\circ$  N-H...N hydrogen bonding results whereas above  $20^\circ$  N-H...O hydrogen bonding results. The probable reason for this preference could be that a  $\theta$  value less than  $20^\circ$  results in steric repulsions between the adjacent pyridyl groups of amide-to-amide supramolecular aggregates (Scheme 7.5). From the point of view of the molecular geometry, most of these derivatives prefer to have a  $\theta$  value above  $20^\circ$  in



**Figure 7.13** Hydrogen bonding patterns in the crystal structures of (a) **9a** and (b) **9b**



**Figure 7.14** Hydrogen bonding patterns in the crystal structures of C-terminal amides: (a) **11b**; (b) **10c**; (c) **11e**; and (d) C-H $\cdots$ N hydrogen bond layer in the crystal structure of **10c**



**Scheme 7.5** (a) Interplanar angle  $\theta$  between the planes of amide and pyridine and (b) possible repulsions between the in-plane aromatic moieties

order to minimize steric repulsions with the  $\alpha$ -CH of the pyridine residue. Therefore, the adopted molecular geometry favoured the formation of amide-to-amide hydrogen bonds over the amide-to-pyridine hydrogen bonds.

## 7.5.2 N-Terminal bis-amidopyridines (reverse amides)

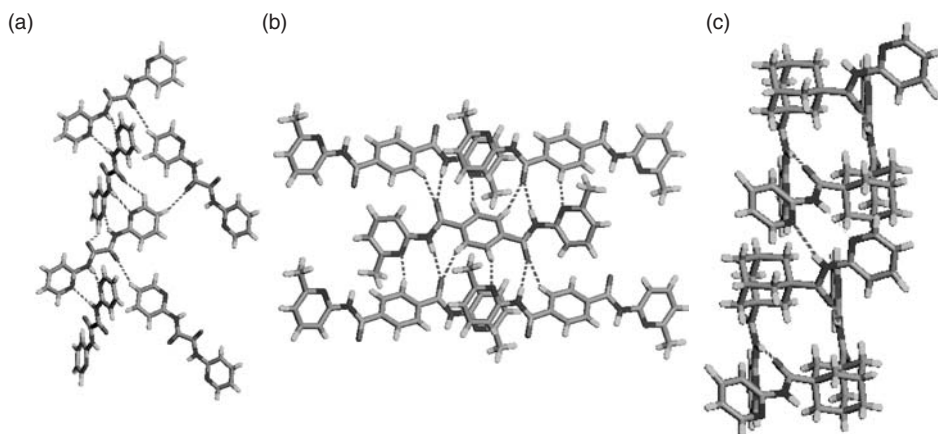
### 7.5.2.1 2-Pyridyl based systems

Three crystal structures come into this category and they are: one without a spacer (**12a**) and two with spacers,  $-\text{C}_6\text{H}_5-$  (**12f**) and adamantyl (**12g**) [24]. All three structures were found to have different features (Figure 7.15). The crystal structure of **12a** contains a 1-D pattern via N-H...N hydrogen bonding with the amide-carbonyl participating in weak C-H...O contacts. In **12f**, the molecules aggregate to form a similar N-H...O hydrogen bonded 2-D layer as found for the C-terminal monoamidopyridines, as shown in Figure 7.11. Finally, the structure containing the adamantyl moiety (**12g**) contains both N-H...O and N-H...N hydrogen bonds. This sampling suggests that the type of spacer is very important in controlling these molecular aggregations.

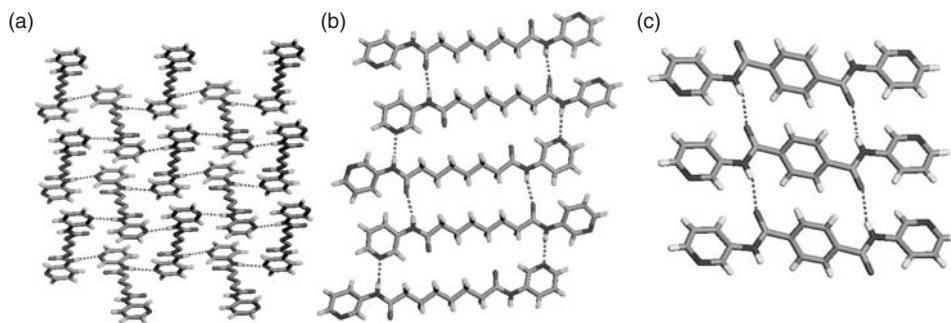
### 7.5.2.2 3-Pyridyl based systems

Four crystal structures of molecules with aliphatic spacers (**13b–e**) and two crystal structures of molecules with the aromatic spacer  $-(\text{C}_6\text{H}_4)-$ , having 1,4- (**13f-1**) and 1,3- (**13f-2**) substitution, have been reported [25]. In the cases of **13b**, **13c** and **13f-2**, amide-to-amide hydrogen bonds are not observed as the amide N-H is involved in N-H...N hydrogen bonds with pyridine-nitrogen to form a corrugated 2-D layer which has similar connectivity to the N-H...O 2-D layer (Scheme 7.4, Figure 7.16). The amide-carbonyl is involved in C-H...O contacts with neighbouring layers.

The crystal structures of molecules **13d** and **13e** were found to form an isostructural hydrogen bonding pattern which resembles a  $\beta$ -sheet. However, the conventional  $\beta$ -sheet contains all N-H...O hydrogen bonds, whereas in the structures observed here,



**Figure 7.15** Hydrogen bonding patterns in N-terminal bis-amido-2-pyridines: (a) **12a**; (b) **12f**; and (c) **12g**

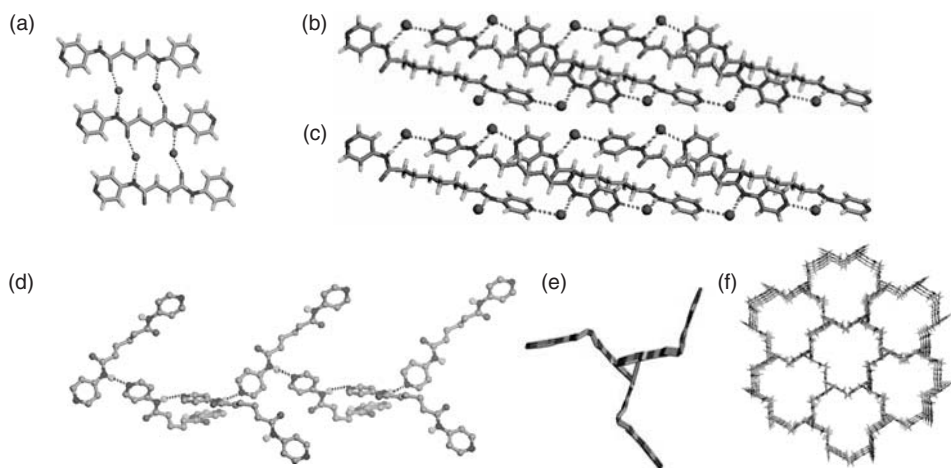


**Figure 7.16** Hydrogen bonding patterns in N-terminal bis-amido-3-pyridines: (a) 2-D layer via N-H...N hydrogen bonds in **13b**; (b)  $\beta$ -sheet pattern via N-H...N and N-H...O hydrogen bonds in **13d**; and (c)  $\beta$ -sheet via N-H...O hydrogen bonds in **13f**

the sheets contain N-H...O as well as N-H...N hydrogen bonds. The N-H...O and N-H...N dimeric motifs are arranged alternately in the new  $\beta$ -sheet. In the crystal structure of **13f-1**, the molecules are assembled via amide-to-amide hydrogen bonding to form a conventional  $\beta$ -sheet. The pyridyl N-atoms were not engaged in any significant hydrogen bonding interactions.

#### 7.5.2.3 4-Pyridyl based systems

The molecules **14b**, **14d** and **14e** have included water molecules in their respective crystal structures. The crystal structure of **14b** contains a  $\beta$ -sheet-like pattern in which amide N-H and carbonyl groups are joined via water molecules (Figure 7.17a–c). However, tubular structures are observed in the crystal structures of **14d** and **14e**. These three structures differ significantly from the rest of the structures owing to the presence of lattice water.



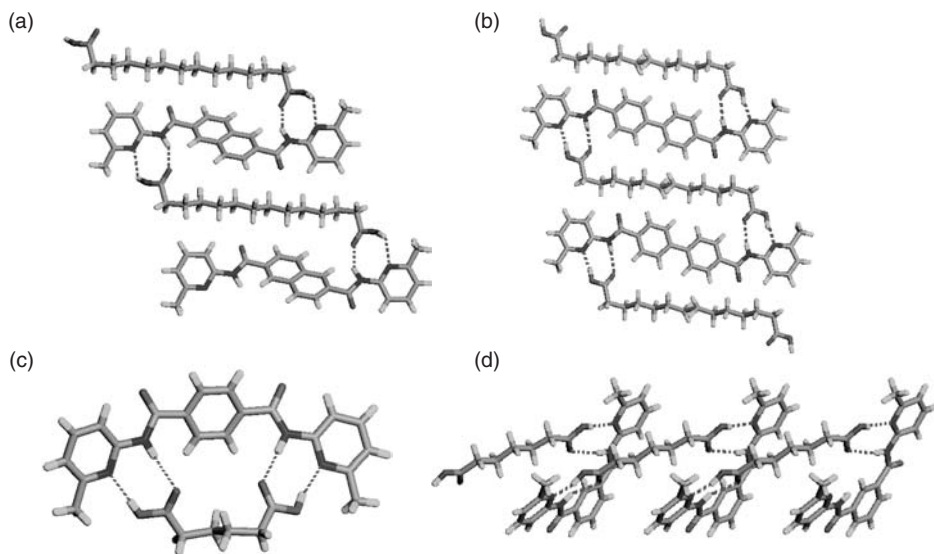
**Figure 7.17** Hydrogen bonding patterns in N-terminal bis-amido-4-pyridines: (a)  $\beta$ -sheet-like pattern in **14b**; tubular structures in (b) **14d** and (c) **14e**. The N-H...N hydrogen bonded helix: (d) side view; (e) top view; and (f) 3-D network in **14e**

The crystal structure of **14e** reveals that the molecular geometry exhibits an arc-like shape and not the usual linear geometry. The amide N-H and pyridine are engaged in N-H...N hydrogen bonds to form threefold helices. These threefold helices are further linked via N-H...N hydrogen bonds into a 3-D network (Figure 17d–f) [25a]. The 3-D network observed here is doubly interpenetrated and has a quartz-like topology [25b–f].

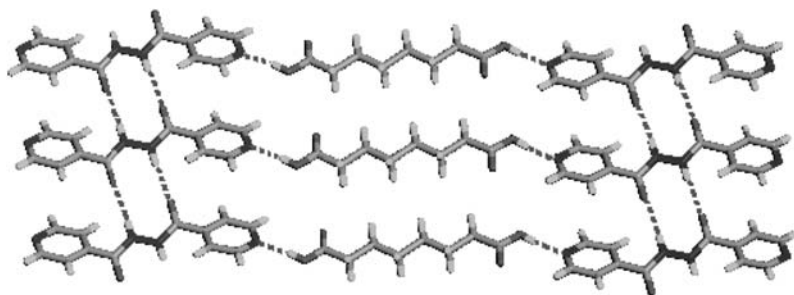
Unlike the amides, the reverse amides are found to prefer  $\theta$  values less than  $20^\circ$  in most of the structures discussed herein. The preference for a planar geometry can be well understood as there is no steric hindrance between amide N-H and  $\alpha$ -CH of the pyridine group. As discussed earlier, the planar geometry of the molecule does not allow amide-to-amide hydrogen bond formation. Accordingly, the reverse amides, except for one structure, all contain N-H...N hydrogen bonding. These studies prove that over and above the hierarchy of the interaction, geometric requirements play a major role in the final outcome of the crystal packing.

## 7.6 Two-component Structures Containing Secondary Amides and Pyridine Derivatives

The homomeric N-H...N synthon observed in the crystal structures of N-terminal amido-2-pyridine derivatives was shown to be disturbed in favour of the hetero dimer with  $\text{-COOH}$ , when present [24b]. Several aliphatic dicarboxylic acids were co-crystallized with the derivatives of **12**. The flexible diacids can adopt either convergent or divergent conformations to form 0-D aggregates or 1-D architectures (Figure 7.18) [24b]. However, the 4-pyridyl-substituted C-terminal amide **11a**, when crystallized with suberic acid, showed the acid linker does not interfere in amide-to-amide hydrogen bonds and it links the  $\beta$ -sheets into 2-D layers (Figure 7.19) [26].



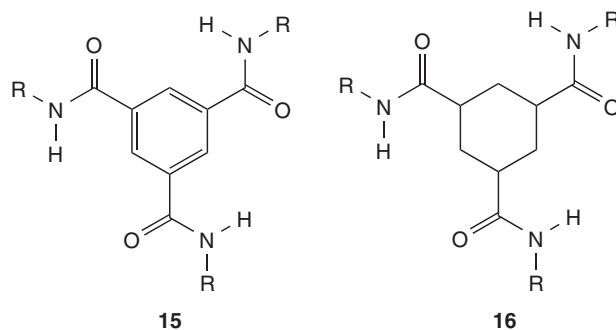
**Figure 7.18** Hydrogen bonding patterns leading to 1-D (a,b,d) and 0-D (c) aggregates in the co-crystals of N-terminal bis-2-amidopyridines with diacids



**Figure 7.19** Hydrogen bonding pattern in the co-crystal of C-terminal bis-amido-4-pyridine **11a** with suberic acid

## 7.7 Triamidopyridine Derivatives

Symmetrically substituted triamidopyridine derivatives will be described in this section (**15** and **16**). The crystal structure of triamide **15** was shown to form a triple helix via N-H...O hydrogen bonds when R =  $-\text{CH}_2\text{CH}_2\text{OMe}$  (Figure 7.20) [27]. However, the possibility of such aggregation may be perturbed when R groups contain pyridyl residues due to the possibility of N-H...N hydrogen bonding.

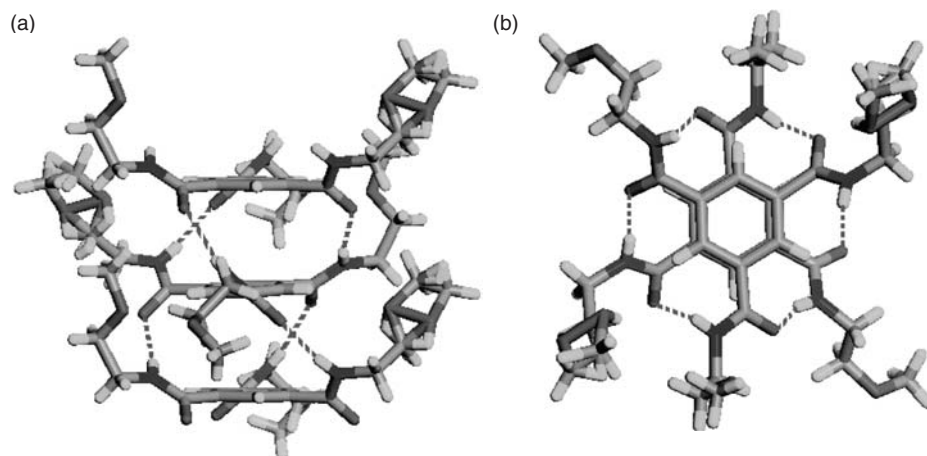


### 7.7.1 2-Pyridyl based systems

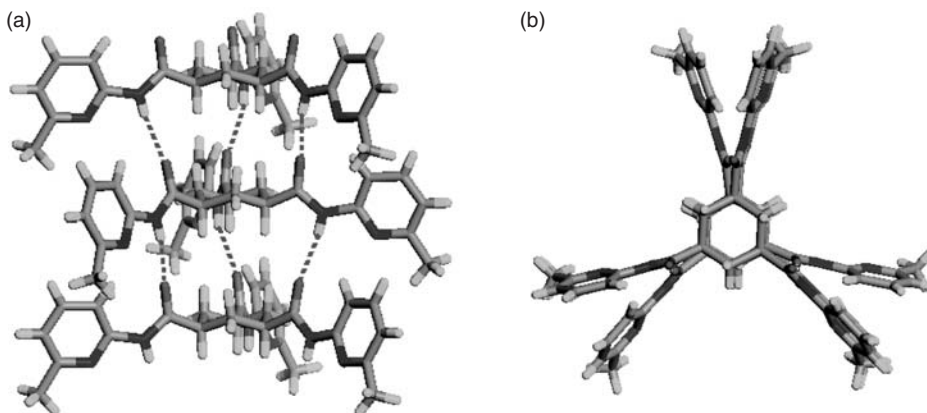
Three crystal structures are known with 2-pyridyl substitution, two with fragment **15** and one with fragment **16** [28]. The crystal structure of **17** (**16** with R = methyl-substituted pyridine) is exclusively constituted by N-H...O hydrogen bonds. The pyridine does not disrupt these. The N-H...N hydrogen bonded homosynthon which is found in related derivatives is not observed here. The N-H...O hydrogen bond pattern leads to the formation of a columnar architecture (Figure 7.21). The pyridine N-atom is involved in weak C-H...N contacts with the aromatic C-H groups of neighbouring molecules. Although this structure is not similar to the one described above (Figure 7.20), the difference is small as one is a helical column and the other is nonhelical.

The two derivatives of **15**, i.e. with R = 5-methyl-2-pyridine (**18**) and R = 3-ethyl-2-pyridine (**19**), totally deviate from the above structure, with the 2-pyridyl group interfering significantly with the N-H...O hydrogen bonds (Figure 7.22a). In **18**, of the





**Figure 7.20** Hydrogen bonding leading to a threefold helix in the crystal structure of **15** with  $R = -OCH_2CH_2OMe$ : (a) side view; and (b) top view; one of the ethoxy groups is disordered



**Figure 7.21** Columnar hydrogen bonding in **17**: (a) side view; and (b) top view

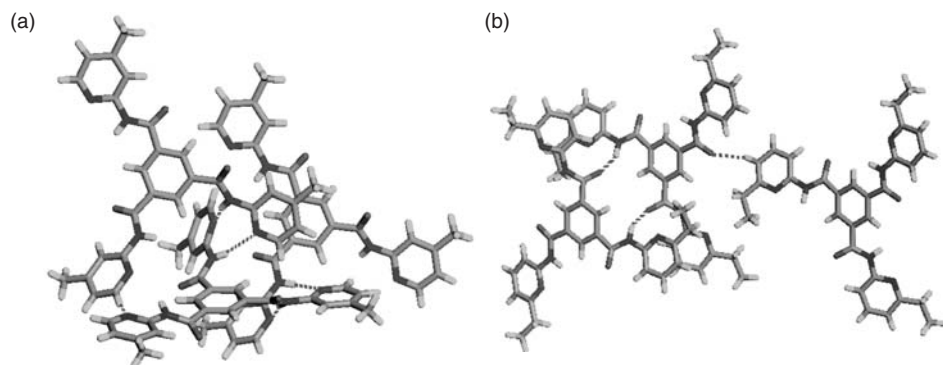
three amide groups two are engaged in N-H...N hydrogen bonds with two different molecules, while the third N-H does not participate in any significant hydrogen bonds. The carbonyl groups form C-H...O interactions with pyridine C-H groups.

In the crystal structure of **19**, amide-to-amide hydrogen bonds are observed between two amide functionalities (Figure 7.22b), no N-H...N hydrogen bonds are observed with the pyridine N-atom. The dramatic difference between **18** and **19** can be rationalized as the  $\alpha$ -position of the pyridine N-atom in **19** was substituted by the bulkier ethyl group thereby blocking the putative N-H...N hydrogen bond formation.

### 7.7.2 3-Pyridyl and 4-pyridyl based systems

These two derivatives are being discussed together as they have some similarities in their crystal structures. It was shown that **20**, i.e. **15** with  $R = 3\text{-pyridyl}$  and **21**, i.e. **15** with





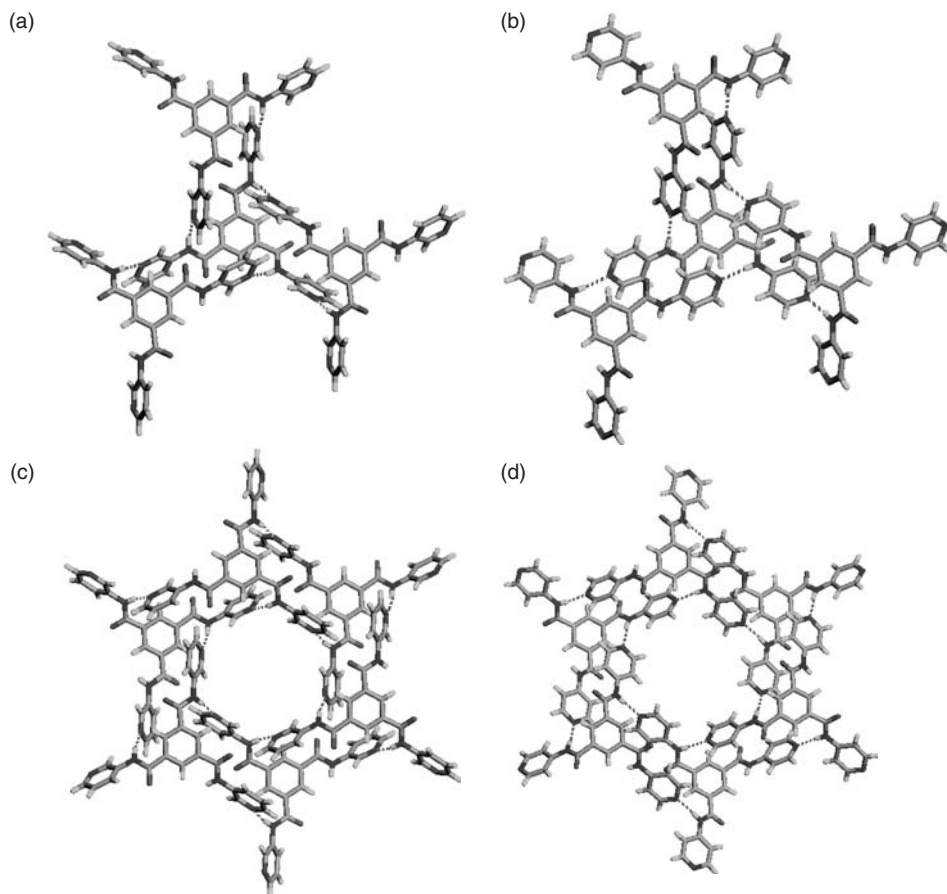
**Figure 7.22** (a) N-H...N hydrogen bonding pattern in **18** and (b) N-H...O hydrogen bonding pattern in **19**

R = 4-pyridyl, retain the threefold molecular symmetry and assemble their respective molecules via N-H...N hydrogen bonds [29a, b]. The 3-pyridyl or 4-pyridyl functionalities in these derivatives interfere in amide-to-amide hydrogen bonds (Figure 7.23). The N-H...N hydrogen bond aggregation generates a honeycomb network with channels. These channels in **20** were shown to be porous and display desorption of MeOH and absorption of water at room temperature. Later, it was found that the 3-pyridyl derivative **20** indeed crystallizes in three crystalline forms (**A**, **B** and **C**) from the same solvent (MeOH) [29c]. The crystallization process was shown to be dependent on the concentration and evaporation rates. Highly concentrated solution resulted in form **A**, moderately concentrated solution resulted in form **B** and dilute solutions resulted in form **C**.

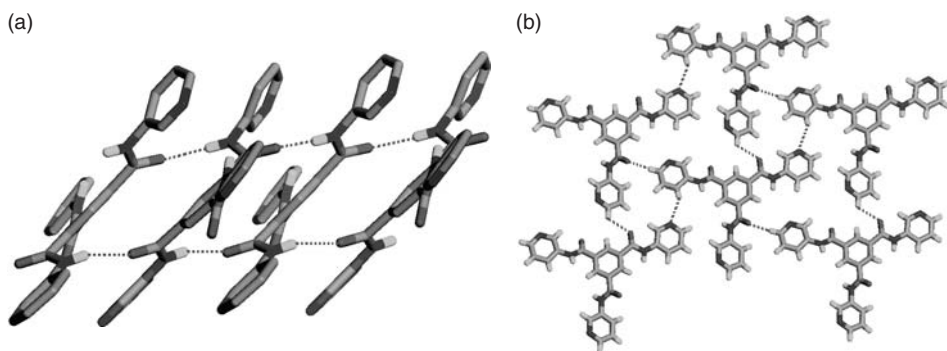
The crystal structure of form **A** ( $20 \cdot 3\text{CH}_3\text{OH}$ ) was basically the same as described above. Form **B** ( $20 \cdot \text{H}_2\text{O}$ ) was also characterized by single crystal X-ray diffraction. The crystal structure of form **B** differs from form **A** both in molecular and supramolecular geometry (Figure 7.24). In form **B** the molecule does not exhibit a threefold symmetry but adopts a T-shaped geometry instead. Out of the three amide groups, two form amide-to-amide N-H...O hydrogen bonds, and the third amide N-H forms an N-H...O hydrogen bond with the included water molecule. No N-H...N hydrogen bonding towards the pyridine-nitrogen was observed as the pyridine forms O-H...N hydrogen bonds with the solvent water molecule. Form **C** was characterized by powder X-ray diffraction, thermogravimetric analysis, differential scanning calorimetry, infrared spectroscopy and elemental analysis as single crystals suitable for structure determination were not obtained. The composition of form **C** was found to be the same as form **B** and therefore, **B** and **C** are polymorphs, while **A** and **B**, and **A** and **C** are pseudo-polymorphs.

## 7.8 Conclusions

The presence of pyridine exhibited interference in amide-to-amide hydrogen bonds in the majority of the cases analysed. In the case of primary amides, the pyridine interrupted one of the amide synthons (**I** or **II**). The survival probability of synthon **I** was found to be greater compared with synthon **II**. The molecular geometry was also found



**Figure 7.23** Hydrogen bonding patterns in **20** (form **A**) and **21**: N-H...N hydrogen bonding synthons in (a) **20**; and (b) **21**. Expansions to honeycomb networks containing channels in (c) **20** and (d) **21**



**Figure 7.24** (a) Amide-to-amide N-H...O hydrogen bonds and (b) C-H...N and C-H...O contacts leading to a layer in form **B** of **20**

to play a crucial role in 2-pyridyl derivatives as they exhibit planar geometry due to intramolecular N-H...N hydrogen bonds. There is a clear distinction observed between C-terminal amidopyridines and N-terminal amidopyridines. C-Terminal amidopyridines show a propensity for N-H...O hydrogen bonds, whereas N-terminal amidopyridines are likely to form N-H...N hydrogen bonds. The value of  $\theta$  seems to be an important factor in deciding between N-H...N and N-H...O hydrogen bonds. Further, crystal structures of co-crystals suggest that the hydrogen bond hierarchy will be followed given the suitable molecular geometry of the components.

## Acknowledgements

We thank DST and CSIR for financial support of our research on crystal engineering studies on derivatives containing amide and pyridine functional groups.

## References

1. (a) J. C. MacDonald and G. M. Whitesides, *Chem. Rev.*, **94**, 2383 (1994); (b) Y. Liu, A. H. W. Lam, F. W. Fowler and J. W. Lauher, *Mol. Cryst. Liq. Cryst. Sci. Technol., Sect. A*, 2002, **389**, 39; (c) X. Wang, M. Simard and J. D. Wuest, *J. Am. Chem. Soc.*, **116**, 12119 (1994); (d) M. Simard, D. Su and J. D. Wuest, *J. Am. Chem. Soc.*, **113**, 4696 (1991); (e) L. S. Reddy, S. Basavoju, V. R. Vangala and A. Nangia, *Cryst. Growth Des.*, **6**, 1161 (2006).
2. (a) G. R. Desiraju, *Crystal Engineering, The Design of Organic Solids*, Elsevier, Amsterdam, 1989; (b) C. B. Aakeröy, *Acta Crystallogr., Sect. B*, **53**, 569 (1997); (c) M. J. Zaworotko, *Chem. Commun.*, 1 (2001); (d) G. R. Desiraju, *Acc. Chem. Res.*, **35**, 565 (2002); (e) K. Biradha, *CrystEngComm*, **5**, 374 (2003); (f) D. Braga and F. Grepioni, *Chem. Commun.*, 3635 (2005); (g) J. D. Wuest, *Chem. Commun.*, 5830 (2005); (h) D. Braga, L. Brammer and N. R. Champness, *Cryst EngComm*, **7**, 1 (2005).
3. (a) G. R. Desiraju, *Angew. Chem. Int. Ed.*, **34**, 2311 (1995); (b) V. R. Vangala, R. Mondal, C. K. Broder, J. A. K. Howard and G. R. Desiraju, *Cryst. Growth Des.*, **5**, 99 (2005); (c) R. Banerjee, R. Mondal, J. A. K. Howard and G. R. Desiraju, *Cryst. Growth Des.*, **6**, 999 (2006); (d) M. Du, Z. H. Zhang, X. G. Wang, H. F. Wu and Q. Wang, *Cryst. Growth Des.*, **6**, 1867 (2006).
4. (a) J. W. Lauher, Y. L. Chang and F. W. Fowler, *Mol. Cryst. Liquid Cryst.*, **211**, 99 (1992); (b) Y. L. Zhao and Y.-D. Wu, *J. Am. Chem. Soc.*, **124**, 1570 (2002).
5. (a) F. H. Herbstein in *Comprehensive Supramolecular Chemistry*, D. D. MacNicol, F. Toda and R. Bishop, Eds, Pergamon, Oxford, 1996, Vol. 6, pp. 61–83; (b) M. C. Etter and G. M. Frankenbach, *Chem. Mater.*, **1**, 10 (1989); (c) P. Vishweshwar, D. A. Beauchamp and M. J. Zaworotko, *Cryst. Growth Des.*, **6**, 2429 (2006); (d) C. V. K. Sharma, K. Panneerselvam, T. Pilati and G. R. Desiraju, *J. Chem. Soc., Perkin Trans. 2*, **11**, 2209 (1993); (e) N. J. Babu, L. S. Reddy and A. Nangia, *Mol. Pharmaceut.*, **4**, 417 (2007); (f) L. S. Reddy, P. M. Bhatt, R. Banerjee, A. Nangia and G. J. Kruger, *Chem. Asian J.*, **2**, 505 (2007).

6. (a) M. C. Etter, *Acc. Chem. Res.*, **23**, 120 (1990); (b) M. C. Etter, *J. Phys. Chem.*, **95**, 4601 (1991).
7. (a) S. Matile, *Chem. Soc. Rev.*, **30**, 158 (2001); (b) G. Cooke and V. M. Rotello, *Chem. Soc. Rev.*, **31**, 275 (2002); (c) N. Sakai, J. Mareda and S. Matile, *Acc. Chem. Res.*, **38**, 79 (2005); (d) S. S. Kuduva, D. Bläser, R. Boese and G. R. Desiraju, *J. Org. Chem.*, **66**, 1621 (2001).
8. F. H. Allen, *Acta Crystallogr., Sect. B*, **58**, 380 (2002).
9. (a) J. Thun, L. Seyfarth, J. Senker, R. E. Dinnebier, P. Doz. and J. Breu, *Angew. Chem., Int. Ed.*, **46**, 6729, (2007); (b) W. I. F. David, K. Shankland, C. R. Pulham N. Blagden, R. J. Davey and M. Song, *Angew. Chem., Int. Ed.*, **44**, 7032, (2005).
10. (a) T. Takano, Y. Sasada and M. Kakudo, *Acta Crystallogr.*, **21**, 514, (1966); (b) M. Gdaniec, M. Jaskolski and Z. Kosturkiewicz, *Pol. J. Chem.*, **54**, 1539, (1980); (c) M. Gdaniec, M. Jaskolski and Z. Kosturkiewicz, *Acta Crystallogr., Sect. B*, **33**, 3558, (1977).
11. (a) Y. Miwa, T. Mizuno, K. Tsuchida, T. Taga and Y. Iwata, *Acta Crystallogr., Sect. B*, **55**, 78, (1999); (b) J. C. Schlenker and R. J. Staples, *Z. Kristallogr. - New Cryst. Struct.*, **217**, 555, (2002).
12. C. B. Aakeroy, A. M. Beatty, B. A. Helfrich and M. Nieuwenhuyzen, *Cryst. Growth Des.*, **3**, 159 (2003).
13. (a) M. B. Hursthouse, P. J. Murphy, D. E. Hibbs and A. Malik, Personal Communication (2003), CSD refcode BATHOU; (b) M. P. Wentland, R. B. Perni, P. H. Dorff, R. P. Brundage, M. J. Castaldi, T. R. Bailey, P. M. Carabateas, E. R. Bacon, D. C. Young, M. G. Woods, D. Rosi, M. L. Drozd, R. K. Kullnig and F. J. Dutko, *J. Med. Chem.*, **36**, 1580 (1993).
14. (a) C. B. Aakeroy, A. M. Beatty and B. A. Helfrich, *Angew. Chem., Int. Ed.*, **40**, 3240 (2001); (b) D. Chakrabarty, H. Nagase, M. Kamijo, T. Endo and H. Ueda, *X-Ray Struct. Anal. Online*, x27 (2006); (c) M. Inoue, T. Sakaki, T. Fujiwara and K. Tomita, *Bull. Chem. Soc. Jpn.*, **51**, 1118 (1978); (d) C. B. Aakeroy, A. M. Beatty and B. A. Helfrich, *J. Am. Chem. Soc.*, **124**, 14425 (2002); (e) P. Vishweshwar, A. Nangia and V. M. Lynch, *CrystEngComm*, **5**, 164 (2003); (f) P. Vishweshwar, A. Nangia and V. M. Lynch, *Cryst. Growth Des.*, **3**, 783 (2003); (g) J. A. McMohan, J. A. Bis, P. Vishweshwar, T. R. Shattok, O. L. McLaughlin and M. J. Zawarotko, *Z. Kristallogr.*, **220**, 340 (2005).
15. (a) S. G. Fleischman, S. S. Kuduva, J. A. McMohan, B. Moulton, R. D. B. Walsh, N. Rodriguez-Hornedo and M. J. Zawarotko, *Cryst. Growth Des.*, **3**, 909 (2003); (b) I. D. H. Oswald, W. D. S. Motherwell and S. Parsons, *Acta Crystallogr., Sect. E*, **61**, o3161 (2005).
16. (a) J. Zhang, Q. Liu, Y. Xu, Y. Zhang, X. You and Z. Guo, *Acta Crystallogr., Sect. C*, **57**, 109 (2001); (b) J. Y. Qi, Q. Y. Yang, K. H. Lam, Z. Y. Zhou and A. S. C. Chan, *Acta Crystallogr., Sect. E*, **59**, o374 (2003).
17. (a) M. Gdaniec, M. Jaskolski and Z. Kosturkiewicz, *Pol. J. Chem.*, **53**, 2563 (1979); (b) R. G. Little and C. Morimoto, *Acta Crystallogr., Sect. B*, **37**, 1637 (1981); (c) T. Srikrishnan and R. Parthasarathy, *Acta Crystallogr., Sect. C*, **46**, 1723 (1990); (d) J. M. Jethmalani, A. G. Camp, N. G. Soman, J. W. Hawley, F. L. Setliff and E. M. Holt, *Acta Crystallogr., Sect. C*, **52**, 438 (1996).

18. D. K. Kumar, D. A. Jose, P. Dastidar and A. Das, *Langmuir*, **20**, 10413 (2004).
19. (a) C. C. Forbes, A. M. Beatty and B. D. Smith, *Org. Lett.*, **3**, 3595, (2001); (b) S. J. Howell, C. S. Day and R. E. Noftle, *Inorg. Chim. Acta*, **358**, 3711 (2005).
20. J. C. Noveron, M. S. Lah, R. E. Del Sesto, A. M. Arif, J. S. Miller and P. J. Stang, *J. Am. Chem. Soc.*, **124**, 6613 (2002).
21. (a) P. A. Palmer and F. Brisse, *Acta Crystallogr., Sect. B*, **36**, 1447 (1980); (b) S. Harkema, G. J. Van Hummel and R. J. Gaymans, *Acta Crystallogr., Sect. B*, **36**, 3182 (1980); (c) C. Pineault and F. Brisse, *Acta Crystallogr., Sect. C*, **39**, 1434 (1983); (d) C. Pineault and F. Brisse, *Acta Crystallogr., Sect. C*, **39**, 1437 (1983); (e) S. Harkema and R. J. Gaymans, *Acta Crystallogr., Sect. B*, **33**, 3609 (1977); (f) W. W. Adams, *Acta Crystallogr., Sect. B*, **34**, 954 (1978); (g) S. Anjum, M. I. Choudhary, S. Ali, H.-K. Fun and Atta-ur-Rahman, *Acta Crystallogr., Sect. E*, **61**, o3001 (2005); (h) S. Harkema, R. J. Gaymans and G. J. van Hummel, *Acta Crystallogr., Sect. C*, **39**, 385 (1983); (i) S. Harkema, R. J. Gaymans, G. J. van Hummel and D. Zylberlicht, *Acta Crystallogr., Sect. B*, **35**, 506 (1979).
22. (a) F. S. Stephens and R. S. Vagg, *Inorg. Chim. Acta.*, **142**, 43 (1988); (b) S. Shao, D. Zhu, Y. Song, X. Z. You, S. S. S. Raj and H.-K. Fun, *Acta Crystallogr., Sect. C*, **55**, 1841 (1999).
23. M. Sarkar and K. Biradha, *Cryst. Growth Des.*, **6**, 202 (2006).
24. (a) Y.-F. Hsu and J.-D. Chen, *Eur. J. Inorg. Chem.*, 1488 (2004); (b) F. G. Tellado, S. J. Geib, S. Goswami and A. D. Hamilton, *J. Am. Chem. Soc.*, **113**, 9265, (1991); (c) I. L. Karle, D. Rangnathan and V. Haridas, *J. Am. Chem. Soc.*, **119**, 2777, (1997).
25. (a) L. Rajput, S. Singha and K. Biradha, *Cryst. Growth Des.*, **7**, 2788 (2007); (b) S. C. Abrahams, L. E. Zyontz and J. L. Bernstein, *J. Chem. Phys.*, **76**, 5458 (1982); (c) J. Sun, L. Weng, Y. Zhou, J. Chen, Z. Chen, Z. Liu and D. Zhao, *Angew. Chem., Int. Ed.*, **41**, 4471 (2002); (d) L. Carlucci, G. Ciani and D. M. Proserpio, *Coord. Chem. Rev.*, **246**, 247 (2003); (e) S. Hu and M.-L. Tong, *Dalton Trans.*, 1165 (2005); (f) B. F. Hoskins, R. Robson and N. V. Y. Scarlett, *Angew. Chem., Int. Ed. Engl.*, **34**, 1203 (1995).
26. M. Du, Z.-H. Zhang, X.-J. Zhao and H. Cai, *Cryst. Growth Des.*, **6**, 114 (2006).
27. M. P. Lightfoot, F. S. Mair, R. G. Pritchard and J. E. Warren, *Chem. Commun.*, 1945 (1999).
28. (a) E. Fan, J. Yang, S. J. Geib, T. C. Stoner, M. D. Hopkins and A. D. Hamilton, *Chem. Commun.*, 1251 (1995); (b) M. Mazik and W. Sicking, *Tetrahedron Lett.*, **45**, 3117 (2004); (c) M. Mazik, W. Radunz and R. Boese, *J. Org. Chem.*, **69**, 7448 (2004).
29. (a) A. R. A. Palmans, J. A. J. M. Vekemans, H. Kooijman, A. L. Spek and E. W. Meijer, *Chem. Commun.*, 2247 (1997); (b) D. K. Kumar, D. A. Jose, P. Dastidar and A. Das, *Chem. Mater.*, **16**, 2332 (2004); (c) L. Rajput and K. Biradha, *J. Mol. Struct.*, **876**, 339 (2008).

# 8

## Urea/Thiourea-Anion Host Lattices, Stabilization of Labile Species, and Designed Construction of Rosette Ribbon and Layers

*Thomas C. W. Mak, Chi-Keung Lam, Jie Han, Qi Li and Feng Xue*

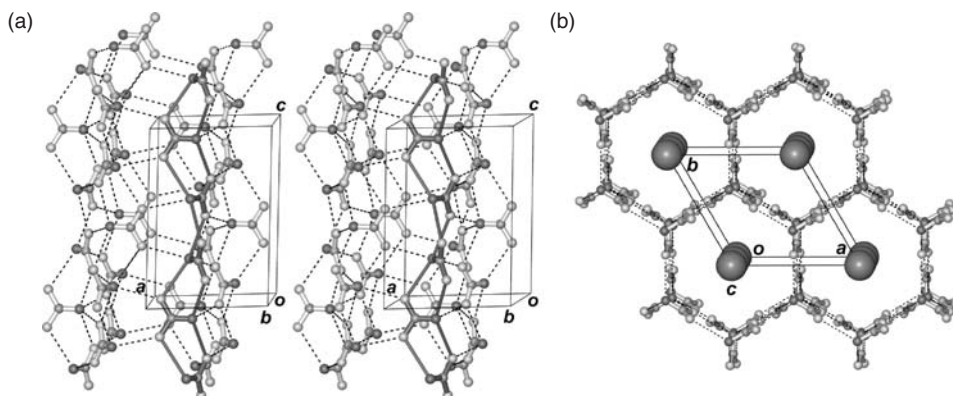
*Dedicated to the memory of Dr Daniel Y. Chang (1906–1982)*

### 8.1 Introduction

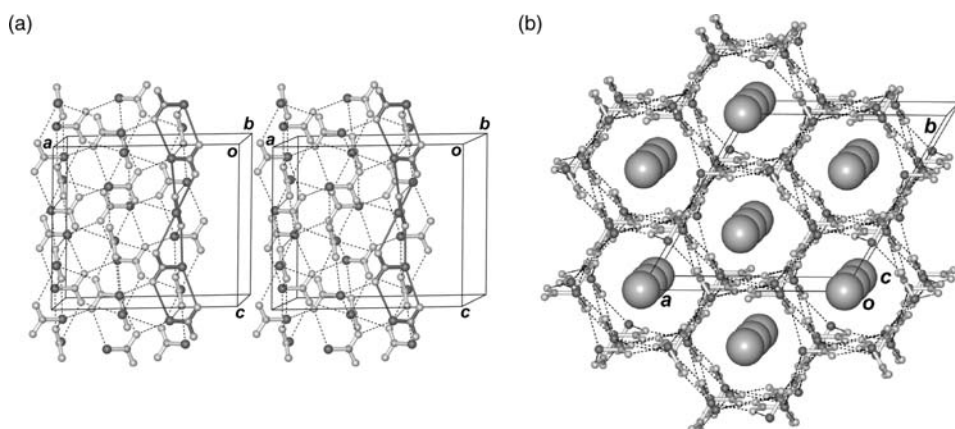
Basically, there are two main classes of urea/thiourea inclusion compounds: the classical channel-type clathrates having a host lattice constructed from urea (Figure 8.1) or thiourea molecules (Figure 8.2) [1] and those whose host lattices contain anions as additional components (Figure 8.3) [2].

Making use of conventional hydrogen bonds formed by urea [5], thiourea [6], or selenourea [7] together with various oxoanions and neutral molecular building blocks, new types of anionic host lattices have been generated and reviewed [8]. The purpose of this chapter is to cover recent advances in supramolecular assembly of host structures composed of urea/thiourea and/or their derivatives and selected polyatomic anions (such as carbonate, trimesate and 5-nitrobarbiturate), with quaternary ammonium cations serving as guest templates. This chapter places particular emphasis on the stabilization of labile species (such as dihydrogen borate, allophanate, thioallophanate, and the valence tautomers of croconate and rhodizonate) as components of an anionic host lattice, as well as

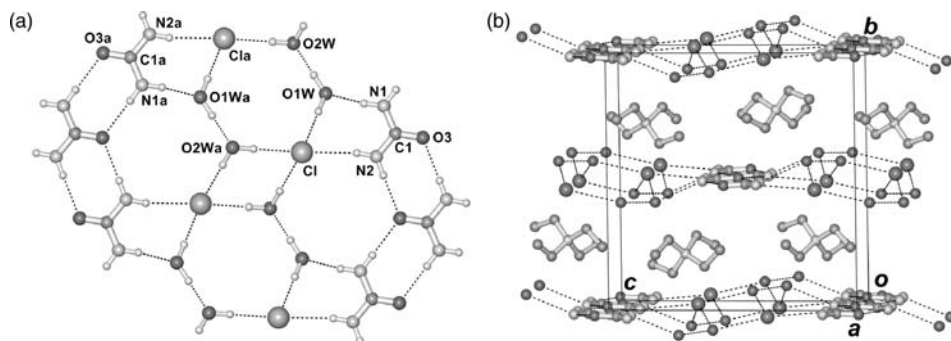




**Figure 8.1** (a) Stereodrawing of the hydrogen-bonded urea host lattice possessing an empty channel extending parallel to the  $c$  axis. Broken lines represent hydrogen bonds, and atom types are distinguished by size and shading. A shoulder-to-shoulder urea ribbon is highlighted by showing its  $\text{N-H}\cdots\text{O}$  hydrogen bonds as thick lines. (b) The channel-type structure of urea clathrates viewed along the  $c$  axis. For clarity hydrogen atoms of urea are omitted and the guest molecules (commonly straight-chain aliphatic compounds) are represented by large spheres [3]



**Figure 8.2** (a) Stereodrawing of the hydrogen-bonded thiourea host lattice possessing an empty channel extending parallel to the  $c$  axis. A shoulder-to-shoulder thiourea ribbon is highlighted by showing  $\text{N-H}\cdots\text{S}$  hydrogen bonds as thick lines. (b) The channel-type structure of thiourea clathrates viewed along the  $c$  axis. For clarity hydrogen atoms of thiourea are omitted and the guest molecules (commonly branched-chain aliphatic compounds) are represented by large spheres [4]



**Figure 8.3** (a) A portion of a urea-water-chloride hydrogen-bonded layer. (b) Layer-type structure of  $(\text{C}_2\text{H}_5)_4\text{N}^+ \text{X}^- \cdot (\text{NH}_2)_2\text{CO} \cdot 2\text{H}_2\text{O}$ , where  $\text{X} = \text{Cl}, \text{Br}$  [2]. Symmetry transformation:  $a (1 - x, 1 - y, -z)$

the designed construction of hydrogen-bonded rosette ribbons and layers employing the guanidinium ion.

