Peter J. Cragg

# Supramolecular Chemistry

## From Biological Inspiration to Biomedical Applications

Springer

Supramolecular Chemistry

Peter J. Cragg

# Supramolecular Chemistry

## From Biological Inspiration to Biomedical Applications

Springer

Dr. Peter J. Cragg
School of Pharmacy and Biomolecular Sciences
University of Brighton
Huxley Bldg, Lewes Road
BN2 4GJ Brighton
UK
P.J.Cragg@bton.ac.uk

ISBN 978-90-481-2581-4 e-ISBN 978-90-481-2582-1
DOI 10.1007/978-90-481-2582-1
Springer Dordrecht Heidelberg London New York

Library of Congress Control Number: 2010932601

© Springer Science+Business Media B.V. 2010
No part of this work may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, microfilming, recording or otherwise, without written permission from the Publisher, with the exception of any material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work.

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

*For James and Alex*

# Preface

From its origins in the last quarter of the 20th Century the field of supramolecular chemistry has expanded to encompass a vast amount of science carried out at the nanoscale yet it is often forgotten that the initial inspiration for supramolecular chemists came from the world of molecular biology. Biological processes construct complex, highly functional molecular assemblies using an array of reversible intermolecular forces. The balance between these forces lies at the heart of enzyme catalysis, DNA replication, the translation of RNA into proteins, transmembrane ion transport, and a wealth of other biological phenomena. Pioneering supramolecular chemists sought to replicate the same complex and subtle interactions in the laboratory so that they could mimic the highly efficient way that chemistry is done in Nature. Key to the success of the field has been the ability of skilled scientists to apply their knowledge of these interactions to the design of unnatural molecules. As a consequence they are able to prepare highly specific sensors, imaging agents and pharmaceuticals, many of which are in widespread use today.

Despite a number of excellent books devoted to supramolecular chemistry there are none that discuss its biological origins and biomedical applications in detail. The aim of this book is to return to the biomimicry and medicinal potential that inspired many of the early supramolecular chemists and to set it in the context of current advances in the field. It starts with an overview, covering the background to the field, the types of molecules and interactions commonly encountered, and methods for investigating the formation of supramolecules. In subsequent chapters parallels are drawn with biological phenomena: the formation of proteins and other biomolecules, self-replication and the origins of life, the evolution of cells, and the design of channel-forming molecules and enzymes. The application of supramolecular principles to sensors and magic bullet therapies is explained and the future of supramolecular therapeutics is considered. The exciting combination of supramolecular chemistry and nanotechnology is discussed together with the likelihood that nanoengineered smart materials could one day circulate in the body, seeking out diseased cells or repairing damaged tissue, so that individuals could receive treatment even before any health problems were apparent.

Brighton, UK
11th May 2010

Peter J. Cragg

# Acknowledgements

Computational results were obtained using Spartan '08 (Wavefunction Inc., Irvine, CA) and software programs from Accelrys Software Inc. with graphical displays generated by the Discovery Studio Visualizer. Where protein structures have been downloaded from the RCSB Protein Data Bank the full references and PDB IDs have been given. I wish to acknowledge the use of the Chemical Database Service at Daresbury for access to other crystal structures. Again, full primary sources can be found in the references.

I would like to thank the University of Brighton for the award of a University Research Sabbatical during the summer of 2009.

Finally, thanks to Margaret, Alex and James for their understanding while I worked on this book.

# Contents


**1 An Introduction to Supramolecular Chemistry** . . . . . 1
1.1 Supramolecular Chemistry . . . . . 1
1.2 Origins . . . . . 3
1.3 Supramolecular Chemistry and Nanotechnology . . . . . 5
1.4 Fundamental Supramolecular Interactions . . . . . 8
1.4.1 Covalent Bonds . . . . . 9
1.4.2 Reversible Covalent Bonds . . . . . 9
1.4.3 Ionic Interactions . . . . . 11
1.4.4 Ion-Dipole Interactions . . . . . 11
1.4.5 Dipole-Dipole Interactions . . . . . 12
1.4.6 Hydrogen Bonds . . . . . 12
1.4.7 Cation-$\pi$ Interactions . . . . . 13
1.4.8 $\pi$–$\pi$ Interactions . . . . . 13
1.4.9 van der Waals Forces . . . . . 13
1.4.10 Hydrophobic Effects . . . . . 13
1.5 Supramolecular Components . . . . . 14
1.5.1 Supramolecular Complexes from Simple Ligands . . . . 14
1.5.2 Macrocycles . . . . . 19
1.6 Supramolecular Entanglements . . . . . 30
1.6.1 Catenanes and Rotaxanes . . . . . 30
1.6.2 Grids . . . . . 37
1.6.3 Dynamic Combinatorial Libraries . . . . . 37
1.7 Observing Supramolecules . . . . . 39
1.7.1 Isolation . . . . . 39
1.7.2 Detection . . . . . 39
1.8 Summary . . . . . 44
References . . . . . 44

**2 Supramolecular Chemistry and the Life Sciences** . . . . . 49
2.1 Life as a Supramolecular Phenomenon . . . . . 49
2.2 Supramolecular Interactions in Biological Systems . . . . . 50
2.2.1 Amino Acids . . . . . 50
2.2.2 Proteins . . . . . 54

2.2.3 Sugars . . . 57
2.2.4 Glycoproteins . . . 58
2.2.5 Lipids . . . 59
2.2.6 RNA and DNA . . . 60
2.2.7 Unusual Structural Forms of DNA . . . 64
2.3 Self-Replication as the Key to Life . . . 66
2.3.1 Replicators . . . 66
2.3.2 Replicator Evolution . . . 68
2.3.3 Orthogonal Translation . . . 69
2.4 Supramolecular Self-Replication . . . 71
2.4.1 Self-Assembling and Self-Replicating Motifs . . . 73
2.5 Supramolecular Chemistry and the Origin of Life . . . 79
2.5.1 Compartmentalization: The Lipid World . . . 80
2.5.2 Catalysis: The Iron-Sulfur World . . . 83
2.5.3 Self-Replication: The RNA World . . . 83
2.6 Supramolecular Biology and Synthetic Biology . . . 86
2.7 Summary . . . 87
References . . . 87

**3 Artificial Cells** . . . 91
3.1 Cells as Capsules . . . 91
3.2 Natural Capsules . . . 91
3.2.1 Clathrins . . . 92
3.2.2 Viral Capsids . . . 93
3.2.3 Coat Proteins . . . 94
3.2.4 Vault Proteins . . . 94
3.3 Unnatural Capsules . . . 95
3.3.1 Self-Complementary Capsules . . . 96
3.3.2 Boxes with Metal Hinges . . . 97
3.3.3 Capsules as Reaction Flasks . . . 100
3.3.4 More Complex Geometries . . . 101
3.4 Synthetic Cells . . . 102
3.4.1 Capsules with Mineral Walls . . . 102
3.4.2 Polymer Based Capsules . . . 104
3.4.3 Lipid Capsules . . . 105
3.4.4 Capsid Virus Mimetics . . . 107
3.5 Towards a Minimal Synthetic Cell . . . 107
3.6 Cellular Aggregation . . . 109
3.7 Summary . . . 110
References . . . 110

**4 Supramolecular Enzyme Mimics** . . . 113
4.1 Enzymes . . . 113
4.2 Metal Complexes as Enzyme Mimics . . . 116
4.3 Enzymes and Their Supramolecular Analogues . . . 118

4.3.1 Haemoglobin, Myoglobin and Their Models . . . . . 119
4.3.2 Cytochromes . . . . . 121
4.3.3 Protection from Radicals: Catalytic Pro- and Antioxidants 122
4.3.4 Copper-Containing Enzymes . . . . . 127
4.3.5 Zinc-Containing Enzymes . . . . . 132
4.3.6 Photosynthesis and Artificial Leaves . . . . . 136
4.3.7 Cyclodextrins as Artificial Enzyme Supports . . . . . 142
4.3.8 Model Enzymes that do not Require Metals . . . . . 143
4.3.9 Molecularly Imprinted Polymers . . . . . 144
4.3.10 Combinatorial Polymers . . . . . 144
4.3.11 Dynamic Combinatorial Libraries . . . . . 145
4.4 De novo Design and Evolutionary Development of Enzymes . . . 146
4.5 Summary . . . . . 148
References . . . . . 148

**5 Natural and Synthetic Transmembrane Channels** . . . . . 153
5.1 Cells and Their Membranes . . . . . 153
5.1.1 Cell Membranes . . . . . 153
5.1.2 Transmembrane Migration: Molecular Shuttles . . . . . 155
5.2 Transmembrane Channels: Selectivity and Gating Mechanisms . 157
5.2.1 Voltage Gating . . . . . 158
5.2.2 Ligand Gating . . . . . 158
5.2.3 Gating by Aggregation . . . . . 158
5.2.4 Gating by pH and Membrane Tension . . . . . 159
5.2.5 Light Gating . . . . . 160
5.3 Channel Architecture . . . . . 161
5.3.1 Channels for Neutral Molecules . . . . . 162
5.3.2 Anion Channels . . . . . 163
5.3.3 Cation Channels . . . . . 163
5.4 Structural Determination . . . . . 167
5.5 Measuring Channel Activity . . . . . 169
5.5.1 Voltage Clamping . . . . . 169
5.5.2 Patch Clamping . . . . . 169
5.5.3 Bilayer Methods . . . . . 170
5.5.4 Dye Release Methods . . . . . 170
5.5.5 NMR Methods . . . . . 170
5.6 Transmembrane Transport by Artificial Systems . . . . . 171
5.6.1 Transporters . . . . . 171
5.6.2 Channel-Forming Systems . . . . . 172
5.7 Summary . . . . . 180
References . . . . . 181

**6 Diagnostic Applications** . . . . . 185
6.1 Applications of Supramolecular Chemistry in Medical Diagnostics 185
6.2 Design Principles . . . . . 185

6.3 Supramolecular Sensors . . . 188
6.3.1 Optical and Fluorescent Biosensors . . . 188
6.3.2 Electrochemical Sensors . . . 195
6.4 Macrocyclic Complexes for Imaging . . . 197
6.5 In vivo Imaging: Magnetic Resonance Imaging Agents . . . 199
6.6 Other Supramolecular Sensors . . . 203
6.7 Summary . . . 204
References . . . 204

**7 Supramolecular Therapeutics** . . . 207
7.1 Therapeutic Applications of Supramolecular Chemistry . . . 207
7.2 Chelation Therapy . . . 208
7.2.1 Desferrioxamine . . . 209
7.2.2 Copper Imbalance: Wilson's Disease and Menke's Syndrome . . . 210
7.3 Macrocyclic Complexes for Radiotherapy . . . 211
7.4 Photodynamic Therapy . . . 211
7.5 Texaphyrins . . . 214
7.6 Targeting Cancer with Peptides . . . 216
7.7 Drug Delivery and Controlled Release . . . 216
7.8 Cyclams as Anti-HIV Agents . . . 217
7.9 A Supramolecular Solution to Alzheimer's Disease? . . . 219
7.10 Calixarenes as Therapeutic Agents . . . 222
7.11 Supramolecular Antibiotics . . . 224
7.12 Summary . . . 227
References . . . 228