## 8.2 Inclusion Compounds with Urea/Thiourea-Anion Host Lattices

### 8.2.1 Isostructurality Based on Interchangeability of Supramolecular Synthons

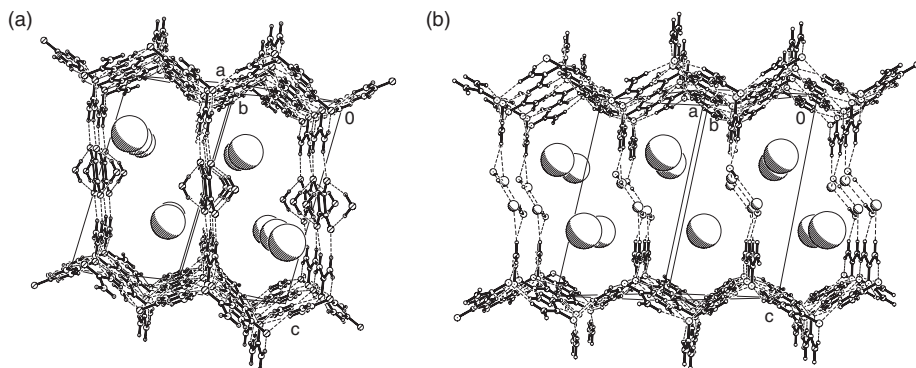
The development of the supramolecular synthon concept from pattern recognition considerably simplifies the analysis and design of complicated crystal structures [9], as topological features and useful empirical rules can be derived by critical examination of the wealth of crystallographic information documented in structural databases [10].

The crystal structures of dibasic squaric acid,  $\text{H}_2\text{C}_4\text{O}_4$  [11], and a wide variety of organic and inorganic hydrogen squarates [12] and squarates [13] are known. Our pilot study started with the preparation and structural characterization of the urea/squarate inclusion compound  $[(n\text{-C}_3\text{H}_7)_4\text{N}^+]_2\text{C}_4\text{O}_4^{2-} \cdot 6(\text{NH}_2)_2\text{CO} \cdot 2\text{H}_2\text{O}$  (**1**), which unexpectedly provides a nice demonstration of the significance of supramolecular synthons in the structural correlation between inclusion compounds that exhibit isostructurality [14] in the solid state.

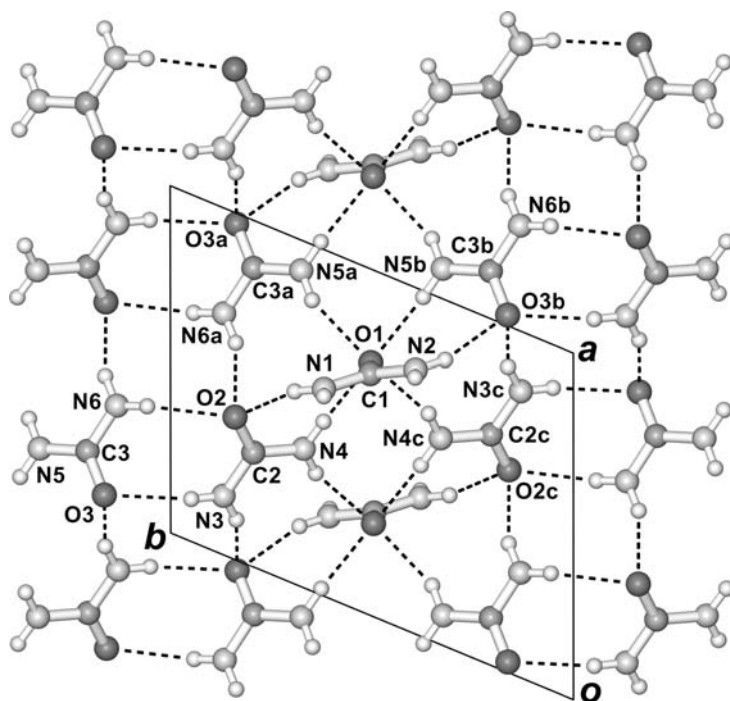
The host structure of **1** [15] is closely related to that of  $(n\text{-C}_3\text{H}_7)_4\text{N}^+\text{Br}^- \cdot 3(\text{NH}_2)_2\text{CO} \cdot \text{H}_2\text{O}$  (**2**) [5d]. Both compounds crystallize in space group  $P\bar{1}$  with similar unit-cell parameters. They both contain a large open-channel framework (Figure 8.4) composed of corrugated urea layers that are inter-linked by spacer moieties each forming four acceptor hydrogen bonds (Figure 8.5). As indicated in Figure 8.6, replacement of the cyclic tetramer  $(\text{H}_2\text{O}-\text{Br}^-)_2$  in **2** by the squarate-water dumbbell moiety  $(\text{H}_2\text{O}-\text{C}_4\text{O}_4^{2-}-\text{H}_2\text{O})$  in **1** causes no significant variation in the size of the channel, but shifts adjacent corrugated urea layers to facilitate linkage of protruding urea molecules by different acceptor linkers.

The building blocks  $(\text{H}_2\text{O}-\text{C}_4\text{O}_4^{2-}-\text{H}_2\text{O})$  and  $(\text{H}_2\text{O}-\text{Br}^-)_2$  in the host structures of **1** and **2**, respectively, can be regarded as equivalent synthons of compatible geometry and

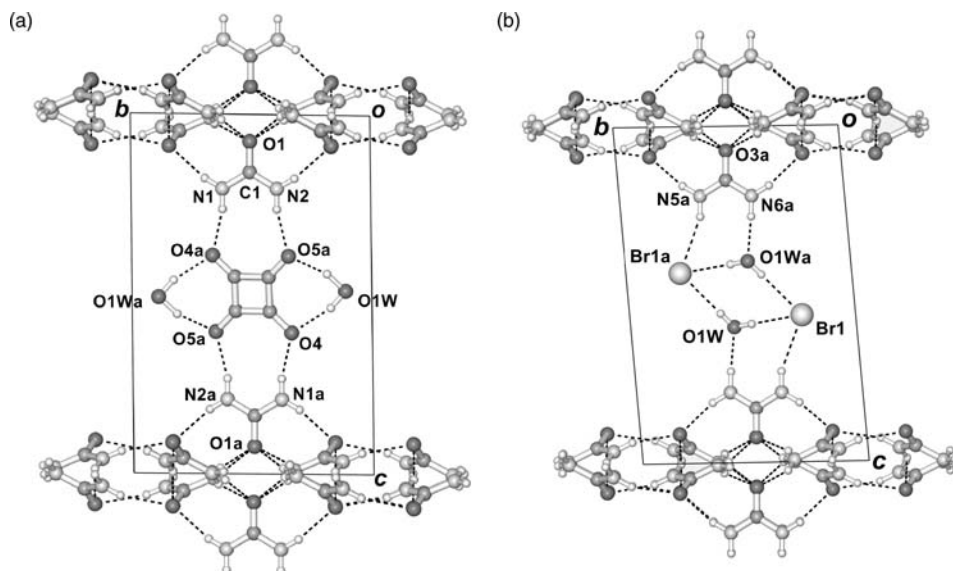




**Figure 8.4** (a) The hydrogen-bonded anionic host lattice of  $[(n\text{-C}_3\text{H}_7)_4\text{N}^+]_2\text{C}_4\text{O}_4^{2-} \cdot 6(\text{NH}_2)_2\text{CO} \cdot 2\text{H}_2\text{O}$  (**1**) contains elongated hexagonal channels extending along the  $[1\bar{2}0]$  direction. (b) Projection along  $[110]$  showing the host lattice of  $(n\text{-C}_3\text{H}_7)_4\text{N}^+\text{Br}^- \cdot 3(\text{NH}_2)_2\text{CO} \cdot \text{H}_2\text{O}$  (**2**) built from corrugated urea layers and bridging cyclic  $(\text{H}_2\text{O}-\text{Br}^-)_2$  tetramers. Note the peanut shape of the cross-section of each channel. For clarity the enclosed  $(n\text{-C}_3\text{H}_7)_4\text{N}^+$  ions are represented by partly shaded large spheres



**Figure 8.5** Hydrogen-bonded corrugated urea layer in  $[(n\text{-C}_3\text{H}_7)_4\text{N}^+]_2\text{C}_4\text{O}_4^{2-} \cdot 6(\text{NH}_2)_2\text{CO} \cdot 2\text{H}_2\text{O}$  (**1**) at  $z = 0$  comprising wide ribbons derived from two independent urea molecules, as well as bridging urea molecules of the third kind that point alternately up and down between ribbons. Broken lines represent hydrogen bonds. Symmetry transformations:  $a (1 - x, 2 - y, -z)$ ,  $b (1 + x, 1 + y, z)$  and  $c (1 - x, 1 - y, -z)$



**Figure 8.6** (a) Projection diagram of a portion of the host lattice of  $[(n\text{-C}_3\text{H}_7)_4\text{N}^+]_2\text{C}_4\text{O}_4^{2-} \cdot 6(\text{NH}_2)_2\text{CO} \cdot 2\text{H}_2\text{O}$  (**1**). The dumbbell-shaped  $(\text{H}_2\text{O}-\text{C}_4\text{O}_4^{2-}-\text{H}_2\text{O})$  unit cross-links two adjacent puckered urea layers by forming two pairs of  $\text{N}-\text{H} \cdots \text{O}^-$  (squarate) hydrogen bonds with two vertically orientated urea molecules to form a side-wall of the channel. The separation between two corrugated urea layers is  $1/c^* = 14.81 \text{ \AA}$ . (b) Projection diagram of a portion of the host structure of  $(n\text{-C}_3\text{H}_7)_4\text{N}^+ \text{Br}^- \cdot 3(\text{NH}_2)_2\text{CO} \cdot \text{H}_2\text{O}$  (**2**) showing that two adjacent puckered urea layers are inter-linked by the tetrameric  $(\text{H}_2\text{O}-\text{Br}^-)_2$  moiety which forms acceptor  $\text{N}-\text{H} \cdots \text{O}$  and  $\text{N}-\text{H} \cdots \text{Br}^-$  hydrogen bonds with two vertically orientated urea molecules. The distance between two corrugated urea layers is  $1/c^* = 14.55 \text{ \AA}$ . Symmetry transformation:  $a (-x, -y, -z)$

hydrogen bonding capability, and are accordingly mutually exchangeable, thus providing an interesting example of isostructurality [14] in the realm of inclusion compounds.

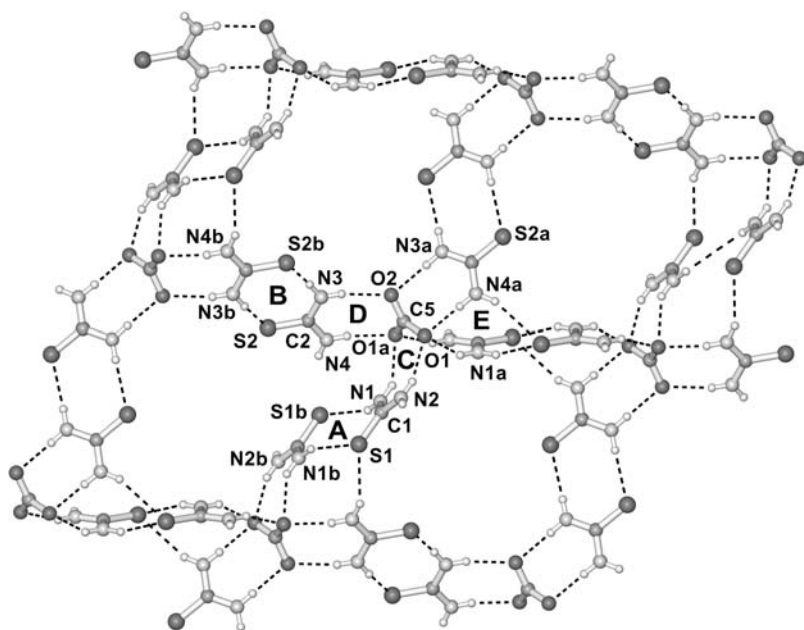
### 8.2.2 Carbonate and Oxalate as Prolific Hydrogen-Bond Acceptors

Successful incorporation of the squarate dianion into the hydrogen-bonded host lattice of **1** prompted us to explore the two simplest oxocarbon dianions, namely carbonate and oxalate, as hubs for supramolecular assembly. Bearing in mind that the carbonyl oxygen atom in the tetragonal phase of urea forms four acceptor hydrogen bonds [16], and that the squarate ion in  $[(n\text{-C}_3\text{H}_7)_4\text{N}^+]_2\text{C}_4\text{O}_4^{2-} \cdot 6(\text{NH}_2)_2\text{CO} \cdot 2\text{H}_2\text{O}$  (**1**) forms eight acceptor hydrogen bonds with two pairs of neighboring urea and water molecules, it would be of interest to see if the smaller acyclic oxocarbon ions can better this record. Experimentation with various peralkylated ammonium ions as structure-inducing agents and different reactant molar ratios eventually led to isolation of the inclusion compounds  $[(\text{C}_2\text{H}_5)_4\text{N}^+]_2\text{CO}_3^{2-} \cdot 7(\text{NH}_2)_2\text{CS}$  (**3**) and  $[(n\text{-C}_3\text{H}_7)_4\text{N}^+]_2\text{C}_2\text{O}_4^{2-} \cdot 4(\text{NH}_2)_2\text{CS}$  (**4**) [17], which display interesting inclusion topologies and demonstrate the high hydrogen-bonding capacity of the respective oxocarbon anions.

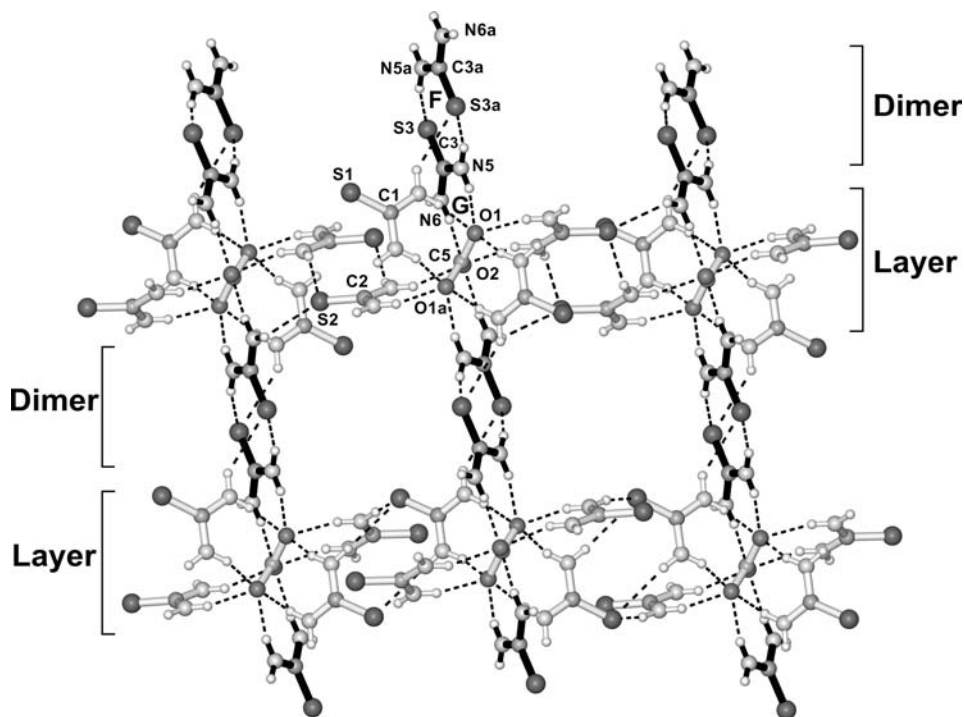
8.2.2.1 Channel-type structure of  $[(C_2H_5)_4N^+]_2CO_3^{2-} \cdot 7(NH_2)_2CS$  (**3**)

In the crystal structure of **3**, the carbonate ion occupies a site of symmetry 2. The independent thiourea molecules **C(1)** and **C(2)** (each conveniently designated by naming only its carbon atom) generate centrosymmetric dimers (labeled **A** and **B** in Figure 8.7) that have the same motif described by the graph-set notation  $N_2 = R_2^2(8)$  [18]. These thiourea dimers are directly connected to the carbonate ion by strong charge-assisted  $N-H \cdots O^-$  hydrogen bonds with ring motifs **C** and **D**,  $N_2 = R_2^2(8)$  to form a highly puckered hydrogen-bonded layer. The layer is further consolidated by additional  $N-H \cdots S$  hydrogen bonds to form a new ring motif **E**,  $N_4 = R_4^2(8)$ . With these thiourea-carbonate layers arranged parallel to the (100) family of planes, interlayer linkage by the thiourea dimers derived from **C(3)** [ring motif **F**,  $N_2 = R_2^2(8)$ ] via ring motif **G**,  $N_2 = R_2^2(8)$ , defines the square cross-section of the channel system extending along the *b* axis (Figure 8.8). The resulting 3-D host lattice also has narrower channels along the [011] and  $[0\bar{1}1]$  directions.

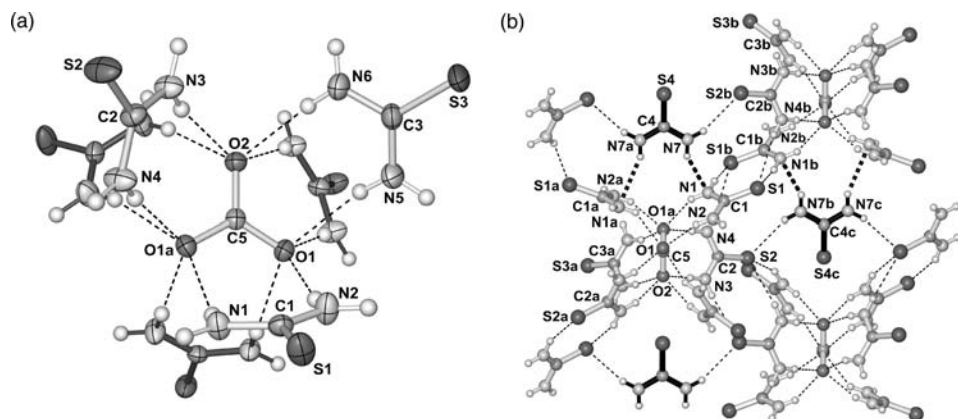
The most prominent feature of the host framework in **3** is the formation of the  $[CO_3^{2-} \cdot 6(NH_2)_2CS]$  aggregate, which may be described as two concave, three-leaved propellers attached to the same core. All three independent thiourea molecules **C(1)**, **C(2)** and **C(3)** are hydrogen-bonded on the same side to the trigonal planar carbonate ion, which has a twofold axis passing through atoms C(5) and O(2), whereas their symmetry-related partners are bound to the opposite face (Figure 8.9a). The stability of the host



**Figure 8.7** In the host structure of **3**, the carbonate ions are linked to centrosymmetric dimers derived from thiourea molecules **C(1)** and **C(2)** to form a hydrogen-bonded layer parallel to (100). Symmetry transformations: *a*  $(-x, y, \frac{1}{2} - z)$  and *b*  $(-x, 1 - y, 1 - z)$



**Figure 8.8** Perspective view along the *b* axis showing the channel-type host framework of **3** built of carbonate-thiourea layers cross-linked by thiourea dimers derived from **C(3)**. Symmetry transformations: *a* ( $\frac{1}{2} - x, \frac{1}{2} - y, 1 - z$ ) and *b* ( $-x, y, \frac{1}{2} - z$ )



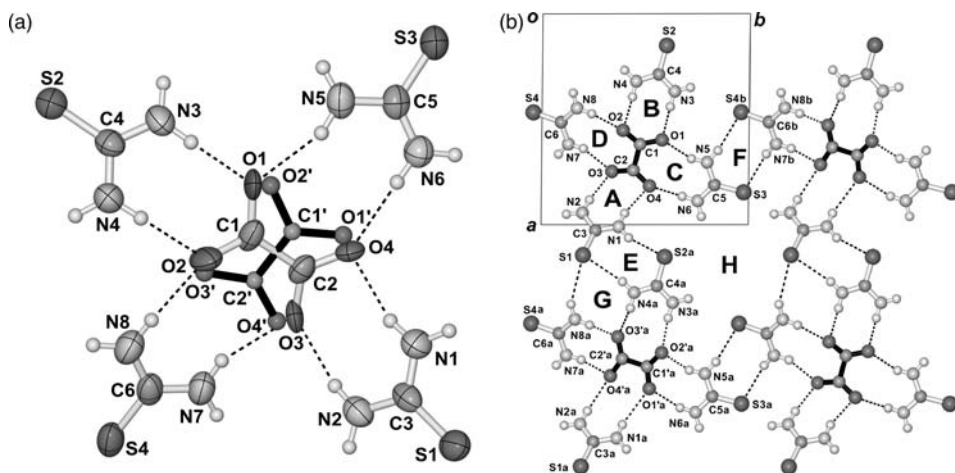
**Figure 8.9** (a) Biconcave, twin propeller-like  $[\text{CO}_3^{2-} \cdot 6(\text{NH}_2)_2\text{CS}]$  molecular aggregate in **3**. Thermal ellipsoids are drawn at the 30% probability level. Symmetry transformation: *a* ( $-x, y, \frac{1}{2} - z$ ). (b) Linkage between  $[\text{CO}_3^{2-} \cdot 6(\text{NH}_2)_2\text{CS}]$  building blocks with thiourea molecules of type **C(4)**. The weak  $\text{N-H} \cdots \pi$  (thioamido) interactions between thiourea molecules are indicated by thick broken lines. Symmetry transformations: *a* ( $-x, y, \frac{1}{2} - z$ ), *b* ( $-x, -y, 1 - z$ ) and *c* ( $x, -y, \frac{1}{2} + z$ )

framework is further strengthened by the presence of the remaining thiourea molecules **C(4)** (Figure 8.9b). Besides pulling a pair of two neighboring  $[\text{CO}_3^{2-} \cdot 6(\text{NH}_2)_2\text{CS}]$  building blocks together by two strong  $\text{N-H} \cdots \text{S}$  hydrogen bonds, each **C(4)** rides on two **C(1)** type thiourea molecules (the dihedral angle is  $79.6^\circ$ ) and forms two weak  $\text{N-H} \cdots \pi(\text{thioamido})$  hydrogen bonds ( $\text{NH} \cdots \text{NH}_2\text{CS}$ ,  $D = 3.367 \text{ \AA}$ ,  $d = 2.601 \text{ \AA}$  and  $\theta = 149.0^\circ$ ) with two thiourea molecules of a third building block. Two columns of well-ordered tetraethylammonium cations are arranged in staggered fashion and accommodated within each channel.

The carbonate and oxalate ions are known to form six [5c] to ten [6b] and eight [19] to twelve [20] acceptor hydrogen bonds, respectively, in their crystalline compounds. The carbonate ion in **3** forms a record number of twelve acceptor hydrogen bonds with the convergent  $\text{N-H}$  donor sites from six thiourea molecules. Furthermore, the presence of the **C(4)** type thiourea molecule in the host lattice of **3** also plays a significant role, as its almost perpendicular orientation to the **C(1)** type thiourea molecule facilitates weak  $\text{N-H} \cdots \pi(\text{thioamido})$  interaction [21] between them (Figure 8.9b).

#### 8.2.2.2 Puckered layer-type structure of $[(n\text{-C}_3\text{H}_7)_4\text{N}^+]_2\text{C}_2\text{O}_4^{2-} \cdot 4(\text{NH}_2)_2\text{CS}$ (**4**)

In the host lattice of **4**, the twofold disordered oxalate anion is bound by strong  $\text{N-H} \cdots \text{O}^-$  hydrogen bonds to two pairs of independent thiourea molecules in nearly orthogonal directions, forming an essentially planar cross-shaped structural building block (Figure 8.10a) that displays two types of ring motifs: **A** and **B**,  $N_2 = R_2^2(8)$ ; **C** and **D**,  $N_2 = R_2^2(9)$ . Adjacent cross-shaped building blocks related by simple axial translation are further interlinked to generate the ring motifs **E** and **F**,  $N_2 = R_2^2(8)$  and **G**,  $N_4 = R_4^2(8)$ , yielding a wavy layer corresponding to the plane  $z = \frac{1}{4}$  (Figure 8.10b).



**Figure 8.10** (a) Twofold disordered oxalate anion in **4**. One of the two possible orientations is represented by spherical atoms labeled with primes. Thermal ellipsoids are plotted at the 50% probability level. (b) Hydrogen-bonding interactions in the host layer of **4**. The twofold disordered oxalate anions are shown in only one orientation in each site. Symmetry transformations:  $a$  ( $1 + x, y, z$ ) and  $b$  ( $x, 1 + y, z$ )

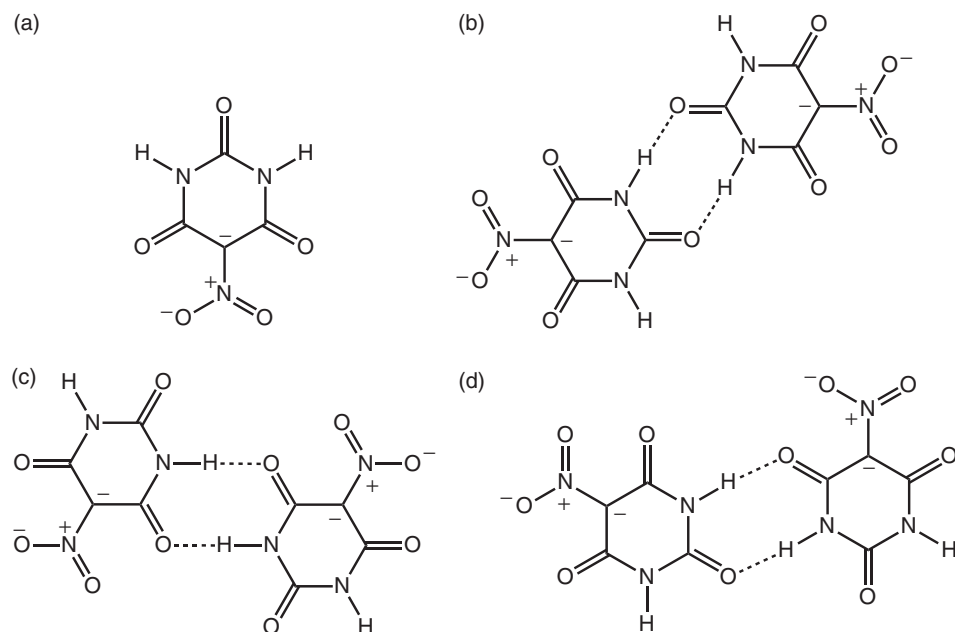
This hydrogen-bonded host layer is characterized by the large and un-symmetrical void **H**,  $N_4 = R_{10}^7(28)$ , and well-ordered tetra-*n*-propylammonium cations are accommodated between these puckered layers with interlayer spacing of  $c/2 = 8.2 \text{ \AA}$ .

Interestingly, the crystal structure of **4** shows that a twofold disordered oxalate ion may be considered as a structural equivalent of the squarate ion in supramolecular assembly via hydrogen bonding, thus providing an elegant example of isostructurality [14].

Previous studies have established that novel polynuclear complexes and coordination polymeric networks can be constructed with the carbonate [22] or oxalate [23] dianion as a multi-dentate building block. The present study complements this by showing that these two ubiquitous species, considered as hubs each bearing a divergent set of hydrogen-bond acceptor sites, may serve as versatile component modules for the *de novo* crystal engineering of supramolecular networks and organic-inorganic hybrid materials [24].

### 8.2.3 Hydrogen-Bonded Networks Constructed with 5-Nitrobarbiturate

5-Nitrobarbituric acid (also named dilituric acid or 5-nitro-2,4,6-trihydroxypyrimidine), formally a cyclic urea derivative, is selected as a host component [25]. In the solid state 5-nitrobarbiturate has the tendency to form a dimer, and three kinds of possible linkage are designated as types **A**, **B** and **C**, as illustrated in Scheme 8.1. Among these,



**Scheme 8.1** (a) Structural formula of the monoanion of 5-nitrobarbituric acid, (b) Type **A** (para-to-para), (c) type **B** (ortho-to-ortho) and (d) type **C** (para-to-ortho) dimer linkage. The designation of para or ortho refers to the carbonyl group that is involved in intermolecular hydrogen bonding, whose position is relative to that of the nitro group

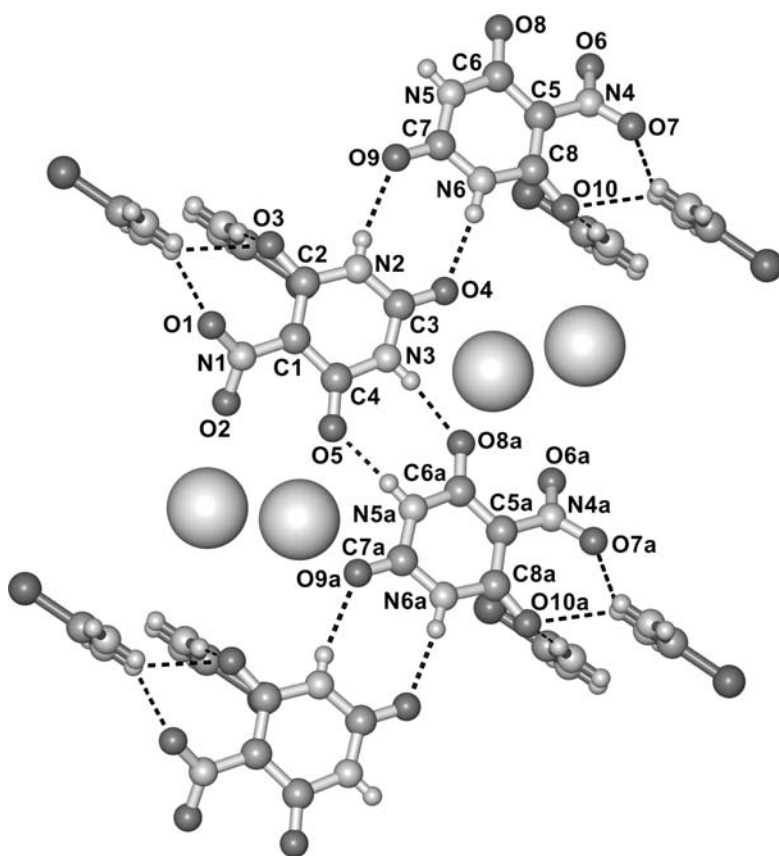


type **A** (para-to-para) appears in neat 5-nitrobarbituric acid [26], type **B** (ortho-to-ortho) is found in hexa-aqua-calcium bis(5-nitrobarbiturate) dehydrate [27], and both types occur in ammonium 5-nitrobarbiturate [28] and 5-nitrobarbituric acid trihydrate [26, 29].

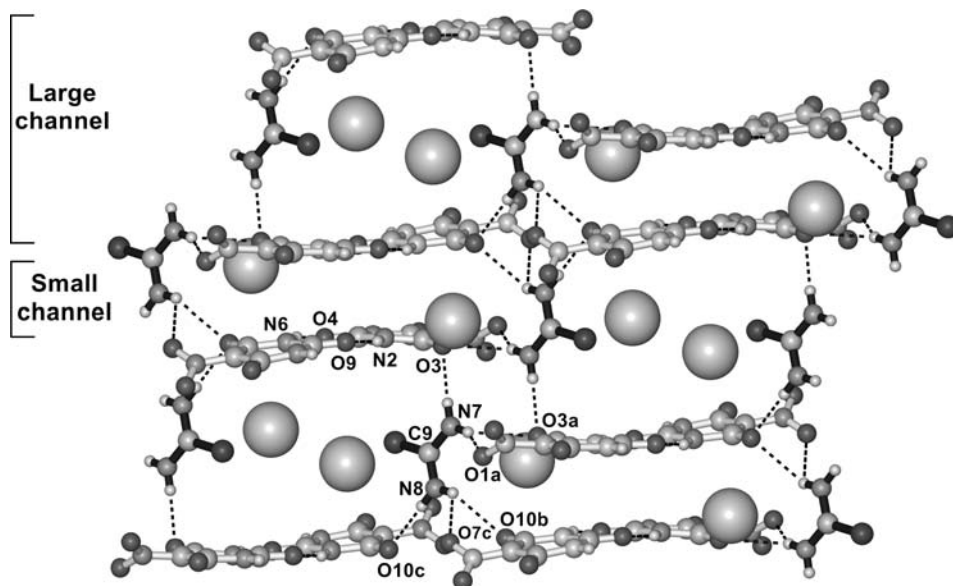
### 8.2.3.1 Crystal structure of $[(CH_3)_4N^+]_2[(C_4H_2N_3O_5)^-]_2 \cdot (NH_2)_2CS$ (**5**)

Two channels of different sizes exist in the crystal structure of inclusion complex **5** [25]. As shown in Figure 8.11, two independent 5-nitrobarbiturate anions are linked in the para-to-para fashion by N–H···O hydrogen bonds to form a dimer, and successive dimers related by the *a* translation associate in the ortho-to-ortho fashion to generate an approximately planar zigzag ribbon.

Adjacent ribbons are bridged by thiourea molecules to form two distinct channels (Figure 8.12). Two *syn* hydrogens of thiourea are connected with two parallel 5-nitrobarbiturate ribbons to form the larger channel with an inter-ribbon spacing of about 7.3 Å. In contrast, two *anti* hydrogens of thiourea are connected with oxygen



**Figure 8.11** Zigzag ribbon formed by 5-nitrobarbiturate ions and the hydrogen-bonded scheme in the thiourea-anion host lattice of **5**, viewed along the *b* axis. The large spheres represent  $(CH_3)_4N^+$  ions. Symmetry transformation:  $a (1 + x, y, z)$



**Figure 8.12** Perspective view of the crystal structure of **5** viewed along the *a* axis. Symmetry transformations: *a*  $(1 - x, 1 - y, -z)$ , *b*  $(x, y, -1 + z)$  and *c*  $(1 - x, 2 - y, 1 - z)$

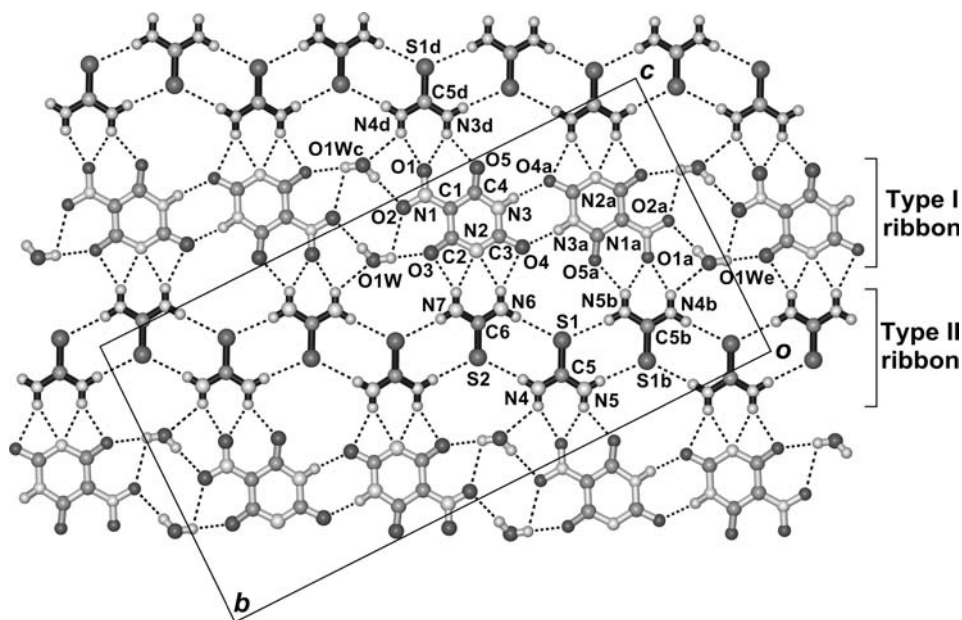
atoms of the nitro group, yielding a smaller channel with an inter-ribbon spacing of 3.6 Å. In this latter case,  $\pi - \pi$  interactions exist between closely spaced ribbons, which reinforce the role of thiourea acting as a clamp. Interestingly, the sulfur atom of the thiourea molecule has no obvious interaction with adjacent molecules, which is a rare phenomenon among the inclusion compounds of thiourea. The closest distance between two sulfur atoms in the crystal structure is about 6.3 Å. The tetramethylammonium cations occupy the larger channels in double columns.

#### 8.2.3.2 Crystal structure of $[(C_2H_5)_4N^+]_2(C_4HN_3O_5)^{2-} \cdot 2(NH_2)_2CS \cdot H_2O$ (**6**)

The structure of complex **6** [25] can be considered as a combination of two major structural components designated as type **I** and type **II**, which are interlinked by N–H···O and N–H···N hydrogen bonds (Figure 8.13).

The type **I** component is a 5-nitrobarbiturate-water ribbon. Centrosymmetric 5-nitrobarbiturate dianions are connected to each other to form a dimer in the para-to-para fashion. Adjacent dimers are further bridged by a pair of inversion-related water molecules **O1w** and **O1wc** to form an infinite ribbon along the [011] direction. The type **II** component consists of thiourea molecules that are alternately linked together in the shoulder-to-shoulder fashion to form an almost planar ribbon extending parallel to the type **I** ribbon. With these two kinds of ribbons arranged side by side, atoms O(1), O(5), O(3), O(4), and N(2) of the 5-nitrobarbiturate-water ribbon are cross-linked with atoms N(4), N(5), N(6), and N(7) belonging to the thiourea ribbons to yield a slightly puckered layer normal to [100]. The interlayer distance is about 7.4 Å and the tetraethylammonium cations are located at  $x \approx 1/4$  and  $3/4$ .





**Figure 8.13** Projection of the crystal structure of complex **6** viewed along the *a* axis. Symmetry transformations: *a* ( $x, \frac{1}{2} - y, \frac{3}{2} + z$ ), *b* ( $x, \frac{1}{2} - y, \frac{1}{2} - z$ ), *c* ( $1 - x, 1 - y, 2 - z$ ), *d* ( $x, y, 1 + z$ ) and *e* ( $1 - x, -\frac{1}{2} + y, -\frac{1}{2} + z$ )

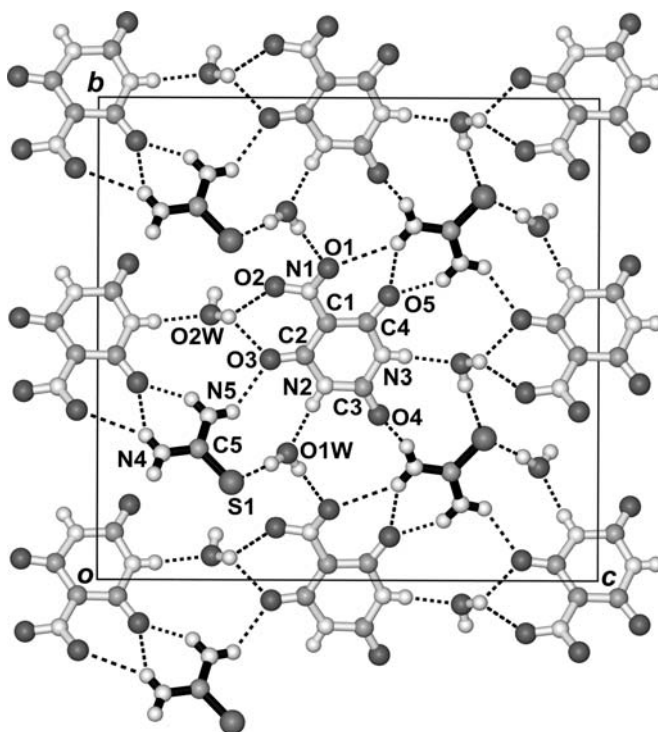
#### 8.2.3.3 Crystal structure of $(n\text{-C}_3\text{H}_7)_4\text{N}^+(\text{C}_4\text{H}_2\text{N}_3\text{O}_5)^- \cdot (\text{NH}_2)_2\text{CS} \cdot 2\text{H}_2\text{O}$ (**7**)

No dimer of nitrobarbiturate occurs in complex **7**. As shown in Figure 8.14, two adjacent nitrobarbiturate ions are separated from each other by bridging thiourea **C(5)** and water molecules **O1w** and **O2w** to form a nitrobarbiturate-thiourea- $(\text{H}_2\text{O})_2$  tetramer motif.

Adjacent tetramers are linked by multiple hydrogen bonds to form infinite chains along the  $[110]$  and  $[1\bar{1}0]$  directions that are further cross-linked to constitute the 3-D channel-type host network shown in Figure 8.15. The well-ordered tetra-*n*-propylammonium cations are accommodated as a double column within each channel.

#### 8.2.3.4 Crystal structure of $[(n\text{-C}_4\text{H}_9)_4\text{N}^+]_2[(\text{C}_4\text{H}_2\text{N}_3\text{O}_5)^-]_2 \cdot (\text{NH}_2)_2\text{CS} \cdot \text{H}_2\text{O}$ (**8**)

In the host layer of complex **8** [25], the two independent 5-nitrobarbiturate ions each form a cyclic hydrogen-bonded centrosymmetric dimer in the para-to-para fashion (Figure 8.16). The two kinds of dimers are alternately interlinked by hydrogen bonds in the ortho-to-ortho fashion to form a zigzag ribbon along the  $[01\bar{1}]$  direction, which is similar to the zigzag chain found in complex **5**. In addition, the ribbons are cross-linked by thiourea **S(1)** and water molecule **O1w** to form a puckered layer with quatrefoil-like channels parallel to the *a* axis. The  $(n\text{-C}_4\text{H}_9)_4\text{N}^+$  ions are sandwiched between adjacent 5-nitrobarbiturate-thiourea anionic layers with an interlayer separation of 7.5 Å. One alkyl chain of the cation extends into the central void of a quatrefoil-like channel to stabilize the wavy layer structure.



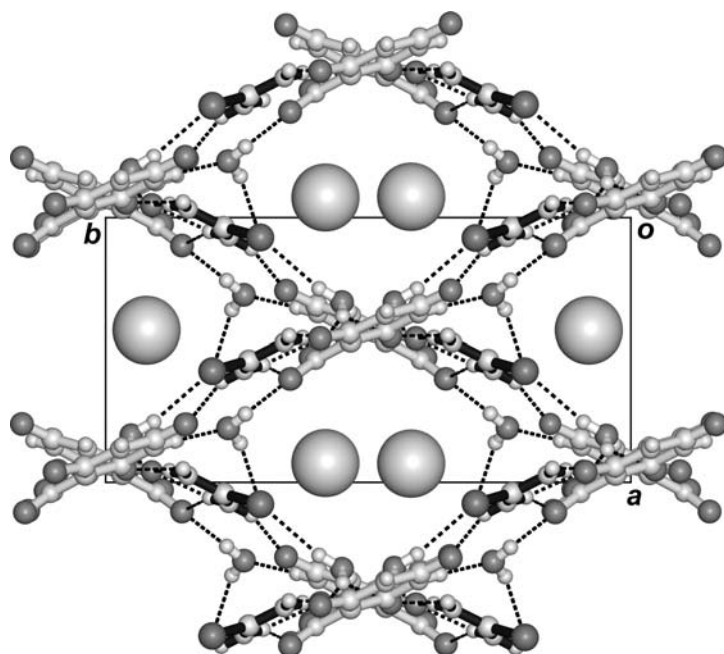
**Figure 8.14** Crystal structure of **7** projected along the *a* axis, showing the hydrogen-bonding environment of a 5-nitrobarbiturate species

Using urea in place of thiourea as a building block led to inclusion compounds **9**, **10** and **11**, which exhibit three connection modes designated as type *D*, type *E* and type *F* in the respective crystal structures (Scheme 8.2).