**8 Bionanotechnology, Nanomedicine and the Future** . . . 231
8.1 Bionanotechnology . . . 231
8.2 The Unnatural Chemistry of DNA . . . 232
8.3 Molecular Muscles . . . 235
8.4 Nanomedicine . . . 238
8.4.1 Labelling with Nanoparticles . . . 239
8.4.2 DNA Fingerprinting . . . 240
8.4.3 Full Genome Sequencing . . . 240
8.4.4 DNA Sequencing in Real Time . . . 242
8.4.5 Therapeutic Multimodal Nanoparticles . . . 244
8.5 Cell Mimics as Drug Delivery Vehicles . . . 246
8.5.1 Polymer Encapsulated siRNA Delivery . . . 246
8.5.2 Drug Delivery by Particle Disintegration . . . 247
8.5.3 Minicells as Drug Delivery Systems . . . 248
8.6 Supramolecular Protein Engineering . . . 249
8.7 Antimicrobial Limpet Mines . . . 249
8.8 Future Directions . . . 251
8.8.1 Medicinal Nanodevices . . . 251
8.8.2 Powering Nanodevices . . . 252

8.8.3 Functional Nanodevices . . . . . 252
8.8.4 Verification of Treatment . . . . . 253
8.8.5 Nanodevice Control . . . . . 254
8.9 Supramolecular Chemistry and Nanomedicine . . . . . 254
References . . . . . 254

**Index** . . . . . 257

# Chapter 1
# An Introduction to Supramolecular Chemistry

## 1.1 Supramolecular Chemistry

Supramolecular chemistry is the branch of chemistry associated with the study of complex molecular systems formed from several discrete chemical components. These multicomponent entities owe their existence to reversible interactions and so may dissociate and reform in response to particular chemical or environmental stimuli. The aggregation of these components gives rise to new entities with different properties that often behave in entirely novel and unexpected ways. The resulting supramolecular phenomena may be as simple as crystal growth from a saturated solution, or as complicated as ribosomal translation of messenger RNA into a protein. Ultimately supramolecular chemists take simple molecules and assemble them using non-covalent forces to make highly functional nanoscale objects. A good example of this is a sensor, illustrated in Fig. 1.1, composed of recognition and signalling elements separated by a short spacer.

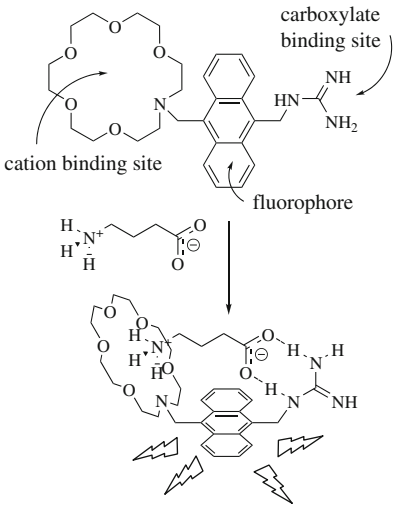



**Fig. 1.1** A sensor based on supramolecular principles

P.J. Cragg, *Supramolecular Chemistry*, DOI 10.1007/978-90-481-2582-1_1,
© Springer Science+Business Media B.V. 2010

A molecule that binds to one target, such as a metal ion or particular amino acid sequence, has no way of signalling its presence. Similarly, a molecule that changes colour or fluorescent intensity, or is electrochemically active, may do so in response to an array of stimuli. Coupling a selective recognition site to a molecule that undergoes an observable response when the recognition event occurs is the basis for a highly specific sensor. The resulting molecule therefore is able to report the formation of a particular supramolecular complex which could signal the presence of a protein associated with cancer proliferation or a metal ion contaminant in drinking water, depending on the recognition element employed.

Many aspects of chemistry, and much of molecular scale biology, may be considered as falling under the 'supramolecular' banner. On the one hand there are phenomena that are observed to result from non-covalent molecular interactions as shown in Fig. 1.2 below. These would include many natural processes and simple chemical behaviour such as precipitation or the formation of oil droplets in water. Then there are examples where several chemical functions have been incorporated into one molecule which then uses the spatial arrangement between those non-covalent interactions to enhance the molecule's properties beyond those of its component functions.

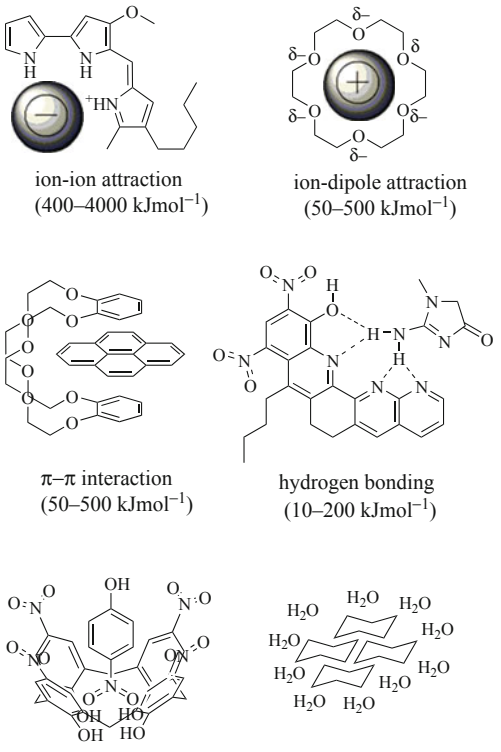



**Fig. 1.2** Supramolecular interactions

## 1.2 Origins

The concept of complex intermolecular interactions being described as 'supramolecular' – literally 'beyond, or transcending, the molecule' – is now associated with Jean-Marie Lehn's definition from the late 1970s:

> Just as there is a field of *molecular chemistry* based on the covalent bond, there is a field of *supramolecular chemistry*, the chemistry of molecular assemblies and of the intermolecular bond. [1]

As Lehn acknowledges, the application of this terminology to chemical species has much to do with Wolf's earlier description of the *übermolecül* – a definition originally designed to cover the self-association of carboxylic acids to form a 'supermolecule' through hydrogen bonding [2], an example of which is illustrated in Fig. 1.3. Indeed, Lehn also used this simpler definition to describe chemical organization in terms of:

> an assembly of two or more molecules, a *supermolecule* [1]

In this sense a 'supermolecule' is defined as a large entity composed of molecular subunits which could be applied equally to a covalently linked polymer as to an assembly held together by weaker interactions. It is in this context that Thomas Pynchon used the word metaphorically in 1973 when describing a character in *Gravity's Rainbow* as:

> a giant supermolecule with so many open bonds available at any given time, and in the drift of things . . . in the dance of things . . . howsoever . . . others latch on [3].

Ultimately the word 'supramolecular' can be traced back as far as the *Century Dictionary* of 1909 [4] where it was given as:

> Composed of an aggregation of molecules; of greater complexity than the molecule.

Other early examples include the 1931 discussion of plant fibres and tendon proteins by Baas-Becking and Galliher [5] who saw no evidence for:

> the presence of supra-molecular discrete and discontinuous units

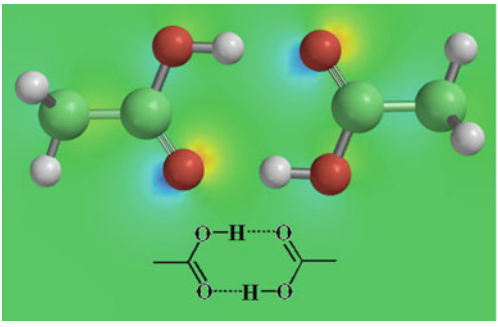



**Fig. 1.3** An acetic acid *übermolekül*

It was later used in the context of biological systems, specifically at the molecular level. In 1955 Palade noted [6] small features in the cytoplasm that had:

> been considered until recently to be devoid of structure at the supramolecular level of organization

Further examples include the description in the journal *Nature* in 1961 [7] of the:

> supramolecular organization of the enzyme systems

Luria, writing from a biologist's perspective in 1970 [8], notes that:

> The transition between molecular structure and morphology is approached by what we may call 'supra-molecular biology'

In this sense it is closer to its modern usage.

Lehn's appropriation of the biological term to cover non-biological chemical entities, and in doing so to supersede the simple *übermolecül*, is entirely appropriate given the complexity of systems with which chemists now work. It also reflects the nature of the dynamically reversible interactions common to both the chemical and biological research fields: hydrogen bonding is essential in many artificial systems as well as secondary protein structure and DNA double helices; metal-ligand interactions are as important in polypropylene catalysis as they are in dioxygen-haemoglobin complexes; hydrophobic effects are seen in both the separation of aliphatic compounds and the formation of transmembrane ion channels. Just as many biological structures are able to form, break up, rearrange, and reform so too can non-biological systems that aggregate through supramolecular interactions. Indeed, in recent years, Lehn has stressed the importance of this type of dynamic interplay between molecules through reversible, non-covalent interactions [9].

Central to much of supramolecular chemistry is Fischer's 'lock and key' analogy of enzyme catalysis [10]. Coupled with later refinements, his concept that a molecular 'host' is somehow an ideal vessel for a smaller 'guest' led to the realization that molecular recognition was dependent upon mutual attractions between host and guest, now known as complementarity. The 'host-guest' concept appears to have its origins in the context of steroid inclusion complexes. Fieser and Fieser's *Steroids* of 1959, discussing inclusion complexes of desoxycholic acid dimers, states that the:

> second component (guest) is not covalently bonded to the enclosing molecules (host) but, if it is of appropriate size, it is completely fenced in and cannot escape [11].

Complementarity may involve the size and shape of guest molecules, the distribution of charged chemical groups on their surfaces, the ability to hydrogen bond through appropriately positioned donor or acceptor groups, the disposition of hydrophobic or hydrophilic chemical groups, or a combination of these.

The concept of supramolecular chemistry gained a wider scientific currency following the award of the 1987 Nobel Prize in chemistry to Donald Cram, Jean-Marie Lehn and Charles Pedersen for:

> their development and use of molecules with structure-specific interactions of high selectivity

Lehn refined his earlier definition of supramolecular chemistry in his Nobel Lecture, calling it:

> the chemistry beyond the molecule bearing on the organized entities of higher complexity that result from the association of two or more chemical species held together by intermolecular forces [12]

He has since made parallels between language and chemistry. As letters are the building blocks of words so atoms become the building blocks of molecules through covalent bonding. By analogy, supramolecules are the chemical equivalent of sentences. The word order of one sentence can be rearranged to make another. 'Are you a supramolecular chemist?' and 'You are a supramolecular chemist!' use identical words but have a different meaning yet they obey the same grammatical rules. Similarly the order in which supramolecular components bind may yield very different results: the order in which two molecular templates are added to a mixture of ligand components will determine which product is formed.

In defining supramolecular chemistry Lehn identified the different levels of molecular complexity: design at the molecular level to synthesize 'hosts' with high affinities for specific 'guest' molecules or ions, molecular assembly (either through self-assembly or self-organization), and dynamic molecular assembly due to the kinetic reversibility of non-covalent interactions between supramolecular components. He also outlined the important noncovalent interactions in supramolecular chemistry. These will be expanded on later but key amongst them are electrostatics, hydrogen bonding, π–π stacking and hydrophobic effects. Individually they are often weaker than formal covalent bonds but their cumulative effects are able drive the formation of supramolecules. The greater the affinity that exists between host and guest through the combination of these forces, the greater will be the selectivity of the host. Exploiting this valuable paradigm through molecular design is at the heart of supramolecular chemistry.