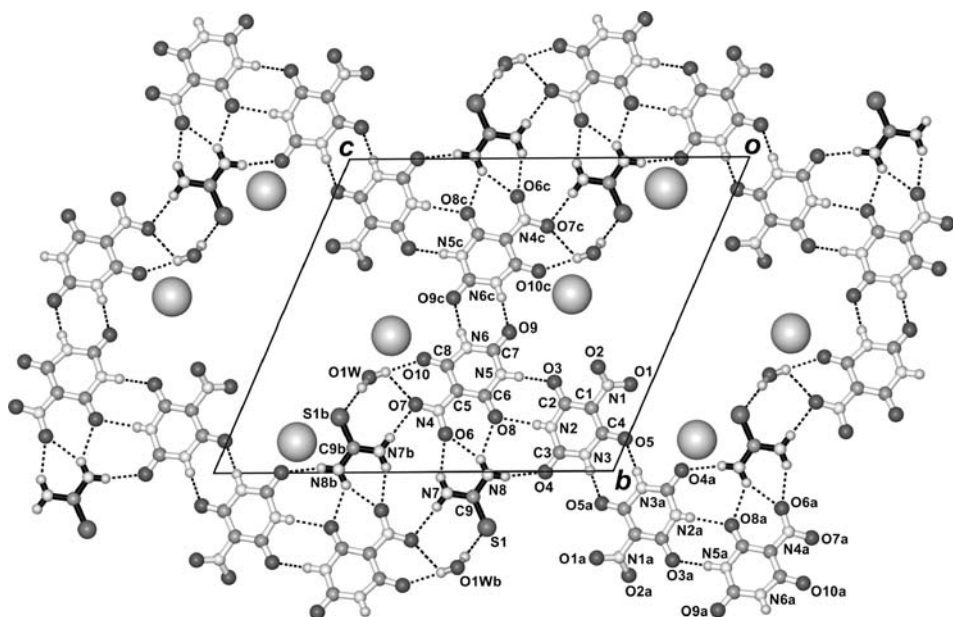
#### 8.2.3.5 Crystal structure of $(C_2H_5)_4N^+(C_4H_2N_3O_5)^- \cdot 2(NH_2)_2CO \cdot 2H_2O$ (**9**)

In this crystal structure, there is no intermolecular interaction between 5-nitrobarbiturate ions [25]. Figure 8.17 shows a portion of a hydrogen-bonded layer viewed along the *a* axis.

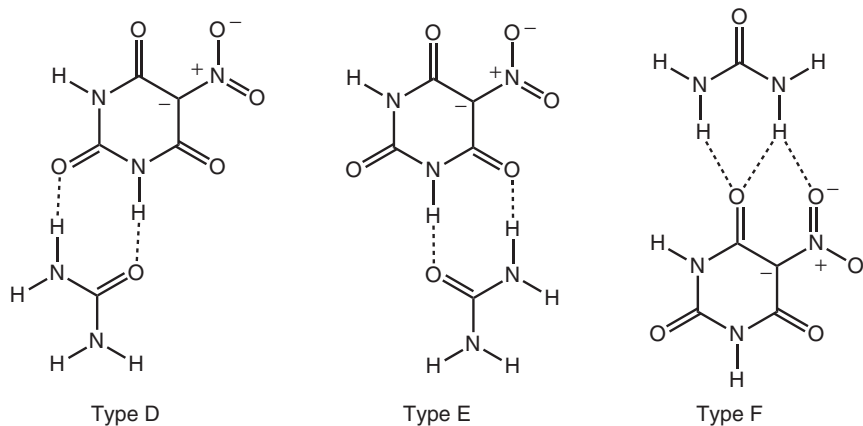
Water molecules play an important role in stabilization of the layer structure. The 5-nitrobarbiturate, urea **C(6)**, and **O2w** molecules are interlinked to form a trimer. Successive nitrobarbiturate-urea-water trimers are further bridged by urea **C(5)** and water molecule **O1w** to form a ribbon running parallel to the *b* axis. The ribbons are cross-linked via three distinguishable modes with urea molecules. Urea **C(6c)** is connected with 5-nitrobarbiturate in type *F* mode (head-to-tail, see Scheme 8.2) by N(7c)–H···O(3), N(6c)–H···O(1) and N(6c)–H···O(3) hydrogen bonds; urea **C(5d)** is linked with 5-nitrobarbiturate **O(1)** in a typical ortho shoulder-to-shoulder mode (type *E*) by N(4d)–H···O(5) and N(3)–H···O(6d) hydrogen bonds, and urea **C(5)** donates its two anti-hydrogen atoms to the nitro group of the nitrobarbiturate ion to form triple hydrogen bonds.



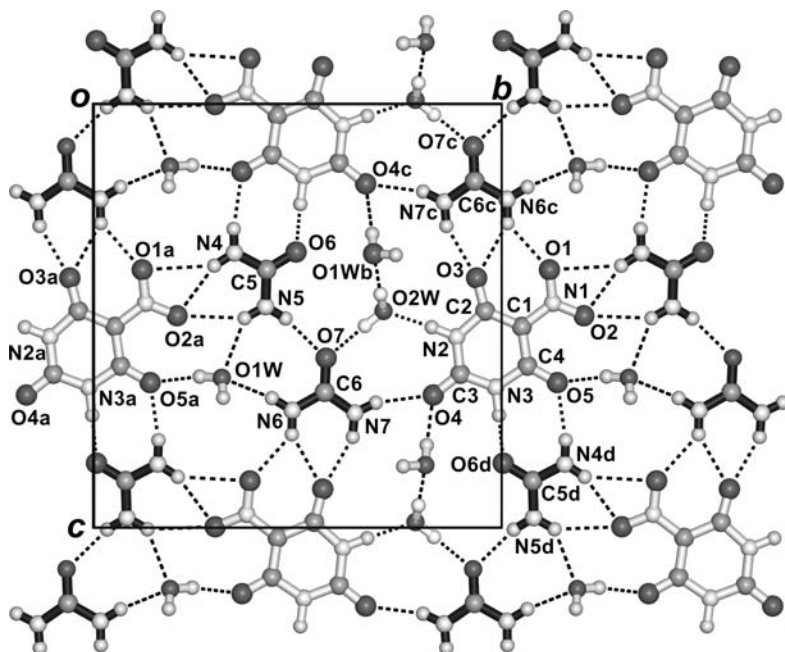
**Figure 8.15** Crystal structure of **7** projected along the *c* axis



**Figure 8.16** Projection of the crystal structure of complex **8** viewed along the *a* axis. Symmetry transformations: *a* ( $2 - x, -y, 2 - z$ ), *b* ( $2 - x, -y, 1 - z$ ) and *c* ( $1 - x, 1 - y, 1 - z$ )



**Scheme 8.2** The nitrobarbiturate-urea connection modes found in hydrogen-bonded complexes: type **D** (para shoulder-to-shoulder), type **E** (ortho shoulder-to-shoulder) and type **F** (head-to-tail)



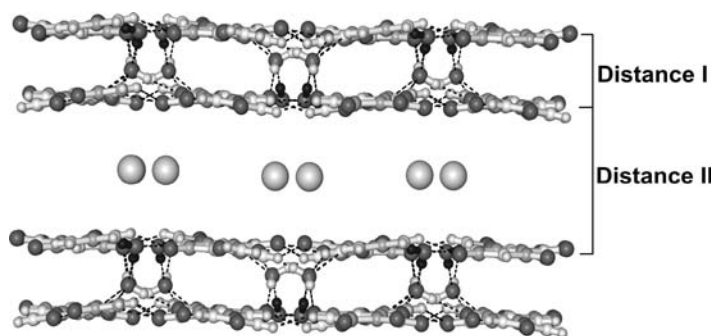
**Figure 8.17** Projection view of the crystal structure of complex **9** viewed along the *a* axis. Symmetry transformations: *a* ( $x, 1 - y, z$ ), *b* ( $2 + x, 1 + y, 1 + z$ ), *c* ( $x, \frac{3}{2} - y, -\frac{1}{2} + z$ ) and *d* ( $x, \frac{3}{2} - y, \frac{1}{2} + z$ )

In addition, water molecule **O1w** and its symmetry equivalents are alternatively connected to two layers by  $\text{O1w-H}\cdots\text{O(5a)}$ ,  $\text{N(5)-H}\cdots\text{O1w}$ ,  $\text{N(6)-H}\cdots\text{O1w}$  (in the same layer) and  $\text{O1wb-H}\cdots\text{O(4c)}$ ,  $\text{O2w-H}\cdots\text{O1wb}$  (in the other layer) hydrogen bonds to generate a double-layer structure (Figure 8.18). The thickness of the double layer is about 3.4 Å (distance I), and the separation between adjacent double layers is about 6.8 Å (distance II). The former distance indicates that  $\pi - \pi$  interaction exists within a double layer.

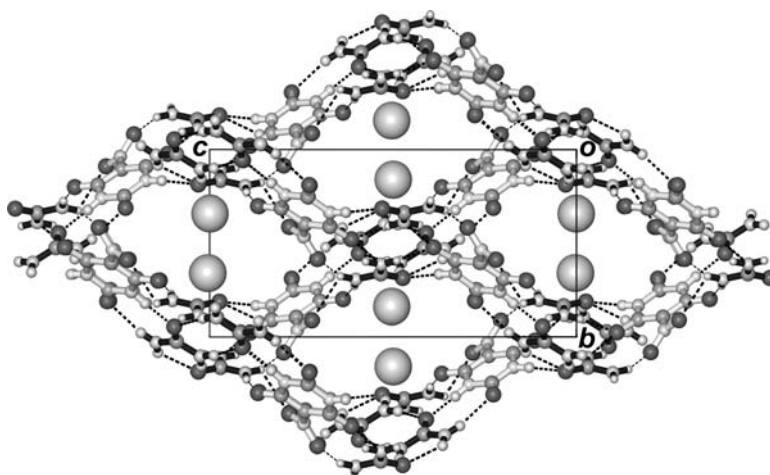
It is notable that in this structure the two independent water molecules play different roles. **O2w** (Figure 8.17) is responsible for connecting adjacent molecules, which lie almost in the same plane, while **O1w** is responsible for bridging two adjacent layers to generate the double-layer inclusion complex.

#### 8.2.3.6 Crystal structure of $(n\text{-C}_3\text{H}_7)_4\text{N}^+(\text{C}_4\text{H}_2\text{N}_3\text{O}_5)^- \cdot 2(\text{NH}_2)_2\text{CO}$ (**10**)

Complex **10** [25] exhibits a 3-D structure with channels extending parallel to the [100], [101], and [011] directions. Figure 8.19 shows the structure viewed along the [100]

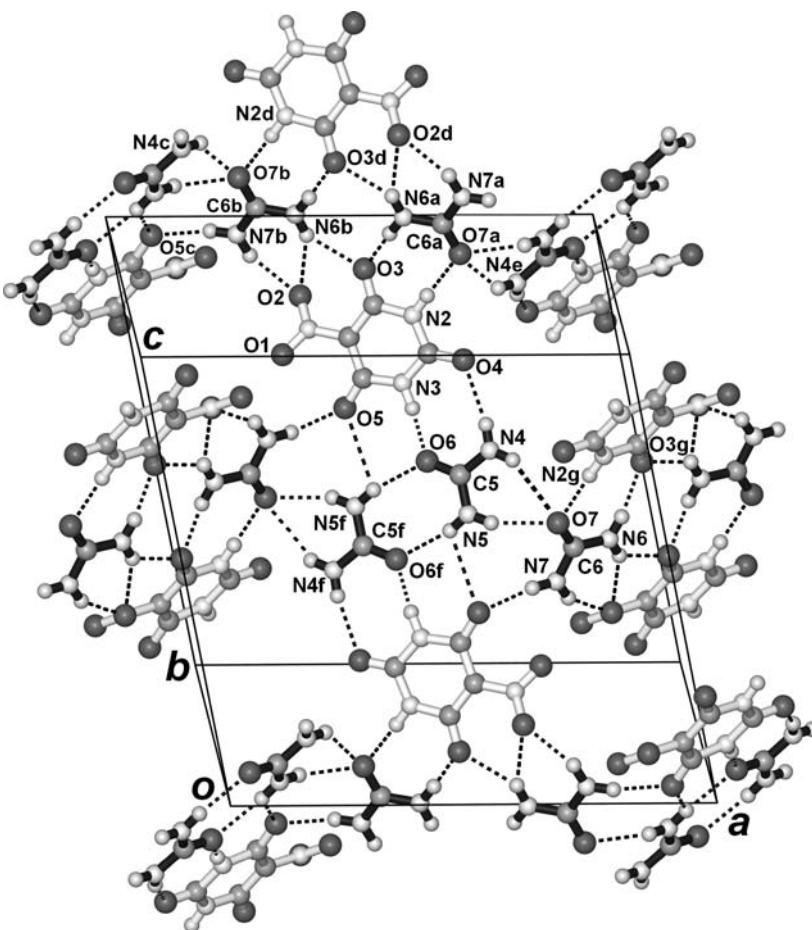


**Figure 8.18** Perspective view of the crystal structure of complex **9** viewed along the *b* axis



**Figure 8.19** Projection view of the crystal structure of complex **10** viewed along the *a* axis



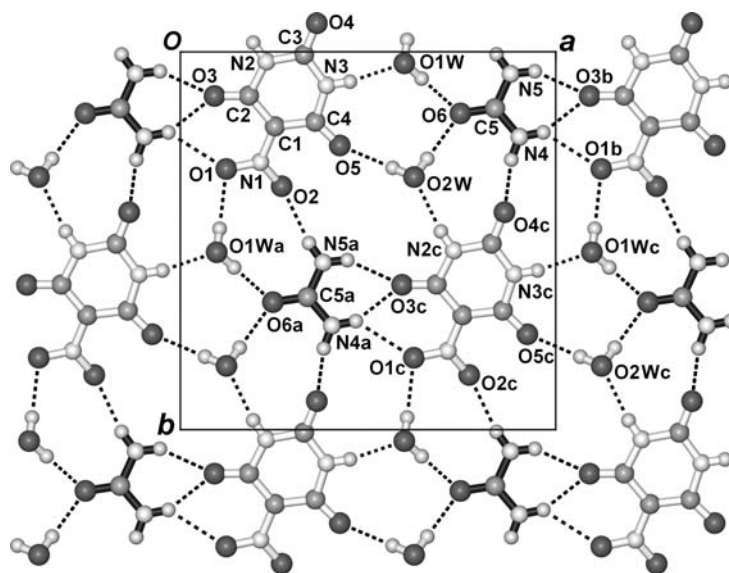


**Figure 8.20** Perspective view of the crystal structure of complex **10** viewed almost along the  $[01\bar{1}]$  direction. Symmetry transformations: a  $(\frac{3}{2} - x, \frac{1}{2} + y, \frac{3}{2} - z)$ , b  $(-\frac{1}{2} + x, \frac{3}{2} - y, 1 + z)$ , c  $(\frac{1}{2} - x, \frac{1}{2} + y, \frac{3}{2} - z)$ , d  $(1 - x, 2 - y, 2 - z)$ , e  $(\frac{3}{2} - x, \frac{1}{2} + y, \frac{3}{2} - z)$ , f  $(1 - x, 1 - y, 1 - z)$  and g  $(\frac{3}{2} - x, -\frac{1}{2} + y, \frac{3}{2} - z)$

direction. The channel framework is built of symmetry-related ribbons which run parallel to the  $[011]$  and  $[0\bar{1}1]$  directions, respectively.

Centrosymmetric pairs of nitrobarbiturates and urea molecules are connected to form a tetramer (Figure 8.20) involving type **E** and **F** connection modes (Scheme 8.2).

These tetramers are further connected by dimeric urea units in type D fashion (para shoulder-to-shoulder) to form wide ribbons along the [011] or  $[0\bar{1}1]$  direction. These equivalent ribbons are alternately linked together to generate a 3-D channel system. The dihedral angle between adjacent planar ribbons which run along the [011] or  $[0\bar{1}1]$  direction is about  $78^\circ$ . Well-ordered tetra-*n*-propylammonium cations are accommodated as a double column within each channel, as illustrated in Figure 8.19.



**Figure 8.21** Perspective view of the crystal structure of complex **11** viewed along the *c* axis. Symmetry transformations: *a*  $(-\frac{1}{2} + x, \frac{1}{2} + y, z)$ , *b*  $(1 + x, y, z)$  and *c*  $(\frac{1}{2} + x, \frac{1}{2} + y, z)$

#### 8.2.3.7 Crystal structure of $(n\text{-C}_4\text{H}_9)_4\text{N}^+(\text{C}_4\text{H}_2\text{N}_3\text{O}_5)^- \cdot (\text{NH}_2)_2\text{CO} \cdot 2\text{H}_2\text{O}$ (**11**)

As shown in Figure 8.21, the 5-nitrobarbiturate ions in **11** [25] are directly linked to independent water molecules **O1w** and **O2w** by donor  $\text{N}(3)\text{--H} \cdots \text{O1w}$  and acceptor  $\text{O2w--H} \cdots \text{O}(5)$  hydrogen bonds, respectively. In addition, each water molecule donates one hydrogen atom to an adjacent urea molecule **C(5)**. The resulting nitrobarbiturate- $(\text{H}_2\text{O})_2$ -urea tetramers are aligned parallel to the *a* axis, and head-to-tail linkage between the nitrobarbiturate and urea moieties (type *F*) yields an infinite chain running in the *a* direction.

The equivalent chains are cross-linked to generate a layer normal to the *c* axis. The well-ordered tetra-*n*-butylammonium cations are alternately arranged and sandwiched between adjacent layers at  $z = 1/4$  and  $3/4$ .

#### 8.2.3.8 Connection mode of 5-nitrobarbiturate

The 5-nitrobarbiturate dimer occurs in three of the four thiourea-anion complexes, but is absent in all three urea complexes. The difference between urea and thiourea in generating complexes **5–11** arises from their relative capability in hydrogen bonding. Normally, a  $\text{N--H} \cdots \text{S}$  hydrogen bond is much weaker than a  $\text{N--H} \cdots \text{O}$  hydrogen bond [30]. The possibility of forming a nitrobarbiturate dimer is closely related to the competition between urea and thiourea for interaction with nitrobarbiturate. In the nitrobarbiturate-urea crystal system, the two components are about equally good as hydrogen-bond formers, and their competition disfavors intermolecular connection between two nitrobarbiturates. However, in the nitrobarbiturate-thiourea system, the  $\text{N--H} \cdots \text{O}$  hydrogen bonds formed between nitrobarbiturate molecules are stronger than  $\text{N--H} \cdots \text{S}$  hydrogen

bonds formed between nitrobarbiturate and thiourea. As a consequence, the dimer of nitrobarbiturate occurs more readily in the thiourea system than in the urea system.

#### 8.2.3.9 Tetra-alkylammonium ions as templates

The large tetra-alkylammonium ions favor the formation of layer-type structures, and the smallest  $\text{Me}_4\text{N}^+$  ion leads to a channel structure. However, using a middle-size tetra-alkylammonium ion, a balance between the two kinds of structures can be achieved by varying the molar ratio of the reactants and the reaction conditions. For complexes **7** and **10**, the channel structure is present, while **6** and **9** adopt a layer structure. In complex **5**, the small tetramethylammonium cation can be placed between two parallel zigzag ribbons. At the same time, the coplanar property of the nitrobarbiturate is changed in order to match the reach of the thiourea molecule, which uses its *syn* or *anti* hydrogens to connect with the oxygen atoms of nitrobarbiturate. In complex **8**, the tetra-*n*-butylammonium cation is large, and hence an almost planar layer is generated.

#### 8.2.3.10 Versatile roles of (thio)urea in the 5-nitrobarbiturate inclusion systems

The series of inclusion compounds **5–8** exhibit thiourea-nitrobarbiturate host lattices that are distinctly different from one another, depending on their stoichiometric ratio in the presence of the tetra-alkylammonium cation and co-crystallized water molecules. Thiourea acts as a strut or clamp in complex **5**, when a small amount of thiourea is used together with a small tetra-alkylammonium ion (molar ratio 5-nitrobarbiturate:thiourea = 2:1). The conventional shoulder-to-shoulder thiourea connection mode occurs in the structure of **6**, which is consistent with the addition of an excess amount of thiourea (5-nitrobarbiturate:thiourea = 1:2) in the synthesis. In complex **7**, the molar ratio of nitrobarbiturate to thiourea is 1:1, and both thiourea and 5-nitrobarbiturate are effectively prevented from forming a dimer or a ribbon. In complex **8**, when a smaller amount of thiourea is present in the system (5-nitrobarbiturate:thiourea = 2:1), all thiourea molecules are separated from one another. The 5-nitrobarbiturates are assembled to form a tetramer, and thiourea acts as a linker to connect the tetramers together with reinforcement by the water molecules. In contrast, in the urea-anion complexes, urea competes with nitrobarbiturate in hydrogen bonding, resulting in three main types of connection modes as shown in Scheme 8.2.

### 8.2.4 Generation and Trapping of Labile Anions

Simple oxo-anions such as  $\text{NO}_3^-$  [6f],  $\text{HCO}_3^-$  [5i],  $\text{HCO}_2^-$  [6d],  $\text{CH}_3\text{CO}_2^-$  [6e] and spirocyclic  $[\text{B}_5\text{O}_6(\text{OH})_4]^-$  [5e] can act as proton donors or acceptors for the construction of novel urea/thiourea host networks. More significantly, some labile anions such as dihydrogen borate, allophanate and 3-thioallophanate anions have been generated *in situ* and stabilized in the host lattices by cooperative hydrogen-bonding interactions [31] with their nearest neighbors.

#### 8.2.4.1 Dihydrogen borate

Boric acid,  $\text{B}(\text{OH})_3$  (also known as orthoboric acid and commonly rewritten as  $\text{H}_3\text{BO}_3$ ) generally behaves not as a Brønsted acid to form the conjugate-base anion  $\text{BO}(\text{OH})_2^-$ , but rather as a Lewis acid to give the tetrahedral anion  $\text{B}(\text{OH})_4^-$  [32]. In dilute aqueous

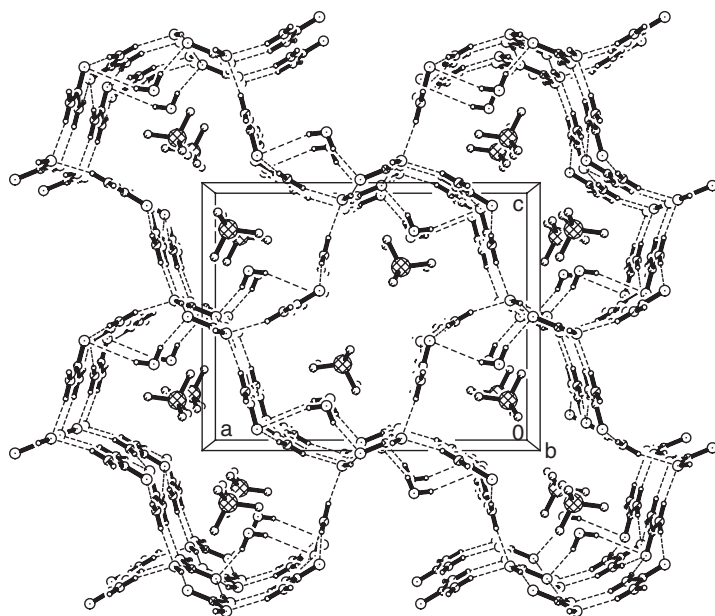


solution this weakly monobasic acid exists almost exclusively as an equilibrated mixture of the undissociated molecule  $\text{B(OH)}_3$  and the tetrahydroxyborate anion  $\text{B(OH)}_4^-$ . At concentrations above 0.1 M, secondary equilibria involving condensation reactions of the two dominant monomeric species give rise to oligomers such as the triborate monoanion  $[\text{B}_3\text{O}_3(\text{OH})_4]^-$ , the triborate dianion  $[\text{B}_3\text{O}_3(\text{OH})_5]^{2-}$ , the tetraborate  $[\text{B}_4\text{O}_5(\text{OH})_4]^{2-}$  and the pentaborate  $[\text{B}_5\text{O}_6(\text{OH})_4]^-$  [33].

The Brønsted-acid behavior of  $\text{B(OH)}_3$  has been observed in the gas phase by mass spectrometric studies [34]. Complete deprotonation of  $\text{B(OH)}_3$  and  $\text{B(OH)}_4^-$  leads to  $\text{BO}_3^{3-}$  and  $\text{BO}_4^{5-}$ , respectively, which occur in a wide range of solid anhydrous metal borates [35]. The natural borate vimsite, which has the stoichiometric formula  $\text{Ca}[\text{BO}(\text{OH})_2]_2$ , does not contain the discrete  $\text{BO}(\text{OH})_2^-$  ion, and the crystal structure features an infinite anionic chain  $[\text{BO}(\text{OH})_2]_n^-$  with ether-like oxygen atoms bridging  $sp^3$  boron centers [36]. The simple dihydrogen borate anion  $\text{BO}(\text{OH})_2^-$  has thus far been reported to exist in  $\text{Cu}_2[\text{BO}(\text{OH})_2](\text{OH})_3$  [37] and  $(\text{Et}_4\text{N}^+)_2[\text{BO}(\text{OH})_2]_2 \cdot \text{B(OH)}_3 \cdot 5\text{H}_2\text{O}$  [38]. Crystal structure analysis of  $\text{Cu}_2[\text{BO}(\text{OH})_2](\text{OH})_3$ , based on powder X-ray diffraction data, did not permit a determination of the hydrogen-atom positions, but a report on  $(\text{Et}_4\text{N}^+)_2[\text{BO}(\text{OH})_2]_2 \cdot \text{B(OH)}_3 \cdot 5\text{H}_2\text{O}$  demonstrated that in a crystalline solid both  $\text{B(OH)}_3$  and its conjugate base  $\text{BO}(\text{OH})_2^-$  could coexist. Using our methodology, the transient  $\text{BO}(\text{OH})_2^-$  ion has been generated *in situ* from  $\text{B(OH)}_3$  and incorporated into the following channel-type inclusion compound.

*Channel-type structure of  $(\text{CH}_3)_4\text{N}^+\text{BO}(\text{OH})_2^- \cdot 2(\text{NH}_2)_2\text{CO} \cdot \text{H}_2\text{O}$  (12).* In compound **12** [5c], both independent urea molecules and the  $\text{BO}(\text{OH})_2^-$  ion in the asymmetric unit occupy special positions of symmetry  $m$  such that only the C–O or B–O bond lies on the mirror plane. A projection of the crystal structure along the [010] direction is presented in Figure 8.22. The host lattice consists of a parallel arrangement of unidirectional channels whose cross-section has the shape of a peanut. The diameter of each spheroidal half is about 7.04 Å, and the separation between two opposite walls at the waist of the channel is about 5.85 Å. The well-ordered tetramethylammonium cations are accommodated in double columns within each channel.

As shown in Figure 8.23, the two independent urea molecules are alternately connected by hydrogen bonds to form a zigzag ribbon running parallel to  $b$ , which has a highly twisted configuration. The  $\text{BO}(\text{OH})_2^-$  ions generated from successive inversion centers aligned parallel to the  $b$  direction are linked into a nearly planar, zigzag ribbon by O–H...O hydrogen bonds. Each  $\text{BO}(\text{OH})_2^-$  anion is further connected to two urea molecules belonging to different neighboring urea ribbons (mutually related by a  $2_1$  axis) by four N–H...O hydrogen bonds that extend outward on the same side of the ribbon (Figure 8.23). The dihedral angles between the planar moieties involved in the cross-linkage are 106.7° for urea with urea and 126.7° for urea with  $\text{BO}(\text{OH})_2^-$ , so that a cross-section of each junction is shaped like the capital letter Y. With these Y-shaped junctions oriented normal to (010) and concentrated at the layers  $y = 1/4$  and  $3/4$ , a unidirectional channel-like host lattice is formed (Figure 8.22). Notably, the water molecule makes no contribution to the construction of the host network, and its role is limited to that of a spectator guest that forms a donor O–H...O hydrogen bond to a  $\text{BO}(\text{OH})_2^-$  oxygen host atom.



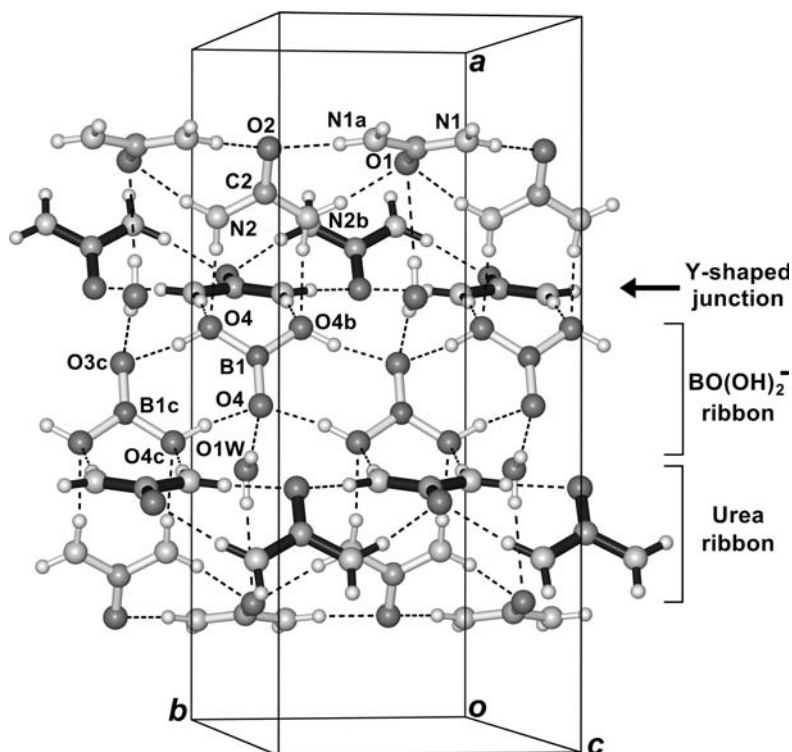
**Figure 8.22** Crystal structure of  $(\text{CH}_3)_4\text{N}^+ \text{BO}(\text{OH})_2^- \cdot 2(\text{NH}_2)_2\text{CO} \cdot \text{H}_2\text{O}$  (**12**) showing the channels extending parallel to the  $b$  axis and the enclosed cations

The elusive species  $\text{BO}(\text{OH})_2^-$  generated from  $\text{B}(\text{OH})_3$  has crystallographically imposed symmetry  $m$ , and the measured dimensions show that structural formula with a formal B–O double bond (Figure 8.24) makes a significant contribution to its electronic structure.

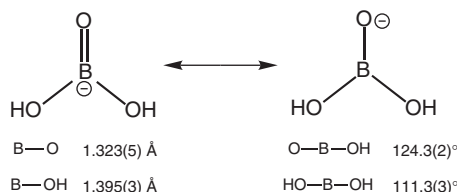
#### 8.2.4.2 Allophanate

Allophanic acid (ureidoformic acid,  $\text{NH}_2\text{CONHCOOH}$ ) is an elusive compound which is known only in the form of its derivatives. Ester-like organic allophanates have been prepared from primary, secondary, and tertiary alcohols and from phenols [39], whereas inorganic allophanate salts are unstable and easily hydrolyzed by  $\text{H}_2\text{O}$  to  $\text{CO}_2$ , urea, and carbonate [40]. The first two crystalline complexes containing host lattices built of urea molecules and allophanate ions were prepared and structurally characterized in 1995 [5h].

*Cage-type structure of  $(\text{CH}_3)_4\text{N}^+ \text{NH}_2\text{CONHCO}_2^- \cdot 5(\text{NH}_2)_2\text{CO}$  (**13**).* In the 3-D network structure of **13** [5h], each tetramethylammonium cation is enclosed in a box-like cage, the separation between individual pairs of opposite walls being 8.3 ( $b/2$ ), 7.8 ( $c/2$ ) and 7.3 Å, respectively (Figure 8.25). The allophanate ion and one independent urea molecule generates a nearly planar double ribbon running in the  $a$  direction (Figure 8.26). Two such double ribbons which are arranged parallel to the (010) plane constitute a pair of opposite walls of the hexahedral cage. Another pair of walls is formed by straight chains of urea molecules extending along the [101] direction, which are derived from two other independent urea molecules that are alternately linked by hydrogen bonds in a



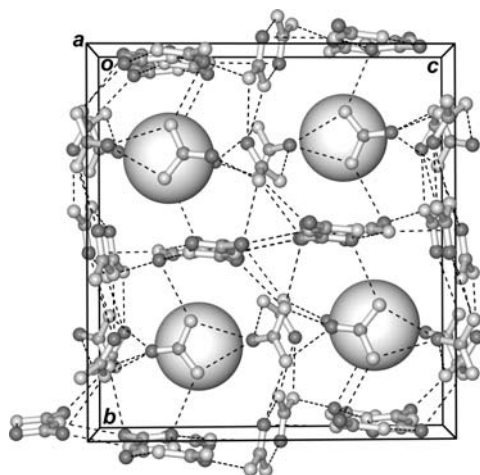
**Figure 8.23** Hydrogen-bonding scheme in the urea- $\text{BO}(\text{OH})_2^-$  host layer of complex **12**. Broken lines represent hydrogen bonds. Symmetry transformations:  $a$  ( $x, \frac{1}{2} - y, z$ ),  $b$  ( $x, \frac{3}{2} - y, z$ ) and  $c$  ( $1 - x, 2 - y, -z$ )



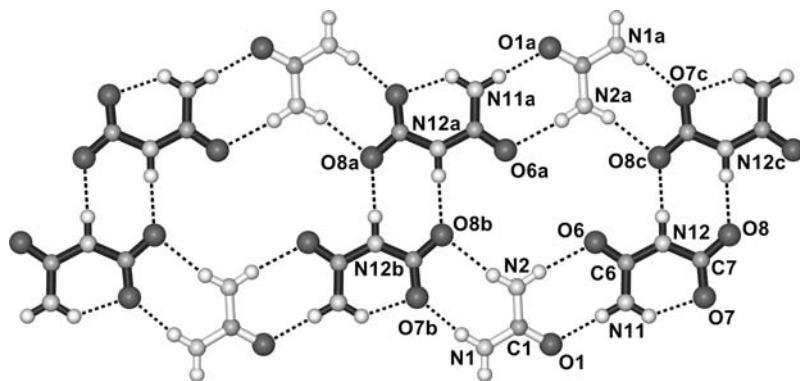
**Figure 8.24** Valence-bond structural formulas for the dihydrogen borate ion,  $\text{BO}(\text{OH})_2^-$ , and its measured dimensions in inclusion compound **12**

head-to-tail mode (Figure 8.27). The third pair of walls of the cage is constructed from the remaining two urea molecules.

The heretofore uncharacterized allophanate ion adopts a planar configuration that facilitates the formation of an intermolecular  $\text{N}-\text{H} \cdots \text{O}$  hydrogen bond (Figure 8.26). The relative instability of the parent allophanic acid is consistent with this hydrogen-bonded ring structure, as protonation at the exocyclic carboxyl oxygen would result in an overall reduction of  $\pi$ -electron delocalization.

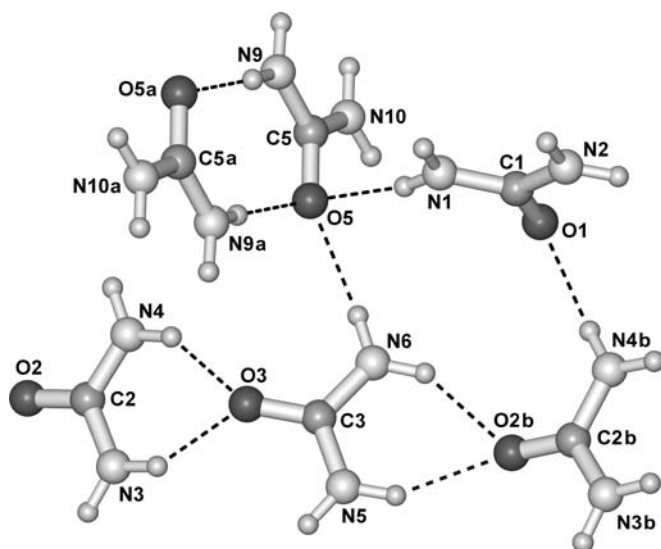


**Figure 8.25** Crystal structure of  $(\text{CH}_3)_4\text{N}^+\text{NH}_2\text{CONHCO}_2^- \cdot 5(\text{NH}_2)_2\text{CO}$  (**13**). Broken lines represent hydrogen bonds and the enclosed  $(\text{CH}_3)_4\text{N}^+$  ions are represented by large shaded spheres

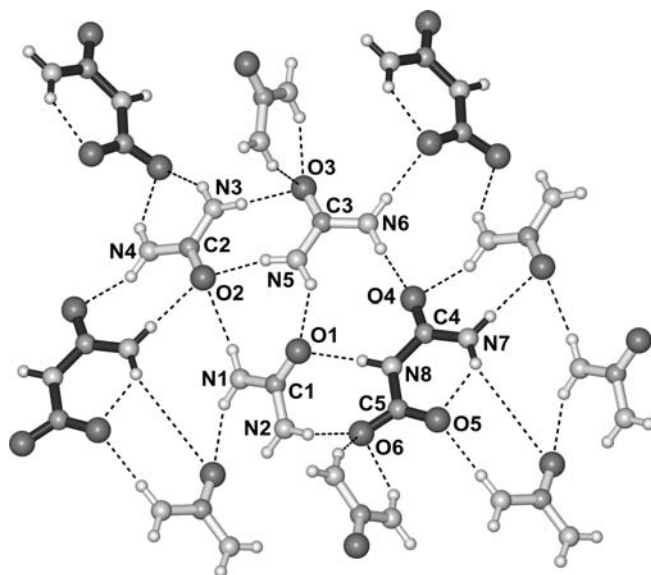


**Figure 8.26** Perspective view of a portion of the planar double ribbon built by urea molecules and allophanate ions in **13**. Symmetry transformation: *a*  $(2 - x, -y, -z)$ , *b*  $(1 + x, y, z)$  and *c*  $(1 - x, -y, -z)$

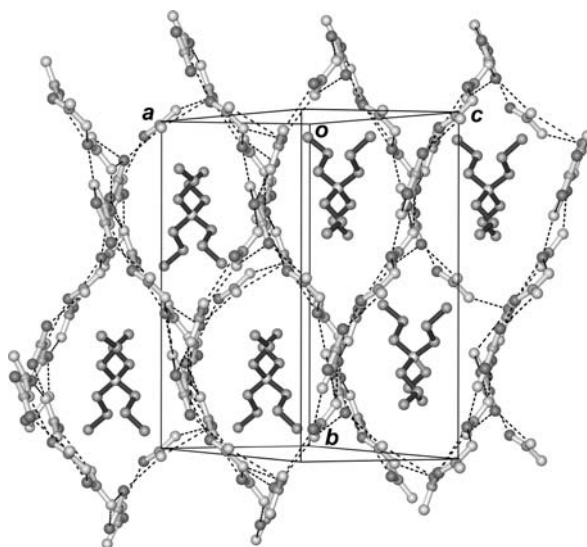
*Intersecting channel-type structure of  $(n\text{-C}_3\text{H}_7)_4\text{N}^+\text{NH}_2\text{CONHCO}_2^- \cdot 3(\text{NH}_2)_2\text{CO}$  (**14**).* The hydrogen-bonding scheme in the crystal structure of **14** [5h] is shown in Figure 8.28. The three independent urea molecules in the asymmetric unit, together with the allophanate ion, generate a wider ribbon running through the structure in the direction of [110]. At a location which is one-half translation away along the *c* axis, there is a similar ribbon which extends in the  $[\bar{1}10]$  direction. Finally, intermolecular hydrogen bonding weaves these two crossed series of ribbons into a 3-D framework containing open unidirectional channels, each of which accommodates a single columnar stack of well-ordered  $(n\text{-C}_3\text{H}_7)\text{N}^+$  cations (Figure 8.29).



**Figure 8.27** Head-to-tail hydrogen-bonded urea chain extending along the [101] direction and the shoulder-to-shoulder dimer of urea in the crystal structure of **13**. Symmetry transformations:  $a$  ( $2 - x, 1 - y, -z$ ) and  $b$  ( $x - \frac{1}{2}, \frac{1}{2} - y, z - \frac{1}{2}$ )



**Figure 8.28** Perspective view of a portion of the wide urea-allophanate ribbon linked by hydrogen bonds in  $(n\text{-C}_3\text{H}_7)_4\text{N}^+\text{NH}_2\text{CONHCO}_2^- \cdot 3(\text{NH}_2)_2\text{CO}$  (**14**)



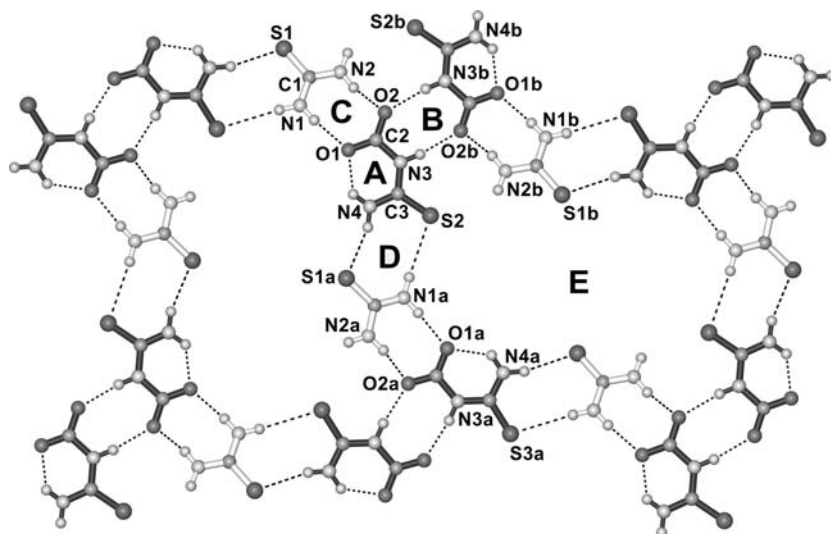
**Figure 8.29** Crystal structure of **14** showing two series of urea-allophanate ribbons running parallel to the  $[110]$  and  $[\bar{1}10]$  directions and their interlinkages by hydrogen bonds to form channels extending parallel to the  $[101]$  direction

#### 8.2.4.3 Thioallophanate

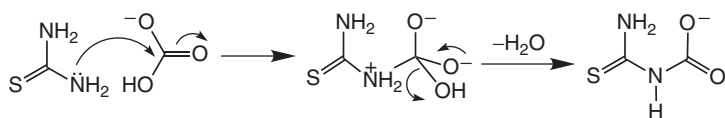
The thio analog of allophanic acid, namely 3-thioallophanic acid,  $\text{H}_2\text{NCSNHCOOH}$ , and its salts are hitherto unknown, although many esters of *N*-substituted 3-thioallophanic acid have been prepared as ligands for transition metal complexes [41]. Based on the methodology employed in the stabilization of dihydrogen borate and allophanate anions, the elusive 3-thioallophanate ion can likewise be trapped in the host lattice of the layer-type thiourea-anion inclusion compound  $(n\text{-C}_4\text{H}_9)_4\text{N}^+ \text{H}_2\text{NCSNHCO}_2^- \cdot (\text{NH}_2)_2\text{CS}$  (**15**) [42].

*Layer-type structure of  $(n\text{-C}_4\text{H}_9)_4\text{N}^+ \text{H}_2\text{NCSNHCO}_2^- \cdot (\text{NH}_2)_2\text{CS}$  (15).* The hydrogen bonding interactions in the host lattice of **15** are best described with the aid of Figure 8.30. The 3-thioallophanate anion forms an intramolecular hydrogen bond (motif **A**), and symmetry-related anions are joined together by pairs of strong  $\text{N-H} \cdots \text{O}^-$  hydrogen bonds to form dimer **B**. The anionic dimers are further linked by thiourea molecules in two different modes. The carboxylate group of 3-thioallophanate is directly joined to the thiourea molecule by a pair of charge-assisted  $\text{N}_{\text{anti}}\text{-H} \cdots \text{O}^-$  hydrogen bonds in motif **C**, while the thiourea end is connected to the other thiourea molecule by a pair of strong  $\text{N}_{\text{syn}}\text{-H} \cdots \text{S}$  hydrogen bonds in a shoulder-to-shoulder manner with ring motif **D**. As a result, an essentially planar host layer with large voids **E** is generated. The mean plane of the host layer matches the (101) family of planes, with an interlayer spacing of about 7.9 Å for accommodation of the organic cation, which has two alkyl legs inserted into the void of an adjacent layer.