## 1.3 Supramolecular Chemistry and Nanotechnology

Supramolecular chemistry has had a major impact on nanotechnology as the two operate on the same scale [13]. The term nanotechnology is often used quite loosely but it specifically refers to the scale of activity that lies between 0.1 nanometre ($10^{-10}$ m, the scale of bonds between atoms) to 100 nanometres ($10^{-7}$m, the size of a small virus). Objects that exist in this range are illustrated in Fig. 1.4. At its core nanotechnology has the idea that matter can be manipulated at the molecular, and even atomic, level in order to produce functional materials. What functions these materials have are dependent upon their design. For example, careful placement of several metal atoms within an organic framework can result in a catalyst that is less easily poisoned than either the bulk metal surface or a complex containing a single metal. In addition, the cluster catalyst retains the activity of a relatively massive particle of the same metal. Similarly the ability to deposit

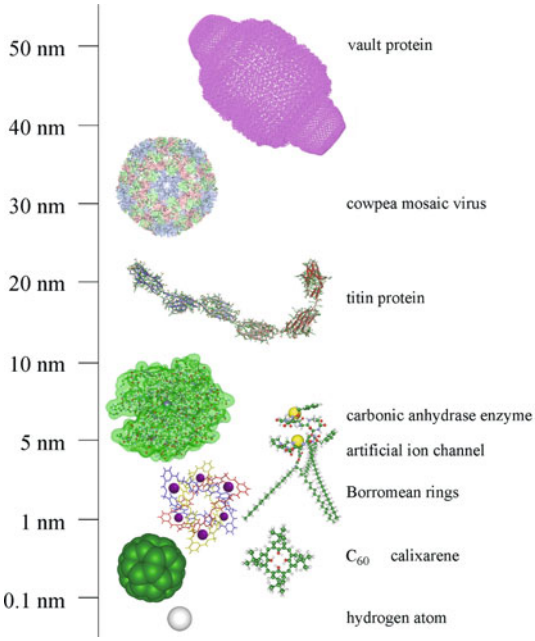



**Fig. 1.4** The nanoscale

single layers of atoms or molecules with precision enables the construction of thin film materials with properties such as high strength, electrical conductivity or a uniformly level surface found in ultraflat screens for computers or televisions. It has even been proposed that multiple properties could be incorporated within single nanoscale objects thus enabling them to perform functions such as movement and turning them into nanomachines [14]. Nanomachines with advanced functions, which are viewed by society as having the potential to produce awe-inspiring medical advances or to cover the Earth in 'grey goo' in approximately equal measure, are far from being feasible to date. Some speculation regarding the potential future of medical nanodevices can be found in final chapter of this book. The main problem with nanomaterials and devices lies in their construction which is hampered by the limitations of accuracy and reproducibility. These limitations become clear when the available methods of nanoscale fabrication are considered. Two approaches are usually taken when constructing nanoscale objects: top down and bottom up.

The top down methods are based on lithography where finely tuned lasers are used to make a pattern in a light-sensitive polymer layer on top of a thin metal sheet that is in turn fixed to a glass plate [15]. The light affected regions of the polymer are chemically removed as is the underlying metal in a separate process. The remaining polymer is then removed to leave the original pattern etched in metal on a glass base to make the mask. When ultraviolet light is shone onto the mask the pattern can

be imprinted onto another light responsive material. As before, the affected sections can be removed chemically to leave an identical pattern. The reverse is also possible: the unaffected material can be removed by a different chemical process to leave a raised pattern. The advantage of this method is that optics can be used to focus the light once it has passed through the pattern to make a smaller version of the original.

While the top down method is widely used, not least in the production of computer chips, there is a limit to which the pattern can be accurately focused leading to problems with the precision and reproducibility of the features produced by this method. As a result lithography becomes less viable when objects below 100 nm need to be manufactured. The alternative bottom up strategy involves chemical deposition at either the atomic or molecular level to build up surface features. While this may appear more accurate there are other limiting factors. First of all it is hard to direct every atom to the desired site. If a monolayer of gold atoms is required on a chromium surface then aspects such as surface roughness and the ability to atomise gold are important. How can we be certain that the gold surface is uniform throughout? If some areas of the base chromium layer have not been covered this will affect any subsequent processes. Alkyl sulfides, long hydrocarbon chains terminating in sulfur, are often deposited on gold surfaces to build in an insulating monolayer. If there are gaps in the underlying gold coating the electrical insulation may be compromised. The more layers of materials that are sequentially deposited the greater the chance that defects will occur and be propagated through the material. Consequently the bottom up approach is usually considered to be useful up to the scale of 10 nm objects.

The limits of the top down and bottom up approaches, illustrated in Fig. 1.5, leave a majority of the nanoworld hard to access. Although constant improvements in technology and chemical synthesis mean that these limits are always shrinking, materials and objects that span the gap between 10 and 100 nm remain hard to fabricate to the level of accuracy and reproducibility expected of most manufacturing techniques. Until recently there was only one way to work on this scale: leave it to Nature.

Protein formation, DNA replication, enzyme catalysis, indeed most biological activities, occur on the scale between 10 and 100 nm. The molecules are prepared rapidly, specifically and with hardly any errors. The only problem is that not all the materials or objects we wish to prepare on this scale are found in the natural world. Here is where supramolecular chemistry steps in.

Using highly specific, reversible bonding interactions that can rearrange until the desired intermolecular geometry is achieved it is possible to orient two or more molecules quite precisely. Furthermore if the molecules themselves are of the order of a nanometre then the resulting supramolecule could easily have linear dimensions above 10 nm. A supramolecular complex synthesized in this manner combines the accuracy of the bottom up approach with a self-checking mechanism, used by Nature to reduce errors, yet is on a scale that reaches well into the regions otherwise unavailable to either conventional fabrication methods.

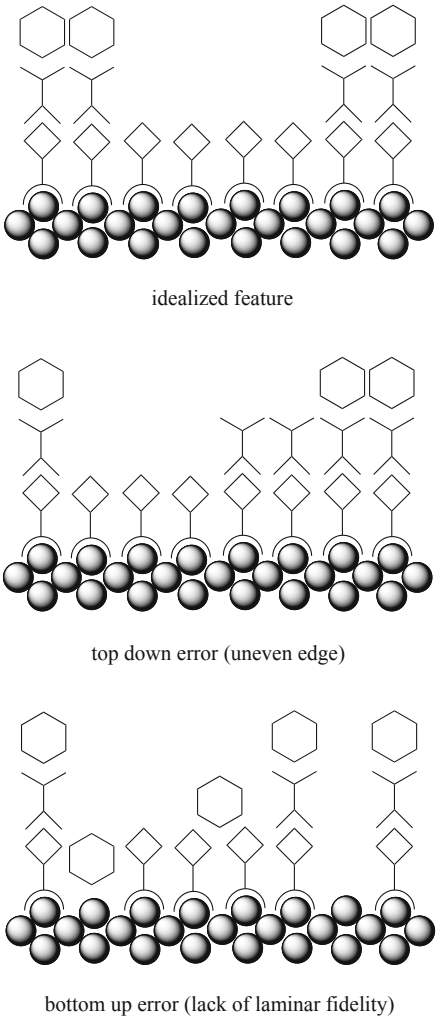



**Fig. 1.5** Imperfections arising during nanofabrication

## 1.4 Fundamental Supramolecular Interactions

A recurring theme in supramolecular chemistry is its appropriation of concepts more usually associated with biological systems. This is particularly true when invoking reversible atomic and molecular interactions in complex formation. In supramolecular chemistry, as in biology, it is common to envisage individual molecular components as the fundamental building blocks from which complex, higher order structures form. While the individual blocks contain strong covalent bonds, the multicomponent aggregate, or supramolecule, is likely to be held together by weaker forces. The overall affinity of the host for the guest is unlikely to be due

to a single intermolecular interaction but will come from a combination of forces. In the design of supramolecular components it is often possible to manipulate the balance of these forces to improve host selectivity. For example, a host that incorporates benzoic acid will have a stronger affinity for hydrogen bond acceptors through judicious choice of *ortho* and *para* substituents that electronically influence the ease with which the acidic proton dissociates.

### *1.4.1 Covalent Bonds*

Covalent bonds, almost by definition, should be of little relevance to supramolecular aggregation. They are the interactions that allow molecules to form through the sharing of electrons between atomic nuclei and include the backbones of all organic compounds which are largely composed of carbon atoms linked by single, double, or triple bonds. Other elements are also incorporated into molecules by covalent bonds either as linking atoms or as part of peripheral groups. An example of this is the peptide bond in proteins where nitrogen forms part of the protein backbone and oxygen extends outwards allowing it to form weaker hydrogen bonds with adjacent amide hydrogen atoms. Although most of the common examples of covalent bonds are strong some are susceptible to attack from acids or other competitors. These 'reversible' covalent bonds are an important class in themselves and are key to several biochemical processes.

### *1.4.2 Reversible Covalent Bonds*

Since the original definition of supramolecular chemistry was coined by Lehn several corollaries have emerged. One that has risen to great importance is the idea of a dynamic combinatorial library of molecular components that self-sort to generate supramolecules with reactive termini which are then predisposed to form covalent bonds. The effects of weak interactions together with geometric and steric constraints lead to the formation of far fewer products than would be predicted by pure statistics. This development will be discussed in greater detail later.

A related observation is that several types of covalent bonds are readily reversible under relatively mild conditions. The importance of this is that, even when certain covalent bonds are formed, other forces may combine to break the bond and send the molecular components back to the pool of available fragments. Such a possibility is essential in any error checking process. Without the ability to undo chemical mistakes any replication process is likely to generate large numbers of errors. The result will be a highly ineffectual method of perpetuating encoded information.

One of the best known examples of reversibility in bond formation is the cross-linking of cysteine, a sulfur-containing amino acid, that affects tertiary structure in proteins and, ultimately, macroscale phenomena such as the degree of curl in hair. Other examples include the imine bond, formed by the reaction of an amine group with an aldehyde, and metal coordinate bonds to atoms such as nitrogen as found in many enzymes.

#### 1.4.2.1 Sulfur-Sulfur Bonds

Thiols, sulfur-containing analogues of alcohols, terminate in a sulfhydryl group. When two of these groups are oxidized they can form a single sulfur-sulfur, or disulfide, bond. Under reducing conditions, as usually exist within cells, the reaction is reversed. The addition of peroxide, containing an oxygen-oxygen single bond, can also break disulfide bonds. The reaction is used widely to break disulfide bridges that exist in proteins that coat hair; subsequent refolding of these proteins leads to the familiar perm effect from the 'permanent wave' that is formed in hair that has little or no natural curl of its own.

#### 1.4.2.2 Imine Formation

The reaction between a primary amine and certain aldehydes or ketones results in the formation of a carbon-nitrogen double bond. Most double bonds involving carbon are extremely stable but the imine bond is susceptible to attack by water leading to hydrolysis and regeneration of the starting carbonyl and amine-containing compounds. Work by Saggiomo and Lüning has shown that simple imine formation is essentially reversible in water but the thermodynamic trap of the product's insolubility is the main driving force for the forward reaction [16]. For this bond to dissociate in a true supramolecular sense the imine must have some solubility in the solvent used.

#### 1.4.2.3 Metal-ligand Coordinate Bonds

Metals are a vital part of any complex biological system. Whether they are used simply to affect osmotic pressure, as centres for catalysis and redox activity, or as structural elements, all must be sequestered, stored, and dispersed to their active sites. The alkali metal cations (predominantly $Na^+$ and $K^+$) and alkaline earth cations (mainly $Ca^{2+}$ and $Mg^{2+}$) are generally encountered either weakly bound to protein carbonyl groups or in a hydrated form surrounded by six or eight water molecules. Transition metals, while they also exist in hydrated forms, are ultimately bound to nitrogen, oxygen or sulfur atoms in proteins through coordinate bonds. These bonds arise because the transition metals have electronic orbitals with specific directional preferences, however, because electrons are technically prohibited from moving between them, certain orbitals are vacant and able to accept electron pairs. Atoms such as nitrogen, oxygen, and sulfur can bind to these metals by donating a lone pair of electrons to a vacant metal orbital thereby forming a stable complex. These interactions, while strong, are also often reversible. It is the relative strength of this coordinate, or dative, bond that allows oxygen gas to bind reversibly to the iron atom at the centre of the haem core in haemoglobin and myoglobin. Where the coordinate bond is stronger, as in the iron-cyanide interaction, the process becomes irreversible and the effect of the coordinate interaction is to poison the metal centre.

### 1.4.3 Ionic Interactions

Complementary cation-anion interactions are usually even stronger than the sharing of electrons in covalent bonds, however, they are easily disrupted by polar solvents. It is for this reason that simple salts often dissolve easily in water and yet have melting points higher than many metals. Each chemical species has attained ionic status through the gain or loss of one or more electrons. This has occurred because the resulting ion pair is more energetically stable than an aggregate of the neutral parent atoms or molecules. For example, sodium metal is uncharged but has a single electron in its valence shell. By giving up this electron, and becoming $Na^+$, it empties its 3s shell and adopts the same extremely stable full shell $2s^2 2p^6$ arrangement as neon, a highly unreactive gas. Similarly, chlorine has seven valence electrons and by gaining one, to become $Cl^-$, adopts the same full shell configuration as argon ($3s^2 3p^6$), another unreactive gaseous element. The resulting ions have opposite charges and are mutually attractive. A large amount of energy is required to overcome the strength of this attraction and separate the ions. This corresponds to a high melting point. However, the fact that the ions are charged means that they attract polar molecules, such as water, which disrupt the strong interactions between oppositely charged ions and eventually solvate each individual ion. This explains the apparent contradiction between their high melting points and ease of dissolution. Being charged also means that small ions can both cause and respond to changes in biological systems.

As well as simple ions that contribute to changes in osmotic pressure and may influence protein binding, we must also consider the impact of complex ionic systems. DNA and RNA are both polymeric anions as the individual nucleosides are linked by charged phosphate groups. Similarly many proteins have side chains that may be ionized, a process that in turn can affect secondary and tertiary structure as well as protein-protein or protein-substrate interactions.

### 1.4.4 Ion-Dipole Interactions

Ions have a permanent charge, positive or negative and can interact with molecules that possess a dipole. A dipole may also be a permanent feature, as in carbon monoxide where the uneven sharing of the bonding electrons leads to the carbon being slightly more electropositive than the oxygen, or it may be a temporary, fluxional feature. A good example of the latter is where a non-polar dioxygen molecule is attracted to iron in haemoglobin. As the dioxygen molecule approaches the positively charged iron a temporary dipole is set up where the oxygen closest to the metal becomes more electronegative. The dative bond formed is weak, and therefore easily reversed, but strong enough for haemoglobin to transport dioxygen from the lungs to the muscles. A similar ion-dipole interaction between magnesium cations and oxygen at the 3′-hydroxyl end of a nucleotide appears to be important in the catalysis of DNA polymerization.

### 1.4.5 Dipole-Dipole Interactions

Polar molecules can interact weakly with other polar molecules through the same mechanism outlined above. As neither 'pole' of the molecule is particularly strongly charged this type of interaction is necessarily weak. Nevertheless, it may play a part in the orientation of polar hydrocarbons that aggregate to form micelles and lipid bilayers, the forerunners of today's biological cells.

### 1.4.6 Hydrogen Bonds

Hydrogen bonds form when a hydrogen atom is covalently bound to an electron rich atom. This leads to a polarization of the covalent bond making the hydrogen electropositive and therefore attractive to nearby electron rich atoms. The resulting hydrogen bond is weak yet several complementary hydrogen bonds can impart enormous stability to molecular interactions. The phenomenon is familiar to anyone who has seen ice form. When freely moving water molecules cool, hydrogen bonds can form between them leading to small clusters and then more rigid three dimensional 'diamondoid' lattices. Each oxygen atom is linked to four hydrogens, two by conventional covalent bonds and two by hydrogen bonds to neighbouring water molecules. The increased stability that this structure brings overcomes the energy available that would allow the molecules free independent movement. At the other end of the scale, water boils when all the hydrogen bonds are broken and the individual molecules can turn to a gas. Unlike many other covalent molecules that boil at much lower temperatures, many weak hydrogen bonds must be disrupted before water molecules become volatile enough to evaporate. A similar effect is observed for other electron rich elements (nitrogen, sulfur, and halides like fluorine, chlorine, bromine and iodine) that form compounds with hydrogen. As noted above, the first use of the term *übermolekül* (supramolecule) was to describe the hydrogen-bonded acetic acid dimer which must dissociate before the molecules can enter the gas phase.

Hydrogen bonds are not just important for small molecules. Duplex strands of DNA and RNA are held together by hydrogen bonds between complementary purine and pyrimidine bases. Because each individual bond is weak it is possible to 'unzip' these large molecules and use the primary sequence in transcription (for sequence copying) or translation (for protein synthesis) and 'zip up' the hydrogen bonds after the information has been accessed by transcription and translation enzymes. Transfer of encoded information can therefore occur without destroying the sequences of the parent compound.

Hydrogen bonds are also invaluable as mechanisms by which secondary and tertiary structure can be imparted to proteins. The most well-known examples are the formation of α-helices and β-sheets. In the former, linear sequences of amino acids form a spiral that is stabilized through multiple interactions between aligned amine hydrogen atoms and carbonyl oxygen atoms. In the latter the same interactions are found between amino acid sequences that are aligned in the same plane.

### 1.4.7 Cation-π Interactions

Aromatic organic molecules, cyclic compounds with a conjugated bonding system such as benzene, are commonly found in both biological and non-biological contexts. The former includes the amino acids phenylalanine, tryptophan and tyrosine. Their interactions with cations appear to be of considerable importance in directing proteins to form correct tertiary structures. The interaction is based on the attraction between a positively charged metal ion and the areas of delocalized electron density that lie above and below the plane of an aromatic ring. There is also some discussion surrounding the nature and importance of anion-π interactions in anion binding ligands and proteins. While the anion-π effect may be shown to be more widespread in the future than is currently the case, the cation-π effect is an essential supramolecular interaction not only in simple host-guest systems but also in protein complexes that incorporate organic or inorganic cations.

### 1.4.8 π–π Interactions

One effect of aromaticity is that, by drawing electron density into orbitals associated with the carbon framework, the hydrogen atoms on the periphery of the molecules are polarized. This creates an electron rich region associated with the π system and an electron poor region associated with the hydrogen atoms. The positively polarized hydrogens are able to interact with the π system of a neighbouring molecule through a perpendicular dipole-dipole interaction. Alternatively they can 'stack' through staggered π–π interactions where one molecule lies above the other, but offset, so that the complementary electron rich and electron poor regions match up.

### 1.4.9 van der Waals Forces

Van der Waals, or London, forces are extremely weak and less easy to control than most others. Although they exist throughout chemistry and biology, it is hard to incorporate functionalization into molecules that specifically introduces these more nebulous intermolecular effects.

### 1.4.10 Hydrophobic Effects

Many molecules do not possess the ability to form hydrogen bonds or other attractive interactions based on complementary charges. These are often compounds composed of carbon and hydrogen, typical examples being linear hydrocarbons and aromatic ring systems. Although some interactions such as π-π stacking can occur, the main effect of these molecules is to interact by excluding charged or polar groups. Given the importance of molecular charge in our largely water

dependent biology it may seem that there is no place for these 'non-interactions' yet the packing arrangement between phospholipid hydrocarbon 'tails' to exclude water is the basis of every cell. Similarly regions of proteins may be composed of amino acids with predominantly hydrophobic side chains. This allows the proteins to bind hydrophobic substrates as the protein-substrate interaction will be energetically preferred to the hydration of either entity. The hydrophobicity of a region of a protein will also determine its position in a lipid membrane which in turn helps to anchor it so that regions designed to function in the extra- or intracellular environment are correctly oriented. Other hydrophobic interactions may promote protein self-aggregation in an analogous manner to the complementary charge-charge self-assembly undergone by other proteins.

## 1.5 Supramolecular Components

Supramolecules, as their very name suggests, are composed of more than one chemical species that interact through non-covalent means. As with much of supramolecular chemistry it can be claimed that the field has merely appropriated known complexes and reclassified them. While this may be true for many existing transition metal coordination complexes and multicomponent protein aggregates, where the supramolecular concept really comes to the fore is in the design and use of novel, multifunctional ligands. For example, metal terpyridyl complexes have existed since Morgan and Burstall first prepared the ligand in 1932 [17] but their potential to mimic the metal centre of photosystem II had to wait until 2001 when, no doubt inspired by earlier work on multicomponent bis(terpyridine) photosensitizers by Sauvage [18] and Balzani [19], the Crabtree group prepared a dioxygen-generating terpyridyl manganese coordination compound [20]. The complexity of the target was coupled to an understanding of both photochemistry and the natural photosystem in a synergistic manner that pervades supramolecular design. Similar examples that push ordinary coordination chemistry into the realm of the supramolecular include the discovery by the group of Fujita [21] that 4,4′-bipyridine reacts with cadmium to form two-dimensional networks that stack on top of each other and catalyse the cyanosilylation of aldehydes, and the metal-assembled tetrahedra, cubes and helices synthesized from linear bifunctional ligands reported by the groups of Raymond [22], Stang [23], Thomas [24] and Hannon [25].

### *1.5.1 Supramolecular Complexes from Simple Ligands*

Many multicomponent complexes form because of the orthogonally divergent bonding preferences of transition metals. Metals favouring square planar geometries, notably those with a $d^8$ electronic configuration such as $Ni^{2+}$, $Pt^{2+}$, $Pd^{2+}$ and $Au^{3+}$, can form linear or grid-like extended structures but only if they are joined by bifunctional ligands. A simple example is the linear, symmetric 4,4′-bipyridine with its

two nitrogen donors set 180° apart. Other transition metals that are more stable with either octahedral or tetrahedral geometries can also employ bifunctional ligands incorporating nitrogen, sulfur or oxygen donors to form complex structures. Other metals, particularly the metal cations in groups 1 and 2 of the Periodic Table and the lanthanide cations, have a greater affinity for oxygen-containing ligands with greater conformational flexibility. This is because, unlike the transition metals, they exhibit spherical electronic density and, with no preferred orbital orientations, the complexes they form are only limited by the steric and electronic repulsions resulting from the proximity of donor and backbone atoms when the ligands encapsulate the cations. One initial aspect of supramolecular design is therefore to match, as well as possible, the properties of the donor atoms in the ligands with the qualities of the metals that are to be targeted. Examples of typical building blocks to bind both metals and hydrogen bonded systems are shown in Fig. 1.6. Fortunately there are some good guidelines available that have arisen through analysis of many coordination complexes. According to Pearson's hard and soft acid and base (HSAB) definition [26] 'hard' elements (e.g. oxygen) will bind metals such as caesium, strontium, chromium, titanium whereas 'soft' elements (e.g. sulfur) will bind metals including mercury, silver, gold, platinum and cadmium. Nitrogen, of 'intermediate' character, preferentially binds cobalt, nickel, iron, rhodium, tin, lead and so forth. This knowledge can be used as a first principle in host design.