The formation of the elusive 3-thioallophanate anion by condensation between thiourea and the bicarbonate anion in methanol is noteworthy, as our attempted synthesis via



**Figure 8.30** Projection showing the hydrogen bonding interactions in the host lattice of **15** constructed by thiourea molecules and 3-thioallophanate anions. The host layer matches the (101) family of planes. Symmetry transformations: *a* ( $\frac{3}{2} - x, y - \frac{1}{2}, \frac{1}{2} - z$ ) and *b* ( $1 - x, -y, 1 - z$ )

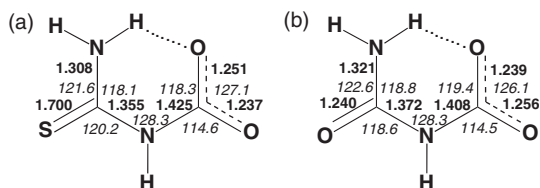


**Scheme 8.3** Possible mechanism of formation of the thioallophanate ion

hydrolysis of the corresponding ester (by analogy to the preparation of the allophanate anion) was unsuccessful [5g]. A plausible mechanism is illustrated in Scheme 8.3. Furthermore, attempts to obtain analogous 3-thioallophanate inclusion compounds using the smaller tetraalkylammonium cations  $(\text{CH}_3)_4\text{N}^+$ ,  $(\text{C}_2\text{H}_5)_4\text{N}^+$  and  $(n\text{-C}_3\text{H}_7)_4\text{N}^+$  were unsuccessful, demonstrating the importance of host-guest selectivity in the formation of the thiourea-anion host framework.

The cyclic structure and molecular dimensions of the 3-thioallophanate ion is comparable with those of the allophanate ion (Figure 8.31). As expected, the C–O bond involved in intramolecular hydrogen bonding in the 3-thioallophanate anion [1.251(3) Å] is longer than that in the allophanate anion [1.239(5) Å]. The previously characterized allophanate anion is able to form hexahedral cage, elliptical channel and hexagonal columnar host lattices with urea molecules in three distinct crystalline inclusion compounds containing tetramethyl- [5h], tetra-*n*-propyl- [5h], and (2-hydroxyethyl)trimethylammonium ions, respectively [5g]. However, owing to the longer N–H···S distance in the thiourea-3-thioallophanate host layer and the use of a more bulky tetra-*n*-butylammonium template, an entirely different layer-type inclusion topology was generated.





**Figure 8.31** Comparison of molecular dimensions of the (a) 3-thioallophanate and (b) allophanate anion [5h]. Bond lengths and angles are shown in different fonts and standard deviations of individual bond lengths and angles are 0.003 Å and 0.2° in (a), and 0.005 Å and 0.3° in (b), respectively

This study further demonstrates that an elusive species such as the 3-thioallophanate anion can be generated *in situ* and stabilized in a crystalline inclusion compound through hydrogen-bonding interactions with neighboring molecules, and here thiourea functions as a supramolecular stabilizing agent.

### 8.3 Stabilization of Cyclic Oxocarbon Dianions by Hydrogen Bonding with Urea/Thiourea

We next turned our attention to oxo-anions bearing divergent hydrogen-bond accepting sites as effective building components, and among the most suitable molecular species for this purpose are the whole series of monocyclic oxocarbon anions.

#### 8.3.1 Cyclic Oxocarbon Acids

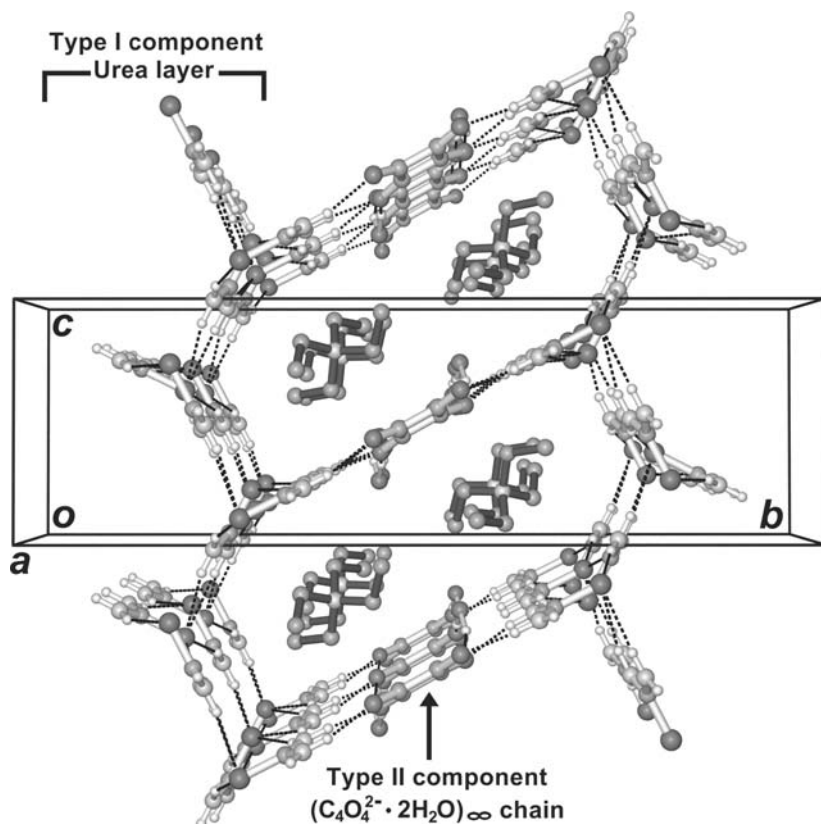
The monocyclic oxocarbon anions  $C_nO_n^{2-}$  ( $n = 3$ , deltate;  $n = 4$ , squarate;  $n = 5$ , croconate;  $n = 6$ , rhodizonate) are formed after full deprotonation of the corresponding parent acids. These anions constitute a series of nonbenzenoid aromatic systems whose historical significance and interesting chemistry are well-documented in several books and reviews [43]. Notably, they possess unusually high molecular symmetry as the two negative charges are delocalized over the ring system so that they can serve as divergent multi-site hydrogen-bond acceptors for the construction of new supramolecular assemblies.

#### 8.3.2 Trapping Cyclic Oxocarbon Dianions in the Crystalline State

##### 8.3.2.1 Squarate Complexes

The squarate ion is utilized in forming the host lattices of the inclusion complexes of  $[(C_2H_5)_4N^+]_2C_4O_4^{2-} \cdot 4(NH_2)_2CS \cdot 2H_2O$  (**16**),  $[(C_2H_5)_4N^+]_2C_4O_4^{2-} \cdot 6(NH_2)_2CS$  (**17**) and  $[(C_2H_5)_4N^+]_2C_4O_4^{2-} \cdot 2(NH_2)_2CS \cdot 2H_2O$  (**18**) [6a]. In this section, a common thiourea linkage pattern is used to correlate the structural characteristics of these inclusion complexes and the previously reported complex  $(n-C_3H_7)_4N^+HCO_2^- \cdot 3(NH_2)_2CS \cdot H_2O$  (**19**) [6d] in the context of solid-state supramolecular assembly.

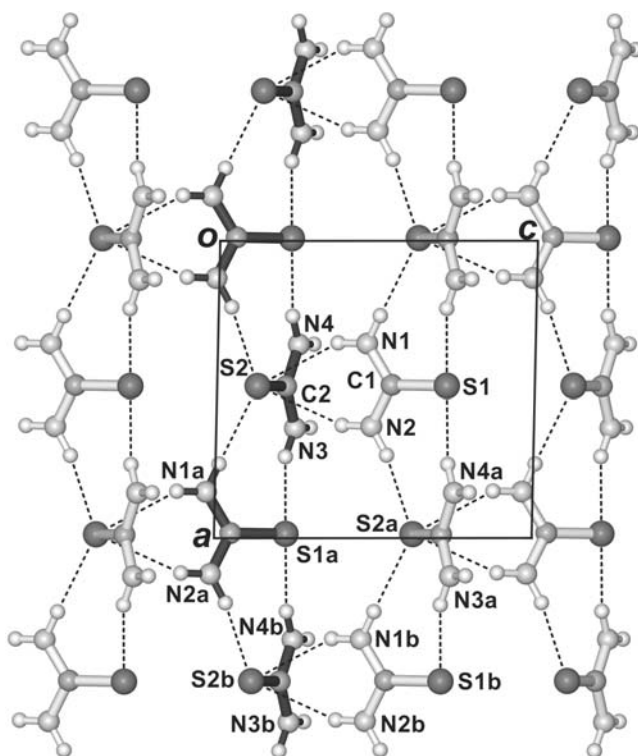




**Figure 8.32** Channel-type crystal structure of  $[(\text{C}_2\text{H}_5)_4\text{N}^+]_2\text{C}_4\text{O}_4^{2-} \cdot 4(\text{NH}_2)_2\text{CS} \cdot 2\text{H}_2\text{O}$  (**16**). Note that the nearly rectangular channels are viewed end-on

Channel-type structure of  $[(\text{C}_2\text{H}_5)_4\text{N}^+]_2\text{C}_4\text{O}_4^{2-} \cdot 4(\text{NH}_2)_2\text{CS} \cdot 2\text{H}_2\text{O}$  (**16**). The host lattice of **16** [6a] contains large 1-D rectangular channels extending parallel to the *a*-axial direction, as illustrated in Figure 8.32.

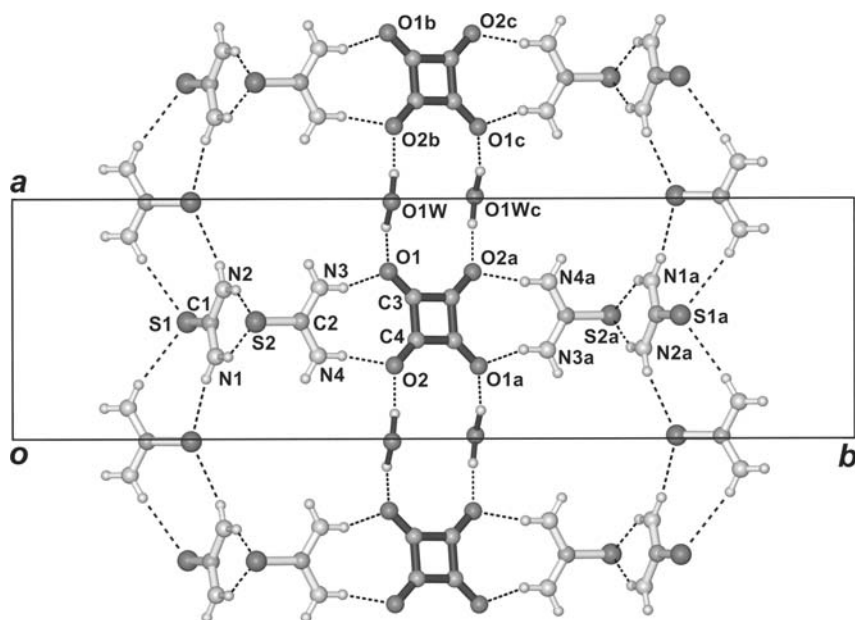
The channel framework is built of a parallel arrangement of puckered thiourea layers that are interlinked by straight chains made of squarate dianions and water molecules via  $\text{N}-\text{H} \cdots \text{O}^-$  (squarate) hydrogen bonds. The host lattice of complex **16** can be considered as a combination of two major structural components (designated type **I** and type **II**) that are interlinked by  $\text{N}-\text{H} \cdots \text{O}^-$  (squarate) hydrogen bonds. The type **I** component is built of two independent thiourea molecules **C(1)** (containing atoms C1, S1, N1 and N2, and hereafter referred to by naming the carbon atom for convenience) and **C(2)**. They are alternately linked together by  $\text{N}-\text{H} \cdots \text{S}$  hydrogen bonds in a shoulder-to-shoulder fashion to form a zigzag and puckered thiourea ribbon extending along the *a*-axial direction, as shown in Figure 8.33. With these zigzag and puckered thiourea ribbons arranged side by side, the sulfur atom of each **C(2)** molecule forms two additional acceptor hydrogen bonds with an adjacent thiourea ribbon through N(1) and N(2) of molecule **C(1)** in a folded head-to-tail fashion, such that this type of cross-linkage gives a puckered layer



**Figure 8.33** Hydrogen-bonded puckered layer in **16** formed by the cross-linkage of adjacent thiourea ribbons; the mean plane of this layer is at  $y = 1/4$ . Broken lines represent hydrogen bonds, and atom types are distinguished by size and shading. Symmetry transformations:  $a (\frac{1}{2} + x, \frac{1}{2} - y, \frac{1}{2} + z)$  and  $b (1 + x, y, z)$

which is normal to the  $b$  axis. These puckered thiourea layers constitute pairs of opposite walls of the large channels.

The type **II** component is of great interest from the structural point of view. Adjacent squarate dianions, each occupying an inversion center, are bridged by a pair of centrosymmetrically related water molecules with a pair of hydrogen bonds to yield a seat-belt-like chain running parallel to the  $a$  axis (Figure 8.34). Each linear  $(C_4O_4^{2-} \cdot 2H_2O)_\infty$  chain contains 14-membered rings each constituted by two pairs of distinct types of hydrogen bonds related by an inversion center. The pattern can be described by the 'chain of rings' notation  $C_2^2(7)[R_4^4(14)]$  according to the graph set topological terminology for the analysis of hydrogen-bonded network patterns [18]. Note that the water molecules do not take part in acceptor hydrogen bonding with other structural units, and the thiourea molecules of type **C(2)** reach out from both sides of each thiourea layer to facilitate their linkage to the squarate-water chains. This squarate-water chain links two adjacent puckered thiourea layers together by forming two pairs of  $N-H \cdots O^-$  (squarate) hydrogen bonds to generate two side walls of a large channel, and their mean planes match the  $(02\bar{1})$  family of planes. Combination of the type **I** and type **II** structural

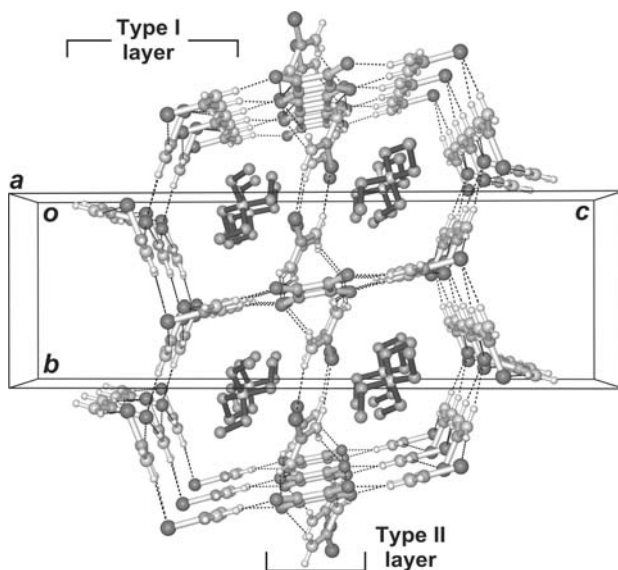


**Figure 8.34** Projection diagram of the hydrogen-bonded squarate-water chains in the crystal structure of  $[(C_2H_5)_4N^+]_2C_4O_4^{2-} \cdot 4(NH_2)_2CS \cdot 2H_2O$  (**16**). Each squarate ion cross-links two adjacent thiourea molecules to form a wide platform matching the  $(02\bar{1})$  family of planes. Broken lines represent hydrogen bonds, and atom types are distinguished by size and shading. Symmetry transformations: *a*  $(1 - x, 1 - y, 1 - z)$ , *b*  $(1 + x, y, z)$  and *c*  $(2 - x, 1 - y, 1 - z)$

components gives channels of nearly rectangular cross-section of  $16.96 \text{ \AA} \times 7.73 \text{ \AA}$  running parallel to the *a* axis. Well-ordered tetraethylammonium cations are accommodated as a double column within each channel, as illustrated in Figure 8.32.

*Channel-type structure of  $[(C_2H_5)_4N^+]_2C_4O_4^{2-} \cdot 6(NH_2)_2CS$  (17).* As illustrated in Figure 8.35, the host lattice of **17** [6a] can be considered as a combination of two different kinds of layers, designated as type **I** and type **II** for discussion. The type **I** layer is built of two independent thiourea molecules in the asymmetric unit while the type **II** layer is constructed from the third independent thiourea molecule and a centrosymmetric squarate dianion. All hydrogen atoms of the thiourea molecules **C(1)** and **C(2)** participate in the construction of the hydrogen-bonded type **I** layer (Figure 8.36). Two pairs of donor and acceptor hydrogen bonds are used to produce an infinitely extended thiourea ribbon while the third pair of donor and acceptor hydrogen bonds is used to cross-link the adjacent thiourea ribbons.

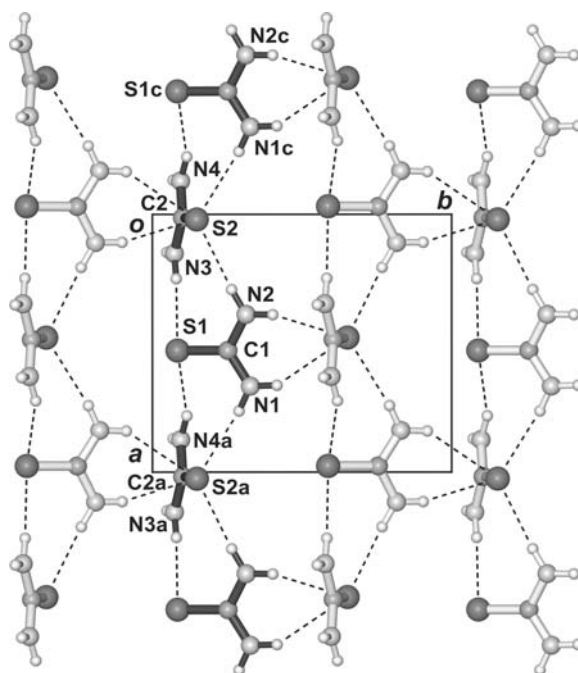
Two independent thiourea molecules **C(1)** and **C(2)** are alternately linked together by pairs of  $N-H \cdots S$  hydrogen bonds in a shoulder-to-shoulder fashion to form zigzag and puckered ribbons directed along the *a* axis. With these ribbons arranged side by side, atom **S(3)** of molecule **C(1)** forms two acceptor hydrogen bonds with atoms **N(2)** and **N(1)** of molecule **C(3)** belonging to the adjacent ribbon, such that this type of



**Figure 8.35** Perspective view of the crystal structure of  $[(\text{C}_2\text{H}_5)_4\text{N}^+]_2\text{C}_4\text{O}_4^{2-} \cdot 6(\text{NH}_2)_2\text{CS}$  (**17**) showing individual columns of well-ordered tetraethylammonium cations enclosed in the channel-type host lattice built of squarate-thiourea species. The origin of the unit cell lies at the upper left corner, with *a* pointing towards the reader, *b* downward and *c* from left to right

cross-linkage yields a puckered layer that is normal to the *c* axis (Figure 8.36). It is noted that this kind of linkage mode of thiourea molecules also exists in **16** and  $(n\text{-C}_3\text{H}_7)_4\text{N}^+\text{HCO}_2^- \cdot 3(\text{NH}_2)_2\text{CS} \cdot \text{H}_2\text{O}$  (**19**) [6d], as it offers two opposite walls to facilitate the direct linkage by a wide variety of inter-layer bridging moieties to form large unidirectional channel systems, provided that such inter-layer linkers have optimal sizes and compatible hydrogen bonding sites.

In the type **II** layer, each thiourea molecule utilizes three donor  $\text{N-H} \cdots \text{O}^-$  hydrogen bonds in building the infinite thiourea-squarate chains while a pair of donor and acceptor hydrogen bonds is used to cross-link the adjacent thiourea-squarate ribbons. The third thiourea molecule **C(3)**, together with its centrosymmetrically related partner **C(3A)**, is hydrogen-bonded to the squarate dianion at  $\bar{1}$  via pairs of charge-assisted  $\text{N-H} \cdots \text{O}^-$  hydrogen bonds to form a twisted bow-tie-like (or two-wing propeller-like) cyclic trimer, which serves as a repeating unit in generating the infinite thiourea-squarate chains. The lower half of this trimer is interlocked with the upper half of the preceding one via an additional pair of  $\text{N-H} \cdots \text{O}^-$  (squarate) hydrogen bonds to form an infinitely extended thiourea-squarate chain along the *a* axis. With these ribbons lying side by side, cross-linkage between adjacent thiourea molecules via pairs of  $\text{N-H} \cdots \text{S}$  hydrogen bonds in a shoulder-to-shoulder fashion generates a puckered layer containing large peanut-shaped voids (Figure 8.37). This thiourea-squarate layer links two adjacent zigzag and puckered thiourea layers together by forming two pairs of donor  $\text{N-H} \cdots \text{O}^-$  (squarate) hydrogen bonds from two centrosymmetrically related **C(2)** type thiourea

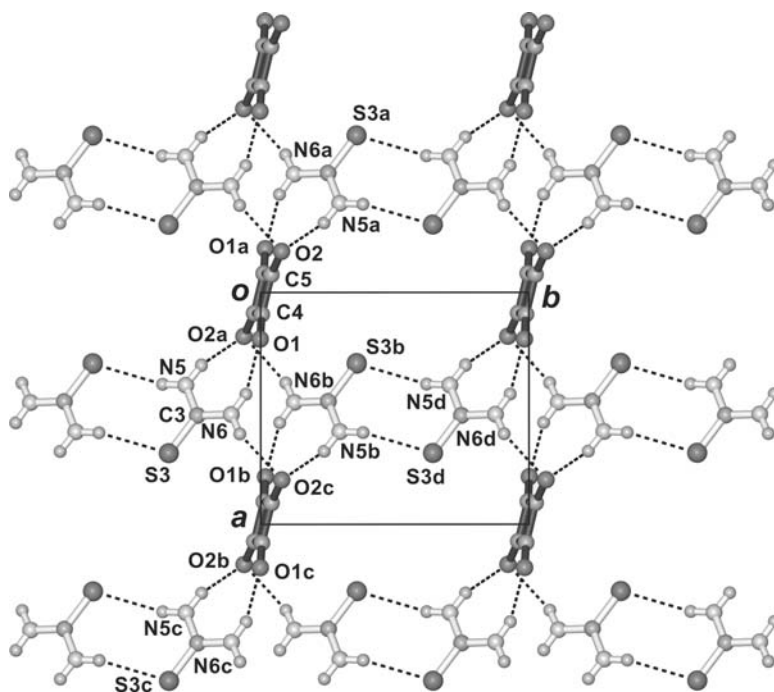


**Figure 8.36** Projection diagram of the type **I** layer in **17** composed of two independent thiourea molecules **C(1)** and **C(2)**. Hydrogen-bonded thiourea ribbons running along the *a* axis cross-link with each other to yield a puckered layer with its mean plane at  $z = 1/4$ . Symmetry transformations: *a* ( $x, y, z$ ), *b* ( $x, 1 + y, z$ ) and *c* ( $-1 + x, y, z$ )

molecules to generate a 3-D host framework with open channels running parallel to the *a* axis. Note that the thiourea molecules of type **C(2)** stretch out from both sides of each thiourea layer to facilitate their linkage to the thiourea-squarate puckered layers. The well-ordered tetraethylammonium cations almost attain idealized 222 molecular symmetry and are stacked densely within each channel with cross-sectional dimensions of about  $9.42 \text{ \AA} \times 8.96 \text{ \AA}$ .

*Layer-type structure of  $[(C_2H_5)_4N^+]_2C_4O_4^{2-} \cdot 2(NH_2)_2CS \cdot 2H_2O$  (**18**).* When a smaller stoichiometric ratio of squaric acid to thiourea was employed, the layer-type inclusion complex  $[(C_2H_5)_4N^+]_2C_4O_4^{2-} \cdot 2(NH_2)_2CS \cdot 2H_2O$  (**18**) was obtained [6a]. In the crystal structure of **18**, all of the hydrogen-bond donating sites on the thiourea molecules participate in the construction of the anionic host lattice which contains puckered layers generated by the cross-linkage of zigzag  $(C_4O_4^{2-} \cdot 2H_2O)_\infty$  chains and thiourea ribbons. The hydrogen-bonding scheme is conveniently described with reference to Figure 8.38.

Neighboring squarate dianions are bridged by two independent water molecules via two pairs of fairly strong hydrogen bonds to generate a zigzag  $(C_4O_4^{2-} \cdot 2H_2O)_\infty$  chain extending along the *b* axis, which contains 11-membered rings each stabilized by four independent hydrogen bonds. The pattern can be described by the ‘chain of rings’ notation  $C_4^4(15)[R_3^4(11)]$ . However, two independent thiourea molecules **C(5)** and **C(6)**

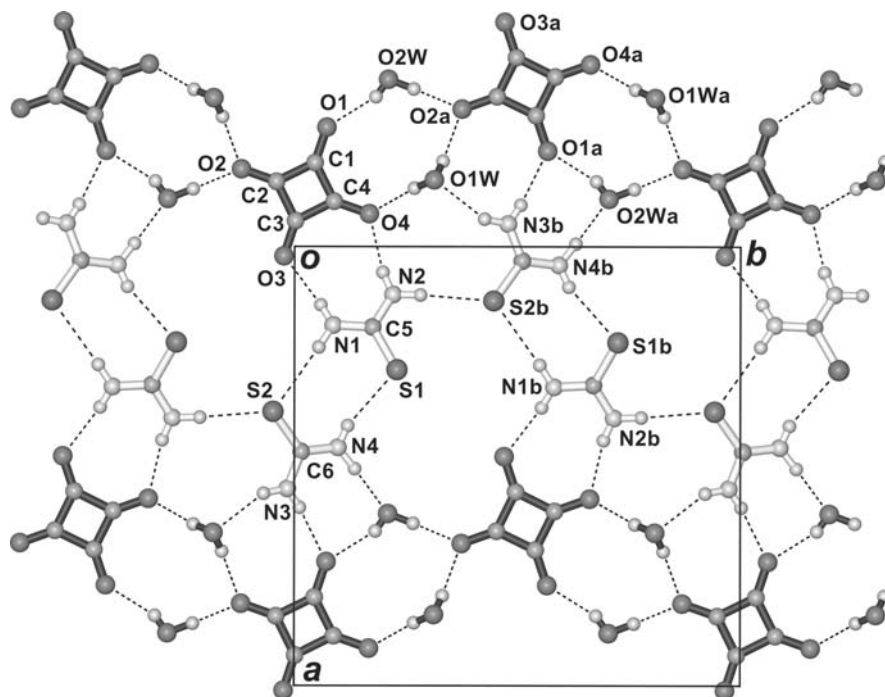


**Figure 8.37** Type II layer in **17** resulting from the cross-linkage between adjacent ribbons built of the centrosymmetric squarate dianion and the third independent thiourea molecule **C(3)**; the mean plane of this layer is at  $z = 0$ . Symmetry transformations:  $a (-x, -y, -z)$ ,  $b (1 - x, -y, -z)$ ,  $c (1 + x, y, z)$  and  $d (x, 1 + y, z)$

are joined together by a pair of  $\text{N-H} \cdots \text{S}$  hydrogen bonds to form a slightly twisted thiourea dimer in a typical shoulder-to-shoulder fashion. Neighboring thiourea dimers are further connected to each other by a single  $\text{N-H} \cdots \text{S}$  hydrogen bond to produce a zigzag 'dimeric thiourea ribbon' running parallel to the  $(\text{C}_4\text{O}_4^{2-} \cdot 2\text{H}_2\text{O})_\infty$  chain. With these zigzag  $(\text{C}_4\text{O}_4^{2-} \cdot 2\text{H}_2\text{O})_\infty$  chains and dimeric thiourea ribbons lying alternately side by side, five additional hydrogen bonds are formed to inter-link them together to yield a puckered layer normal to the  $c$  axis. The puckered layers lie close to the planes  $z = 1/8, 3/8, 5/8$  and  $7/8$ , and the well-ordered tetraethylammonium cations are alternately arranged and sandwiched between them, as shown in Figure 8.39.

*Structural features and topological correlation of the thiourea-squarate host lattices.* The host structure of  $[(\text{C}_2\text{H}_5)_4\text{N}^+]_2\text{C}_4\text{O}_4^{2-} \cdot 4(\text{NH}_2)_2\text{CS} \cdot 2\text{H}_2\text{O}$  (**16**) is closely related to that of another inclusion complex  $(n\text{-C}_3\text{H}_7)_4\text{N}^+\text{HCO}_2^- \cdot 3(\text{NH}_2)_2\text{CS} \cdot \text{H}_2\text{O}$  (**19**) [6d]. Both contain a uni-directional open-channel system. Although the channels are completely different in shape, the replacement of the squarate-water chain with the large cyclic linker composed of two pairs of centrosymmetric thiourea dimers cyclized by another pair of centrosymmetrically related formate and water molecules modifies the cross-section of the channel from lunch-box (about  $16.96 \text{ \AA} \times 7.73 \text{ \AA}$ ) to peanut (about  $25.38 \text{ \AA} \times 8.77 \text{ \AA}$ ) shape. As a result of this expansion in channel size, two columns



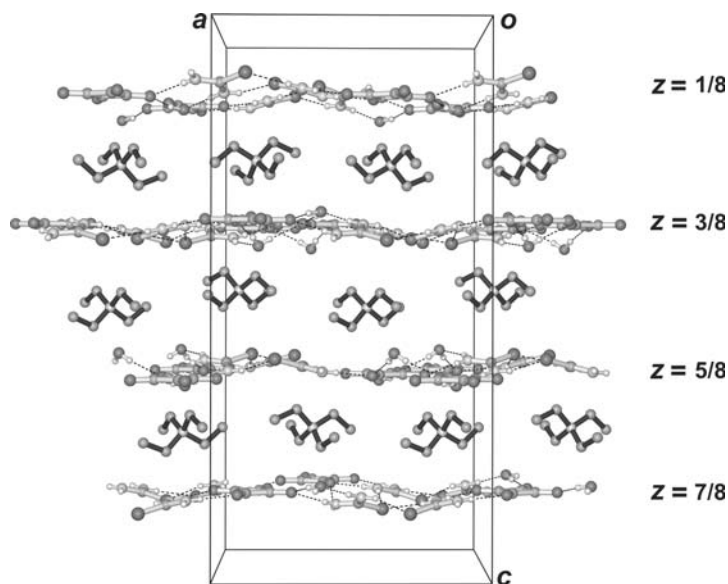


**Figure 8.38** Projection diagram along the *c* axis showing the hydrogen bonding interaction within a host layer in the crystal structure of **18**; the mean plane of this layer is at  $z = 3/8$ . Symmetry transformations: *a*  $(-\frac{1}{2} - x, \frac{1}{2} + y, z)$  and *b*  $(\frac{1}{2} - x, \frac{1}{2} + y, z)$

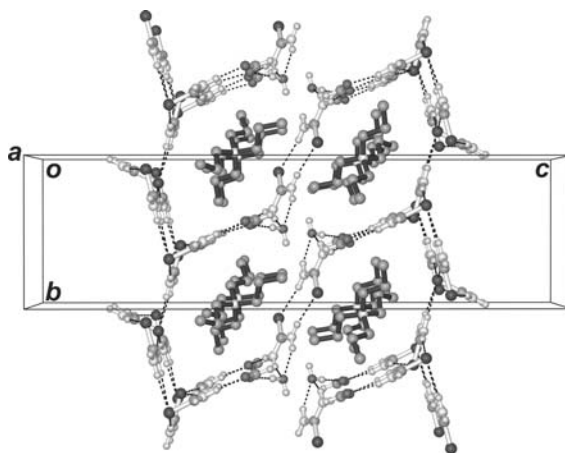
of larger and well-ordered  $(n\text{-C}_3\text{H}_7)_4\text{N}^+$  cations can be accommodated within each peanut-shaped channel in **19** (Figure 8.40). Figure 8.41 shows the hydrogen bonding interactions and structural characteristics of the chain of cyclic linkers in **19**, which plays the same role as the squarate-water chain in **16** (Figure 8.34).

Although the squarate dianion is an electron-rich chemical species with high molecular symmetry, it does not always utilize its hydrogen bonding acceptor sites to the maximum extent in the construction of host lattices. In the present three crystal structures, each squarate forms a different number of acceptor hydrogen bonds with its surrounding donor molecular species. In the host lattice of **16**, each squarate dianion forms eight acceptor hydrogen bonds. Four of them are used to generate a linear  $(\text{C}_4\text{O}_4^{2-} \cdot 2\text{H}_2\text{O})_\infty$  chain, while another two pairs of hydrogen bonds are used to cross-link two adjacent puckered thiourea layers to produce a large uni-directional rectangular channel system.

Through the substitution of bridging water molecules in **16** by thiourea molecules in **17**, each squarate dianion in the latter host lattice forms an unusually large number of ten acceptor hydrogen bonds with six neighboring thiourea molecules, as shown in Figure 8.42. The twisted dimeric thiourea linkers connect the squarate dianions in an end-to-end fashion, hence facilitating the formation of thiourea-squarate ribbons and the cross-linkage between similar ribbons on adjacent floors, as illustrated in



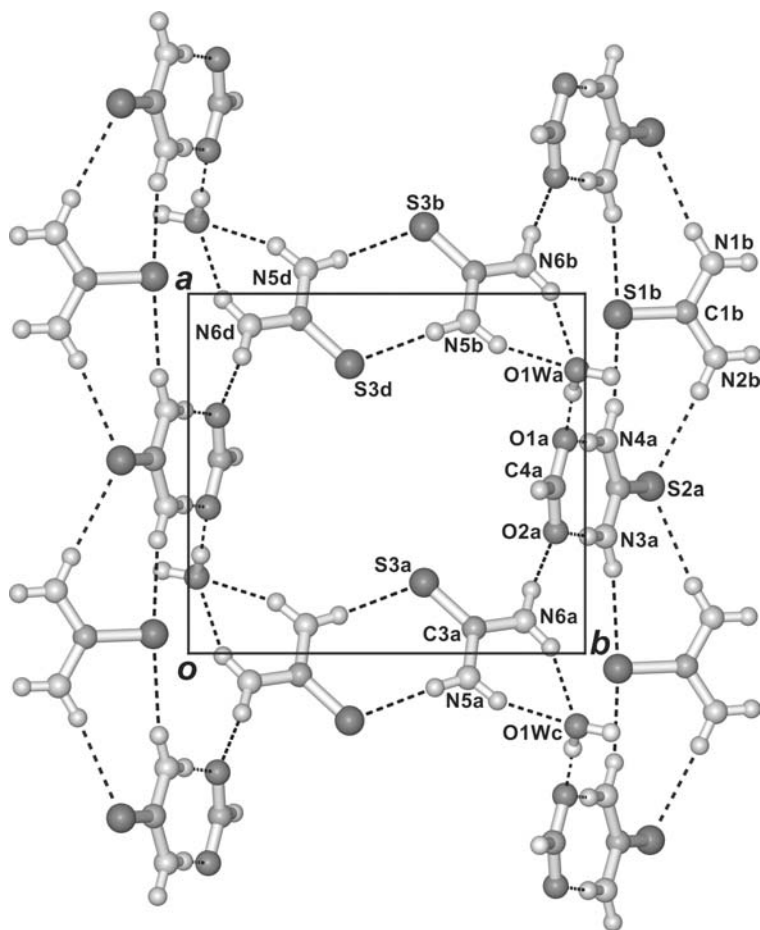
**Figure 8.39** Perspective view of the crystal structure of **18** showing that the well-ordered  $(\text{C}_2\text{H}_5)_4\text{N}^+$  cations are sandwiched between adjacent puckered squarate-thiourea-water layers



**Figure 8.40** Channel-type crystal structure of  $(n\text{-C}_3\text{H}_7)_4\text{N}^+\text{HCO}_2^- \cdot 3(\text{NH}_2)_2\text{CS} \cdot \text{H}_2\text{O}$  (**19**) [6d]. The peanut-shaped channels are viewed end-on

Figure 8.37. Moreover, this substitution effectively slices each large rectangular channel in **16** that accommodates a double column of tetraethylammonium cations (see Figure 8.32) lengthwise into two equal parts in **17** (see Figure 8.35), as the additional adjacent thiourea molecules are cross-linked by pairs of  $\text{N-H} \cdots \text{S}$  hydrogen bonds to create the partition. Each smaller ‘halved’ channel in **17**, with a cross-section



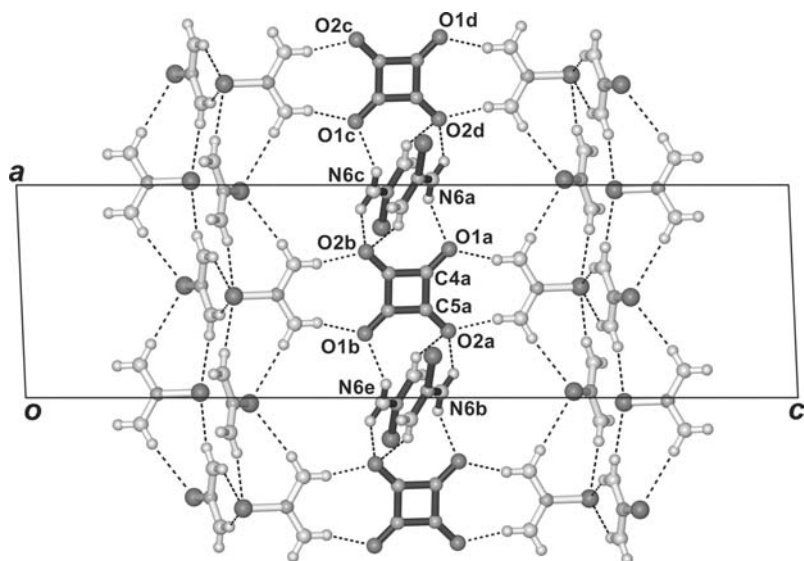


**Figure 8.41** Projection diagram of a portion of a host lattice in  $(n\text{-C}_3\text{H}_7)_4\text{N}^+\text{HCO}_2^- \cdot 3(\text{NH}_2)_2\text{CS} \cdot \text{H}_2\text{O}$  (**19**) [6d] showing the cyclic linker that interlinks two adjacent puckered thiourea layers via two pairs of  $\text{N-H} \cdots \text{O}^-$  (formate) hydrogen bonds. Symmetry transformations: *a*  $(x, 1 + y, z)$ , *b*  $(1 + x, 1 + y, z)$ , *c*  $(-1 + x, 1 + y, z)$  and *d*  $(1 - x, -y, -z)$

of about  $9.42 \text{ \AA} \times 8.96 \text{ \AA}$ , thus accommodates a single column of well-ordered tetraethylammonium cations with nearly 222 molecular symmetry.

In contrast to **16** and **17**, each squarate forms only seven acceptor hydrogen bonds in the host layer of **18** (see Figure 8.38). Four of them involve two pairs of bridging water molecules to form a zigzag  $(\text{C}_4\text{O}_4^{2-} \cdot 2\text{H}_2\text{O})_\infty$  ribbon. Another pair of acceptor hydrogen bonds is used to directly join with an adjacent thiourea dimer, and the seventh acceptor hydrogen bond is used to consolidate the squarate-thiourea-water anionic host layer.

The most prominent structural feature of these three channel-type inclusion complexes is their analogous linkage mode of thiourea molecules which leads to the same type of



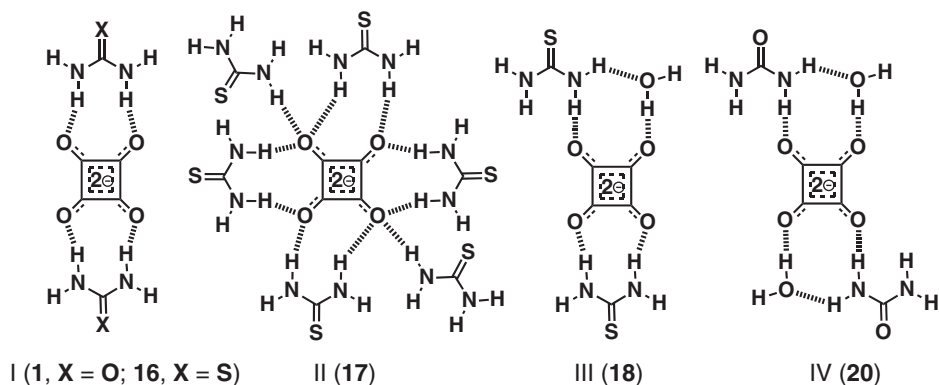
**Figure 8.42** Projection diagram of the hydrogen-bonded squarate-thiourea chain in **17**. Each squarate ion cross-links two thiourea molecules protruding from two adjacent puckered thiourea layers to form a wide platform, while another two pairs of thiourea molecules bridge the squarate dianions and inter-link parallel squarate-thiourea chains. Symmetry transformations: *a* ( $\frac{1}{2} + x, \frac{1}{2} - y, \frac{1}{2} + z$ ), *b* ( $\frac{1}{2} - x, \frac{1}{2} + y, \frac{1}{2} - z$ ), *c* ( $\frac{3}{2} - x, \frac{1}{2} + y, \frac{1}{2} - z$ ), *d* ( $\frac{3}{2} - z, \frac{1}{2} - y, \frac{1}{2} + z$ ) and *e* ( $-\frac{1}{2} + x, \frac{1}{2} - y, \frac{1}{2} + z$ )

hydrogen-bonded puckered thiourea layer (compare Figure 8.33 with Figure 8.36). Two independent thiourea molecules are joined together in a shoulder-to-shoulder fashion to form a twisted thiourea ribbon. Ribbons that lie side-by-side in different orientations are further cross-linked by forming pairs of N–H···S hydrogen bonds in a head-to-tail fashion. In this way, a pair of N–H hydrogen-bond donating sites is made available for direct connection to the linkers to generate large uni-directional channel systems with distinct sizes and shapes.

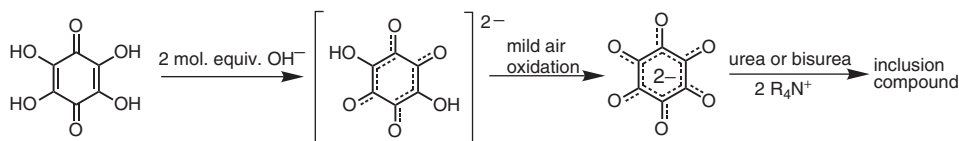
As demonstrated in the structures of the thiourea inclusion complexes **16–18** and those reported previously [15, 44], the squarate dianion plays an important role in providing hydrogen bonding acceptor sites for other molecular species, such as urea, thiourea and water molecules, in the construction of a rich variety of hydrogen bonding patterns. These three examples further demonstrate that the geometry of the thiourea-anion host lattice can be modified in interesting ways by substitution of appropriate hydrogen-bonded synthons through variation of molecular building blocks, thus providing another interesting example of isostructurality [14] in the context of supramolecular architecture.

### 8.3.2.2 Rhodizonate Complexes

In the series of inclusion complexes  $[(n\text{-C}_3\text{H}_7)_4\text{N}^+]_2\text{C}_4\text{O}_4^{2-} \cdot 6(\text{NH}_2)_2\text{CO} \cdot 2\text{H}_2\text{O}$  (**1**) [15],  $[(\text{C}_2\text{H}_5)_4\text{N}^+]_2\text{C}_4\text{O}_4^{2-} \cdot 4(\text{NH}_2)_2\text{CS} \cdot 2\text{H}_2\text{O}$  (**16**) [6a],  $[(\text{C}_2\text{H}_5)_4\text{N}^+]_2\text{C}_4\text{O}_4^{2-} \cdot 6(\text{NH}_2)_2\text{CS}$  (**17**) [6a],  $[(\text{C}_2\text{H}_5)_4\text{N}^+]_2\text{C}_4\text{O}_4^{2-} \cdot 2(\text{NH}_2)_2\text{CS} \cdot 2\text{H}_2\text{O}$  (**18**) [6a] and  $[(\text{C}_2\text{H}_5)_4\text{N}^+]_4\text{C}_4\text{O}_4^{2-}$



**Figure 8.43** Hydrogen bonding interaction between squarate ions and urea/thiourea molecules in some inclusion complexes

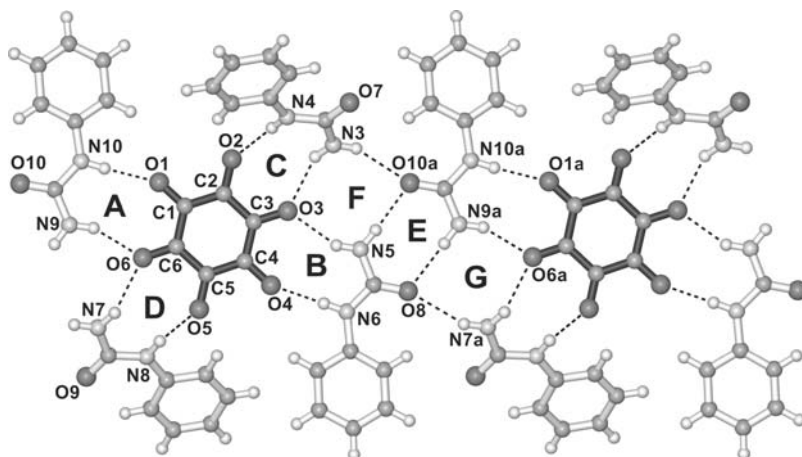


**Scheme 8.4** Trapping the rhodizonate ion in an inclusion compound

( $\text{HCO}_3^-$ )<sub>2</sub> · 4( $\text{NH}_2$ )<sub>2</sub>CO · 6H<sub>2</sub>O (**20**) [44], direct linkages between squarate and urea/thiourea exhibit the hydrogen bonding patterns I–IV shown in Figure 8.43, which are further connected with the commonly occurring infinite urea/thiourea chains or ribbons [8] to generate 2- or 3-D host networks.

Attempts to obtain analogous croconate and rhodizonate inclusion complexes with urea (and thiourea) have so far been unsuccessful. We then proceeded on the hypothesis that an oxocarbon dianion could be stabilized by surrounding it with the maximum number of convergent N–H donor sites, and to do so it would be necessary to disrupt the intermolecular association between urea molecules. We therefore decided to replace urea with phenylurea, whose bulky and hydrophobic substituent is expected to prevent, or at least suppress, the formation of urea ribbons. Accordingly we developed a simple method (see Scheme 8.4) starting with tetrahydroxy-1,4-quinone rather than rhodizonic acid to generate the rhodizonate ion *in situ* and incorporate it into the host lattices of the new inclusion compound [(*n*-C<sub>4</sub>H<sub>9</sub>)<sub>4</sub>N<sup>+</sup>]<sub>2</sub>C<sub>6</sub>O<sub>6</sub><sup>2-</sup> · 4(C<sub>6</sub>H<sub>5</sub>NHCONH<sub>2</sub>) (**21**) [45].

In the crystal structure of **21**, four crystallographically independent phenylurea molecules are directly connected to the rhodizonate ion via separate pairs of strong N–H<sub>anti</sub> ···O<sup>−</sup> hydrogen bonds in the range of 2.834–2.933 Å [ring motifs **A**, **B**, **C** and **D**; graph set  $N_2 = R_2^2(9)$ ] [18] to form a pseudo-centrosymmetric, slightly twisted X-shaped pentamer (Figure 8.44). One pair of opposite phenylurea molecules (designated as type A) is nearly coplanar with the rhodizonate ion, while the other pair (type B) is inclined to it at about 45°. Adjacent pentamers are joined together by pairs of strong N–H<sub>syn</sub> ···O hydrogen bonds between urea fragments in a complementary

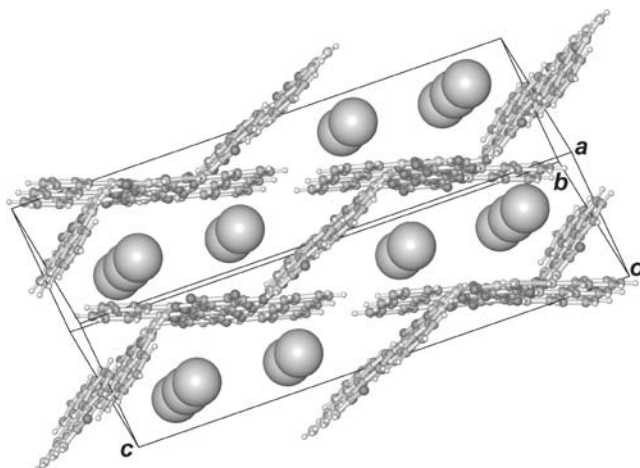


**Figure 8.44** Hydrogen bonding scheme of **21** showing a portion of the rhodizonate-phenylurea wide ribbon consolidated by strong  $\text{N-H} \cdots \text{O}$  hydrogen bonds. The phenyl rings lie on opposite sides of the zigzag ribbon; half of them are nearly coplanar with the ribbon, and the other half are inclined at  $45^\circ$ . The orientation of these two kinds of phenyl rings is further illustrated in Figure 8.45. Symmetry transformation:  $a (-x, -y, -z)$

manner [**E**,  $N_2 = R_2^2(8)$ ] to generate a wide rhodizonate-phenylurea ribbon, which is further strengthened by additional strong  $\text{N-H}_{\text{syn}} \cdots \text{O}$  hydrogen bonds to produce two other ring motifs [**F** and **G**,  $N_2 = R_2^2(8)$ ]. This wide ribbon runs parallel to the  $[1\bar{1}0]$  direction, and the hydrophilic amido fragment of the type A phenylurea molecule participates in the construction of its zigzag ‘pseudo rhodizonate-urea’ backbone in a ‘chain of rings’ pattern  $C_3^3(12)[R_2^2(9)R_2^2(8)R_2^2(9)]$ . The phenyl rings of type A urea molecules and whole urea molecules of type B protrude outward from the backbone, and stacking of the wide ribbons leads to a system of grated, broken-walled channels each accommodating two parallel columns of well-ordered tetra-*n*-butylammonium cations (Figure 8.45).

The well-ordered tetra-*n*-butylammonium cations (represented by large shaded spheres for clarity) are arranged in two parallel columns in the  $[1\bar{1}0]$  direction. Half of the phenylurea molecules reach out of the plane of each rhodizonate-phenylurea ribbon, functioning as spacers which effectively separate the double cationic columns. The cross-section of the parallelogram-shaped channel has dimensions of about  $11.6 \text{ \AA} \times 19.2 \text{ \AA}$ .

This study provides the first reasonably precise molecular dimensions of the rhodizonate dianion, which lends support to the aromaticity of this nonbenzenoid cyclic oxocarbon. Notably, the measured C–C bond lengths [ $1.421(5) - 1.458(5) \text{ \AA}$ ] of the rhodizonate in **21**, which exhibits approximate  $D_{6h}$  molecular symmetry, are significantly shorter than the corresponding values ( $1.488$  and  $1.501 \text{ \AA}$ ) in  $\text{Rb}_2\text{C}_6\text{O}_6$  [46] and the calculated values ( $1.500$  and  $1.501 \text{ \AA}$ ) for the  $C_2$  structure of this dianion [47]. Compound **21** provides yet another example of the use of urea and its derivatives for stabilizing elusive molecular anions such as allophanate [5h] and dihydrogen borate [5c] in a hydrogen-bonded host lattice.



**Figure 8.45** Crystal structure of **21** showing the 1-D 'broken-walled' channel system

Recently Robson's group has shown that each rhodizonate ion in the crystal structure of  $[\text{C}(\text{NH}_2)_3^+]_2\text{C}_6\text{O}_6^{2-}$  (**22**) is directly hydrogen-bonded to six neighboring guanidinium cations to form an infinite (3,6)-connected 3-D network instead of the expected 2-D network [48] (Figure 8.46). This deviation can be rationalized as a result of reducing steric repulsion between the neighboring guanidinium cations by twisting out of the mean plane of the rhodizonate.

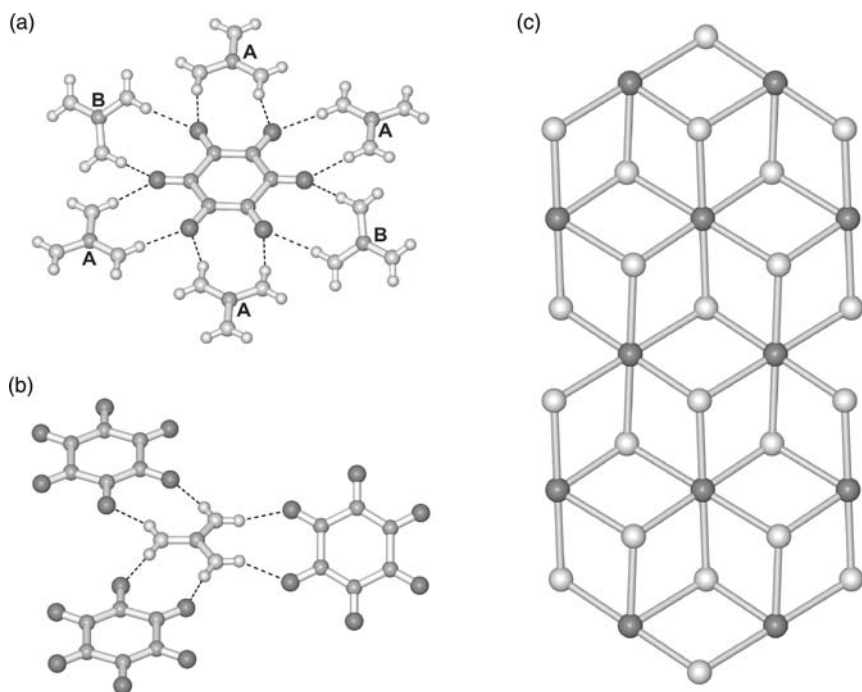
### 8.3.2.3 Croconate Complex

The rhodizonate dianion readily undergoes oxidative ring contraction through an  $\alpha$ -oxo alcohol rearrangement [49], which is related to the well-known benzil-benzylic acid rearrangement [50], to produce the croconate ion. Using this convenient approach, the croconate complex  $[(n\text{-C}_4\text{H}_9)_4\text{N}^+]_2\text{C}_5\text{O}_5^{2-} \cdot 4(\text{C}_6\text{H}_5\text{NHCONH}_2)$  (**23**) [45] can be obtained by warming the reaction mixture of **21** in air.

Interestingly, the hydrogen bonding pattern in the crystal structure of **23** (Figure 8.47) is very similar to that in **21** (Figure 8.44), and the cyclic pentameric structural unit with an oxocarbon core is basically retained, so that the pair of inclusion compounds exhibits an isostructurality relationship [14]. However, the croconate ion is disordered about an inversion center, and it adopts two equally populated orientations. Consequently, there are only two independent phenylurea molecules in the asymmetric unit of **23**, and the length of the *c* axis is about half of that in **21** (Figure 8.48).

### 8.3.3 Charge-localized and -delocalized Valence Tautomeric Forms of Cyclic Oxocarbon Dianions

For over four decades, interest in the family of cyclic oxocarbon dianions has focused on their aromaticity and molecular structure. Theoretical studies have indicated that aromaticity decreases with increasing ring size, and  $\text{C}_6\text{O}_6^{2-}$  possesses only a small degree of aromatic character [47, 51]. Furthermore, while it is firmly established that the



**Figure 8.46** Crystal structure of  $[C(NH_2)_3^+]_2C_6O_6^{2-}$  (**22**). (a) Six guanidinium cations are hydrogen-bonded to each  $C_6O_6^{2-}$  ion. Guanidinium cations **A** are nearly coplanar with the  $C_6O_6^{2-}$  ion and cations **B** are significantly deviated from the plane. (b) Each guanidinium cation is hydrogen-bonded to three  $C_6O_6^{2-}$  anions, two of which are almost coplanar with the guanidinium ion, the third being twisted out of plane. (c) The (3,6)-connected topological network underlying the hydrogen-bonded structure

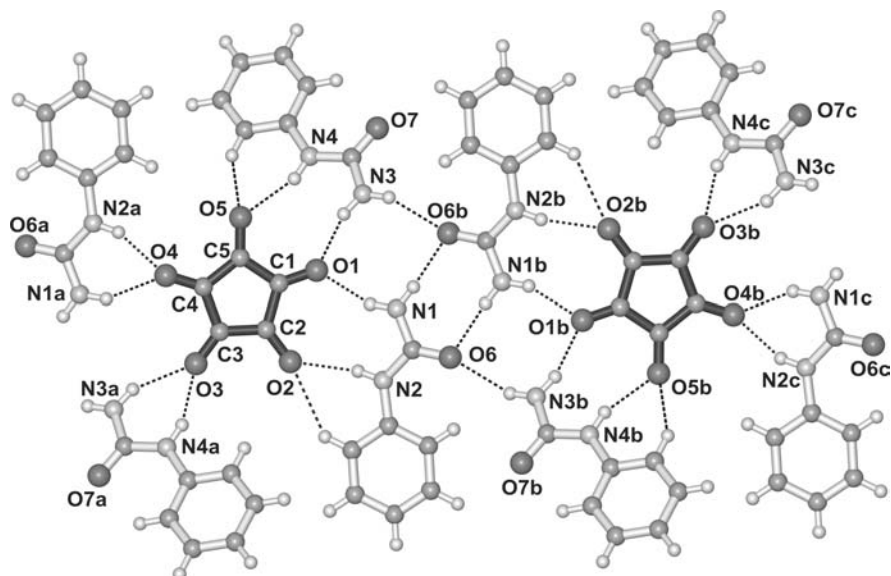
squarate and croconate ions have planar  $D_{nh}$  structures, infrared spectroscopy [43c] and X-ray analysis showed that the rhodizonate ion conforms to idealized  $D_{6h}$  geometry in  $Rb_2C_6O_6$ , but nonplanar  $D_{3d}$  and  $C_2$  forms coexist in  $K_2C_6O_6$  [43, 46]. A recent study indicated that the electron distribution on the rhodizonate ring depends largely on the chemical environment of the rhodizonate anion, which can display a large variety of molecular symmetry in different supramolecular structures [52].

#### 8.3.3.1 Valence Tautomers of Rhodizonate Dianion

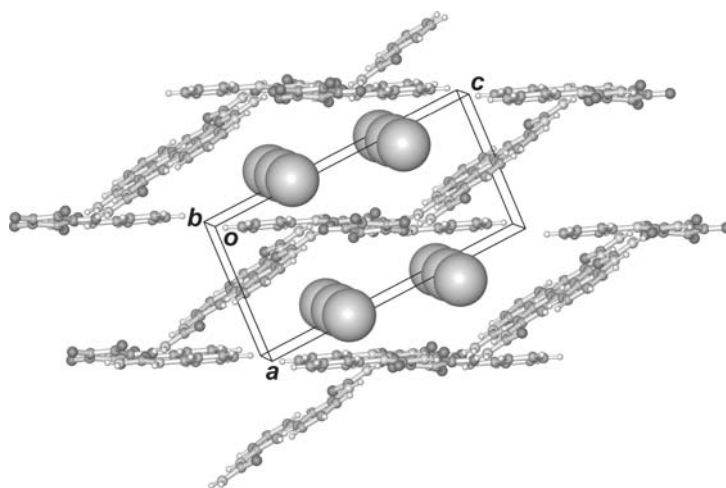
We attempted to alter the electron distribution on the rhodizonate ring by changing the hydrogen-bonding environment and planned the synthesis of two air-stable crystalline inclusion compounds of tetra-*n*-butylammonium rhodizonate with urea derivatives, namely  $[(n-C_4H_9)_4N^+]_2C_6O_6^{2-} \cdot 2(m-OHC_6H_4NHCONH_2) \cdot 2H_2O$  (**24**) and  $[(n-C_4H_9)_4N^+]_2C_6O_6^{2-} \cdot 2(NH_2CONHCH_2CH_2NHCONH_2) \cdot 3H_2O$  (**25**) [53].

In the host lattice of **24**, the rhodizonate dianion resides at an inversion center, being directly linked to a pair of *N*-(3-hydroxyphenyl)urea molecules through pairs of  $N-H \cdots O^-$  [**A**,  $N_2 = R_2^2(9)$ ] and phenyl  $C-H \cdots O^-$  [**B**,  $N_2 = R_2^1(6)$ ] hydrogen bonds



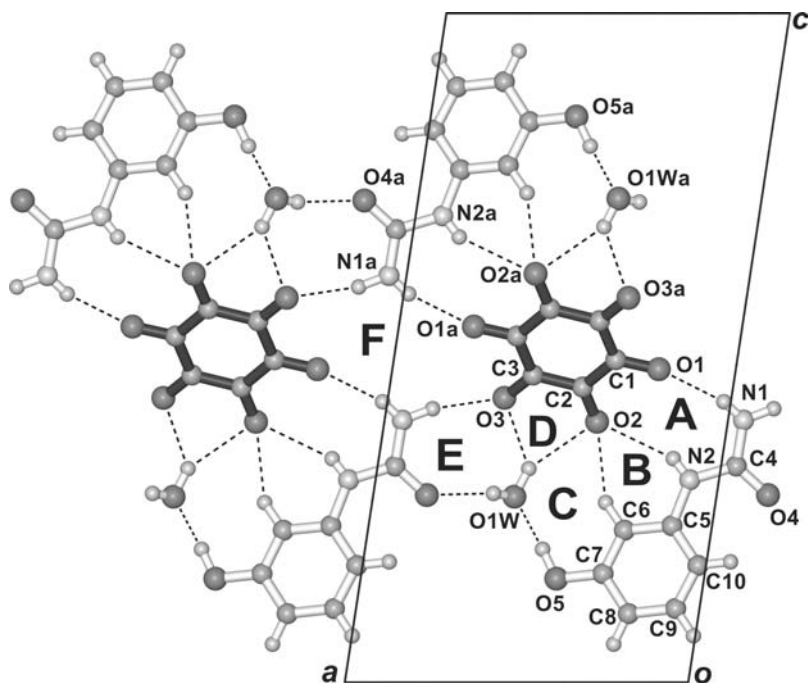


**Figure 8.47** Hydrogen bonding scheme of **23**, showing a wide croconate-phenylurea ribbon running parallel to the *b* axis. The disordered croconate dianion lies at an inversion center; for clarity, one orientation is shown on the left and the other on the right. Symmetry transformations: *a*  $(1 - x, -y, 1 - z)$ , *b*  $(1 - x, 1 - y, 1 - z)$  and *c*  $(x, 1 + y, z)$



**Figure 8.48** The crystal structure of **23** showing the 1-D 'broken-walled' channel system. The well-ordered tetra-*n*-butylammonium cations (represented by large shaded spheres for clarity) are arranged in two parallel columns in the  $[010]$  direction. This crystal structure is very similar to that of **21** (see Figure 8.45)

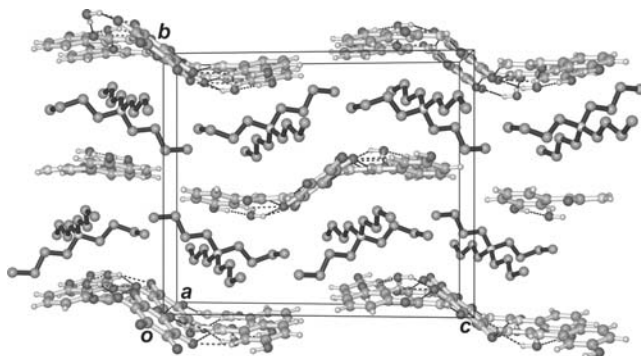




**Figure 8.49** Projection along the *b* axis showing extensive hydrogen bonding interactions around the centrosymmetric rhodizonate dianion in the host lattice of  $[(n\text{-C}_4\text{H}_9)_4\text{N}^+]_2\text{C}_6\text{O}_6^{2-} \cdot 2(m\text{-OHC}_6\text{H}_4\text{NHCONH}_2) \cdot 2\text{H}_2\text{O}$  (**24**). Symmetry transformation:  $a(1-x, 1-y, 1-z)$

to form a puckered trimer (Figure 8.49). The trimer is consolidated by a pair of bridging water molecules that form strong  $\text{O-H} \cdots \text{O}^-$  hydrogen bonds [**C**,  $N_3 = R_3^2(8)$  and **D**,  $N_2 = R_1^2(5)$ ] to generate a pentamer. Adjacent pentamers are further connected together by strong  $\text{O-H} \cdots \text{O}$  and  $\text{N-H} \cdots \text{O}^-$  hydrogen bonds [**E**,  $N_3 = R_3^2(8)$  and **F**,  $N_4 = R_4^4(14)$ ] to produce a wide, puckered ribbon running parallel to the *a* axis. Well-ordered  $(n\text{-C}_4\text{H}_7)_4\text{N}^+$  ions are accommodated between broken sinusoidal layers formed by the alternately stacked puckered ribbons, with an interlayer spacing of about 8.0 Å (Figure 8.50).

In the host lattice of **25**, two independent 1,1'-ethylenediurea molecules are directly bound to the rhodizonate by four pairs of strong  $\text{N-H} \cdots \text{O}^-$  hydrogen bonds [**A**, **B**, **C** and **D**,  $N_2 = R_2^2(9)$ ] to yield a trimer (Figure 8.51). Adjacent trimers related by simple translation along the *a* axis are connected together by pairs of amido  $\text{N-H} \cdots \text{O}$  hydrogen bonds in the usual shoulder-to-shoulder manner [8] [**E**,  $N_2 = R_2^2(8)$ ] to produce a zigzag ribbon, which is further consolidated by three independent water molecules through additional strong  $\text{O-H} \cdots \text{O}$  hydrogen bonds to generate two new motifs [**F**,  $N_4 = R_4^2(8)$  and **G**,  $N_2 = R_5^4(13)$ ]. Two antiparallel ribbons related by  $\bar{1}$  are cross-linked by pairs of  $\text{N-H} \cdots \text{O}$  [**H**,  $N_2 = R_2^2(8)$ ] hydrogen bonds to yield a double ribbon with large voids [**I**,  $N_8 = R_8^8(36)$ ]. Adjacent double ribbons related by the *c*-translation are interwoven with three additional ring motifs [**J**,  $N_4 = R_4^4(22)$ , **K**,  $N_3 = R_3^3(8)$  and **L**,



**Figure 8.50** Perspective diagram viewed along the *a* axis showing the crystal structure of  $[(n\text{-C}_4\text{H}_9)_4\text{N}^+]_2\text{C}_6\text{O}_6^{2-} \cdot 2(m\text{-OHC}_6\text{H}_4\text{NHCONH}_2) \cdot 2\text{H}_2\text{O}$  (**24**). The spacing between the broken sinusoidal layers is about 8.0 Å

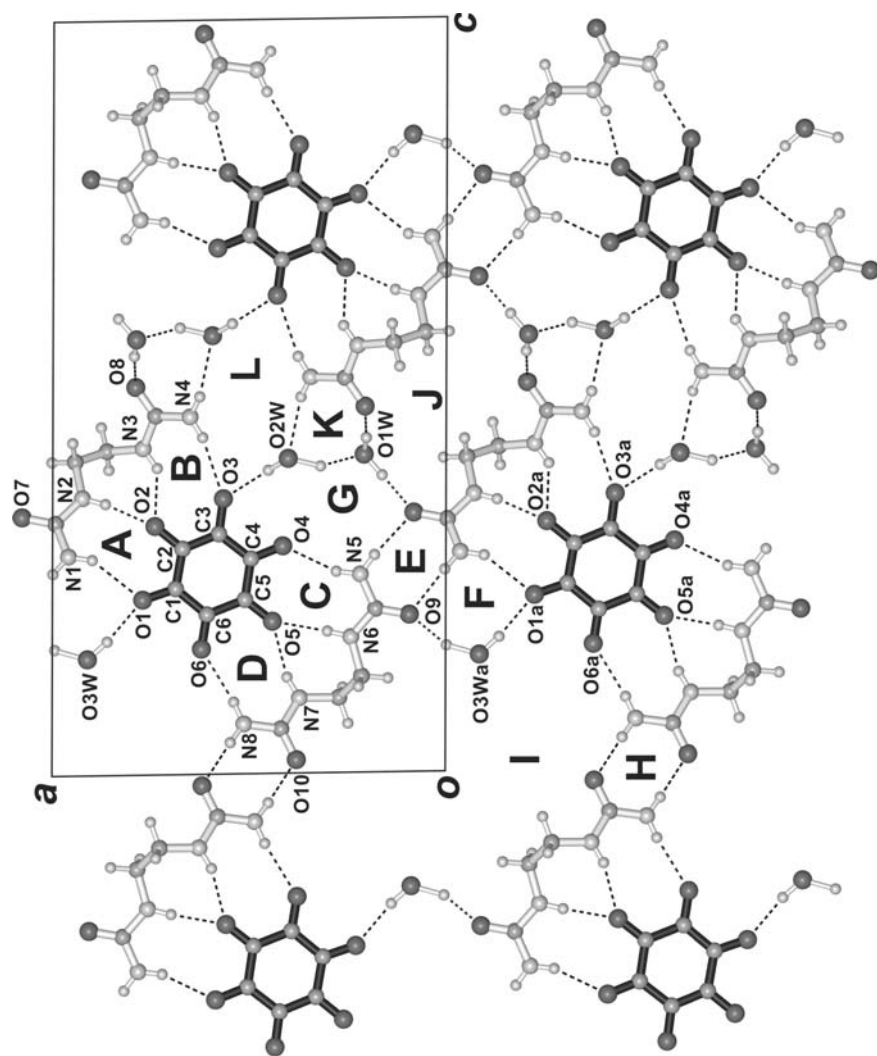
$N_6 = R_6^4(12)]$  by strong  $\text{N-H} \cdots \text{O}$  and  $\text{O-H} \cdots \text{O}$  hydrogen bonds to generate a highly corrugated layer. Well-ordered  $(n\text{-C}_4\text{H}_7)_4\text{N}^+$  ions are sandwiched between adjacent layers with an interlayer spacing of about 7.8 Å (Figure 8.52).

A theoretical study in 2000 concluded that while the oxocarbon dianions  $\text{C}_n\text{O}_n^{2-}$  ( $n = 3\text{--}5$ ) all favor  $D_{nh}$  symmetry, the  $\text{C}_6\text{O}_6^{2-}$  ground state has  $C_2$  symmetry, and all isomeric structures are very close in energy, e.g. the energy difference is 0.34 kcal mol<sup>−1</sup> between the  $D_{6h}$  and nonplanar  $C_2$  forms [47]. Notably the measured dimensions of the relatively unstable  $\text{C}_6\text{O}_6^{2-}$  species in **24** and **25** conform to idealized  $D_{6h}$  and  $C_{2v}$  molecular symmetry, corresponding to respective valence tautomeric structures that manifest nonbenzenoid aromatic and enediolate character (Figure 8.53).

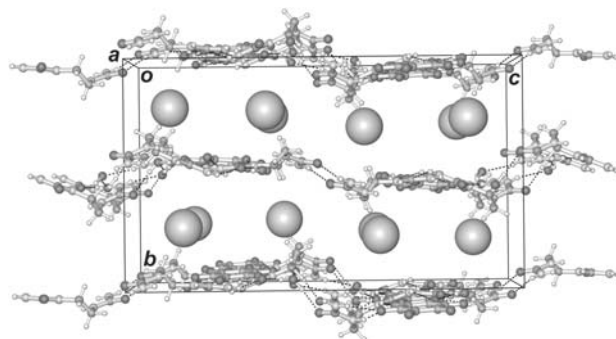
The measured C–C bond lengths in **24** [1.440(5)–1.442(5) Å] and **25** [1.404(5)–1.472(5) Å] are markedly shorter than those (1.488 and 1.501 Å) in  $\text{Rb}_2\text{C}_6\text{O}_6$  [43] and the computed values (1.500 and 1.501 Å) for the  $C_2$  structure [47], implying that  $\pi$ -electron delocalization in the aromatic ring of  $\text{C}_6\text{O}_6^{2-}$  is much more extensive than hitherto recognized. The strong negative charge-assisted hydrogen bonding interaction between rhodizonate and 1,1'-ethylenediurea causes the latter to assume a less stable *gauche* conformation, in preference to its favored *anti* conformation. Occurrence of the charge-localized  $C_{2v}$  structure of  $\text{C}_6\text{O}_6^{2-}$  in **25** can be attributed to unequal hydrogen bonding interaction with its two neighboring bisurea donors (Figure 8.51).

### 8.3.3.2 Valence Tautomers of Croconate Dianion

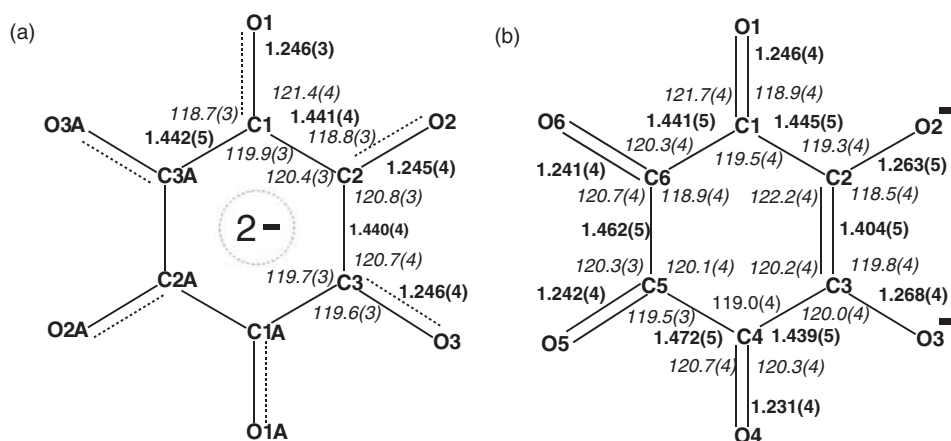
As previous theoretical [54] and experimental [55] studies on the  $\text{C}_n\text{O}_n^{2-}$  system indicated that the extent of electron delocalization increases as ring size decreases [56], we attempted to accomplish the presumably more challenging task of trapping analogous  $D_{5h}$  and  $C_{2v}$  valence tautomers of the lighter homolog  $\text{C}_5\text{O}_5^{2-}$ . After much experimentation, the valence tautomers of the croconate anion have been stabilized in two new inclusion compounds  $[(n\text{-C}_3\text{H}_7)_4\text{N}^+]_2\text{C}_5\text{O}_5^{2-} \cdot 3(\text{NH}_2)_2\text{CO} \cdot 8\text{H}_2\text{O}$  (**26**) and  $[(n\text{-C}_4\text{H}_9)_4\text{N}^+]_2\text{C}_5\text{O}_5^{2-} \cdot 3(\text{CH}_3\text{NH})_2\text{CO}$  (**27**), respectively, and their crystal structures were determined by low-temperature (123 K), single-crystal X-ray diffraction [57].



**Figure 8.51** Projection along the *b* axis showing the hydrogen bonding interactions within the puckered rhodizonate-bisurea-water layer of  $[(n-C_4H_9)_4N^+I_2C_6O_6^{2-} \cdot 2(NH_2CONHCH_2CH_2NHCONH_2) \cdot 3H_2O]$  (**25**). Symmetry transformation:  $a(x-1, y, z)$



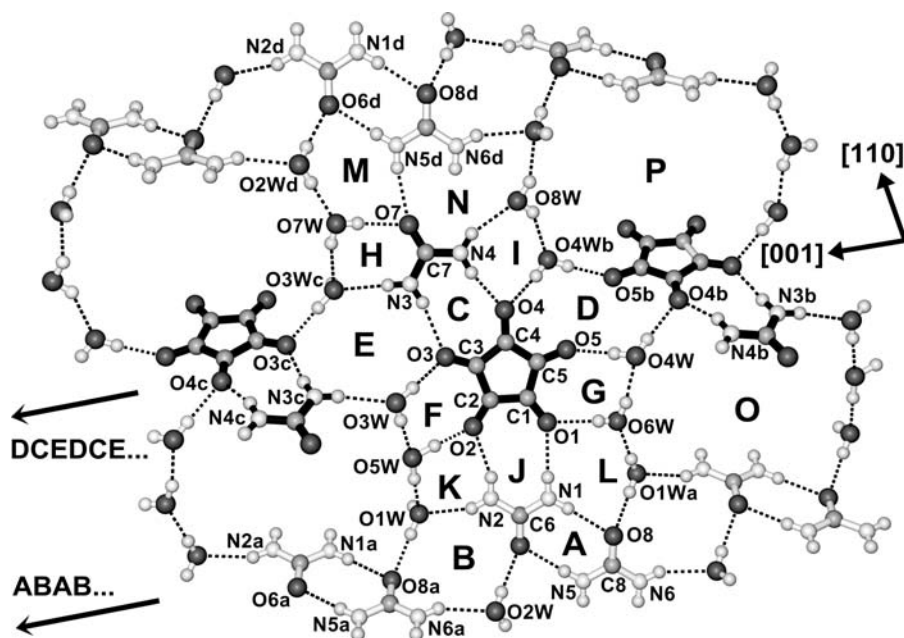
**Figure 8.52** A packing diagram of **25** viewed down the *a* axis shows the sandwich-type structure. Two independent  $[(n\text{-C}_4\text{H}_9)_4\text{N}]^+$  cations are accommodated between the corrugated layers, each being represented by large shaded circles



**Figure 8.53** Bond lengths and angles of the (a)  $D_{6h}$  and (b)  $C_{2v}$  valence tautomers of the rhodizonate dianion in **24** and **25**, respectively

*Nonbenzenoid aromatic  $D_{5h}$  valence tautomer of croconate dianion.* In the crystal structure of **26**, two of the three independent urea molecules constitute a hydrogen-bonded dimer, labeled motif **A** (Figure 8.54). Adjacent pairs of urea dimers lying parallel to the *c*-axial direction are twisted by  $60^\circ$  and bridged by pairs of water molecules derived from **O1w** and **O2w** via pairs of strong  $N_{\text{syn}}\text{-H}\cdots\text{O}$  and  $\text{O-H}\cdots\text{O}$  hydrogen bonds [58a] to form a twisted [urea dimer-( $\text{H}_2\text{O}$ ) $_2$ ] $_\infty$  ribbon with a new ring motif **B**. The third independent urea molecule is attached to the croconate anion directly via a pair of strong, charge-assisted  $N_{\text{anti}}\text{-H}\cdots\text{O}^-$  hydrogen bonds with ring motif **C**.

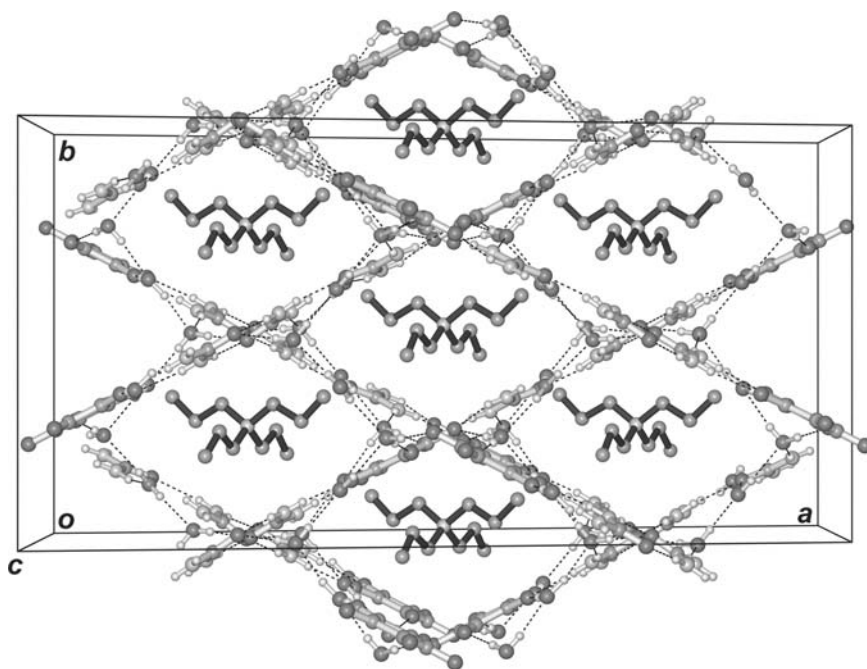
Neighboring croconate-urea building blocks rotated by about  $60^\circ$  along the *c* axis are alternately bridged by pairs of two distinct types of water molecules derived from **O4w** and **O3w** via pairs of strong  $\text{O-H}\cdots\text{O}^-$  hydrogen bonds to form another twisted ribbon [croconate-urea-( $\text{H}_2\text{O}$ ) $_4$ ] $_\infty$  characterized by two additional hydrogen-bonded rings **D** and



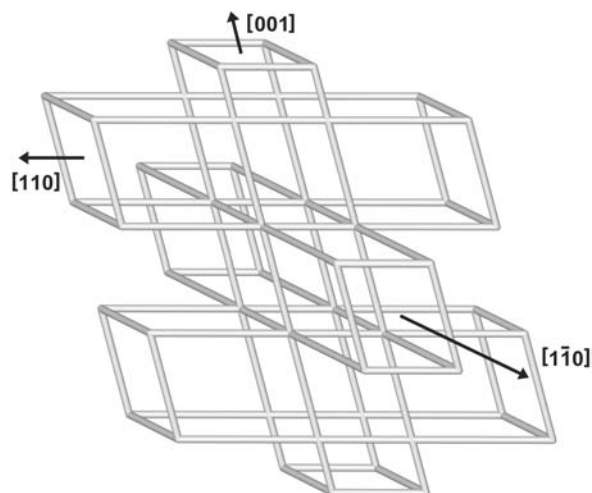
**Figure 8.54** Projection diagram on the  $(\bar{1}\bar{1}0)$  plane showing the hydrogen-bonding scheme for a portion of the host lattice in **26**. The slanted vertical (urea dimer-croconate-urea) $_{\infty}$  chain constitutes a side wall of the  $[110]$  channel system. The parallel [urea dimer-(H<sub>2</sub>O)<sub>2</sub>] $_{\infty}$  ABAB... and [croconate-urea-(H<sub>2</sub>O)<sub>4</sub>] $_{\infty}$  DCEDCE... ribbons define the channel system in the  $c$  direction. Symmetry transformations:  $a$  ( $x, -y, -\frac{1}{2} + z$ ),  $b$  ( $1 - x, y, \frac{3}{2} - z$ ),  $c$  ( $1 - x, y, \frac{1}{2} - z$ ) and  $d$  ( $1 - x, \frac{1}{2} + y, z$ ).

**E.** The ribbon is further consolidated by independent water molecules **O5w**, **O6w**, **O7w** and **O8w** at four corners of the croconate-urea building block by forming pairs of strong negative-charge assisted O-H...O<sup>-</sup>, O-H...O and N-H...O hydrogen bonds to yield four new motifs **F**, **G**, **H** and **I**. These two types of twisted hydrogen-bonded ribbons are arranged alternately side by side parallel to the  $c$  axis and constitute the  $[001]$  channel system (see Figure 8.55). They are further cross-linked together by pairs of strong N-H<sub>anti</sub>...O<sup>-</sup> (ring motif **J**), O-H...O and N-H...O hydrogen bonds to produce a wide platform containing new ring patterns **K**, **L**, **M** and **N** for the construction of a side-wall of the  $[110]$  channel system. Two large voids characterized by ring motifs **O** and **P** define the cross-section of the channel with dimensions of about 10.8 Å × 7.4 Å, and a similar channel is directed along  $[1\bar{1}0]$ . The host lattice contains three systems of intersecting channels (Figure 8.56). The two independent, well-ordered tetra-*n*-propylammonium cations, which occupy general and  $\bar{1}$  positions, are accommodated in a single column along the  $[001]$  channel and in a double column along each of the other two channels.

*Enediolate C<sub>2v</sub> valence tautomer of croconate.* To create an asymmetric hydrogen-bond donor environment that stabilizes the C<sub>2v</sub> valence tautomer of C<sub>5</sub>O<sub>5</sub><sup>2-</sup>, we

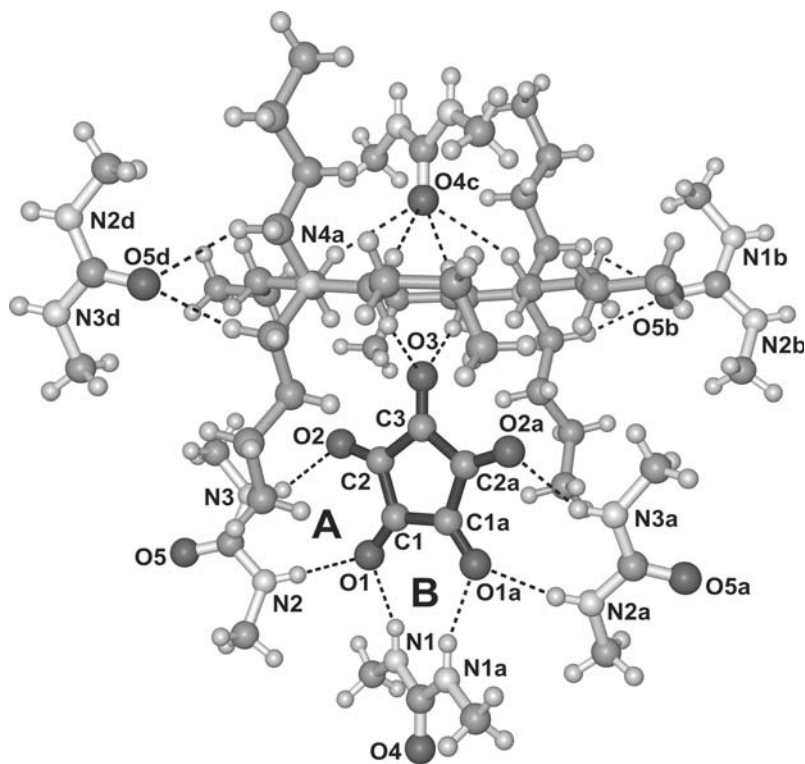


**Figure 8.55** Rhombic channel system in  $[(n\text{-C}_3\text{H}_7)_4\text{N}^+]_2\text{C}_5\text{O}_5^{2-} \cdot 3(\text{NH}_2)_2\text{CO} \cdot 8\text{H}_2\text{O}$  (**26**) with cross-section  $12.8 \text{ \AA} \times 9.6 \text{ \AA}$  running parallel to the  $c$  axis. Well-ordered tetra- $n$ -propylammonium cations are accommodated in a single column along each channel



**Figure 8.56** Schematic diagram showing the arrangement and orientation of three intersecting channel systems in  $[(n\text{-C}_3\text{H}_7)_4\text{N}^+]_2\text{C}_5\text{O}_5^{2-} \cdot 3(\text{NH}_2)_2\text{CO} \cdot 8\text{H}_2\text{O}$  (**26**)

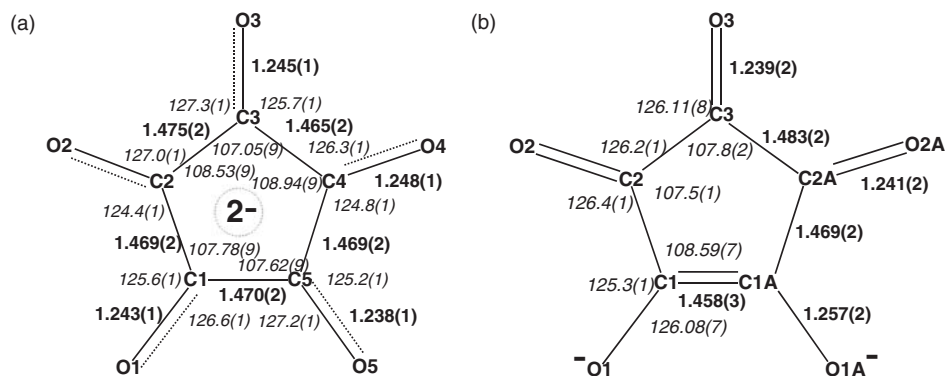




**Figure 8.57** A portion of the hydrogen-bonded layer in **27** showing extensive C–H···O hydrogen bonding between the  $C_5O_5^{2-} \cdot 3(CH_3NH)_2CO$  aggregate and its neighboring tetra-*n*-butylammonium ions. Symmetry transformations: *a*  $(1 - x, y, \frac{1}{2} - z)$ , *b*  $(1 - x, y - \frac{1}{2}, z)$ , *c*  $(x, y - 1, z)$  and *d*  $(\frac{3}{2} - x, y - \frac{1}{2}, \frac{1}{2} - z)$

replaced urea with its 1,3-dimethyl derivative to generate the inclusion compound  $[(n-C_4H_9)_4N^+]_2C_5O_5^{2-} \cdot 3(CH_3NH)_2CO$  (**27**). In the resultant host lattice, the croconate dianion resides on a twofold axis, being directly linked to three 1,3-dimethylurea molecules through pairs of N–H···O<sup>−</sup> (ring motifs **A** and **B**) hydrogen bonds to form a semicircular structural unit (Figure 8.57). In particular, each type O(1) oxygen atom forms two strong acceptor hydrogen bonds with the N–H donors from a pair of 1,3-dimethylurea molecules, while each type O(2) oxygen atom forms only one N–H···O<sup>−</sup> hydrogen bond. In contrast, the solitary type O(3) oxygen atom is stabilized by two weak C–H···O hydrogen bonds [58b] with neighboring tetra-*n*-butylammonium cations. Such a highly unsymmetrical environment engenders a sharp gradient of hydrogen-bonding donor strength around the croconate ion, which is conducive to stabilization of its  $C_{2v}$  valence tautomer with significantly different C–C and C–O bond lengths around the cyclic system. The  $C_5O_5^{2-} \cdot 3(CH_3NH)_2CO$  aggregates are further connected together by weak C(sp<sup>3</sup>)–H···O donor hydrogen bonds from the quaternary ammonium cations to form a highly puckered layer. The well-ordered tetra-*n*-butylammonium ions are sandwiched between anionic host layers.





**Figure 8.58** Bond lengths (Å) and angles (°) of the  $D_{5h}$  (a) and  $C_{2v}$  (b) valence tautomers of the croconate dianion as observed in **26** and **27**, respectively, with standard deviations in parentheses

The croconate ion is a very stable species that can coordinate to a transition metal in the terminal bidentate mode [59], in which case the C–C and C–O bond lengths involving atoms in the resulting five-membered chelate ring are in agreement with their formal bond orders of two and one, respectively, in the enediolate form of the ligand. When  $C_5O_5^{2-}$  functions as a multidentate ligand that bridges two or more metals to form a dinuclear, polynuclear or polymeric coordination compound [60], the bond-length variation exhibits a more complex pattern. The measured bond lengths and bond angles of  $C_5O_5^{2-}$  in the crystal structure of **26** (Figure 8.58a) are consistent with its expected charge-delocalized  $D_{5h}$  structure in the ground state. In contrast, the molecular dimensions of  $C_5O_5^{2-}$  in **27** are indicative of a vicinal triketo-enediolate structure (Figure 8.58b), which corresponds to the less stable, partially charge-localized  $C_{2v}$  valence tautomer, whose stabilization can be attributed to its highly unsymmetrical hydrogen-bonding environment within the host layer.