#### 1.5.1.1 Pyridine Derivatives

Pyridine is widely used as a solvent that also ligates a number of transition metals through an aromatic nitrogen donor. From this perspective it has little importance in facilitating supramolecular assembly but fortunately the pyridine morif can be incorporated in numerous derivatives. A single pyridine molecule can form the basis

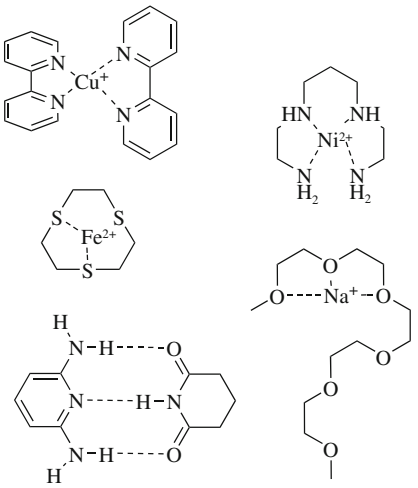



**Fig. 1.6** Some simple supramolecular binding motifs

of a hydrogen bond acceptor. Two pyridines can be connected to form bipyridine where the nitrogen atoms can adopt divergent or convergent positions. One of them, 4,4′-bipyridine, can act as a rigid bifunctional spacer and is often encountered as a structural element in molecular arrays, boxes, and other similar structures. The other, 2,2′-bipyridine, has convergent donor sites and can act as a bidentate ligand. Because the sites are constrained by the rigid nature of the molecule only certain metals are able to be bound. Rigidity can be increased if the two bipyridine rings are linked by a second bridge leading to the planar aromatic 1,10-phenanthroline. When three of these ligands encapsulate iron in the +2 oxidation state a distinctive red complex is formed which can be oxidized to a blue +3 complex. The latter slowly decomposes, turning khaki over time, where the former remains unaffected. This is in stark contrast to the usual behaviour of iron which readily oxidizes from the +2 to the +3 state which is more stable in an aerobic environment. Extending 2,2′-bipyridine by a further pyridine group leads to the tridentate 2,2′:6′,2″-terpyridine. This compound has been widely used in photosystem mimics through coordination to a metal, often ruthenium, and functionalization of the central pyridine in the 4-position. The three nitrogen donor atoms can converge on a metal that usually exhibits an octahedral coordination geometry to occupy three sites in the same plane. This meridional binding mode leaves three metal coordination sites available for a second terpyridine ligand to bind at right angles to the first. The motif lends itself to the design of supramolecular materials that assemble only in the presence of an appropriate metal.

It is possible to increase the number of pyridines leading to multidentate ligands that form helical structures around a number of metals as has been elegantly demonstrated by the groups of Constable [27] and Bell [28]. A convergent synthetic approach based on octahydroacridine, which contains a central pyridine, leads to cyclic sexipyridines known as torands, after their flattened, toroidal shape. The Bell group has shown that these compounds bind simple cations with extremely high binding constants and can form unusual stacked dimers through which a water-cation-water-cation-water chain can be threaded [29].

#### 1.5.1.2 Schiff Bases

Schiff bases are formed by the condensation reaction between an aldehyde and primary amine that generates an imine bond. The entire class of these compounds is named after Hugo Schiff who published papers on the reaction in the 1850s and 1860s [30]. The nitrogen in a Schiff base has the capability to coordinate to many transition metals but the main use of imine bond formation is to connect two molecules that have useful donor properties in themselves and, in doing so, create a multifunctional product. This is most evident in the synthesis of the salen and related ligands. The parent compound is formed by the reaction of ethylenediamine and two equivalents of salicylaldehyde to yield a symmetric ligand with two imine bonds and two phenolic functional groups. Although the ethylene link is flexible the introduction of a transition metal initiates a conformational rearrangement that allows both nitrogen and both phenolic oxygen donors to converge

on the metal. Other derivatives can introduce greater rigidity, for example by the substitution of ethylenediamine by *ortho*-phenylenediamine, or extend the distance and angle between the terminal phenolic oxygens, as seen in the Hannon group's 4,4′-methylenedianiline derivative [25].

The Schiff base condensation is generally a very efficient coupling method and has therefore found applications in self-replicating systems where high yields and fidelity are required.

#### 1.5.1.3 Polyamines

Polyamines, compounds with the $H_2N(CH_2CH_2NH)_nCH_2CH_2NH_2$ or $H_2N(CH_2)_nNH_2$ structure, are well known in Nature and include putrescine, cadaverine, spermine and spermidine. Putrescine is synthesized by the action of arginine decarboxyalase on arginine; cadaverine results from the action of lysine decarboxylase on lysine. Spermidine is formed by the spermidine synthase catalysed transfer of an aminopropyl group from S-adenosyl-L-methionine to putrescine. Spermidine can be extended through the same reaction, catalysed by spermine synthase, to form spermine. As amines are well known metal chelating agents it is unsurprising that similar polyamines have been prepared with the intention of forming metal complexes. The simplest compound, ethylenediamine, forms complexes with many metals due to its two donor atoms and flexible spacer. Extending this leads to diethylenetriamine and triethylenetetramine both of which are highly flexible and able to encapsulate metals. A related compound, 1,2-bis(3-aminopropylamino)ethane is the staring material for the cyclam macrocycle that has a number of clinical applications.

#### 1.5.1.4 Polyethers

The acid or base catalysed reaction of ethylene oxide, a highly strained triangular molecule, with ethylene glycol or water results in a linear polymer with a $–CH_2CH_2O-$ repeat unit and hydroxyl termini. The compounds, known as polyethyelene glycols or polyethers, are amphiphilic and dissolve in a vast range of solvents from water to chloroform and benzene. Consequently they have found use in a number of phase transfer applications, where salts need to be extracted into nonaqueous solvents, and clinical uses, where they can be attached to large drug molecules to improve their distribution in the body. Their amphiphilic behaviour also makes these compounds good lubricants.

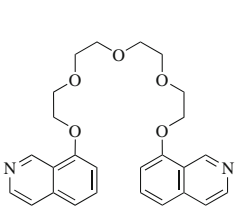
As with any polymerization the length of the resulting polymer varies with the method used. Base-catalysed polymerization yields lower weight polymers such as the tri-, tetra-, penta- and hexaethylene glycols that are often incorporated in supramolecular components. Higher molecular weight polymers are usually referred to by their average molecular mass, thus PEG 2000 would have a mass of 2 kDa (2000 atomic mass units, or Daltons) and an average composition of $HO(CH_2CH_2O)_{45}H$. Studies have shown that low molecular mass polyethers and

their derivatives are highly toxic and teratogenic but that toxicity drops off rapidly with increasing mass.

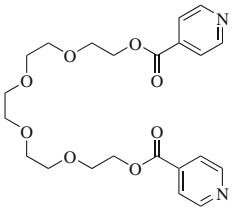
On their own polyethers have had limited use in supramolecular chemistry. Low molecular mass compounds have been shown to bind a number of metal ions, notably those in the lanthanide series that can accommodate ligands with large numbers of oxygen donor atoms. One of the main problems is the lack of diversity in the compounds' functional groups which limits the range of their ligating opportunities. Where polyethers have been highly successful is as substituents to other molecules, such as calixarenes, and in their cyclic forms as the crown ethers, which will be described later.

### 1.5.1.5 Podands

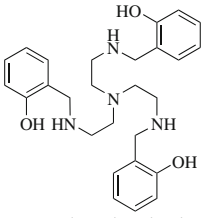
At the same time that crown ethers were first being investigated in the 1960s some simple reactions were used to functionalize polyethers. The terminal hydroxyl groups were found to be amenable to tosylation and subsequent reaction with nucleophiles could introduce variations in ligand donor groups. Vögtle [31] and co-workers, notably Weber [32], used this route to incorporate a number of benzoic acid and quinoline groups. The new polyether derivatives were classed as podands in recognition of their two ligating 'feet'. Numerous variations, as illustrated in Fig. 1.7, have since been prepared with the aim of broadening the range of metals that could be bound. The earliest compounds could coordinate to oxophilic



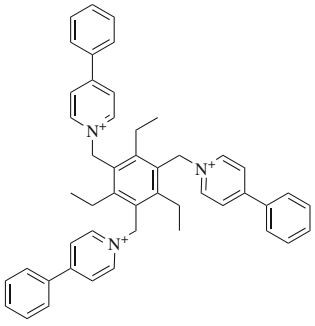
6-hydroxyquiniline podand



isonicotinyl podand



tren-based podand



1,3,5-triethylbenzene-based podand

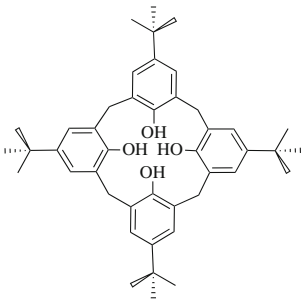
**Fig. 1.7** Podands

metals, those classed as hard acids to use Pearson's terminology, while further development opened up the possibility of ligation to metals with borderline and soft characteristics through donors such as pyridine and sulfur. One interesting multifunctional podand was reported by the Hosseini group in 2001 which demonstrated that even compounds with simple structures can generate surprising supramolecular architectures [33]. Hexaethylene glycol was terminated in isonicotinic groups and reacted with silver salts to form a linear polymer in which each silver cation was held in a polyether loop around its equatorial plane and coordinated axially to the nitrogen atoms from the isonicotinyl termini of two adjacent ligands. An elegant interpenetrating linear structure resulted in which one linear network interpenetrated another running in the opposite direction: the same motif that runs through self-complementary double strand DNA and RNA. Despite some interesting results podands with two feet remain limited in their utility, particularly those created from polyethers, as most examples merely wrap up metals by coordinating to all available binding sites.

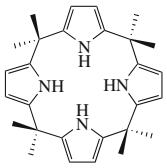
Multifunctional podands can also be prepared from nonlinear parent compounds. Two of the most accessible routes are based on molecules that have threefold symmetry. Derivatives of 1,3,5-tri(bromomethyl)benzene and *N*,*N*,*N*-tri(2-aminoethyl)amine ('tren) have been attractive as tripodal ligands because of their ease of functionality. Trisubstituted benzene podands are prepared by reaction of formaldehyde and bromine on trimethyl- or, more commonly, triethybenzene. The substituents alternate in their orientation from the aromatic plane so that the most stable conformer has all three bromomethyl groups emerging from the same face. The compound is thus preorganized in a cone geometry so that further reactions, usually with pyridine derivatives to form the tri(*N*-alkylpyridinium) salt, extend the cavity. The synthesis can be used to introduce a number of secondary features that respond to guest encapsulation by the tipodal host such as the electroactive ferrocene groups introduced by the Steed group [34]. The tren route to tripodal ligands has also been explored, notably by the groups of Orvig [35] and Bowman-James [36], as the terminal amines lend themselves to Schiff base condensations, outlined above. Reaction of tren and three equivalents of an aldehyde, especially an aromatic aldehyde, results in a flexible podand. In addition to the bridgehead nitrogen, the three imine nitrogen atoms can act as donors for metals, as shown in numerous examples by the Orvig group, or can be reduced to the amine for anion binding. The terminal aromatic groups are not restricted to benzene but could contain phenolic, benzoic or pyridyl functionality by analogy to Vögtle's original podands. The important difference is that the affinities of tripodal ligands for trigonal planar and tetrahedral guests should be enhanced over those with other geometries.

### *1.5.2 Macrocycles*

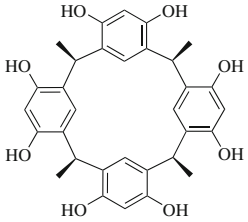
Although supramolecular chemistry is a relatively recent concept many of the compounds associated with the field have much earlier origins. The first appearance of synthetic cyclic molecules capable of encapsulating guests was documented in the




4-t-butylcalix[4]arene




octamethylcalix[4]pyrrole



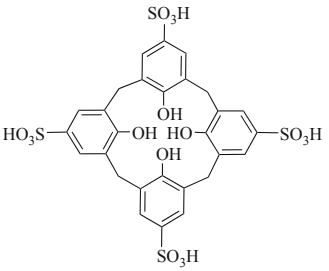

tetramethyl[4]resorcinarene

**Fig. 1.8** Some of the macrocycles discovered by von Baeyer

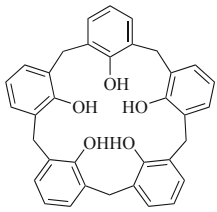
late 19th century by Adolf von Baeyer. In particular it was the analysis of the products from the reactions between phenols and aldehydes that led to the discovery of calixarenes and resorcinarenes as shown in Fig. 1.8 that are still in constant use today.

### 1.5.2.1 Calixarenes

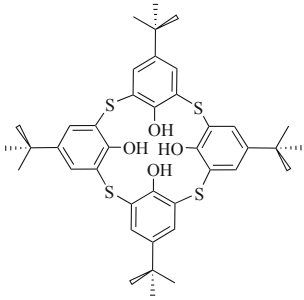
Calixarenes have had a great impact on supramolecular chemistry. They were probably prepared by von Baeyer in the early 1870s; his paper of 1872 [37] describes a number of reactions involving phenols and aldehydes. Interest in the Bakelite$^{TM}$ process, through which the reaction of phenol with formaldehyde results in the formation of a brittle polymer, undoubtedly led to the first reproducible calixarene syntheses that appeared in the 1940s [38] and 1950s [39]. The calixarenes are cyclic compounds comprised of phenols linked to each other through the 2- and 6-positions by methylene (–$CH_2$-) spacers, as can be seen in Figs. 1.8 and 1.9, and are prepared by the high temperature reaction between phenolic compounds and formaldehyde. The name was coined by Gutsche who likened the shapes of the molecules to a Greek vase known as the *calix crater* [40]. Calixarenes can be functionalized at the 'upper rim' (4-position) and 'lower rim' through the formation of phenolic derivatives. Compounds containing different numbers of phenols are known, from the easily prepared tetramer, calix[4]arene, to calix[12]arene and beyond. Reaction of bis(methylol)phenols with sulfur yields the thiacalixarenes, in which sulfur replaces the methylene bridge, and oxa- and azacalixarenes where the bridge is –$CH_2OCH_2$- or -$CH_2N(R)CH_2$-.



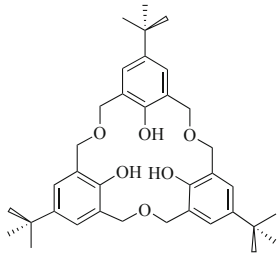

4-sulfonatocalix[4]arene



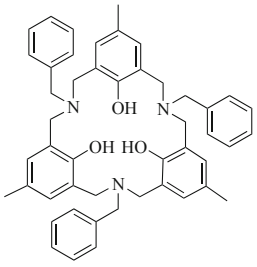

calix[5]arene




4-t-butyltetrathiacalix[4]arene




4-t-butyloxacalix[3]arene




4-methyl-(*N*-benzyl)azacalix[3]arene

**Fig. 1.9** Examples of macrocycles in the calixarene family

The reason that calixarenes have become so important in supramolecular chemistry lies in their cyclic structures. The smaller members of the group, in particular the calix[4]arenes, form several different conformers including a rigid cone where the phenol rings all have the same orientation. The central void is capable of encapsulating ions or small molecules while the upper and lower rims of the calixarene can be modified to inorporate metal binding groups. This suggested to those working in the field in the 1970s and 1980s that the compounds had the potential to make simple enzyme mimics. Indeed it was the search for an artificial aldolase mimic that sparked Gutsche's interest in calixarenes. To mimic this enzyme, responsible for the biosynthesis of fructose-1,6-diphosphate, he needed a scaffold that could support a

positively charged centre, a proton abstracting group and a hydrogen bond donor in a particular spatial arrangement. Polymers and peptides were linear and did not offer the desired disposition of functionalities. The crown ethers, first prepared by Pedersen in the 1960s, were too flexible and derivatization of cyclodextrins too challenging. By sheer luck Gutsche had been a consultant for the Petrolite Corporation with a particular focus on surfactants for the oil industry. One such compound that he was asked to investigate was a 4-t-butylphenol-formaldehyde polymer that precipitated from bulk solution. Noting a similarity between this material and one prepared by Zinke in the 1940s Gutsche believed that he had stumbled upon a macrocyle that could provide the scaffold for his aldolase mimic. Later work by his group would lead to a rational synthesis of 4-t-butylcalix[4]arene and other calixarenes which opened up the field to generations of synthetic chemists. While the goal of a calixarene-based aldolase mimic has yet to be realized many other applications for calixarenes and their derivatives have been found.

In addition to many other applications the calixarenes have been investigated as potential therapeutics: in the mid 1950s Cornforth investigated the apparent antitubercular activity of calixarenes with long alkyl chains attached. At the time the compounds were thought to be cyclotetramers, or calix[4]arenes, but later crystallographic determinations revealed them to be the larger calix[8]arenes. Crucially, these derivatives were able to span a cell membrane and had a central cavity that allowed the passage of water and small molecules through the membrane. Cornforth published his original work in 1955 [41] but the compounds he prepared, still stored in his laboratory, were recently reinvestigated with more modern analytical methods to determine their composition and mode of action [42]. Other applications of calixarenes as facilitators of transmembrane ion transport can be found in a later chapter.

Much of the attraction that supramolecular chemists have for calixarenes is due to their conformational mobility. The calix[4]arenes are particularly widely used as their conformers, illustrated in Fig. 1.10, interconvert until large groups are attached

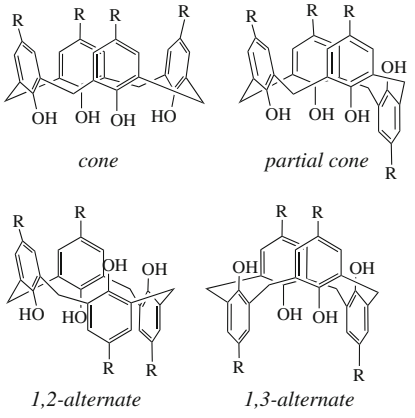
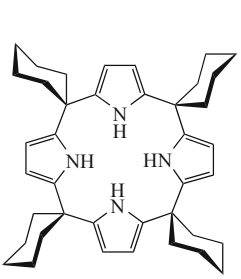



**Fig. 1.10** The four calix[4]arene conformers

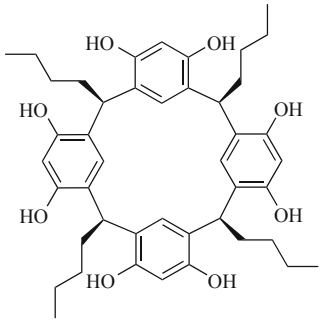
to the lower rim phenolic positions. This allows for geometrically varied products to be isolated from the same reaction.

### 1.5.2.2 Calixpyrroles and Resorcinarenes

In 1872 von Baeyer reported that the reaction between pyrroles and aldehydes generated cyclic compounds, now known as calixpyrroles by analogy to the calixarenes, and also that the reaction of resorcinol and range of aldehydes formed cyclotetramers, now known as resorcinarenes, shown in Fig. 1.11. Calixpyrroles have been of considerable interest of late as they are able to bind anions, with a particular preference for fluoride [43].




tetraspirocyclohexylcalix[4]pyrrole




tetrabutyl[4]resorcinarene

**Fig. 1.11** A calixpyrrole and resorcinarene

The resorcinarenes were investigated by Niederl and Vogel [44] before Högberg [45] developed the most reproducible synthetic routes. Subsequently the Cram group expanded the chemistry of these compounds by cross-linking phenolic oxygens with groups to make 'cavitands' that extended the molecules' 'walls' giving them greater depth [46]. By linking two resorcinarenes together it is possible to make a single molecule with a large cavity that can function as a molecular scale reaction flask.

### 1.5.2.3 Cyclotriveratrylene

Cyclotriveratrylenes are made by the condensation of veratrole with formaldehyde and were a curiosity when they were first prepared by Robinson in 1915 [47]. At the time she believed the compound to be a dimer, 2,3,6,7-tetramethoxy-9,10-dihydroxyanthracene. The reason for this was that the composition of the original compound was determined by elemental analysis. This technique determines the percentage of hydrogen and carbon in the sample which would be identical for a dimer, trimer or any other cyclic product with the same proportion of carbon to

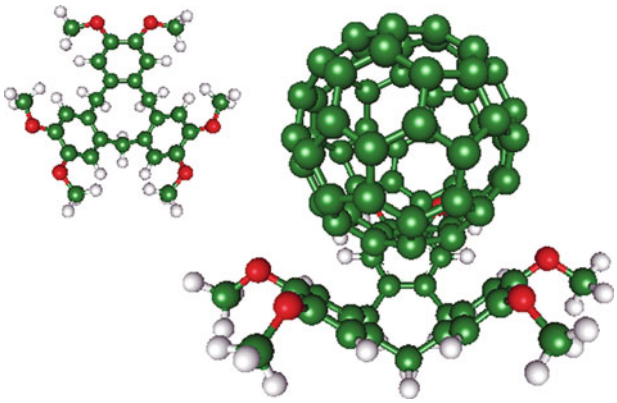


**Fig. 1.12** Cyclotriveratrylene (*top left*, not to scale) and its complex with $C_{60}$

hydrogen. Remarkably the true nature of the compound was only revealed following NMR analysis in 1965 [48] and the determination of its crystal structure in 1979 [49].

Cyclotriveratrylene synthesis is a poorly understood process, as with many reactions between phenols and aldehydes, and yields are as variable as the methods to prepare them: a review lists fifteen different conditions that give between 21 and 89% yield. Despite this the compounds are worth preparing as they have an interesting affinity for buckminsterfullerene, $C_{60}$, and are cited in papers and patents that describe methods to isolate pure $C_{60}$ from a mixture of fullerenes [50]. It transpires that the threefold symmetry of cyclotriveratrylene is complementary to the threefold axis of $C_{60}$ and that the two form very stable complexes in toluene as shown in Fig. 1.12, which precipitate leaving other fullerenes in solution. If the precipitate is isolated and taken up in chloroform the complex dissociates leaving cyclotriveratrylene in solution and precipitates $C_{60}$. The purity of the $C_{60}$ treated in this way is significantly enriched and can approach 100%. At the time of this discovery fullerene research was very much in its infancy, and the material available was of variable purity, making the purification technique an important milestone in the history of fullerene chemistry.