*Ab initio* and density functional theory calculations at various correlation levels [61] employing basis sets of increasing complexity have been employed to study the geometry and charge distribution of the croconate dianion in isolated and hydrogen-bonded situations. For a starting model consisting of an isolated charge-localized  $C_5O_5^{2-}$  system with imposed  $C_{2v}$  symmetry, the calculated bond lengths converge to the values that correspond to idealized  $D_{5h}$  molecular structure, showing that the latter geometry adequately represents the ground state of the dianion. In other words, the  $C_{2v}$  form of the croconate dianion cannot exist as a single entity. For the charge-delocalized  $D_{5h}$  form of  $C_5O_5^{2-}$ , the computed bond lengths at either the rather primitive HF/3-21G(d) level (C–C = 1.462 Å and C–O = 1.249 Å) or the much more sophisticated CASSCF(6,4)/3-21G(d) level (1.460, 1.252 Å) agree reasonably well with the measured metrics of **27** and previously reported values (1.451–1.465, 1.241–1.254 Å) [45]. Our calculated results compare favorably with those (1.488, 1.247 Å) optimized at the B3LYP/6-311+G(d) level in a recent theoretical study [47]. In model calculations on a 1:1 hydrogen-bonded aggregate of urea with  $C_5O_5^{2-}$  constrained to  $C_s$ ,  $C_2$  or  $C_{2v}$  symmetry (which converged to the same results), the  $C_5O_5^{2-}$  moiety exhibits the enediolate

characteristics similar to those observed in **27**. Notably the best agreement with the measured bond lengths (Figure 8.58b) is obtained at the HF/3-21G(d) level (C1–C2 = 1.433, C2–C3 = 1.457, C3–C4 = 1.477, C1–O1 = 1.267, C3–O3 = 1.238, C4–O4 = 1.232), as higher-level computations invariably yield significantly longer C–C bond distances. Natural bond orbital (NBO) analysis [62] lends credence to nonbenzenoid aromaticity in the isolated  $\text{C}_5\text{O}_5^{2-}$  system, with equal negative charge (−0.716) on each O atom and equal positive charge (+0.316) on each C atom. However, in the 1:1 adduct of  $\text{C}_5\text{O}_5^{2-}$  with urea, the pair of O atoms in the hydrogen-bonded ring carries a higher negative charge of −0.759, whereas those on the remaining three O atoms of  $\text{C}_5\text{O}_5^{2-}$  range from −0.644 to −0.671; when urea is replaced by 1,3-dimethylurea, the corresponding charges are changed to −0.790 and −0.618 to −0.691, respectively.

These findings clearly show that crystal engineering provides a viable route to breaking the degeneracy of canonical forms of a molecular species (Figures 8.51, 8.53b, 8.57 and 8.58b), furnishing reliable experimental data for comparison with results from theoretical models.

## 8.4 Supramolecular Assembly Based on the Rosette Motif

In the realm of crystal engineering [63] of organic and inorganic hybrid materials, the *de novo* self-assembly of hydrogen-bonded networks [64] that are topologically equivalent to the planar (6,3) honeycomb (or graphene) net [65] is a shining achievement in chemical creativity. The known examples are hexagonal or quasihexagonal, single- or two-component, layer-type networks constructed with molecular building units possessing inherent threefold axial symmetry [66]. Notably, essentially planar rosette layers exist in all three crystalline phases of guanidinium nitrate, and the two structural phase transitions involve a delicate balance between  $\text{N}^+\text{--H}\cdots\text{O}^-$  hydrogen bonds and electrostatic interactions [67]. In order to introduce the rosette element and increase the complexity of our previously studied urea/thiourea-anion host lattices, we incorporate the trigonal planar guanidinium cation in the hydrogen-bonded network. This idea is motivated by analogy to the neutral melamine-cyanuric acid rosette layer [66e, 68] and the guanidinium sulfonate puckered layer [69].

### 8.4.1 Anionic Rosette Ribbon Built of Guanidinium and Hydrogen Carbonate Dimer

Basically, the design and construction of hydrogen-bonded ‘supramolecular rosettes’ from guanidinium/organic sulfonate [69], trimesic acid [70], or cyanuric acid/melamine [71] depend on utilization of their topological equivalence, i.e. equal numbers of donor and acceptor hydrogen bonding sites and  $\text{C}_3$  symmetry of the component moieties. As a modification of the above strategies, a new kind of ‘fused-rosette ribbon’ can be constructed with the guanidinium cation ( $\text{GM}^+$ ) [69] and hydrogen carbonate dimer [5i, 6h, 8] ( $\text{HC}^-$ )<sub>2</sub> in the ratio of 1:1 (Figure 8.59).

Each supramolecular rosette comprises a quasihexagonal assembly of two  $\text{GM}^+$  and four  $\text{HC}^-$  units connected by strong  $\text{N}_{\text{GM}}\text{--H}\cdots\text{O}_{\text{HC}}$  and  $\text{O}_{\text{HC}}\text{--H}\cdots\text{O}_{\text{HC}}$  hydrogen bonds. The ( $\text{HC}^-$ )<sub>2</sub> dimer is shared as a common edge of adjacent rosettes and makes full use of its remaining acceptor sites in linking with  $\text{GM}^+$ . However, each  $\text{GM}^+$  in

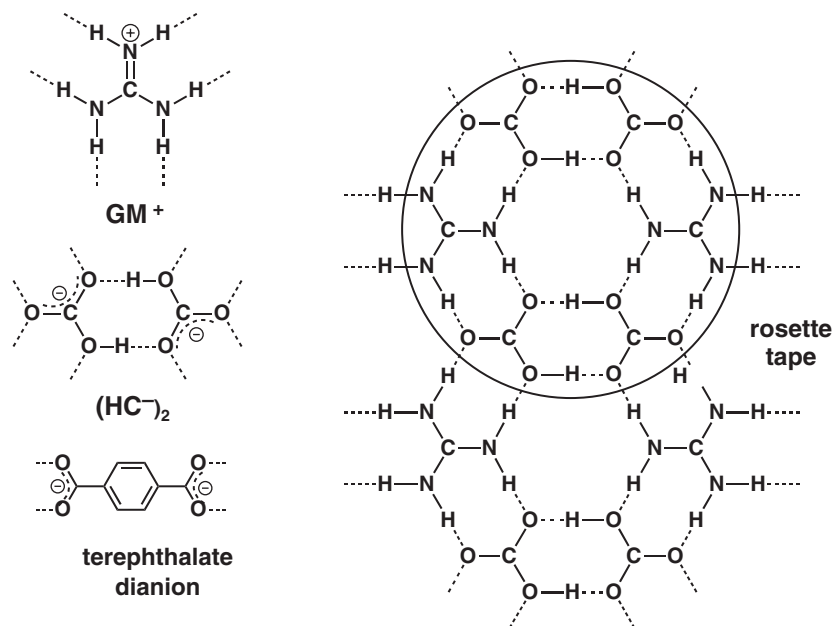
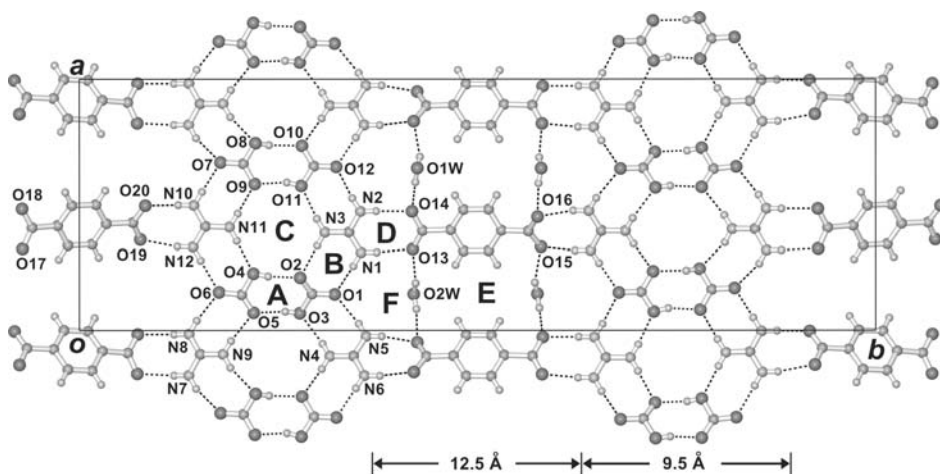


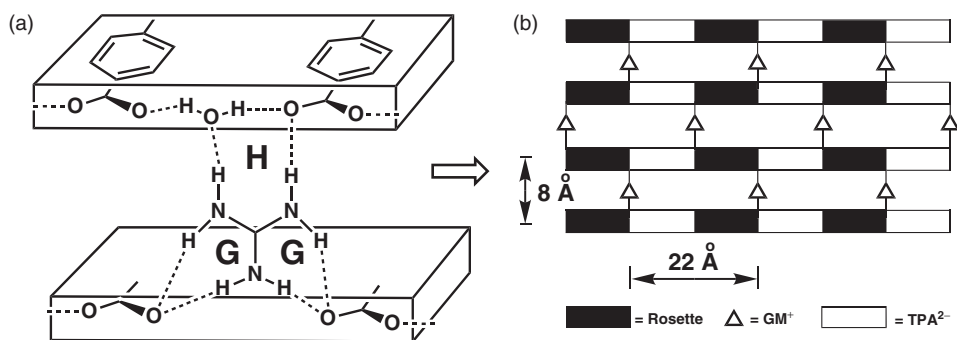
Figure 8.59 Design of supramolecular rosette tape and linker [72]

the resulting linear ribbon (or tape) still possesses a pair of free donor sites, and it is anticipated that some ‘molecular linker’ with suitable acceptor sites may be used to bridge an array of parallel ribbons to form a sheet-like network. This design objective has been realized in the synthesis and characterization of the inclusion compounds  $[(n\text{-C}_4\text{H}_9)_4\text{N}^+]_3[\text{C}(\text{NH}_2)_3]^+[\text{HCO}_3^-]_4[1,4\text{-C}_6\text{H}_4(\text{COO}^-)_2]_2 \cdot 2\text{H}_2\text{O}$  (**28**) and  $(n\text{-C}_4\text{H}_9)_4\text{N}^+[\text{C}(\text{NH}_2)_3]^+[\text{HCO}_3^-][1,2\text{-C}_6\text{H}_4(\text{COO}^-)(\text{NO}_2)] \cdot \text{H}_2\text{O}$  (**29**) with the terephthalate ( $\text{TPA}^{2-}$ ) or the 4-nitrobenzoate ( $\text{NB}^{4-}$ ) anion functioning as a linker [72].

As shown in Figure 8.60, each  $\text{HC}^-$  provides one donor and one acceptor site to form a planar dimer motif [A,  $R_2^2(8)$ ] [18]. The remaining eight acceptor sites of each  $(\text{HC}^-)_2$  dimer are topologically complemented by four  $\text{GM}^+$  units, such that each  $\text{GM}^+$  connects two  $(\text{HC}^-)_2$  dimers through two pairs of  $\text{N}^+-\text{H}_{\text{syn}} \cdots \text{O}^-$  hydrogen bonds [B,  $R_2^2(8)$ ]. Thus two  $\text{GM}^+$  units and two  $(\text{HC}^-)_2$  dimers constitute a planar, pseudocentrosymmetric, quasi-hexagonal supramolecular rosette [C,  $R_6^4(12)$ ] with inner and outer diameter of ca. 0.55 and 0.95 nm, respectively. In the resulting fused-rosette ribbon of **28**, the remaining two *exo*-orientated donor sites of each  $\text{GM}^+$  unit form a pair of  $\text{N}^+-\text{H}_{\text{anti}} \cdots \text{O}^-$  hydrogen bonds [D,  $R_2^2(8)$ ] with a terephthalate  $\text{TPA}^{2-}$  carboxylate group. Thus two types of ladders are developed: type (I) [ $\text{TPA}^{2-}$  containing O(13) to O(16)] is consolidated by two independent water molecules that alternately bridge carboxylate oxygen atoms of neighboring steps by pairs of donor  $\text{O}_w\text{-H} \cdots \text{O}^-$  hydrogen bonds, generating a centrosymmetric ring motif [E,  $R_4^4(22)$ ] and pentagon pattern [F,  $R_6^3(12)$ ]; in the type (II) ladder [ $\text{TPA}^{2-}$  containing O(17) to O(20)], carboxylate oxygen atoms O(18a) and O(19) belonging to adjacent steps are connected by the remaining



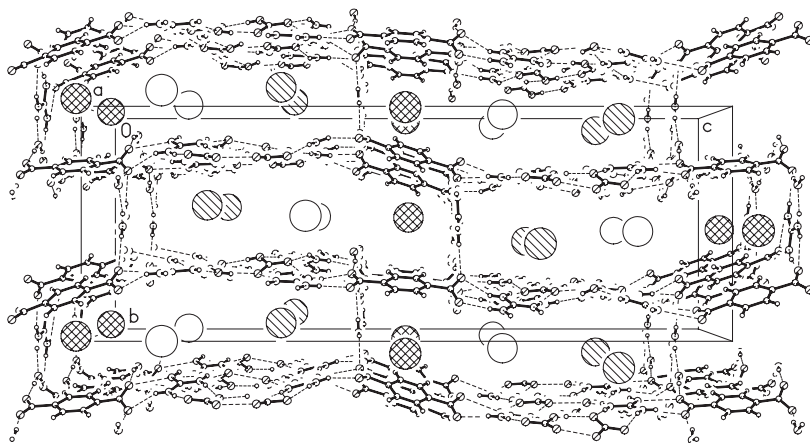
**Figure 8.60** Projection on (010) showing the 2-D network in **28**. Various hydrogen-bonding motifs as discussed in the text are labeled by bold capital letters **A–F**



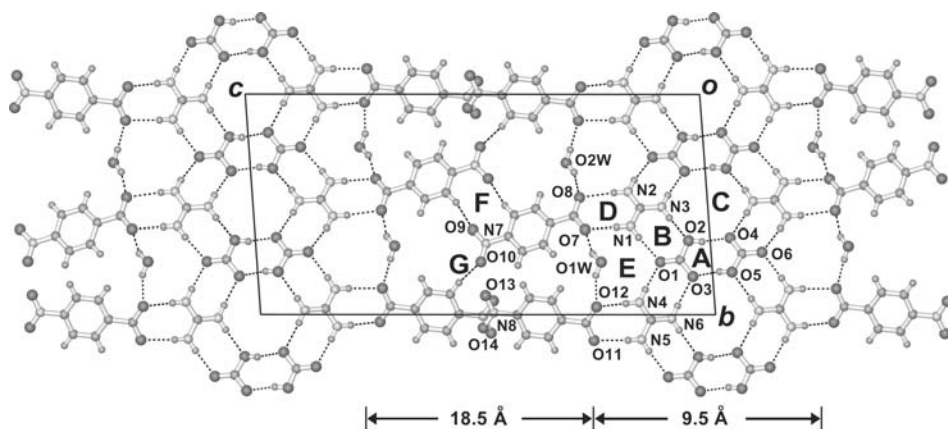
**Figure 8.61** (a) Hydrogen-bonding motifs involving linkage of the free guanidinium ion to neighboring rosette ribbon-terephthalate layers viewed along the *c* axis. (b) Schematic presentation of the pillared layer structure

$\text{GM}^+$  ions, which are not involved in rosette formation, via two pairs of donor hydrogen bonds [**G**,  $R_2^1(6)$ ] (Figure 8.61a). The remaining two pairs of donor sites of each free  $\text{GM}^+$  ion are linked to a carboxylate oxygen atom and a water molecule of  $\text{TPA}^{2-}$  column (**I**) of an adjacent layer to form a pentagon motif [**H**,  $R_3^2(8)$ ], thus yielding a 3-D pillared layer structure (Figure 8.61b). The large voids in the pillar region generate nanoscale channels extending along the [100] direction. The dimensions of the cross-section of each channel are about  $8 \text{ \AA} \times 22 \text{ \AA}$ , within which three independent  $[(n\text{-C}_4\text{H}_9)_4\text{N}]^+$  cations are aligned in separate columns in a well-ordered manner (Figure 8.62).

The rosette ribbon of **29** bears a strong resemblance to that of **28**, except that the hydrogen-bonding pattern **C** is centrosymmetric and the spacing between neighboring parallel ribbons is increased to ca.  $18.5 \text{ \AA}$  (Figure 8.63). The independent  $\text{GM}^+$  cations,



**Figure 8.62** Crystal structure of **28** viewed along the *a* axis showing the pillared layer motif depicted in Figure 8.61(b). The nanoscale voids in the pillared region with dimensions of  $8 \text{ \AA} \times 22 \text{ \AA}$  are capable of accommodating three columns of independent  $[(n\text{-C}_4\text{H}_9)_4\text{N}]^+$  cations which are represented by large open, shaded and cross-hatched circles



**Figure 8.63** Perspective view of the hydrogen-bonded layer in **29** along the  $[110]$  direction. Linear ribbons of fused rosettes are interlinked by 4-nitrobenzoate ions through strong  $\text{N}^+\text{-H}\cdots\text{O}^-$  (carboxylate) and weak  $\text{C-H}\cdots\text{O}$  (nitro) hydrogen bonding

**C(1)** and **C(2)**, form a pair of  $\text{N}^+\text{-H}_{anti}\cdots\text{O}^-$  hydrogen bonds [**D**,  $N_2 = R_2^2(8)$ ] with the carboxylate groups of **NB4**<sup>−</sup> anions **N(7)** and **N(8)**, respectively. As in the case of **28**, two independent water molecules secure the hydrogen-bonded network by bridging the oxygen atoms of neighboring **NB4**<sup>−</sup> carboxylate groups alongside the rosette ribbon [**E**,  $N_6 = R_6^3(12)$ ]. The resulting hemmed rosette ribbon-nitrobenzoate chains are organized into an undulating layer structure with the formation of weak  $\text{C-H}\cdots\text{O}$  interactions [73], leading to a centrosymmetric ring motif [**F**,  $N_1 = R_2^2(10)$ ] plus a discrete motif [**G**,  $N_1 = D$ ] in the interdigitated region. The two independent  $[(n\text{-C}_4\text{H}_9)_4\text{N}]^+$  cations

are sandwiched between adjacent layers. The interlayer separation is 8.2 Å, which is comparable with those in **28** and in other anionic host structures accommodating  $[(n\text{-C}_4\text{H}_9)_4\text{N}]^+$  cations as guest moieties [5a, 5e, 6c, 6d].

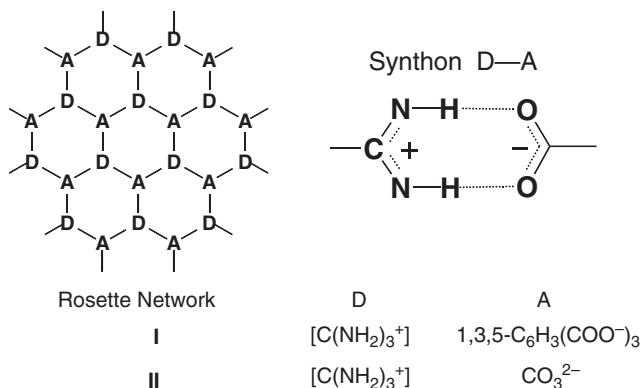
#### 8.4.2 Anionic Rosette Networks Assembled with Guanidinium and $C_3$ -symmetric Oxo-anion Building Blocks

Drawing upon the design of a linear ‘fused-rosette ribbon’ assembled from the hydrogen carbonate dimer and guanidinium ion in 1:1 molar ratio [72], as well as the generation of urea/thiourea-anion host layers and their stabilization with bulky organic cations as interlayer spacers [5c, 5h, 8, 57], we set about the hydrogen-bond-mediated construction of two premeditated anionic rosette-layer architectures using guanidinium and ubiquitous  $C_3$ -symmetric oxo-anions that carry *unequal* charges, namely, guanidinium-trimesate **I** and guanidinium-carbonate **II**, as illustrated in Scheme 8.5 [74].

In principle, the negatively charged, presumably planar guanidinium-carbonate network can be combined with one molar equivalent of tetraalkylammonium ion  $R_4N^+$  of the right size as interlayer template to yield a crystalline inclusion compound of idealized stoichiometric formula  $(R_4N^+)[C(NH_2)_3]^+CO_3^{2-}$  that is reminiscent of the graphite intercalates [75]. Anionic guanidinium-trimesate network, however, needs twice as many monovalent cations for charge balance, and furthermore possesses honeycomb-like host cavities of diameter  $\sim 7$  Å that have to be filled by suitable guest species. The expected formula of the corresponding inclusion compound is  $(R_4N^+)_2[C(NH_2)_3]^+[1,3,5-C_6H_3(COO^-)_3] \cdot G$ , where G is an entrapped guest moiety with multiple hydrogen-bond donor sites to match the nearly planar set of six carboxylate oxygens that line the inner rim of each cavity.

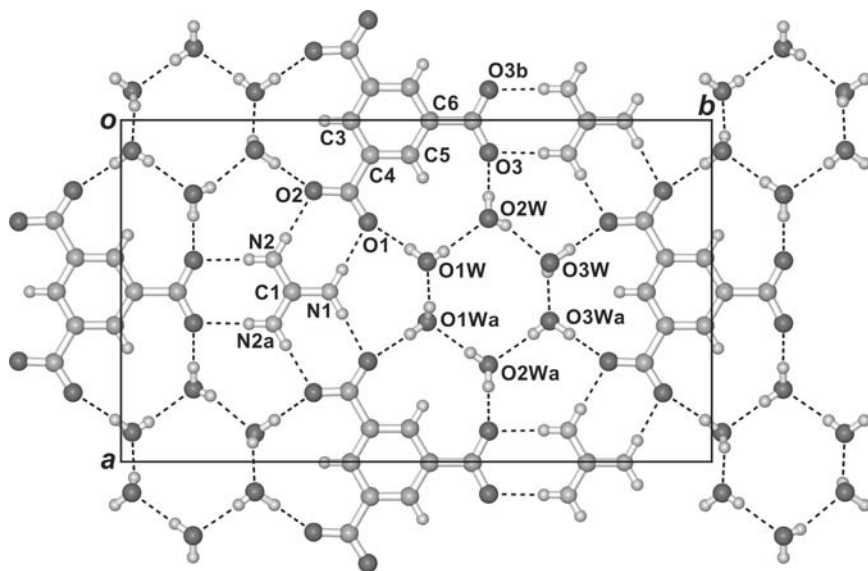
#### 8.4.2.1 Planar Rosette Layer

The predicted assembly of guanidinium-trimesate network **I** (Scheme 8.5) was achieved through the crystallization of  $[(\text{C}_2\text{H}_5)_4\text{N}^+]_2[\text{C}(\text{NH}_2)_3]^+[\text{1,3,5-}\text{C}_3\text{H}_3(\text{COO}^-)_3] \cdot 6\text{H}_2\text{O}$  (**30**) [74]. The guanidinium and trimesate ions are connected together by pairs of strong,



**Scheme 8.5** *Designed construction of two-component rosette layer*



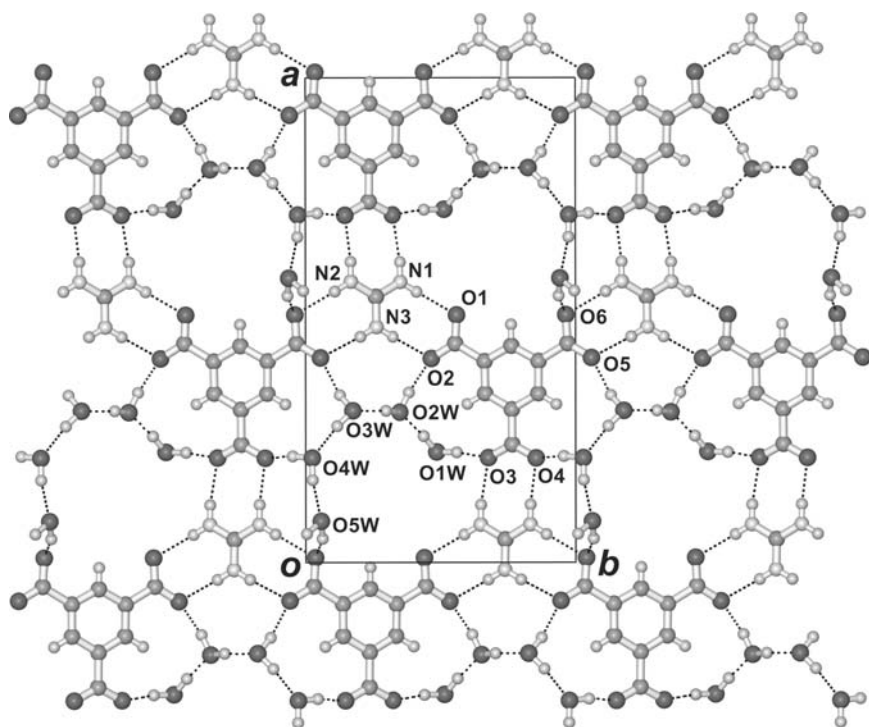


**Figure 8.64** Projection diagram showing the hydrogen-bonding scheme in the infinite rosette layer **I** of **30**. Only one of the two cyclic arrangements of disordered hydrogen atoms lying on the edges of each  $(\text{H}_2\text{O})_6$  ring is displayed. Symmetry transformations: *a*  $(1 - x, y, \frac{1}{2} - z)$  and *b*  $(-x, y, \frac{1}{2} - z)$

charge-assisted  $\text{N}^+-\text{H}\cdots\text{O}^-$  hydrogen bonds to generate an essentially planar rosette layer with large honeycomb cavities (Figure 8.64). Three independent water molecules constitute a cyclic  $(\text{H}_2\text{O})_6$  cluster of symmetry 2, which is tightly fitted into each host cavity by adopting a flattened-chair configuration in an out-of-plane orientation, with  $\text{O}\cdots\text{O}$  distances comparable with 2.759 Å in deuterated ice  $I_h$  [76]. Each water molecule has its ordered hydrogen atom pointing outward to form a strong  $\text{O}-\text{H}\cdots\text{O}^-$  hydrogen bond with a carboxylate oxygen on the inner rim of the cavity. The well-ordered  $(\text{C}_2\text{H}_5)_4\text{N}^+$  guests are sandwiched between anionic rosette host layers with an interlayer spacing of  $\sim 7.5$  Å.

The robustness of the planar rosette layer in **30** can be tested by varying the size of the cationic template. A series of four new isostructural crystalline inclusion compounds, namely  $[\text{CH}_2 = \text{CHCH}_2(\text{C}_2\text{H}_5)_3\text{N}^+]_2[\text{C}(\text{NH}_2)_3^+][1,3,5\text{-C}_6\text{H}_3(\text{COO}^-)_3] \cdot 6\text{H}_2\text{O}$  (**31**),  $[\text{n-C}_3\text{H}_7(\text{C}_2\text{H}_5)_3\text{N}^+]_2[\text{C}(\text{NH}_2)_3^+][1,3,5\text{-C}_6\text{H}_3(\text{COO}^-)_3] \cdot 6\text{H}_2\text{O}$  (**32**),  $[\text{n-C}_4\text{H}_9(\text{C}_2\text{H}_5)_3\text{N}^+]_2[\text{C}(\text{NH}_2)_3^+][1,3,5\text{-C}_6\text{H}_3(\text{COO}^-)_3] \cdot 6\text{H}_2\text{O}$  (**33**) and  $[(\text{n-C}_3\text{H}_7)_3\text{N}(\text{CH}_2)_4\text{N}(\text{n-C}_3\text{H}_7)_3]^{2+}[\text{C}(\text{NH}_2)_3^+][1,3,5\text{-C}_6\text{H}_3(\text{COO}^-)_3] \cdot 5\text{H}_2\text{O}$  (**34**) have been prepared and shown to contain the same kind of planar guanidinium-trimesate-water network [77]. Isomorphous complexes **30–33** belong to space group *Pbcn*, and gradual extension of one leg of the  $(\text{C}_2\text{H}_5)_4\text{N}^+$  reference template does not affect the inter-rosette layer separation. In contrast, **34** generated with a dicationic  $[(\text{n-C}_3\text{H}_7)_3\text{N}(\text{CH}_2)_4\text{N}(\text{n-C}_3\text{H}_7)_3]^{2+}$  guest template crystallizes in space group *Pca*2<sub>1</sub>, and its honeycomb network is distorted to a pleated sheet while each host cavity accommodates a helical  $(\text{H}_2\text{O})_5$  fragment (Figure 8.65).

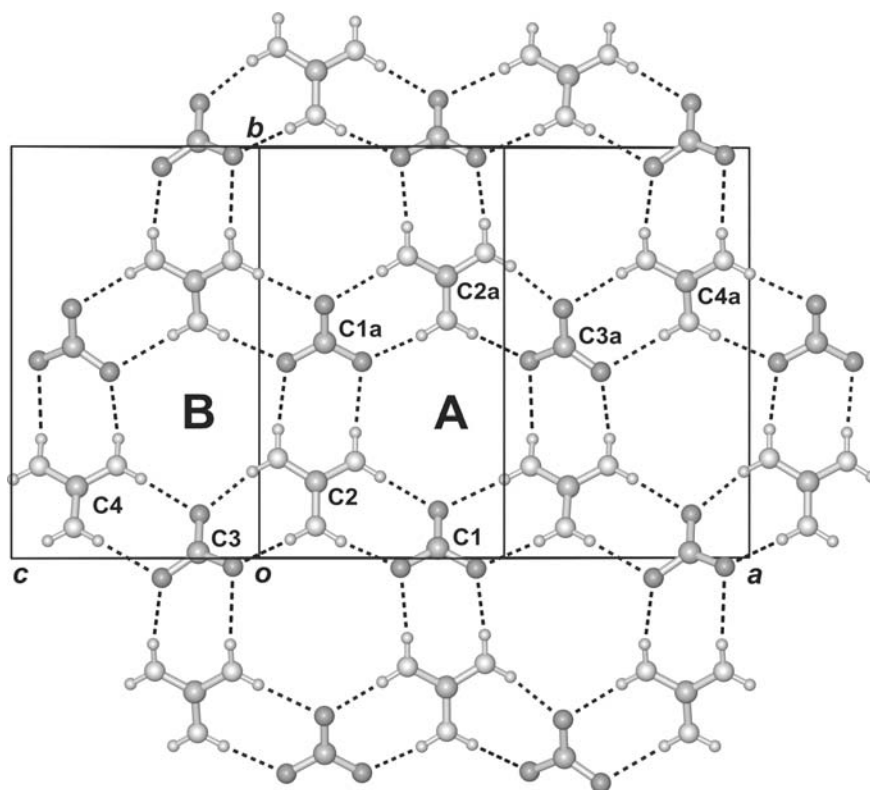




**Figure 8.65** Projection diagram showing the hydrogen-bonding scheme in the puckered rosette layer of **34**. The well-resolved helical  $(\text{H}_2\text{O})_5$  chain is embedded inside a distorted hexagonal cavity

In the crystal structure of  $(\text{C}_2\text{H}_5)_4\text{N}^+[\text{C}(\text{NH}_2)_3]^+_{17}[(\text{CO}_3)^{2-}]_3[\text{C}_3\text{N}_2\text{H}_2(\text{COO}^-)_2]$  (**35**) [78], which contains the dianion of 1*H*-imidazole-4,5-dicarboxylic acid, there are seven guanidinium and three carbonate ions in the asymmetric unit, in which two guanidinium cations [**C(2)**, **C(4)**] (designated by their central carbon atoms for convenience) and two carbonate anions [**C(1)**, **C(3)**] are interconnected via strong  $\text{N}^+-\text{H}\cdots\text{O}^-$  charge-assisted hydrogen bonds to form a quasi-hexagonal supramolecular rosette layer (network **II**, Scheme 8.5) exhibiting the  $R_6^3(12)$  motif (Figure 8.66). It is notable that this guanidinium-carbonate (1:1) rosette layer has an approximately planar configuration with a mean atomic deviation of 0.44 Å, in sharp contrast to the sinusoidal wavy layer of the same composition in  $[(\text{C}_2\text{H}_5)_4\text{N}^+]_4[\text{C}(\text{NH}_2)_3]^+_{18}[(\text{CO}_3)^{2-}]_3[3(\text{C}_2\text{O}_4)^{2-}] \cdot 2\text{H}_2\text{O}$  (**36**) [74].

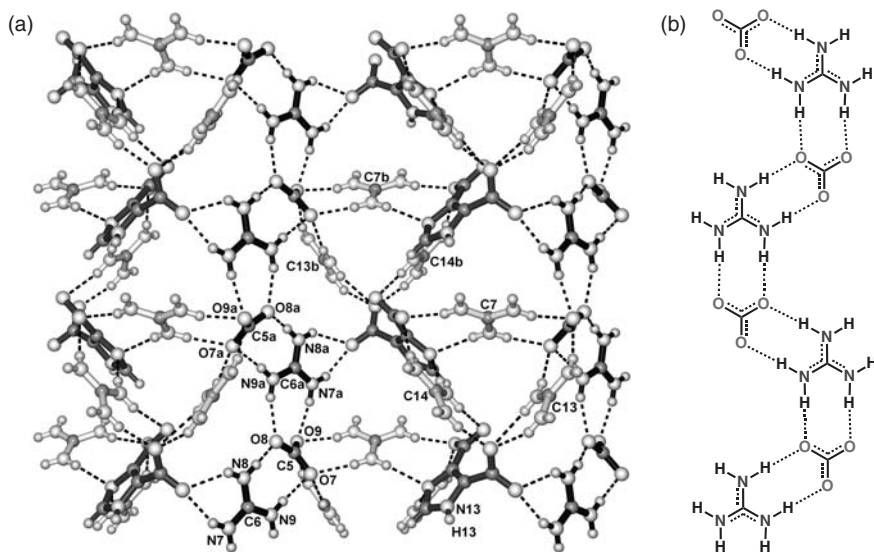
Remarkably, there is an entirely different kind of sinusoidal guanidinium-carbonate-1*H*-imidazole-4,5-dicarboxylate (4:1:1) layer (Figure 8.67a) that functions as a lamina between the essentially planar guanidinium-carbonate (1:1) rosette layers. A perspective view of this wavy layer along  $c^*$  reveals that carbonate **C(5)** and guanidinium **C(6)** are joined together through pairs of  $\text{N}^+-\text{H}\cdots\text{O}^-$  hydrogen bonds to create an infinite crinkled tape  $[\text{C}(\text{NH}_2)_3^+\text{CO}_3^{2-}]_\infty$  (colored in red) that extends along the  $b$  direction (Figure 8.67b). Such tapes are further cross-linked through three independent



**Figure 8.66** Projection along  $c^*$  showing the nearly planar guanidinium-carbonate (1:1) rosette network **II** at  $z = 1/2$  in the crystal structure of **35**. The atom types are differentiated by size and different grey scales, and  $N-H \cdots O$  hydrogen bonds are represented by dotted lines. The layer is composed of independent carbonate ions [labeled by their central carbon atoms **C(1)** and **C(3)**] and guanidinium ions [**C(2)** and **C(4)**] in equal numbers. {Note: Mean atomic deviation from plane of one rosette =  $0.36 \text{ \AA}$ , and mean atomic deviation from plane of ribbon fragment [**C(1)**, **C(2)**, **C(3)**, **C(4)**] =  $0.44 \text{ \AA}$ . Symmetry transformation:  $a(1 - x, \frac{1}{2} + y, 1 - z)$ }

guanidinium cations and one dicarboxylate. As illustrated in Figure 8.67(a), guanidinium cations **C(7)**, **C(13)**, and **C(14)** (depicted in blue) constitute a ‘tripod’ linker at the crest of the wavy layer, providing a total of six hydrogen-bond donor sites that bind to the carbonate anions in the rosette layer above at  $N^+-H \cdots O^-$  distances ranging from  $2.82$  to  $2.99 \text{ \AA}$ .

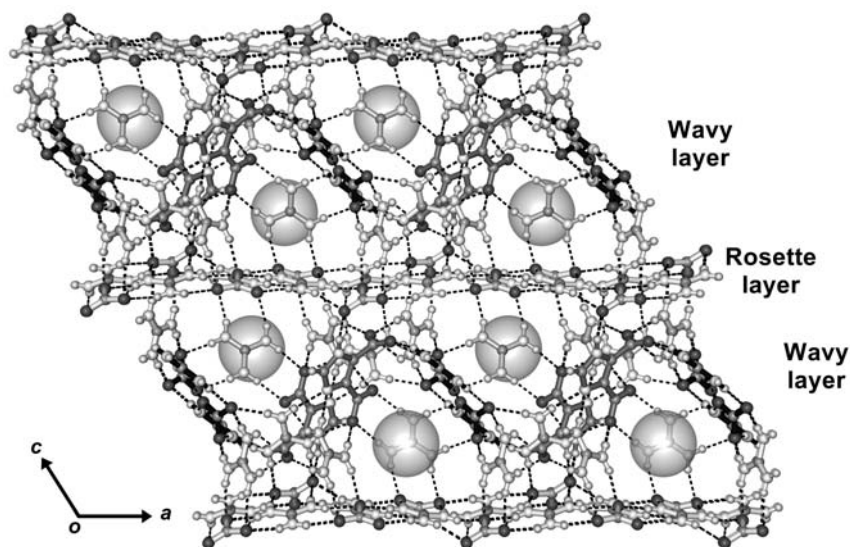
Concomitantly, the symmetry-related ‘tripod’ [**C(7b)**, **C(13b)**, and **C(14b)**] at the guanidinium cations exhibits different hydrogen-bonding environments. Guanidinium **C(7)** and **C(13)** both form a pair of  $N^+-H \cdots O^-$  hydrogen bonds to the same carbonate ion from the crinkled tape; at the other side, **C(7)** is attached to a dicarboxylate through  $N^+-H \cdots O^-$ (carboxylate) and  $N^+-H \cdots N$ (imidazole ring) hydrogen bonds, and **C(13)** is linked to a symmetry-related 1*H*-imidazole-4,5-dicarboxylate through a pair of chelating  $N^+-H \cdots O^-$ (carboxylate) hydrogen bonds.



**Figure 8.67** (a) Perspective view of the wavy guanidinium-carbonate-1*H*-imidazole-4,5-dicarboxylate (4:1:1) layer of **35** lying in the (001) plane. The embedded crinkled tape  $[C(NH_2)_3^+ CO_3^{2-}]_\infty$  (colored in red) is composed of an alternate arrangement of guanidinium **C(6)** and carbonate **C(5)** ions running parallel to the *b* axis. Three guanidinium cations [**C(7)**, **C(13)**, and **C(14)**] constitute a 'tripod' that connects to carbonate anions in the proximal rosette layer. Symmetry transformations: *a*  $(2 - x, \frac{1}{2} + y, -z)$  and *b*  $(1 - x, \frac{1}{2} + y, -z)$ . (b) Chemical structure of the crinkled tape (see colour plate section)

Guanidinium **C(14)** is connected to a dicarboxylate through  $N^+ - H \cdots O^-$  (carboxylate) and  $N^+ - H \cdots N$  (imidazole ring) hydrogen bonds on one side and by a pair of  $N^+ - H \cdots O^-$  (carboxylate) hydrogen bonds to a symmetry-related dicarboxylate on the other. In addition to the 'tripod' linkages above and below each wavy layer, there is an extra (imidazole) $N13 - H \cdots O6$  [carbonate **C(3)**] hydrogen bond with  $d_{N \cdots O} = 2.84$  Å from 1*H*-imidazole-4,5-dicarboxylate in the wavy layer to carbonate **C(3)** which protrudes prominently out of the rosette layer.

Adjacent guanidinium-carbonate (1:1) rosette layers are thus connected by a multitude of hydrogen bonds to a sandwiched guanidinium-carbonate-1*H*-imidazole-4,5-dicarboxylate (4:1:1) layer, which is folded and consolidated by the independent guanidinium cation **C(15)** (Figure 8.68). This guanidinium ion plays a significant role in cementing the two kinds of layers together, as it forms two pairs of strong  $N^+ - H \cdots O^-$  hydrogen bonds to carbonate and carboxylate groups of the wavy layer and a third pair of  $N^+ - H \cdots O^-$  hydrogen bonds to a carbonate group in the rosette layer, with  $d_{N \cdots O}$  distances ranging from 2.84 to 3.07 Å. These quasi- $C_3$  symmetric interactions essentially hold neighboring layers in position in a wafer-like fashion, such that each infinite column of approximately triangular cross-section, as viewed along *b*, is partitioned by the guanidinium ions **C(15)** into cavities for accommodating the hydrophobic tetraethylammonium cations.



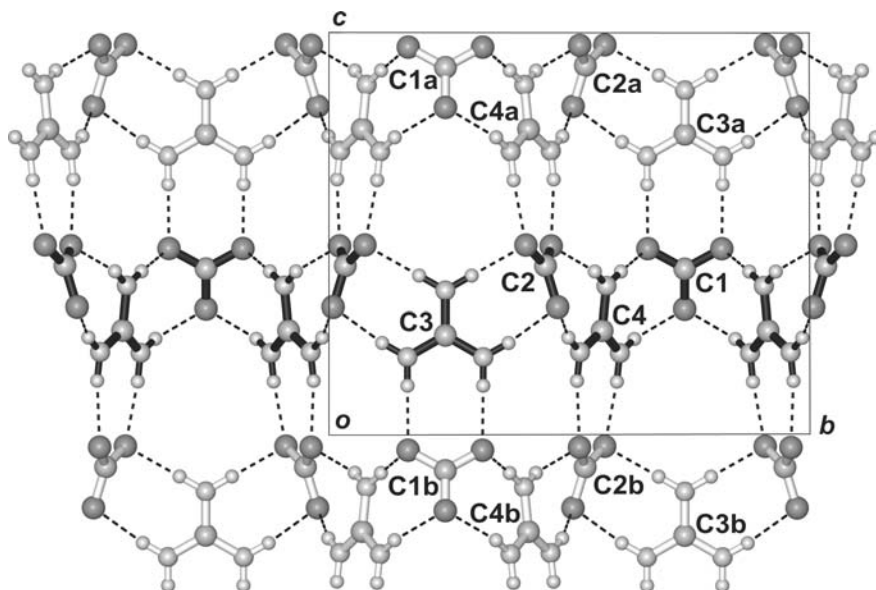
**Figure 8.68** Projection along the *b* axis showing the crystal structure of **35**, in which the nearly planar rosette layers are sandwiched by the wavy layers. The large spheres representing the ordered  $(\text{C}_2\text{H}_5)_4\text{N}^+$  cations are each accommodated inside a pocket bounded by adjacent rosette and wavy layers and closed by ‘windows’ composed of independent guanidinium cations **C(15)** that stabilize the hydrogen-bonded cage work (see colour plate section)

#### 8.4.2.2 Nonplanar Rosette Layer

Persistent attempts to crystallize  $(\text{R}_4\text{N}^+)[\text{C}(\text{NH}_2)_3]^+\text{CO}_3^{2-}$  by variation of R, based on the conceptual network **II**, were not successful. Mindful of the fact that the guanidinium ion can function as a pillar between layers [72] and the carbonate ion is capable of forming up to twelve acceptor hydrogen bonds, as observed in crystalline bis-guanidinium carbonate [79] and  $[(\text{C}_2\text{H}_5)_4\text{N}^+]_2\text{CO}_3^{2-} \cdot 7(\text{NH}_2)_2\text{CS}$  [17], we modified our synthetic strategy by incorporating a second guanidinium salt  $[\text{C}(\text{NH}_2)_3]\text{X}$  as an extra component. After much experimentation with various combinations of R and X, the targeted construction of network **II**, albeit in undulating form, was realized through the isolation of crystalline  $[(\text{C}_2\text{H}_5)_4\text{N}^+]_4[\text{C}(\text{NH}_2)_3]^+_{18}[(\text{CO}_3)^{2-}]_3[3(\text{C}_2\text{O}_4)^{2-}] \cdot 2\text{H}_2\text{O}$  (**36**) [74].

In the asymmetric unit of **36**, there are two independent carbonate anions and five independent guanidinium cations, which are henceforth conveniently referred to by their carbon atom labels in bold type. Carbonate **C(1)** and guanidinium **C(3)** each has one bond lying in a crystallographic mirror plane; together with carbonate **C(2)** and guanidinium **C(4)**, they form a nonplanar zigzag ribbon running parallel to the *b* axis, neighboring units being connected by a pair of strong  $\text{N}^+-\text{H} \cdots \text{O}^-$  hydrogen bonds (Figure 8.69).

Adjacent antiparallel  $\{[\text{C}(\text{NH}_2)_3]^+ \cdot \text{CO}_3^{2-}\}_\infty$  ribbons are further cross-linked by strong  $\text{N}^+-\text{H} \cdots \text{O}^-$  hydrogen bonds to generate a highly corrugated rosette layer, which is folded into a plane-wave pattern by guanidinium **C(5)** (Figure 8.70). Guanidinium **C(6)** and **C(7)** protruding away from carbonate **C(1)** and **C(2)** are hydrogen-bonded to the twofold disordered oxalate ion containing **C(8)** and **C(9)**, forming a pouch that cradles the



**Figure 8.69** Projection diagram showing a portion of the nonplanar anionic rosette network **II** concentrated at  $a = \frac{1}{4}$  in the crystal structure of **36**. Adjacent antiparallel  $\{[C(NH_2)_3]^+ \cdot CO_3^{2-}\}_\infty$  ribbons are running parallel to the  $b$  axis. Symmetry transformations:  $a (\frac{1}{2} - x, 1 - y, \frac{1}{2} + z)$  and  $b (\frac{1}{2} - x, y - \frac{1}{2}, z - \frac{1}{2})$  (see colour plate section)

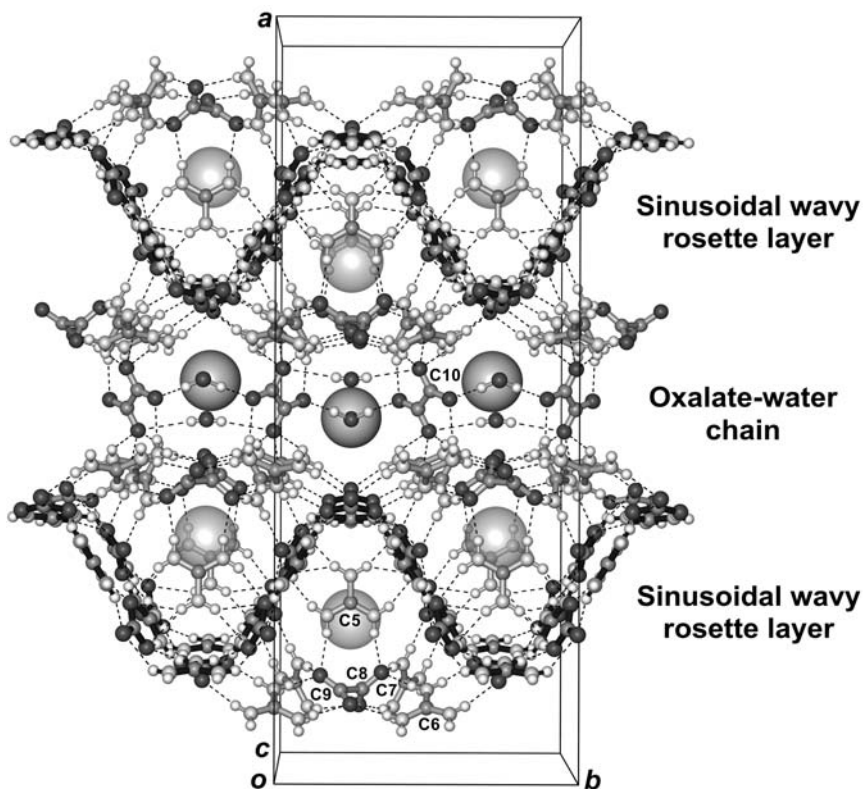
disordered  $(C_2H_5)_4N^+$  ion. The carbonate ions **C(1)** and **C(2)** each forms eleven acceptor hydrogen bonds, only one fewer than the maximum number [17]. The resulting composite hydrogen-bonded layers at  $a = 1/4$  and  $3/4$  are interconnected by  $[C_2O_4^{2-} \cdot (H_2O)_2]_\infty$  chains derived from centrosymmetric oxalate **C(10)** and water molecules **O1w** and **O2w** via strong  $N^+ - H \cdots O^-$  hydrogen bonds to generate a complex 3-D host framework, within which the second kind of disordered  $(C_2H_5)_4N^+$  ions are accommodated in a zigzag fashion within channels extending along the  $[010]$  direction.

#### 8.4.2.3 Three-component Planar Rosette Layer

In the crystal structure of  $(n-C_3H_7)_4N^+ [C(NH_2)_3]^+ CO_3^{2-} \cdot [B(OH)_3]_2$  (**37**), three differently charged molecular components are used to construct a hydrogen-bonded rosette layer: guanidinium **C(1)**, boric acid **B(1)**, and carbonate **C(2)** in the ratio of 1:2:1 [78]. The centrosymmetrically related boric acid molecules **B(1a)** and **B(1c)** are joined together through a pair of  $O-H \cdots O$  hydrogen bonds of about 2.72 Å to form dimer **[C]** (Figure 8.71).

Carbonate **C(2)** and guanidinium **C(1)** both have site symmetry  $m$  and are connected pairwise through charge-assisted  $N^+ - H \cdots O^-$  hydrogen bonds of about 2.80 Å to form motif **[D]**. Notably, motifs **[C]** and **[D]** are linked together in two ways by pairs of  $O_{BA} - H \cdots O_{CB}^-$  (2.56 and 2.59 Å) or  $N_{GM}^+ - H \cdots O_{BA}$  (2.89 and 2.93 Å) hydrogen bonds to generate motif **[E]** and **[F]**, respectively. Thus three  $C_3$ -symmetric but differently charged molecular components, namely the guanidinium cation, boric acid, and the carbonate dianion, coexist in a 1:2:1 ratio in one rosette layer with the first and last serving as

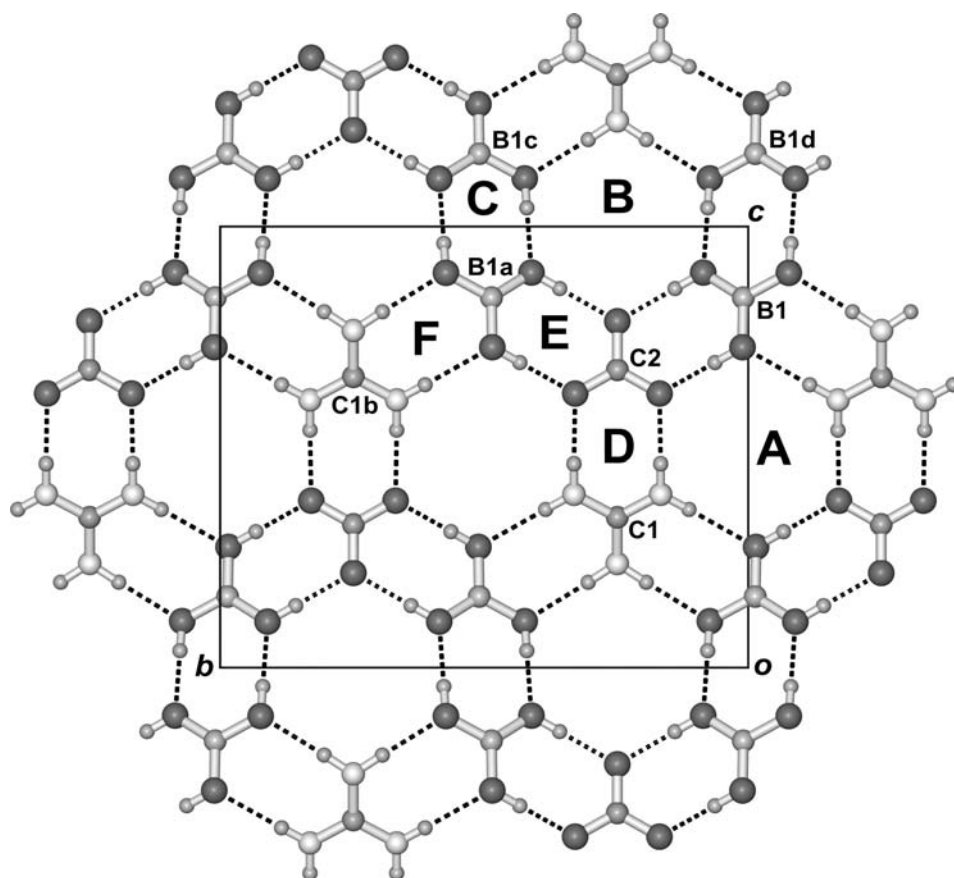




**Figure 8.70** Perspective view of the crystal structure of **36** along [001]. The undulating guanidinium-carbonate rosette network **II**, seen as a sinusoidal cross-section, is highlighted in yellow. Adjacent composite hydrogen-bonded layers are inter-connected by a  $[\text{C}_2\text{O}_4^{2-} \cdot (\text{H}_2\text{O})_2]_\infty$  chain. The disordered oxalate containing **C(8)** and **C(9)** is shown in one possible orientation, and the two different types of disordered  $(\text{C}_2\text{H}_5)_4\text{N}^+$  ions (represented by large semi-transparent spheres of two different colors) are included in the pouches and the zigzag channels running parallel to the *b* axis, respectively (see colour plate section)

hydrogen-bonding donor and acceptor, respectively. However, each boric acid molecule functions as both donor and acceptor, such that a pair of adjacent hydroxyl groups is orientated syn-syn, syn-anti, and anti-anti, respectively, with respect to its proximal carbonate, boric acid, and guanidinium neighbors (Figure 8.71). In this tertiary guanidinium-boric acid-carbonate (1:2:1) system, dimeric motifs **[C]**, **[D]**, **[E]**, and **[F]** are interlinked to generate two distinguishable rosette motifs **[A]** and **[B]** in the same layer. Scheme 8.6(a)–(c) compares the centrosymmetric assembly around rosette **[A]** designated as  $[\text{GM}_2 \cdot \text{CB}_2 \cdot \text{BA}_2]$ , the  $[\text{GM}_1 \cdot \text{CB}_1 \cdot \text{BA}_4]$  assembly around rosette **[B]** having mirror symmetry and the  $[\text{GM}_3 \cdot \text{CB}_3]$  assembly in a conventional two-component system.

The mean atomic deviation from the mean plane is 0.12 and 0.29 Å for rosettes **[A]** and **[B]**, respectively. The dihedral angle between **[A]** and **[B]** is about  $24.3^\circ$ , which reflects the wavy characteristics of the rosette layer.



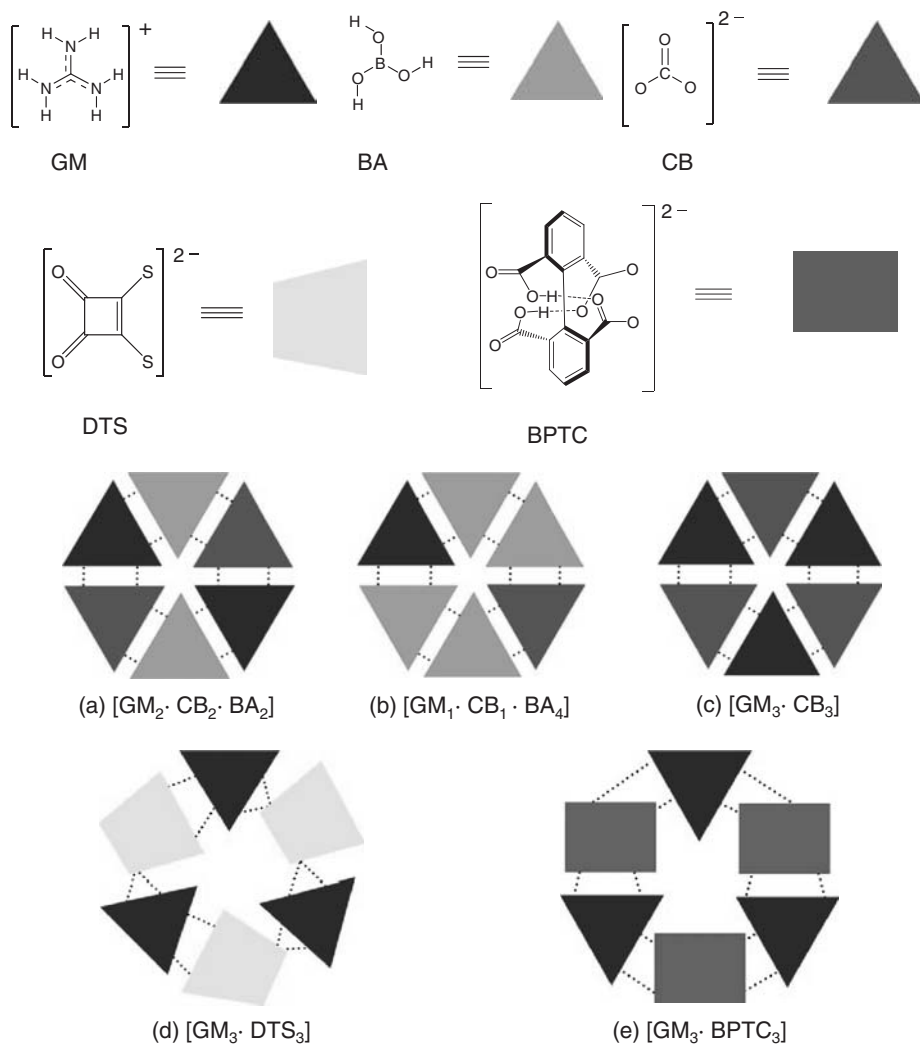
**Figure 8.71** Projection diagram showing a portion of the guanidinium-boric acid-carbonate 1:2:1 rosette network viewed along the  $a$  axis at  $x = 1/2$  in the crystal structure of **37**. The atom types are differentiated by size and color, and  $N-H \cdots O$  and  $O-H \cdots O$  hydrogen bonds are represented by broken lines. The layer is composed of independent guanidinium ion **C(1)** (labeled by the central carbon atom), boric acid **B(1)**, and carbonate ion **C(2)**. Symmetry transformations:  $a$  ( $x, \frac{1}{2} - y, z$ ),  $b$  ( $1 - x, 1 - y, 1 - z$ ),  $c$  ( $1 - x, \frac{1}{2} + y, 2 - z$ ) and  $d$  ( $1 - x, -y, 2 + z$ ). Mean atomic deviation from rosette plane: [**A**], 0.12 Å; [**B**], 0.29 Å (see colour plate section)

In the crystal lattice of **37**, the well-ordered tetra- $n$ -propylammonium cation occupies a special position of symmetry  $m$ . The hydrophobic cations located at  $x \approx 1/4$  and  $3/4$  are sandwiched between the sinusoidal rosette layers.

#### 8.4.3 Anionic Rosette Layers Formed by Guanidinium and Non- $C_3$ Molecular Building Blocks

In our attempt to extend the conventional topological design of supramolecular hydrogen-bonded rosette layers by relaxing the requirement of exact or near  $C_3$  symmetry of





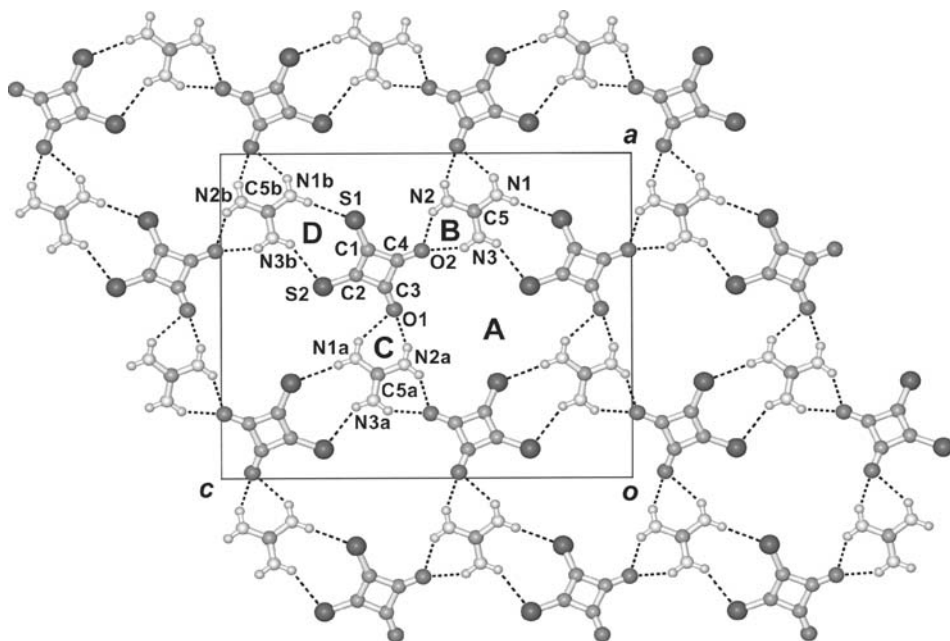
**Scheme 8.6** Diagrammatic representation of different rosette motifs and their graph-set notations. (a) Two-component rosette built of  $C_3$ -symmetric components. (b) and (c) Different three-component rosettes built of  $C_3$ -symmetric components. (d) and (e) Two-component rosettes involving different non- $C_3$ -symmetric components (see colour plate section)

the molecular building blocks, 1,2-dithiosquaric acid ( $H_2C_4O_2S_2$ ) and 1,1'-biphenyl-2,2',6,6'-tetracarboxylic acid ( $H_2BPTC$ ) were used as suitable starting materials. We selected 1,2-dithiosquarate  $C_4O_2S_2^{2-}$  in anticipation that, owing to the weakness of the (guanidinium) $N-H \cdots S$  hydrogen bond compared with  $N-H \cdots O$ , it may form acceptor hydrogen bonding in three principal directions for the generation of a rosette network. The 1,1'-biphenyl-2,2',6,6'-tetracarboxylate (BPTC) dianion is known to possess a rigid nonplanar molecular skeleton held by a pair of intramolecular hydrogen bonds, so that the two phenyl rings are nearly orthogonal to each other.

## 8.4.3.1 1,2-Dithiosquarate as Component

The complex  $(n\text{-C}_4\text{H}_9)_4\text{N}^+[\text{C}(\text{NH}_2)_3]^+[\text{C}_4\text{O}_2\text{S}_2^{2-}]$  (**38**) [78] has an anionic host network exhibiting a distorted quasi-hexagonal supramolecular rosette motif [**A**] =  $R_6^6(21)$  (Figure 8.72). The dithiosquarate dianion carries a negative charge on each sulfur atom. It is connected to guanidinium cations **C(5)** and **C(5a)** by chelated hydrogen bonds with the respective carbonyl oxygen atom serving as a bifurcated acceptor to form motifs [**B**] and [**C**], both being  $R_2^1(6)$ , and also to **C(5b)** via a pair of  $\text{N}^+\text{--H}\cdots\text{S}^-$  hydrogen bonds to form  $R_2^2(9)$  motif [**D**]. The measured distances of  $\text{N}^+\text{--H}\cdots\text{S}^-$  hydrogen bonds involving atoms S(1) and S(2) are 3.36 and 3.46 Å, respectively, and those of geminal  $\text{N}^+\text{--H}\cdots\text{O}$  hydrogen bonds are 2.86 and 3.05 Å at O(1), and 2.81 and 2.87 Å at O(2), respectively. Thus the guanidinium cations and dithiosquarate dianions are alternately linked to form the highly asymmetrical quasi-rosette motif [**A**] =  $R_6^6(21)$ , and the mean atomic deviation from its least-squares plane is 0.31 Å (see Scheme 8.6d).

The slightly wavy rosette layer in the crystal structure of **38** has an interlayer spacing of 8.45 Å ( $= b/2$ ). The well-ordered tetra-*n*-butylammonium cations have the terminal methyl carbon atom of one butyl group pointing towards the central void of rosette motif [**A**].

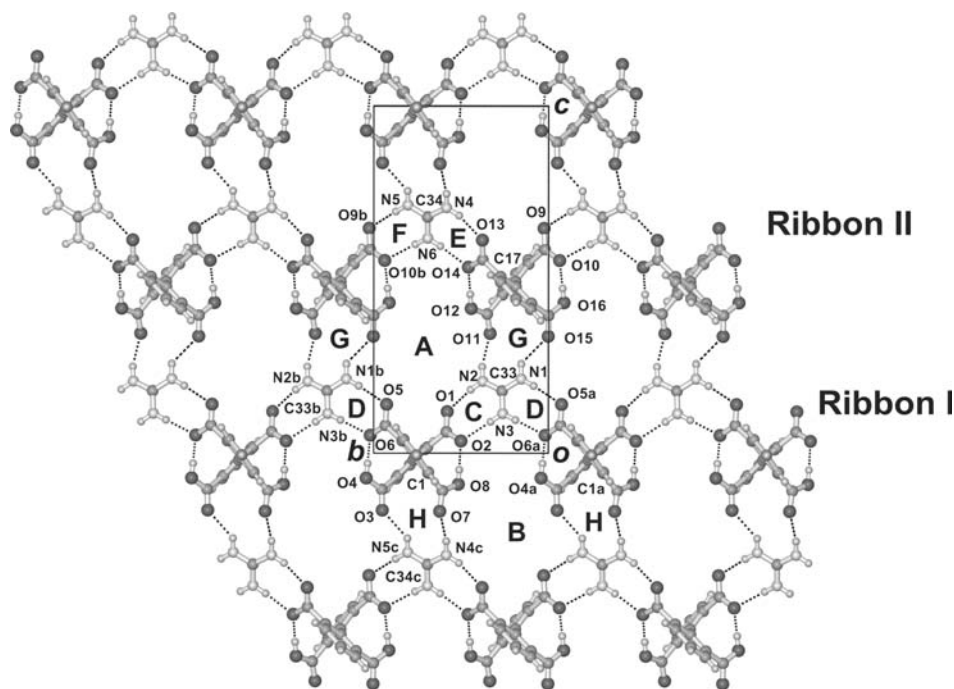


**Figure 8.72** Projection diagram along the *b* axis showing a portion of the anionic rosette network at  $y = 1/4$  in the crystal structure of **38**. Symmetry transformations: *a* ( $-\frac{1}{2} + x, \frac{1}{2} - y, 1 - z$ ) and *b* ( $x, \frac{1}{2} - y, \frac{1}{2} + z$ ). Mean atomic deviation from least-squares plane calculated from the coordinates of three GM carbon atoms plus those of one O–C–C–O fragment and two O–C–C–S fragments that constitute the interior boundary of the deformed rosette unit = 0.31 Å

## 8.4.3.2 Dianion of 1,1'-Biphenyl-2,2',6,6'-tetracarboxylic Acid as Component

In the supramolecular rosette layer structure of  $(\text{C}_2\text{H}_5)_4\text{N}^+[\text{C}(\text{NH}_2)_3]^+[(\text{C}_6\text{H}_3)_2(\text{COOH})_2(\text{COO}^-)_2]$  (**39**) [78], the nonplanar BPTC dianion replaces the planar dithiosquarate dianion as a building block (Figure 8.73). The asymmetric unit contains pairs of crystallographically independent BPTC dianions [**C(1)**, **C(17)**], guanidinium cations [**C(33)**, **C(34)**] and tetraethylammonium cations. Intramolecular linkage occurs between carboxyl and carboxylate groups attached to different phenyl rings in each tetracarboxylate dianion. In Figure 8.73, carboxylate oxygen atoms O(1) and O(2) come from the lower phenyl ring (view along *a* axis) of BPTC **C(1)**, and carboxylate oxygen atoms O(5) and O(6) belong to the upper phenyl ring. The pair of carboxylate groups of BPTC **C(1)** is linked with adjacent guanidinium cations **C(33)** and **C(33b)** by charge-assisted hydrogen bonds to form  $R_2^2(8)$  motifs [**C**] and [**D**], yielding an infinite twisted ribbon **I** running parallel to the *b* axis. Similarly, motifs [**E**] and [**F**] are generated by connecting guanidinium cations **C(34)** with BPTC **C(17)** and **C(17b)**, and further extension gives rise to ribbon **II** (Figure 8.73).

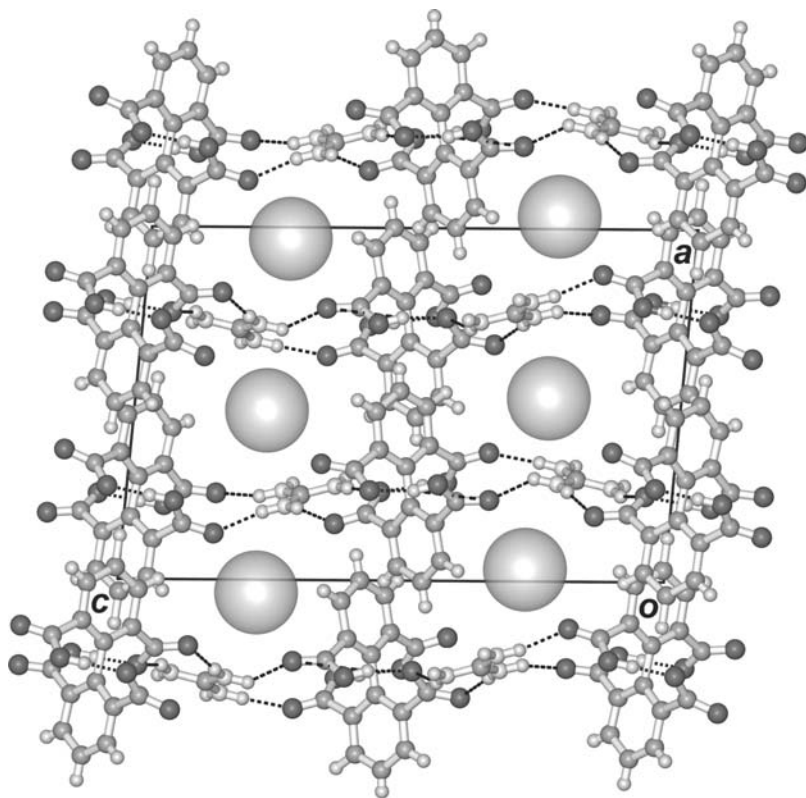
Ribbons of type **I** and **II** are cross-linked by pairwise  $\text{N-H}\cdots\text{O}$  hydrogen bonds ranging from 2.80 to 3.02 Å between the carbonyl group of carboxylic acid and the guanidinium cation via  $R_2^2(13)$  motifs [**G**] and [**H**] to generate similar distorted quasi-rosettes designated as [**A**] and [**B**] that exhibit the  $R_6^8(27)$  motif (see Scheme 8.6e).



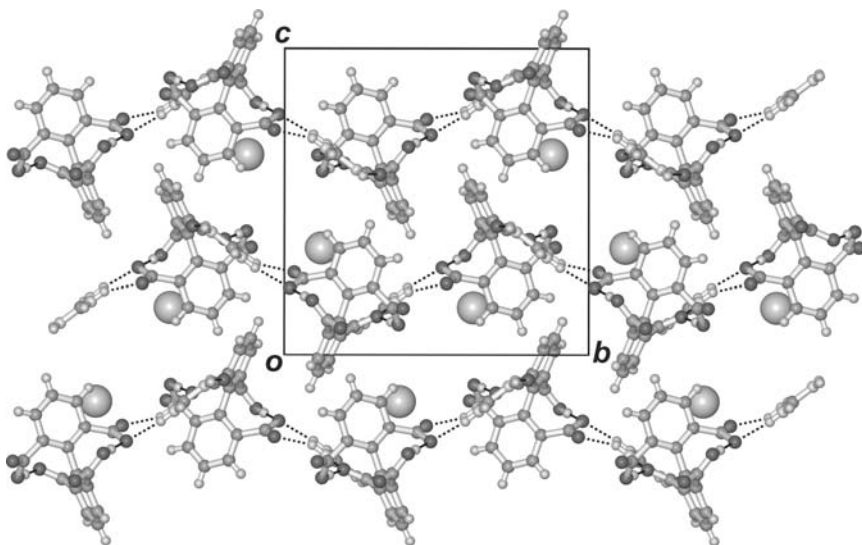
**Figure 8.73** Projection diagram showing a portion of the anionic rosette network at  $x = 1/4$  in the crystal structure of **39**. Symmetry transformations: *a* ( $x, 1 - y, z$ ), *b* ( $x, 1 + y, z$ ) and *c* ( $x, y, 1 - z$ )

Rosette [A] is composed of two GM C(33), one GM C(34), and three BPTCs [one C(1) and two C(17)], while rosette [B] is composed of one C(33), two C(34), and three BPTCs [two C(1) and one C(17)]. Figure 8.74 shows the packing of the approximately planar rosette layers in the crystal structure of **39**. In fact, there is no significant  $\pi - \pi$  interaction between phenyl rings belonging to adjacent rosette layers. Both independent tetraethylammonium cations occupy the interlayer region; the well-ordered one is located at  $x \approx 0$ , and the other is disordered and located at  $x \approx 1/2$ .

The distorted rosette layer of  $(n\text{-C}_3\text{H}_7)_4\text{N}^+[\text{C}(\text{NH}_2)_3]^+[(\text{C}_6\text{H}_3)_2(\text{COOH})_2(\text{COO}^-)_2]$  (**40**) [78] is similar to that in complex **39**, but it takes the form of a pleated sheet. Such sheets are concentrated at  $z \approx 1/4$  and  $3/4$  with the well-ordered tetra-*n*-propylammonium cation sandwiched in between (Figure 8.75). One alkyl leg of the tetra-*n*-propylammonium cation extends into the central void of a rosette motif to stabilize the pleated sheet structure.



**Figure 8.74** Approximately planar layer structure of **39** viewed along the *b* axis. No significant  $\pi - \pi$  interaction occurs between the phenyl rings belonging to adjacent rosette layers. Both independent tetraethylammonium cations represented by the large spheres occupy the interlayer region; the well-ordered  $(\text{C}_2\text{H}_5)_4\text{N}^+$  is located at  $x \approx 0$ , and the disordered one is located at  $x \approx 1/2$



**Figure 8.75** Crystal structure of **40**, with large spheres representing the ordered  $(n\text{-C}_3\text{H}_7)_4\text{N}^+$  cations that are accommodated between the pleated sheets

## 8.5 Conclusion and Outlook

This review shows that the hydrogen-bonded urea-anion host lattices can be modified in interesting ways by substitution of appropriate synthons through the variation of molecular building blocks, thus providing several interesting examples of isostructurality (i.e. structural similarity) in the context of supramolecular architecture. In addition, the designed synthesis of crystalline molecular solids may lead to the stabilization, by way of crystal engineering, of a reactive anionic covalent species, such as dihydrogen borate, allophanate and thioallophanate, which can be generated *in situ* and trapped through hydrogen-bonding interactions with its donor neighbors. This strategy also provides a viable route to break the degeneracy of canonical forms of a molecular species, such as the valence tautomers of rhodizonate and croconate dianions. Furthermore, by employing organic cations as templates and suitable counter anions as ancillary host materials, with or without neutral molecules such as water as an additional component, the ‘lattice engineering’ of new urea and thiourea inclusion compounds by self-assembly may be further explored.

Inclusion complexes **28** and **29** are prototypes of a new class of guanidinium-hydrogen carbonate host lattice featuring an uncharged, linear fused-rosette ribbon hemmed with multiple hydrogen-bonding donor sites, leading to nanoscale voids for the enclathration of guest species of moderate size. The designed construction of anionic host networks in **30–34** and **38–40** are unlike previously reported neutral honeycomb lattices of the same (6,3) topology, thus expanding the scope of *de novo* engineering of charge-assisted hydrogen-bonded networks using modular components, from which discrete molecular aggregates bearing the rosette motif [71c] may be derived. The flattened-chair  $(\text{H}_2\text{O})_6$  guest species, filling the cavity within the robust guanidinium-trimesate layer in **30**

and resembling that in the host lattice of bimesityl-3,3'-dicarboxylic acid [80], remains intact by varying the size of the quaternary ammonium template to obtain isomorphous **31–33**. In contrast, the guanidinium-carbonate (1:1) rosette network has been shown to be malleable, taking the planar and sinusoidal wavy configurations in **35** and **36**, respectively; this is rendered possible by the prolific hydrogen-bond accepting capacity of the carbonate building block, the employment of suitably chosen auxiliary anions as interlayer spacers, and the presence of tetraalkylammonium ions as templates. Complex **37** provides the first example of a three-component rosette layer, which turns out to exhibit a sinusoidal wavy configuration. Deviating from conventional topological design, the generation of new rosette layers, albeit highly distorted, has been accomplished with guanidinium ions and anionic building blocks (1,2-dithiosquarate in **38** and dianionic form of 1,1'-biphenyl-2,2',6,6'-tetracarboxylic acid in **39** and **40**) that do not conform to  $C_3$  molecular symmetry, thus widening the scope for further exploration of supramolecular network assemblies.

## Acknowledgements

This work is supported by Hong Kong Research Grants Council General Research Fund CUHK No. 456/95P, 420699, 426800, 402003 and 402405, as well as by the Wei Lun Foundation. C.-K. Lam would like to acknowledge the award of a Postdoctoral Fellowship by The Chinese University of Hong Kong and financial support by the Natural Science Foundation of China (No. 20503041).

## References

1. (a) K. D. M. Harris, *Supramol. Chem.*, **19**, 47–53 (2007); (b) M. D. Hollingsworth and K. D. M. Harris, in *Comprehensive Supramolecular Chemistry* (Eds J. L. Atwood, J. E. D. Davies, D. D. MacNicol and F. Vögtle), Elsevier, Oxford, Vol. 6, pp. 239–280, 1996.
2. T. C. W. Mak and R. K. McMullan, *J. Incl. Phenom.*, **6**, 473–481 (1988).
3. (a) S.-O. Lee, B. M. Kariuki and K. D. M. Harris, *New J. Chem.*, **29**, 1266–1271 (2005); (b) L. Yeo and K. D. M. Harris, *Can. J. Chem.*, **77**, 2105–2118 (1999).
4. (a) Z. Pan, A. Desmedt, E. J. MacLean, F. Guillaume and D. M. Harris, *J. Phy. Chem. C*, **112**, 839–847 (2008); (b) L. Yeo, K. D. M. Harris and B. M. Kariuki, *J. Solid State Chem.*, **156**, 16–25 (2001).
5. (a) F. Xue and T. C. W. Mak, *Acta Crystallogr., Sect. B*, **56**, 142–154 (2000); (b) F. Xue and T. C. W. Mak, *J. Phy. Org. Chem.*, **13**, 405–414 (2000); (c) Q. Li, F. Xue and T. C. W. Mak, *Inorg. Chem.*, **38**, 4142–4145 (1999); (d) Q. Li and T. C. W. Mak, *Acta Crystallogr., Sect. B*, **54**, 180–192 (1998); (e) Q. Li and T. C. W. Mak, *Supramol. Chem.*, **8**, 147–156 (1997); (f) Q. Li and T. C. W. Mak, *J. Inclusion Phenom.*, **28**, 151–161 (1997); (g) Q. Li and T. C. W. Mak, *Supramol. Chem.*, **8**, 73–80 (1996); (h) T. C. W. Mak, W. H. Yip and Q. Li, *J. Am. Chem. Soc.*, **117**, 11995–11996 (1995); (i) Q. Li, W. H. Yip and T. C. W. Mak, *J. Inclusion Phenom.*, **23**, 233–244 (1995).



6. (a) C.-K. Lam and T. C. W. Mak, *Tetrahedron*, **56**, 6657–6665 (2000); (b) Q. Li and T. C. W. Mak, *Supramol. Chem.*, **10**, 49–61 (1998); (c) Q. Li and T. C. W. Mak, *Acta Crystallogr., Sect. B*, **53**, 252–261 (1997); (d) Q. Li and T. C. W. Mak, *J. Inclusion Phenom.*, **27**, 319–340 (1997); (e) Q. Li and T. C. W. Mak, *J. Inclusion Phenom.*, **28**, 183–204 (1997); (f) Q. Li and T. C. W. Mak, *Acta Crystallogr., Sect. B*, **52**, 989–998 (1996); (g) Q. Li and T. C. W. Mak, *Acta Crystallogr., Sect. C*, **52**, 2830–2832 (1996); (h) Q. Li and T. C. W. Mak, *J. Inclusion Phenom.*, **20**, 73–88 (1994).
7. Q. Li and T. C. W. Mak, *Acta Crystallogr., Sect. B*, **53**, 262–271 (1997).
8. T. C. W. Mak and Q. Li, in *Advances in Molecular Structure and Research* (Eds M. Hargittai, I. Hargittai), JAI Press, Stamford, Connecticut, Vol. 4, pp. 151–225, 1998 and references cited therein.
9. G. R. Desiraju, *Angew. Chem., Int. Ed. Engl.*, **34**, 2311–2327 (1995).
10. A. Nangia and G. R. Desiraju, *Top. Curr. Chem.*, **198**, 57–95 (1998).
11. D. Semmingsen, *Acta Chem. Scand.*, **27**, 3961–3972 (1973).
12. (a) B. B. Koleva, T. Kolev, R. W. Seidel, H. Mayer-Figge, M. Spiteller and W. S. Sheldrick, *J. Phys. Chem. A*, **112**, 2899–2905 (2008); (b) B. Zheng, H. Dong, J. Bai, Y. Li, S. Li and M. Scheer, *J. Am. Chem. Soc.*, **130**, 7778–7779 (2008); (c) O. Z. Yesilel, M. Odabasoglu and O. Buyukgungor, *J. Mol. Struct.*, **874**, 151–158 (2008); (d) D. Wolstenholme and T. S. Cameron, *J. Phys. Chem. A*, **110**, 8970–8978 (2006).
13. (a) M. G. Fisher, P. A. Gale, M. E. Light and R. Quesada, *CrystEngComm*, **10**, 1180–1190 (2008); (b) P. M. T. Piggot, S. Seenarine and L. A. Hall, *Inorg. Chem.*, **46**, 5243–5251 (2007); (c) P. A. Gale, M. E. Light and R. Quesada, *CrystEngComm*, **8**, 178–188 (2006); (d) N. Das, A. Ghosh, A. M. Arif and P. J. Stang *Inorg. Chem.*, **44**, 7130–7137 (2005).
14. (a) M. R. Caira, *Encyclopedia of Supramolecular Chemistry*, Marcel Dekker, Inc., New York, pp. 767–775, 2004; (b) A. Kálmán and L. Párkányi, in *Advances in Molecular Structure Research* (Eds M. Hargittai and I. Hargitai), JAI Press, Stamford, Connecticut, Vol. 3, pp. 189–226, 1997; (c) A. Kálmán, L. Párkányi and Gy. Argay, *Acta Crystallogr., Sect. B*, **49**, 1039–1049 (1993).
15. C.-K. Lam and T. C. W. Mak, *Cryst. Engng.*, **3**, 33–40, 225–226 (2000).
16. D. Mullen and E. Hellner, *Acta Crystallogr., Sect. B*, **34**, 1624–1627 (1978) and references therein.
17. C.-K. Lam and T. C. W. Mak, *Chem. Commun.*, 2660–2661 (2003).
18. (a) J. Bernstein, R. E. Davis, L. Shimon and N.-L. Chang, *Angew. Chem., Int. Ed. Engl.*, **34**, 1555–1573 (1995); (b) M. C. Etter, J. C. MacDonald and J. Bernstein, *Acta Crystallogr., Sect. B*, **46**, 256–262 (1990); (c) M. C. Etter, *Acc. Chem. Res.*, **23**, 120–126 (1990).
19. (a) J. C. Barnes and T. J. R. Weakley, *Acta Crystallogr., Sect. C*, **54**, i, IUC9800025/1–2 (1998); (b) Q. Lu, R. J. Motekaitis, J. J. Reibenspies and A. E. Martell, *Inorg. Chem.*, **34**, 4958–4964 (1995).
20. J. M. Adams and R. G. Pritchard, *Acta Crystallogr., Sect. B*, **32**, 2438–2440 (1976).
21. (a) R. F. Baggio, C. Manzanares and S. Baggio, *Acta Crystallogr., Sect. B*, **31**, 2359–2361 (1975); (b) G. F. Gasparri, A. Mangia, A. Musatti and M. Nardelli, *Acta Crystallogr., Sect. B*, **25**, 203–213 (1969).



22. (a) B. F. Abrahams, M. G. Haywood, R. Robson and D. A. Slizys, *Angew. Chem., Int. Ed.*, **42**, 1112–1115 (2003); (b) M. Rodriguez, A. Llobet, M. Corbella, P. Muller, M. A. Uson, A. E. Martell and J. Reibenspies, *J. Chem. Soc., Dalton Trans.*, 2900–2906 (2002); (c) T. E. Gier, X. Bu, S.-L. Wang and G. D. Stucky, *J. Am. Chem. Soc.*, **118**, 3039–3040 (1996).
23. (a) R. Vaidhyanathan, S. Natarajan, A. K. Cheetham and C. N. R. Rao, *Chem. Mater.*, **11**, 3636–3642 (1999); (b) T. M. Reineke, M. Eddaoudi, M. Fehr, D. Kelly and O. M. Yaghi, *J. Am. Chem. Soc.*, **121**, 1651–1657 (1999) and references therein.
24. (a) U. Schlickum, R. Decker, F. Klappenberger, G. Zoppellaro, S. Klyatskaya, W. Auwaerter, S. Neppel, K. Kern, H. Brune, M. Ruben and J. V. Barth, *J. Am. Chem. Soc.*, **130**, 11 778–11 782 (2008); (b) R. Custelcean, P. Remy, P. V. Bonnesen, D.-E. Jiang and B. A. Moyer, *Angew. Chem., Int. Ed.*, **47**, 1866–1870 (2008); (c) R. Custelcean, *Chem. Commun.*, 295–307 (2008); (d) M. Du, Z.-H. Zhang, L.-F. Tang, X.-G. Wang, X.-J. Zhao and S. R. Batten, *Chem. Eur. J.* **13**, 2578–2586 (2007).
25. J. Han, L. Zhao, C.-W. Yau, and T. C. W. Mak, *Cryst. Growth Des.*, 308–319 (2009).
26. W. Bolton, *Acta Crystallogr.*, **16**, 950–956 (1963).
27. O. Simonsen, *Acta Chem. Scand.*, **51**, 861–864 (1997).
28. O. Simonsen, *Acta Crystallogr., Sect. C*, **41**, 1258–1260 (1985).
29. B. M. Craven, S. Martinez-Carrera and G. A. Jeffrey, *Acta Crystallogr.*, **17**, 891–903 (1964).
30. J. A. Platts, S. T. Howard and B. R. F. Bracke, *J. Am. Chem. Soc.*, **118**, 2726–2733 (1996).
31. (a) T. Steiner, *Angew. Chem., Int. Ed.*, **41**, 48–76 (2002); (b) L. J. Prins, D. N. Reinhoudt and P. Timmerman, *Angew. Chem., Int. Ed.*, **40**, 2382–2426 (2001); (c) P. D. Beer and P. A. Gale, *Angew. Chem., Int. Ed.*, **40**, 486–516 (2001).
32. (a) J. M. Coddington and M. J. Taylor, *J. Coord. Chem.*, **20**, 27–38 (1989); (b) J. B. Farmer, *Adv. Inorg. Chem. Radiochem.*, **25**, 187–237 (1982).
33. (a) R. J. Brotherton, in *Encyclopedia of Inorganic Chemistry* (Ed. R. B. King), John Wiley & Sons, Ltd, New York, Vol. 1, pp. 357–374, 1994; (b) N. Ingri, *Sven. Kem. Tidskr.*, **75**, 199–230 (1963).
34. (a) M. Attinà, F. Cacace, G. Occhiucci and A. Ricci, *Inorg. Chem.*, **31**, 3114–3117 (1992); (b) M. Attinà, F. Cacace, A. Ricci, F. Grandinetti and G. Occhiucci, *J. Chem. Soc., Chem. Commun.*, 66–68 (1991).
35. (a) D. A. Keszler, in *Encyclopedia of Inorganic Chemistry* (Ed. R. B. King), John Wiley & Sons, Ltd, New York, Vol. 1, pp. 318–327, 1994; (b) G. Heller, *Top. Curr. Chem.*, **131**, 39–98 (1986).
36. M. A. Simonov, Yu. K. Egorov-Tismenko and N. V. Belov, *Kristallografiya*, **21**, 592–594 (1976).
37. H. Behro and Ch. Baerlocher, *Acta Crystallogr., Sect. C*, **41**, 5–7 (1985).
38. C. C. Freyhardt and M. Wiebcke, *J. Chem. Soc., Chem. Commun.*, 1675–1676 (1994).
39. H. W. Blohm and E. I. Becker, *Chem. Rev.*, **51**, 471–504 (1952).
40. J. Bougault and J. Leboucq, *Bull. Soc. Chim. Fr.*, **47**, 594–605 (1930).