### 1.5.2.4 Cyclophanes

Compounds known as cyclophanes technically include the calixarenes and their relatives, as they consist of aromatic rings linked by short bridges. It is, however, a family of simpler molecules that is usually envisioned when using the cyclophane nomenclature. The first compound of this type, comprising two benzene rings linked through the 1,3-positions by ethyl bridges, was reported in 1899 [51] although it took a further five decades before a rational cyclophane synthesis was published. The 1949 method of Brown and Farthing [52] used to prepare the 1,4-linked [2,2]paracyclophane was followed, two years later, by an improved

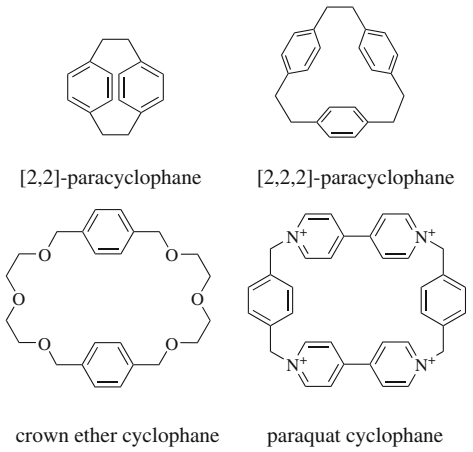



**Fig. 1.13** The cyclophane class of macrocycles

synthesis by Cram and Steinberg [53]. Numerous examples, such as those illustrated in Fig. 1.13, are now known.

### 1.5.2.5 Porphyrins and Phthalocyanines

The porphyrin motif, four pyrrole rings linked by carbon atoms to give a planar cyclic compound, is found in many biomolecules. It is present in haemoglobin and myoglobin and its derivatives are also at the core of vitamin $B_{12}$, hydrogenases, cytochrome c, the chlorophyll photosystem and similar essential proteins.

Phthalocyanines are analogues of the porphyrins in which the linking carbons have been replaced by nitrogens. While it is likely that one was first prepared by Braun and Tcherniac in 1907 [54], the first metal complexes were only reported in 1927 [55]. Despite this, their rediscovery by Dandridge, as described in a later chapter, is usually taken as the starting point for the field. As well as their original important role as pigments phthalocyanines have also been found to be valuable photosensitising agents with clinical applications in the treatment of skin cancer [56]. Examples of these compounds are illustrated in Fig. 1.14.

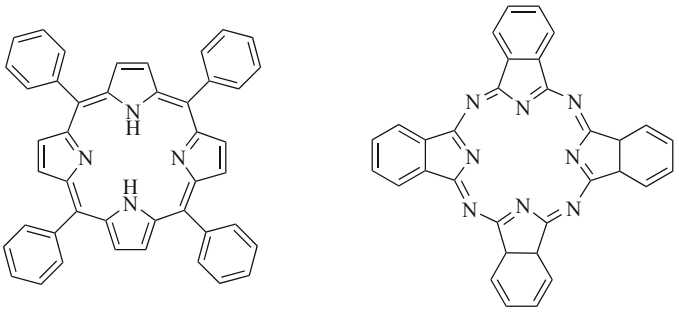


**Fig. 1.14** The porphyrin and phthalocyanine structural motifs

### 1.5.2.6 Crown Ethers

Crown ethers are cyclic molecules with a name to match their appearance. In their simplest form they are macrocycles consisting of repeating $–CH_2CH_2O-$ motifs. The 'crown' nomenclature describes how many atoms are linked in a ring and how many of these are atoms other than carbon. It is assumed that the heteroatoms are oxygens so, for example, a crown ether of 18 linked atoms, six of which are oxygen, would be [18]crown-6. An analogue of this compound incorporating benzene in place of a $–CH_2CH_2-$ link would be benzo[18]crown-6. If, instead, one of the oxygen atoms was replaced by nitrogen the compound would be aza[18]crown-6.

The discovery of crown ethers by Pedersen is a well known story in the world of supramolecular chemistry [57]. As with many great discoveries it resulted from an experiment that went wrong but, again in common with similar events, it took an extremely astute individual to understand the significance of the discovery and to exploit it.

Pedersen was working at Du Pont in the 1960s when he attempted the synthesis of bis[2-(hydroxyphenoxy)ethyl]ether from 2-(hydroxyphenoxy)tetrahydropyran. The starting material contained a trace of catechol which reacted to form a cyclic compound that incorporated two benzene rings and six oxygen atoms: dibenzo[18]crown-6. Pedersen was aware of earlier work on similar cyclic compounds by Lüttringhaus and others and also noticed the compound's ability to assist in the dissolution of simple salts in non-aqueous solvents. This application intrigued Pedersen who spent several years investigating the phenomenon before publishing his groundbreaking paper on the crown ethers, as he called them, in 1967 [58]. Common and systematic names are used to illustrate the usefulness of Pedersen's nomenclature in the examples shown in Fig. 1.15.

Although crowns containing benzene or cyclohexane were relatively easy to prepare, Pedersen was only able to isolate small amounts of the simpler crowns. Fortunately other researchers were fascinated by these new compounds and in a short period of time a high yielding synthesis of [18]crown-6 was reported by Greene [59] and the nitrogen containing analogue, aza[18]crown-6, was reported by Gokel and Garcia [60]. Since then crown ethers with variable ring sizes, heteroatoms and substituents have entered the literature.

### 1.5.2.7 Lariat Ethers

As their name implies the lariat ethers are crown ethers with pendent substituents and take their nomenclature from America's Old West. When a single side arm is introduced, either at a carbon or through the nitrogen atom of an azacrown, the resulting structure resembles the loop of a lasso, or lariat. The sidearm resembles the rope that would be held by a nanosized cowboy. Lariat ethers were pioneered by the group of Gokel who had realized that the attachment of a polyether chain to the crown ether would enhance guest binding [61]. In an elegant series of experiments

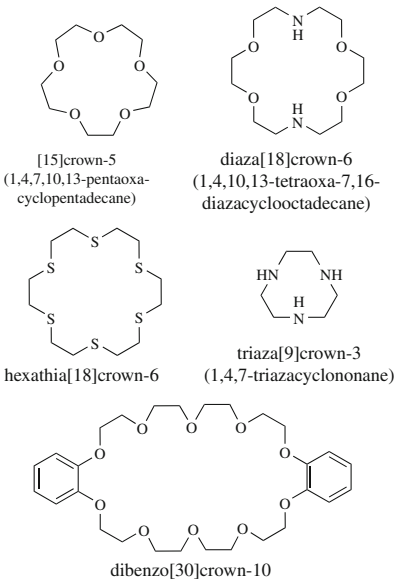



**Fig. 1.15** Examples of crown ethers

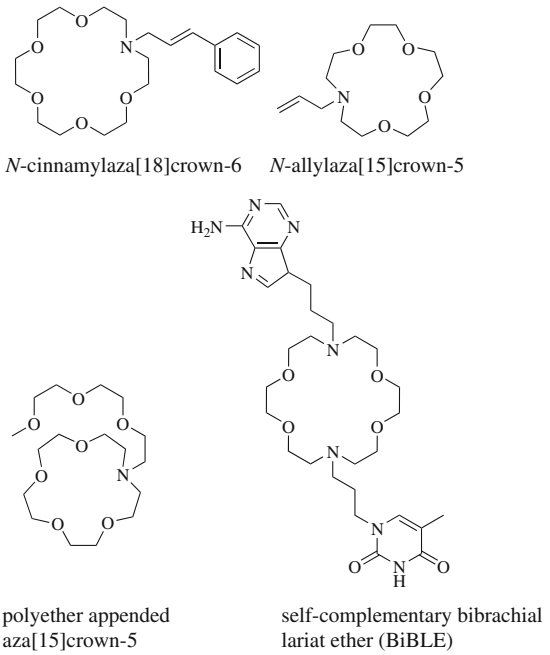
the group determined that the optimum length of the pendent polyether necessary to enhance cation binding was dependent on both accessibility to the metal and the steric effects of the side arm as shown in Fig. 1.16.

### 1.5.2.8 Cryptands, Sepulchrates and Sarcophagenes

If crown ethers can be thought of as two dimensional rings then it seems obvious to extend into a third dimension and synthesize macrocycles, or strictly speaking macrobicycles, that can encapsulate guests with greater selectivity. Such a feat was achieved by Simmons and Park [62] who prepared a 'katapinand' shortly after Pedersen had reported his crown ethers. Unlike the crowns, however, the katapinands comprised three alkyl chains extending from a nitrogen bridgehead to a second nitrogen atom. The three stranded compounds lacked the coordinating oxygen atoms present in crown ethers and consequently had little affinity for guest ions. The Lehn group made the obvious connection between the crowns and katapinands and reported the synthesis of 'les cryptates' in 1969 [63]. Specifically the authors introduced a group of compounds based on diaza[18]crown-6 in which the two nitrogen atoms were bridged by an additional polyether strand. The compounds are flexible enough to allow a guest cation to enter but the preferential orientation of oxygen and nitrogen atoms sets up an environment for the guest that makes it, like a crypt, almost impossible to leave.

The cryptates have since been joined by the equally gruesomely named sepulchrates and sacophagenes. Both were developed in the Sargeson laboratory in the

**Fig. 1.16** Lariat ethers




1970s [64] and are essentially all-nitrogen analogues of cryptands: the latter differ only in that the bridgeheads, where the three strands of the molecules meet, are carbon rather than nitrogen [65]. Examples of encapsulating molecules encountered in supramolecular chemistry are shown in Fig. 1.17.

In an interesting aside, many years after these synthetic capsules were first prepared, it was found that cells produce protein vaults that have similar encapsulating abilities but on a much larger scale. Whereas cryptands bind single alkali metal cations, vault proteins bind RNA composed of over 140 bases and other similarly sized guests. These 'biological crypts' are discussed later.

### 1.5.2.9 Cyclodextrins

The cyclodextrins are a group of cyclic sugars synthesized by enzymes in *Bacillus macerans*. The main members of the group are α-, β- and γ-cyclodextrin, comprising six, seven or eight sugars respectively as shown in Fig. 1.18, though smaller and larger homologues are known. All are water soluble but are able to secrete hydrophobic guests within their central cavities. The upper and lower rims are decorated with hydroxyl groups, some or all of which can be derivatized, which accounts for their solubility. The presence of a hydrophobic cavity, which must be entered through amphiphilic regions at each end, has led to numerous uses from timed release of fragrance molecules to absorption of volatile odoriferous molecules and other hydrophobic species [66].