41. (a) X. L., T.-B. Wei and Y.-M. Zhang, *Coord. Chem.*, **57**, 453–457 (2004); (b) X. Shen, B.-S. Kang, Y. Liu, L.-Q. Gu, X.-Y. Huang, J. Sun and Q.-T. Liu, *J. Coord. Chem.*, **46**, 397–408 (1999).
42. C.-K. Lam, T.-L. Chan and T. C. W. Mak, *CrystEngComm*, **6**, 290–292 (2004).
43. (a) G. Seitz and P. Imming, *Chem. Rev.*, **92**, 1227–1260 (1992); (b) F. Serratos, *Acc. Chem. Res.*, **16**, 170–176 (1983); (c) R. West (Ed.), *Oxocarbons*, Academic Press, New York, 1980.
44. C.-K. Lam and T. C. W. Mak, *Russ. J. Chem. (Zh. Strukt. Khim.)*, **40**, 714–720 (1999).
45. C.-K. Lam and T. C. W. Mak, *Chem. Commun.*, 1568–1569 (2001).
46. M. A. Neuman, University of Wisconsin, Madison, Available from University Microfilms (Ann Arbor, MI), Order No. 66-1300, 1966; Dissertation Abstr. 26(11), 6394 (1966).
47. P. v. R. Schleyer, K. Najafian, B. Kiran and H. Jiao, *J. Org. Chem.*, **65**, 426–431 (2000).
48. B. F. Abrahams, M. G. Haywood and R. Robson, *CrystEngComm*, **7**, 629–632 (2005).
49. (a) J. D. Dunitz, P. Seiler and W. Czechtizky, *Angew. Chem., Int. Ed.*, **40**, 1779–1780 (2001); (b) R. Nietzki, *Ber.*, **20**, 1617–1620 (1887).
50. J. F. Eastham and S. Selman, *J. Org. Chem.*, **26**, 293–296 (1961).
51. J. Aihara, *J. Am. Chem. Soc.*, **103**, 1633–1635 (1981).
52. C.-C. Wang, C.-T. Kuo, P.-T. Chou and G.-H. Lee, *Angew. Chem., Int. Ed.*, **43**, 4507–4510 (2004).
53. C.-K. Lam and T. C. W. Mak, *Angew. Chem., Int. Ed.*, **40**, 3453–3455 (2001).
54. (a) G. M. A. Junqueira, W. R. Rocha, W. B. de Almeida and H. F. Dos Santos, *Phys. Chem. Chem. Phys.*, **5**, 437–445 (2003); (b) A. Ranganathan and G. U. Kulkarni, *J. Phys. Chem. A*, **106**, 7813–7819 (2002); (c) L. R. Martins, M. C. C. Ribeiro and M. S. Skaf, *J. Phys. Chem. B*, **106**, 5492–5499 (2002); (d) D. Quinonero, C. Garau, A. Frontera, P. Ballester, A. Costa and P. M. Deya, *Chem. Eur. J.*, **8**, 433–438 (2002).
55. (a) D. Braga, G. Cojazzi, L. Maini and F. Grepioni, *New J. Chem.*, **25**, 1221–1223 (2001); (b) D. Quinonero, A. Frontera, P. Ballester and P. M. Deya, *Tetrahedron Lett.*, **41**, 2001–2005 (2000).
56. (a) G. M. A. Junqueira, W. R. Rocha, W. B. De Almeida and H. F. Dos Santos, *Phys. Chem. Chem. Phys.*, **3**, 3499–3505 (2001); (b) R. West and D. L. Powell, *J. Am. Chem. Soc.*, **85**, 2577–2779 (1963).
57. C.-K. Lam, M.-F. Cheng, C.-L. Li, J.-P. Zhang, X.-M. Chen, W.-K. Li and T. C. W. Mak, *Chem. Commun.*, 448–449 (2004).
58. (a) P. Vishweshwar, A. Nangia and V. M. Lynch, *Chem. Commun.*, 179–180 (2001); (b) T. Steiner and G. R. Desiraju, *Chem. Commun.*, 891–892 (1998).
59. (a) D. Deguenon, G. Bernardinelli, T. P. Tuchagues and P. Castan, *Inorg. Chem.*, **29**, 3031–3037 (1990); (b) I. Castro, J. Setten, J. Faus and M. Julve, *J. Chem. Soc., Dalton Trans.*, 2271–2275 (1992).
60. (a) T. K. Maji, S. Konar, G. Mostafa, E. Zandrando, T.-H. Lu and N. R. Chaudhuri, *J. Chem. Soc., Dalton Trans.*, 171–175 (2003); (b) C.-C. Wang, C.-H. Yang and G.-H. Lee, *Inorg. Chem.*, **41**, 1015–1018 (2002); (c) F. Dumestre, B. Soula, A. M.

- Galibert, P. L. Fabre, G. Bernardinelli, B. Donnadieu and P. Castan, *J. Chem. Soc., Dalton Trans.*, 4131–4138 (1998).
61. M. J. Frisch et al., *GAUSSIAN 98 (Revision A.11)*, Gaussian Inc., Pittsburgh, PA, 1998.
  62. A. E. Reed, L. A. Curtiss and F. Weinhold, *Chem. Rev.*, **88**, 899–926 (1988).
  63. (a) L. R. MacGillivray, G. S. Papaefstathiou, T. Friscic, T. D. Hamilton, D. K. Bucar, Q. Chu, D. B. Varshnev and I. G. Georgiev, *Acc. Chem. Res.*, **41**, 280–291 (2008); (b) T. P. Radhakrishnan, *Acc. Chem. Res.*, **41**, 367–376 (2008); (c) M. O’Keeffe, M. A. Peskov, S. J. Ramsden and Yaghi, O. M., *Acc. Chem. Res.*, **41**, 1782–1789 (2008); (d) P. Metrangolo, G. Resnati, T. Pilati and S. Biella, *Struct. Bonding*, **126**, 105–136 (2008); (e) F. Nouar, J. F. Eubank, T. Bousquet, L. Wojtas, M. J. Zawaorotko and M. Eddaoudi, *J. Am. Chem. Soc.*, **130**, 1833–1835 (2008); (f) K. E. Plass, A. L. Grzesiak and A. J. Matzger, *Acc. Chem. Res.*, **40**, 287–293 (2007); (g) S. J. Dalgarno, P. K. Thallapally, L. J. Barbour and J. L. Atwood, *Chem. Soc. Rev.*, 236–245 (2007); (h) X.-M. Chen and M.-L. Tong, *Acc. Chem. Res.*, **40**, 162–170 (2007); (i) R. Glaser, *Acc. Chem. Res.*, **40**, 9–17 (2007); (j) G. R. Desiraju, *Crystal Design: Structure and Function*, John Wiley & Sons, Ltd, Chichester, 2003.
  64. (a) Y.-B. Men, J. Sun, Z.-T. Huang and Q.-Y. Zheng, *Angew. Chem., Int. Ed.*, **48**, 2873–2876 (2009); (b) C. Silien, M. T. Raisanen and M. Buck, *Angew. Chem., Int. Ed.*, **48**, 3349–3352 (2009); (c) R. Madueno, M. T. Ræisaenen, C. Silien and M. Buck, *Nature*, **454**, 618–621 (2008); (d) E. S. Barrett, T. J. Dale and J. Rebek, Jr, *J. Am. Chem. Soc.*, **130**, 2344–2350 (2008); (e) K. E. Maly, E. Gagnon, T. Maris and J. D. Wuest, *J. Am. Chem. Soc.*, **129**, 4306–4322 (2007).
  65. A. F. Wells, *Three-Dimensional Nets and Polyhedra*, John Wiley & Sons, Ltd, New York, 1977.
  66. (a) J. A. Theobald, N. S. Oxtoby, M. A. Phillips, N. R. Champness and P. H. Beton, *Nature*, **424**, 1029–1031 (2003); (b) B.-Q. Ma and P. Coppens, *Chem. Commun.*, 2290–2291 (2003); (c) M. Tadokoro, T. Shiomi, K. Isobe and K. Nakasuji, *Inorg. Chem.*, **40**, 5476–5478 (2001); (d) S. V. Kolotuchin, P. A. Thiessen, E. E. Fenlon, S. R. Wilson, C. J. Loweth and S. C. Zimmerman, *Chem. Eur. J.*, **5**, 2537–2547 (1999); (e) A. V. Ranganathan, R. Pedireddi and C. N. R. Rao, *J. Am. Chem. Soc.*, **121**, 1752–1753 (1999); (f) F. H. Herbstein, in *Comprehensive Supramolecular Chemistry* (Eds J. L. Atwood, J. E. D. Davies, D. D. MacNicol and F. Vögtle), Elsevier, Oxford, Vol. 6, pp. 61–83, 1996 and references therein.
  67. A. Katrusiak and M. Szafranski, *J. Mol. Struct.*, **378**, 205–223 (1996).
  68. G. M. Whitesides, E. E. Simanek, J. P. Mathias, C. T. Seto, D. N. Chin, M. Mammen and D. M. Gordon, *Acc. Chem. Res.*, **28**, 37–44 (1995).
  69. (a) M. J. Horner, K. T. Holman and M. D. Ward, *J. Am. Chem. Soc.*, **129**, 14640–14660 (2007); (b) K. T. Holman, A. M. Pivovar, J. A. Swift and M. D. Ward, *Acc. Chem. Res.*, **34**, 107–118 (2001); (c) V. A. Russell, M. C. Etter and M. D. Ward, *J. Am. Chem. Soc.*, **116**, 1941–1952 (1994).
  70. R. E. Melendez, C. V. K. Sharma, M. J. Zaworotko, C. Bauer and R. D. Rogers, *Angew. Chem., Int. Ed. Engl.*, **35**, 2213–2215 (1996).
  71. (a) S. Yagai, T. Nakajima, T. Karatsu, K.-I. Saitow and A. Kitamura, *J. Am. Chem. Soc.*, **126**, 11500–11508 (2004); (b) M. Arduini, M. Crego-Calama, P. Timmerman and D. N. Reinhoudt, *J. Org. Chem.*, **68**, 1097–1106 (2003); (c) D. N. Chin,

- J. A. Zerkowski, J. C. MacDonald and G. M. Whitesides, in *Organised Molecular Assemblies in the Solid State* (Ed. J. K. Whitesell), John Wiley & Sons, Ltd, Chichester, pp. 185–253, 1999.
72. T. C. W. Mak and F. Xue, *J. Am. Chem. Soc.*, **122**, 9860–9861 (2000).
73. (a) G. R. Desiraju and T. Steiner, *The Weak Hydrogen Bond in Structural Chemistry and Biology*, Oxford University Press, Oxford (1999); (b) T. Steiner, *Chem. Commun.*, 313–314 (1999).
74. C.-K. Lam, F. Xue, J.-P. Zhang, X.-M. Chen and T. C. W. Mak, *J. Am. Chem. Soc.*, **127**, 11536–11537 (2005).
75. W. Rüdorff and E. Z. Schulze, *Anorg. Allg. Chem.*, **277**, 156–171 (1954).
76. S. W. Peterson and H. A. Levy, *Acta Crystallogr.*, **10**, 70–76 (1957).
77. L. Lin, MS Thesis, *Synthesis and Crystal Structure Determination of Inclusion Compounds Containing Guanidinium Cation*, School of Chemistry and Chemical Engineering, SunYat-Sen University, 2007 (in Chinese).
78. J. Han, C.-W. Yau, C.-K. Lam and T. C. W. Mak, *J. Am. Chem. Soc.*, **130**, 10315–10326 (2008).
79. J. M. Adams and R. W. H. Small, *Acta Crystallogr., Sect. B*, **30**, 2191–2193 (1974).
80. J. N. Moorthy, R. Natarajan and P. Venugopalan, *Angew. Chem., Int. Ed.*, **41**, 3417–3420 (2002).

# Index

References to figures are given in italic type. References to tables are given in bold type.

- 3DSEARCH 13–17,
- acridine 109
- active pharmaceutical ingredients (API)  
67–68  
crystal form types 68–72  
*see also* pharmaceuticals
- ad hoc software 17–18
- alkyl groups 122–125
- alkylammonium surfactants  
aromatic complexes 103–108  
biphenyl complexes 110–114  
heteroaromatic complexes 109–110  
mortar grinding 117–122  
perfumes and 130–134  
pharmaceuticals and 127–130
- alkynols 184–186
- allophanate 259–260
- AMG517 **70**, 71, 78–81
- amides 154–156, 216–217  
bis-amidopyridines 225  
nitrophenyl 160–166  
primary 217–220  
pyridyl 156–159  
bis-amidopyridines 225–231  
C-terminal 227–229  
N-terminal 229–231  
co-crystals 220–225  
secondary 224–225, 231  
triminopyridines 232–234, 235
- API *see* active pharmaceutical ingredients
- aromatic compounds 109–110  
complexes with surfactants 103–108  
anionic 125–126  
supramolecular interactions 122–125
- ASER format 2–3
- aspirin 57
- atom-atom modelling 47–48
- benzamide 217–218, 218
- benzyl(hexadecyl)dimethylammonium chloride (BCDAC) 137, 138–141, 138, 143–147
- $\beta$ -lactams **70**, 82
- biphenyl 110–114
- 1–1'-biphenyl-2,2',6,6'-tetracarboxylic acid 304–305
- 1,1-bis(hydroxyphenyl)cyclohexane (BHPC) 152
- blind tests 48–52  
success rates **51**  
target molecules 50
- C-H... $\pi$  interactions 122–125
- Cambridge Crystallographic Data Centre 1, 1–2, 2, 36–37
- Cambridge Structural Database 1–2  
3DSEARCH tool 13–17  
ad hoc search software 17–18  
data format 2–3  
future directions 36–37  
organisation and management 2–3  
polymorphism and 30–32

## Cambridge Structural Database

*(continued)*

search algorithms

3D structures 13–18

functional group exchange 18–20

solvated structures 20–22

search result classification 23–25

simulated PXRD profiles 31

structural similarity tools 22–23

*d*SNAP 23–25

validation of structural discoveries

5–6

*see also* CSDContact; IsoStar;*Mogul*; powder X-ray diffraction  
structurescarbamazepine **70**, **71**

acetic acid co-crystallization 62

nicotinamide co-crystallization 223

packing 59–61, 85

polymorphism 59–61

stabilization 83–87

structure prediction 58–64

supramolecular structures **33**, 34–35

carbazole 123

carbimazole 192, **193**

carbonates 243–246

carboxylic acid dimer homosynthon

191–192

CBZ *see* carbamazepinecefradoxil **71**celecoxib **70**, **71**, 79–81

chloro/methyl interchange 18

4-chloro-*m*-cresol 127, **130**

chloroform 124

CIF file format 2–3

co-crystallization

carbamazepine 62, 83–86, 223

celecoxib 80–81

itraconazole 81

nicotinamide 80–81, 84–85, **86**, 221

norfloxacin 78

primary amidopyridines 220–225

secondary amides and pyridines **231**,  
**232**

theophylline:phenobarbital 87–90

theophylline 90

thiocarbamides 206, **210**

computational studies 43–44

structure prediction *see* structure  
prediction*ConQuest* 13, 37

GRX search and 18–19

*p*-cresol 105–106, 119–120, **119**

critical micelle concentration (cmc)

101–102

Crixivan 75–77

croconate 278, 284–289

valence tautomers 282–283

enediolate 285–289

nonbenzenoid aromatic 284–285

crystal form types 68–72

crystal structure prediction *see* structure  
prediction

CRYSTALS program 5

CSDcontact 8–10

CSDsymmetry 10–13

CTAB 103–107, **118**, 119–120, 130–131

2-methindole complex 130–131

aromatic compound complexes

103–107, **104**, 109, 109–110

biphenyl complex 110–112

mortar grinding 119–120

skatole complex 132–133

void spaces **136***m*-cyanophenol 105

Darunavir ethanolate 75

DASH program 5

databases

advantages 9–10

Cambridge Structural *see* Cambridge  
Structural Database

CSDContact 8–10

CSDsymmetry 10–13

future role 36–38

IsoStar 6–8

*Mogul* 3–6

search procedures

3DSEARCH 13–17

ad hoc 17–18

functional group exchange 18–20

solvated structures 20–22

- decyltrimethylammonium bromide *see* DTAB  
 delaviridine **70**, 71, 90–91  
 dendrograms 23–25  
 diarylsulfonamide 33  
 N–N-diarylureas **172–181**  
 dibenzofuran 123  
 dihydrogen borate 257–259  
 dimethyl sulfoxide (DMSO) 162–164, 168  
 dimethylformamide (DMF) 162–164, 168  
 dinitrophenyl urea (DNPU) 160–162  
*m*-dipyridylurea (MDPU) 156–159, 160, 161  
 1,2-dithiosquarate 303  
 DNPU 162–164  
 dodecyltrimethylammonium bromide *see* LTAB  
 Doxycycline HCl 75  
 drugs *see* pharmaceuticals  
*d*SNAP 23–25  
 DTAB **104**, 130–131  
  
 electrostatic modelling 53–56, 167–168  
 ephedrine 71  
 ethionamide 192, 193  
 ethyl(hexadecyl)dimethylammonium bromide (ECDAB) 137, 138  
     hydroquinone complexes 138–141  
  
 flopropione 127, 129  
 formamide 223  
 functional group exchange, searches 18–20  
 functional groups, effect on supramolecular structures 151–154  
  
*gem*-alkynols 184–187, 185  
 GRX 18–20  
 guaiacol 130–131, 131, 134  
 guanidinium urea complexes  
     C<sub>3</sub>-symmetric oxo-anions 293–301  
     hydrogen carbonate dimer 289–293  
     non-C<sub>3</sub>-symmetric building blocks 301–305  
  
 halogen-substituted *p*-nitrophenylureas 162–166, 167  
 heteroaromatic compounds 109–110  
 hexadecyltrimethylammonium bromide *see* CTAB  
 1-hexadecylpyridinium bromide 134–136  
 high-throughput screening 72–75  
     composition and process variables **74**  
 HSAB theory 182–184  
 HT *see* high-throughput screening  
 hydrates 20–22, 82–83  
     β-lactams 82  
     indinavir sulfate ethanoate 75–77  
 hydrogen bonding 8–9, 48  
     acceptor strength 168–169  
     benzamide 218  
     *gem*-alkynols 184–186, 184–187, 185  
     HSAB interactions 182–184  
     molecular conformation and 166–170  
     multi-functional compounds 151–154  
     nicotinamides 219, 221, 222, 223  
     nitrobarbiturate-urea 253  
     picolinamides 218–219, 219  
     squarates to urea/thiourea 265–275  
         bond geometry 272–274  
     structure prediction and 56–57  
     thiocarbamides 192–194  
     triamidopyridines 233, 234, 235  
     urea and thiourea host-lattices  
         5-nitrobarbiturate 247–257  
         carbonate and oxalate 243–247  
     ureas and amides 154–156  
     pyridyl 156–159  
     *see also* C–H... $\pi$  interactions  
 hydroquinone 106–107, 140  
     complexation with surfactants, applications as whitening agent 142–147  
     complexation with surfactants 136–142  
     isonicotinamide co-crystallization 222  
     oxidation 144  
 3-hydroxybenzoic acid 221



- p*-hydroxybenzoic acid 103, 119, 121–122
- in silico* studies *see* computational studies
- indinavir **70**
- indinavir sulfate ethanoate 75–77
- indole acetic acid 222
- inherent crystal structures 44
- interaction energy 47–48
- intermolecular reactions
- search approaches 6–13
  - surfactant-aromatic complexes 122–125
  - see also* hydrogen bonding
- iodo compounds 164–165
- o*-iodophenol 107–108, 117–122, 118, 120–121
- IsoQuest 26–28
- IsoStar 6–8
- isotropic models 48
- itraconazole *71*
- co-crystallization 81
- lattice energy minimization 45–46
- predictive efficacy 53–56
- LTAB **104**, 107–108, 130–131
- complexes with aromatic compounds **104**
- MDPU 156–159, 160, 161
- meropenem *71*
- methimazole 192
- 2-methylindole 130–131, 131–132, **132**
- micelle concentration 101–102
- molecular geometry, database searches 3–5
- Mogul* 3–5
- structural discovery validation 5–6
  - see also* Cambridge Structural Database
- mortar grinding 117–122, **120**
- MTAB **104**, 107, 110, 130–131
- multi-functional molecules 152
- multidimensional scaling (MMDS) 23–25
- nicotinamide 218, 219
- carbamazepine co-crystal 84–85, 86
  - co-crystallization 221
- iso*-nicotinamide 221
- co-crystallization 222
- 5-nitrobarbiturate 247–257
- nitrophenyl ureas and amides 160–166
- norfloxacin **70**, *71*, 77–78
- nylon 216
- oxalate 246–247
- oxazolidine-2,5-dione 53–55
- oxocarbon anions
- croconate complexes 278
  - rhodizonate complexes 275–278
  - squarate complexes 265–275
  - trapping, squarates 265–275
- packing
- carbamazepine 59–61, 83–87, 85
  - effect of functional groups 151–153
  - surfactant-aromatic complexes 114–117
- peer review 37
- pentadecyltrimethylammonium bromide (PTAB) 112
- peptides 29
- perfumes 130–134
- pharmaceuticals 67–68, 91–93
- crystal form types 69–72, 69
  - formulation *see* drug formulations
  - formulations
    - AMG517 78–81
    - carbamazepine 83–87
    - celecoxib 78–82
    - delaviridine 90–91
    - norfloxacin 77–78
    - theophylline and phenobarbital 87–90
- high-throughput screening 72–75
  - thiocarbamide 192, 193
  - water-solubility 127–130
- phenobarbital 57, **70**, *71*, 87–88
- dissolution profiles **89**
  - theophylline co-crystallisation 87–91

- phenothiazine 123
- phenoxazine 123
- phenyl compounds 127
- phenyl urea, rhodizonate complex 275–278
- N-X-phenyl-*N'*-*p*-nitrophenyl ureas 160–166
- 2-phenylpyridine 111–112
- phenylurea 154–155, 157
- phloroglucinol 222
- picolinamide 218
- piracetam 57
- PNPU-X 160–162, 163, 167
- polymorphism 44
  - Cambridge Structural Database and 30–32
  - carbamazepine 58–59
  - flexible molecules 57–58
  - lattice energy differences 47
- powder X-Ray diffraction structures 25–26
  - IsoQuest 26–28
  - self-organising maps 28–30
  - simulated profiles 31
- 6-*n*-propylthiouracil 192, 193
- Protein Data Bank 6, 7
- PTAB 112
- PXRD *see* powder X-ray diffraction
- pyrazinamide 156
- pyridines
  - amide 217–220
    - co-crystals 220–225
  - amido
    - bis-amidopyridines 227–231
    - secondary 224–225
    - triamidopyridines 232–234, 235
    - two-component structures 231
    - ureas and amides 156–159
- 2-pyridyl 224–225, 229–230
- 3-pyridylurea 156–159, 157
- quinoline 218–219
- quinone 136–137
- remacemide 71
- rhodizonate
  - urea/thiourea complexes 275–278, 282
  - valence tautomers 279–282
- rosettes
  - guanidinium and C3-symmetric oxo-anion building blocks 293–301
  - guanidinium and hydrogen carbonate dimer 289–293
  - guanidinium and non-symmetrical building blocks 301–307
- saccharin, carbamazepine co-crystals 84–86, 87
- scatter plots, IsoStar 6–7, 7
- search procedures
  - 3D structure 13–18
  - functional group exchange 18–20
  - solvated and unsolvated structures 20–22
- sertraline 71
- similarity measures
  - IsoQuest 26–28
  - self-organizing maps 28–30
- skatole 131, 132–133, 133
- sodium dodecyl sulfate (SDS) 125–126
- sodium octyl sulfate (SOS) 125–126
- solid form screening 63
- solvates
  - carbamazepine 61–63
  - database search procedures 20–22
  - indinavir 75–77
  - see also* hydrates
- Solvates* program 20–22
- space group symmetry 46
- squaric acid 241–242
- thiourea complexes 265–275
  - see also* dithiosquarate
- steroids 29
- structure prediction 43–44, 45–46
  - assessment 52–57
    - blind tests 48–52
  - carbamazepine 58–63
  - flexible molecules 57–58
  - goals 44
  - hydrogen bonding and 56–57
  - search methods 46–47

- structure prediction (*continued*)
  - structure evaluation 47–48
- structure production, likelihood of
  - minimum lattice energy structure being observed structure 53–56
- succinic acid 221
- supramolecular constructs 32
  - carbamazepine 33, 34–35
  - see also* supramolecular synthons; XPac
- supramolecular interactions
  - alkyl-aryl interactions 121–125
  - HSAB 182–184
  - see also* hydrogen bonding
- supramolecular structures
  - databases 6–10
    - search procedures 13–17, 32–36
  - interaction energy 47
  - search procedures, XPac 32–36
  - searching 6
- supramolecular synthons 151–152, 151–154, 152
  - amides and pyridines 216
  - thiocarbamides 191–192
  - urea tape 166–167
  - urea/thioureas 241–243
- surfactants 101–102
  - anionic 125–126
  - complexes with aromatic compounds 103–109
    - mortar grinding 117–122
    - packing 114–117
  - complexes with biphenyl 110–112
  - critical micelle concentration 101–102
  - 1-hexadecylpyridinium bromide 134–136
  - hydroquinone complexes 136–142
  - thermal stability of perfumes 130–134
  - water-insoluble drugs and 127–130
- symmetry 46
  - databases 10–13
- tetradecyltrimethylammonium bromide
  - see* MTAB
- theophylline **70**, 71, 87
- co-crystallization 90
  - phenobarbital 87–90
  - disolution profiles **89**
- thermal stability 130–134
- thioallophanate 263–265
- thiocarbamides 192–194
  - acyclic
    - E configuration 196–200, 200, **201**, 202–204
    - Z configuration 201–206
- co-crystals 206, 210
- cyclic
  - 5-membered 194–195, **196**, 197
  - 6-membered 196, 197, 198, **199**
  - Z configuration **206**, 207–208, 209, 210
- thioureas
  - croconate complexes 278, 282–289
  - host-lattice structures 241–243
    - 5-nitrobarbiturate 247–251, 256–257
  - rhodizonate complexes 275–278
  - squarate complexes
    - channel-type 266–270
    - layer-type 270–271
    - structural features 271–275
- tridecyltrimethylammonium bromide (TTAB) 112–114
- triphenylchlorosilane **27**
- ureas 90, 154–156
  - croconate complexes 278, 282–289
  - diaryl **172–181**
  - host-lattice inclusion compounds 241–243
    - 5-nitrobarbiturate 247–257
    - allophanate 259–261
    - carbonate and oxolates 243–247
    - dihydrogen borate 257–259
    - thioallophanate 263–265
  - hydrogen-bonding and molecular conformation 166–170
  - nitrophenyl 160–166

- pyridyl 156–159
- rhodizonate complexes 275–281
- rosette motif complexes, guanidinium  
and hydrogen carbonate dimer  
289–293
- squarate complexes 265–275
  
- validation, structural discoveries  
5–6
- void spaces, CTAB and CTB complexes  
**136**
  
- water-solubility 127–130
- weighted cross-correlation (WCC)  
functions 26  
IsoQuest 26–28  
self-organizing maps 28–30
- whitening agents 142–147, **146**, 147
  
- X-ray structures *see* powder X-ray  
diffraction structures
- XPac* 32–36  
*see also* supramolecular constructs

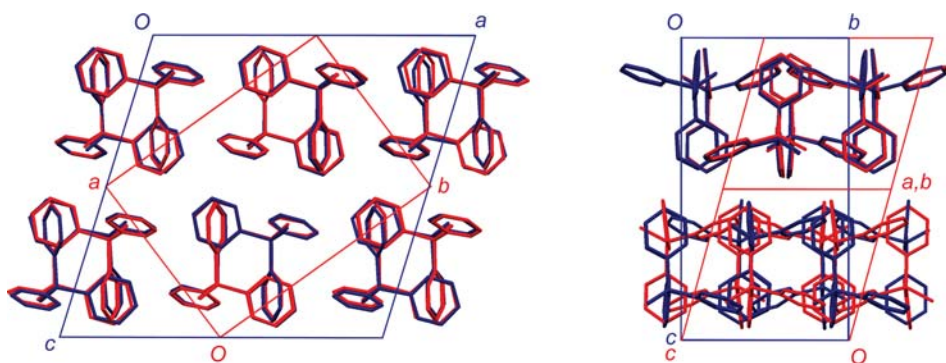


Figure 1.13

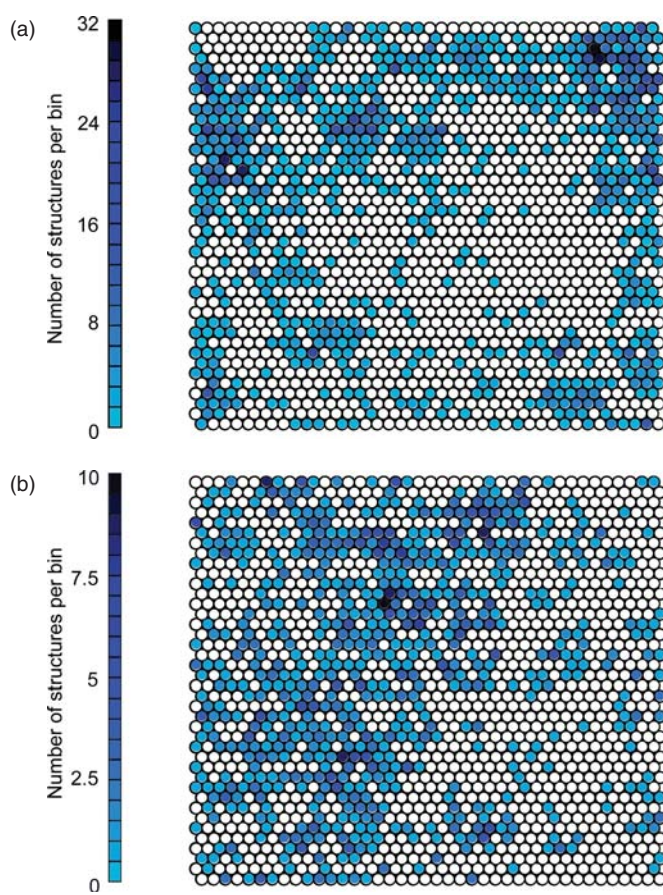
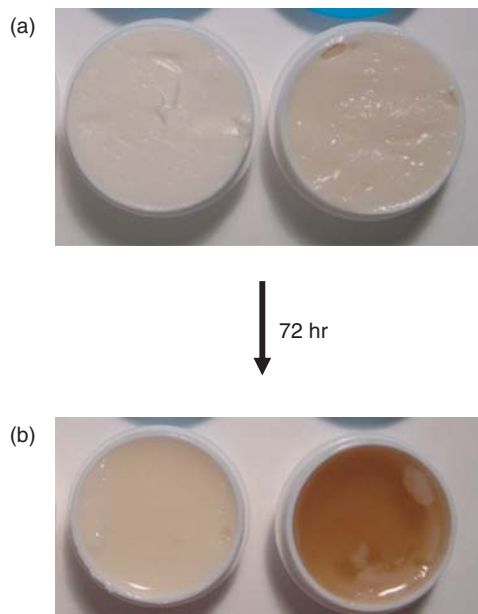


Figure 1.15



**Figure 4.48**

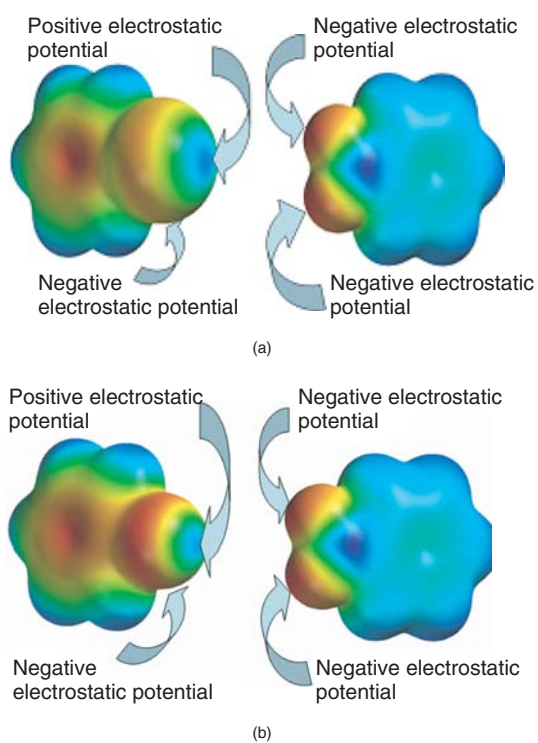


**Figure 4.58**





**Figure 4.60**



**Figure 5.21**



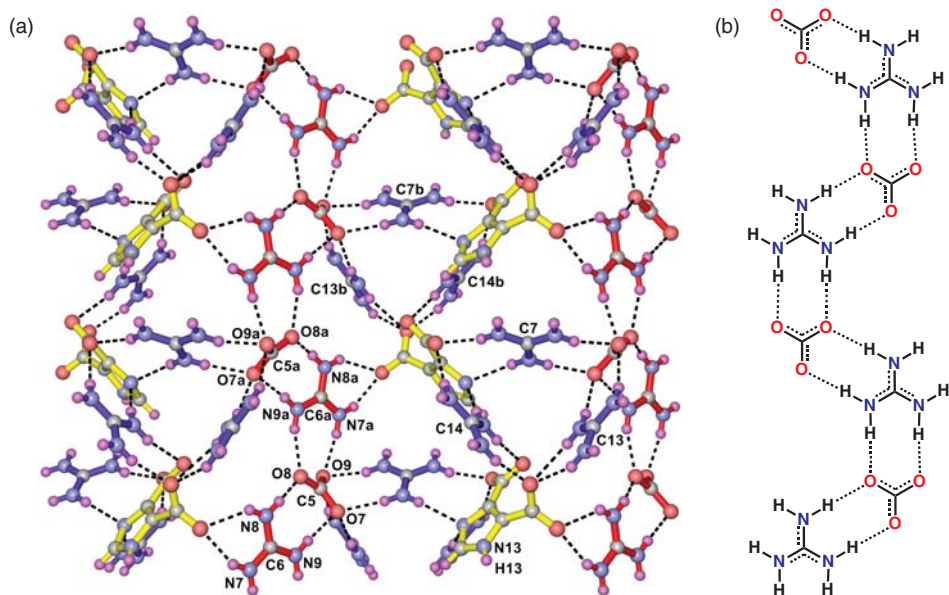


Figure 8.67

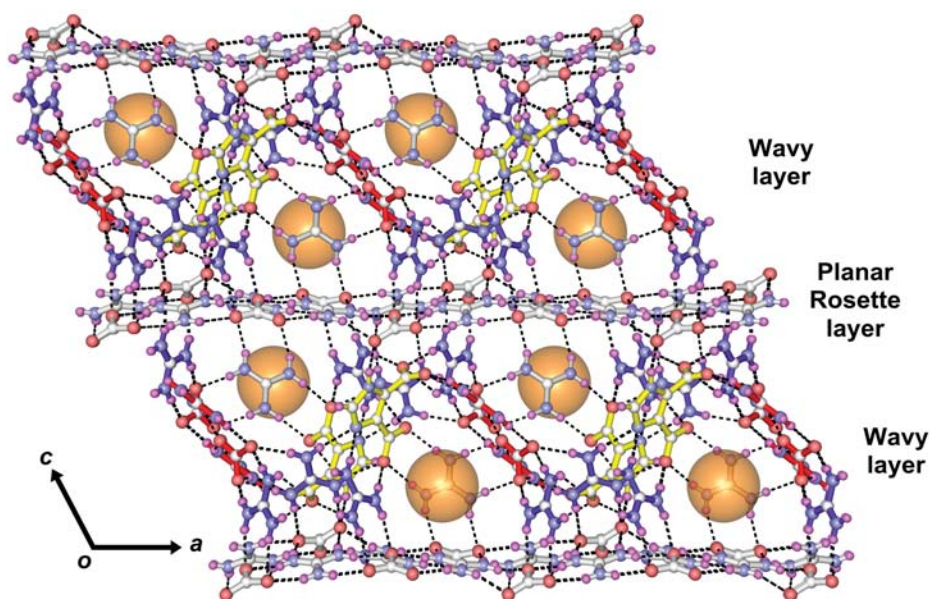


Figure 8.68

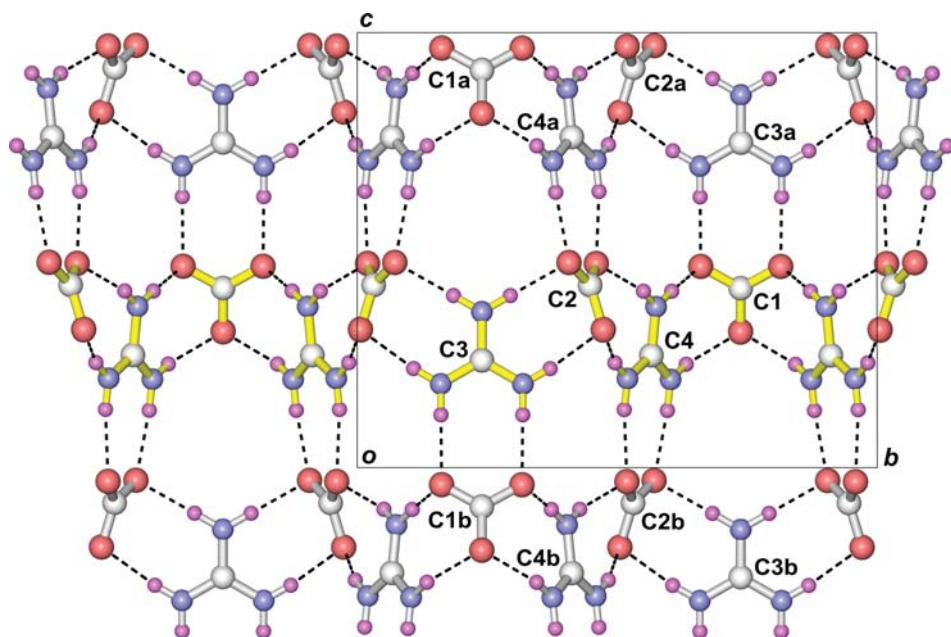


Figure 8.69

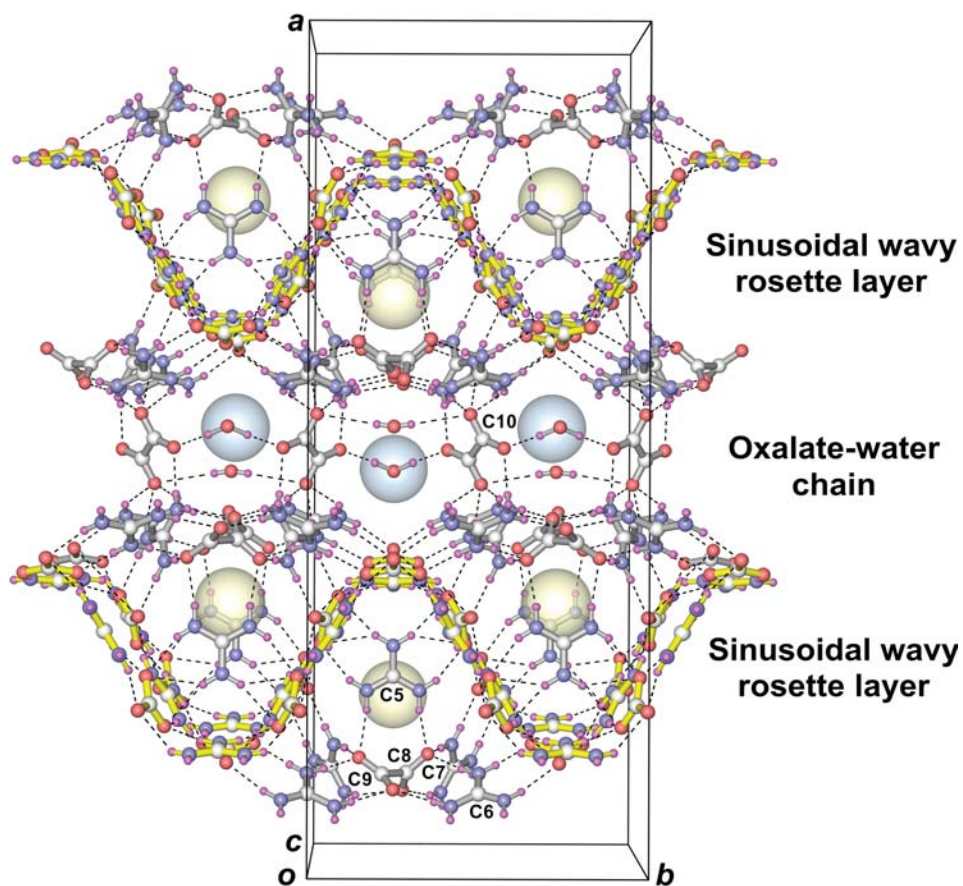


Figure 8.70

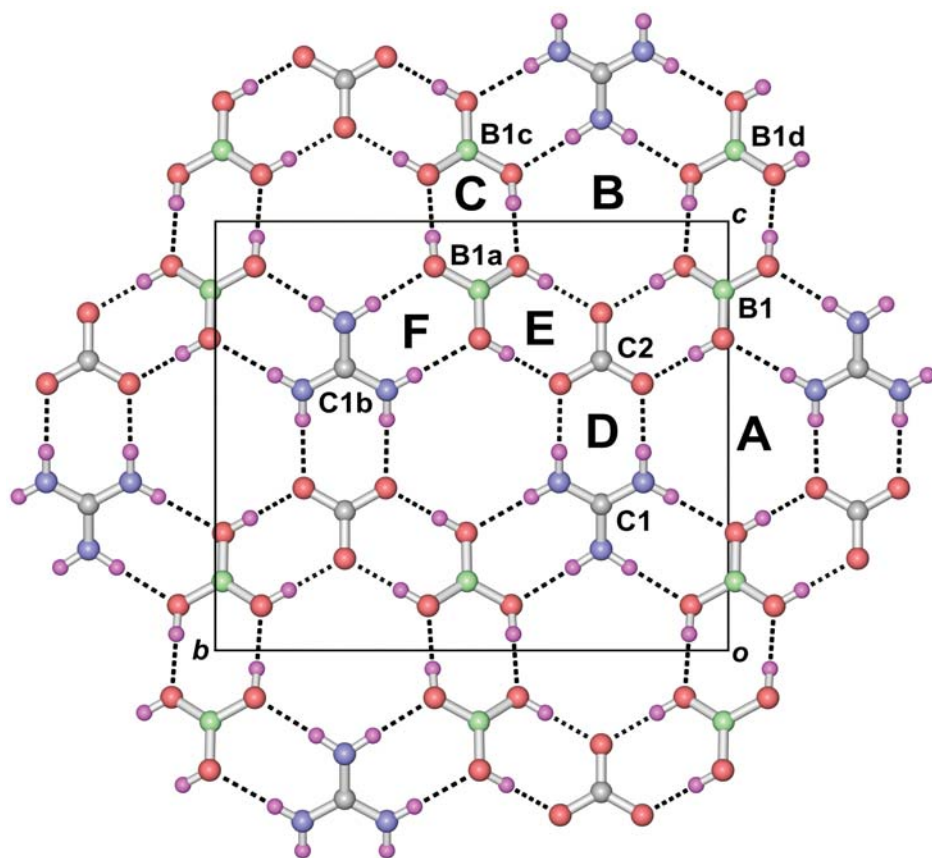
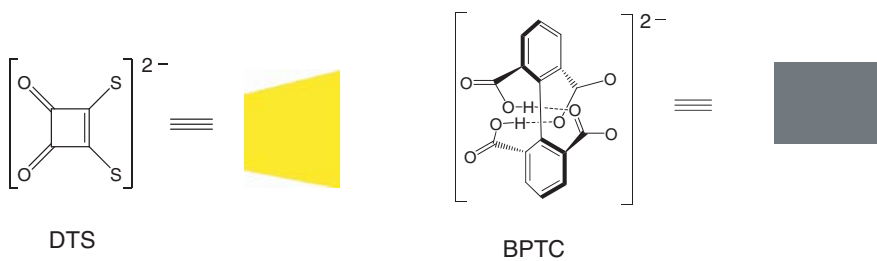
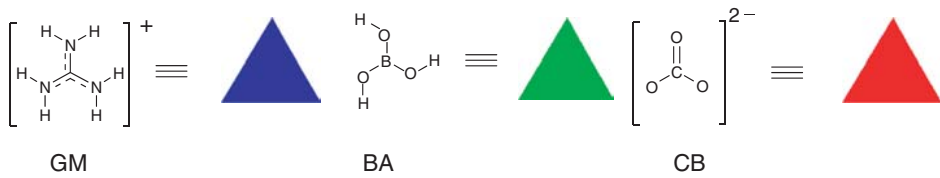


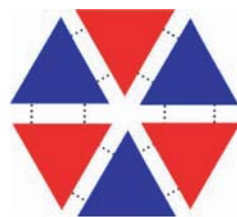
Figure 8.71



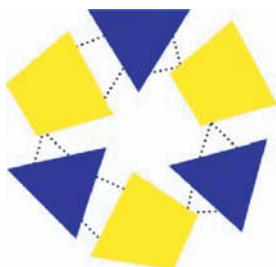
(a)  $[\text{GM}_2 \cdot \text{CB}_2 \cdot \text{BA}_2]$



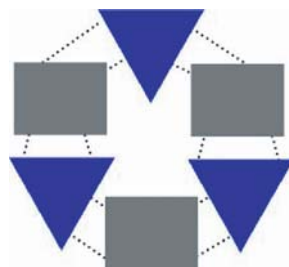
(b)  $[\text{GM}_1 \cdot \text{CB}_1 \cdot \text{BA}_4]$



(c)  $[\text{GM}_3 \cdot \text{CB}_3]$



(d)  $[\text{GM}_3 \cdot \text{DTS}_3]$



(e)  $[\text{GM}_3 \cdot \text{BPTC}_3]$

**Scheme 8.6**