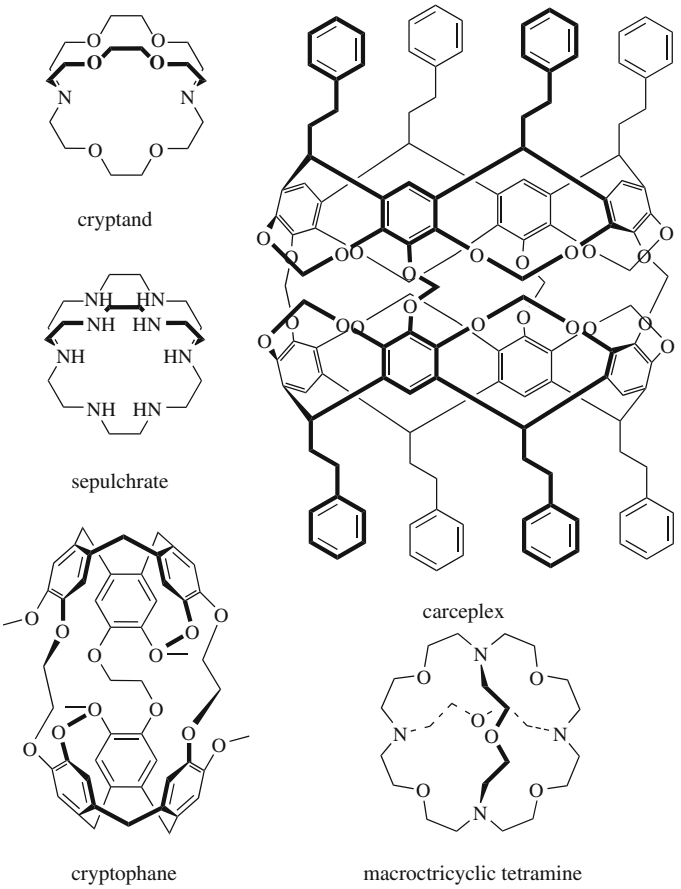



**Fig. 1.17** Encapsulating molecules

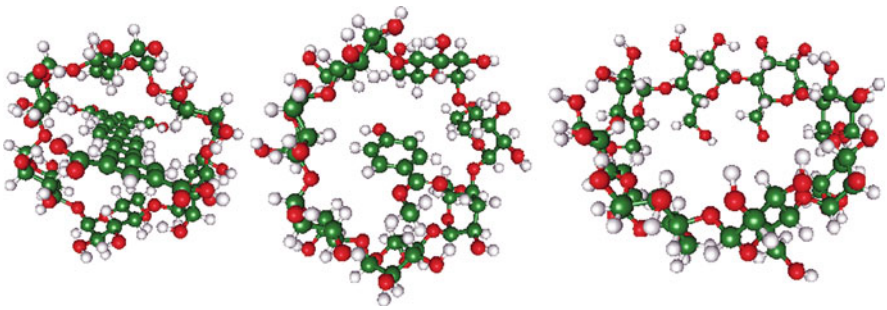


**Fig. 1.18** The cyclodextrins: (*left* to *right*) an α-cyclodextrin rotaxane [67], a β-cyclodextrin complex with methylparaben [68] and γ-cyclodextrin [69]

## 1.6 Supramolecular Entanglements

The undoubted strength of supramolecular chemistry is that it draws inspiration from the complex multicomponent structures observed in natural systems. One goal of supramolecular chemists is therefore to create models of biological entities that involve several interacting species. Many areas of biology are amenable to this reductionist approach though some are more complex than others. Simple ion recognition, as seen in natural ionophores, can be replicated quite well by crown ethers whereas inroads are just being made into the transduction of sunlight into electrical impulses, and it is likely to take a major concerted research effort for analogues of the protein-based molecular motors that power bacterial flagellae to be constructed by supramolecular chemists.

### *1.6.1 Catenanes and Rotaxanes*

In 1960 Wasserman reported the preparation of interlocked organic rings and named the complex a 'catenane' from the Latin for chain, *catena* [70]. In his paper he noted that other workers, notably Lüttringhaus [71], had attempted similar molecular entanglements but had been unsuccessful. Such entanglements are only possible because many of the key compounds used in supramolecular chemistry, such as polyethers, polyamines and crown ethers, are flexible. Furthermore they may also incorporate potential donor atoms, hydrogen bonding motifs, aromatic or hydrophobic groups that have affinities for complementary species. The resulting combination of flexibility and functionality can generate inter- and intramolecular attractions leading to entangled molecules such as the catenanes, and related rotaxanes, which are illustrated in Fig. 1.19.

The topological description of a catenane is an interlinked system where two or more cyclic molecules interpenetrate each other. Using this nomenclature a [2]catenane has two interpenetrating rings, a [3]catenane has three and so on. In an extreme example the Stoddart group synthesized a [5]catanane in 5% yield which was reported in 1994 following the Lillehammer winter Olympic games [72]. It was named Olympiadane following an earlier suggestion for the name of a [5]catenane by van Gulick who noted the resemblance to the symbol that had been used since the 1920 summer Olympic games held in Antwerp [73].

There are broadly two approaches to catenane synthesis. In both cases there needs to be a complementarity between the molecules being interlinked. Sauvage applied some basic coordination chemistry to the problem by modifying 1,10-phenanthroline to incorporate divergent phenolic groups which could be linked with a polyether to form a macrocycle. To bring two molecules together Sauvage introduced copper(I) to form the 2:1 ligand:copper complex in which two 1,10-phenanthroline derivatives bound in an orthogonal arrangement. Cyclization then occurred to give the bimetallic catenate in which a pair of macrocycles were irreversibly interlocked and bound to the copper centre. Copper could be removed with cyanide whereupon a conformational rearrangement occurred so that the

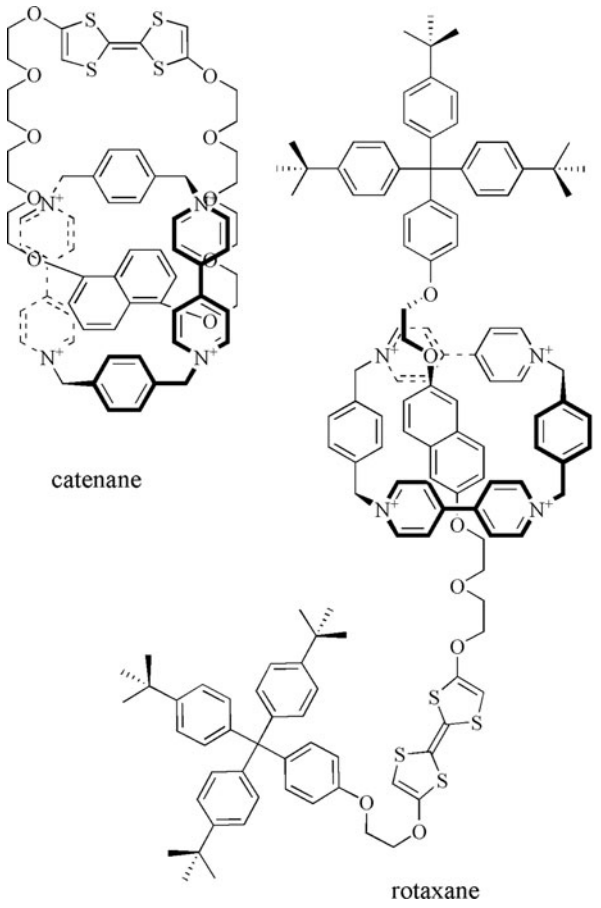



**Fig. 1.19** A catanane and rotaxane

phenanthrolines could separate. A more complex entanglement could be realized by linking two of the phenanthrolines together. Addition of copper(I) leads to a self-penetrating system, which can be unwound once the termini had been covalently linked, to reveal a molecular trefoil knot as shown in Fig. 1.20 [74]. The use of tetrahedral metals and extended bidentate ligands was also exploited by Lehn to form double helical complexes from linked bipyridyl ligands.

As a contrast to inorganic metal coordination the Stoddart group's approach used organic complementarity to hold components in place prior to cyclization. Large macrocycles were synthesized that incorporated polyethers and aromatic groups with π-stacking potential. An electron poor 4,4′-bipyridinium derivative, attracted to the latter, was held in place so that a small macrocycle could be formed around the polyether, through the reaction with 1,4-di(bromomethyl)benzene. The cyclization reaction results in a rigid electron deficient macrocycle that can spin around,

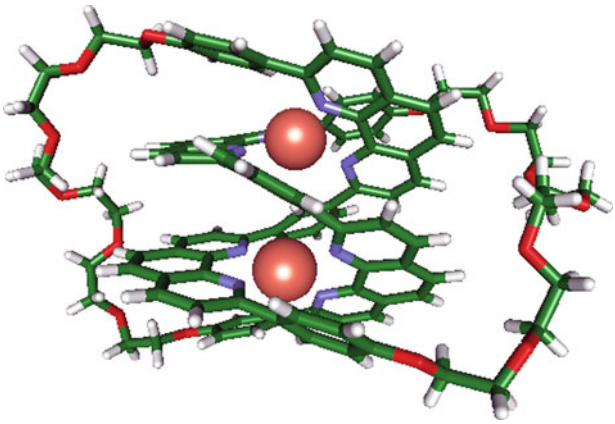


**Fig. 1.20** A supramolecular trefoil knot [74]

and move along, the benzocrown framework but cannot be removed as can be seen in Fig. 1.21.

Stoddart likened this to a train that travels around the crown ether track and stops when it enters an aromatic station because of the complementary π–π-interactions between the pyridinium groups in the train and the electron-rich aromatic stations [75]. While it was not always clear at the outset what use could be made of such esoteric molecules, interesting possibilities have since emerged. Electrochromic catenanes have been prepared that change colour when oxidized or reduced and can be programmed into one of several coloured electronic states. Because the molecules are flexible and form monolayers on the nanoscale it is possible to incorporate them into thin films that are red, green and blue. When the three films are layered together they respond to an applied voltage to give a colour display. The

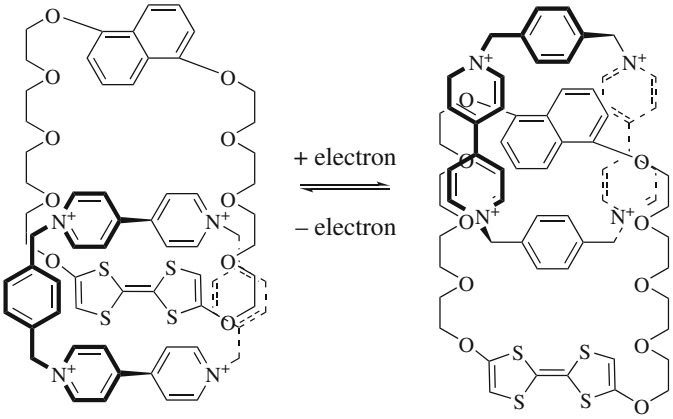



**Fig. 1.21** Catenane interconversions

display can be subdivided to contain addressable arrays of catenanes so that the resulting device can be used as an electronic page that can be refreshed whenever the external voltage changes [76].

Unlike the cyclic catenanes, rotaxanes are simpler species originally proposed by Wasserman and first demonstrated by Harrison [77]. In these systems a cyclic molecule is threaded on a rigid or flexible molecular axle, attracted by complementary binding sites, to form a pseudorotaxane. Under normal entropically driven supramolecular chemistry the cyclic component would eventually slip off one or other end of the central axle, however, it can be kept in place if both ends of the axle react with bulky groups while the macrocycle is still threaded. Alternatively a macrocycle can be formed around an axle molecule that already possesses bulky termini, as shown in Fig. 1.22. Either method leads to an entanglement in which the cycle species can move along the 'thread' without ever coming off.

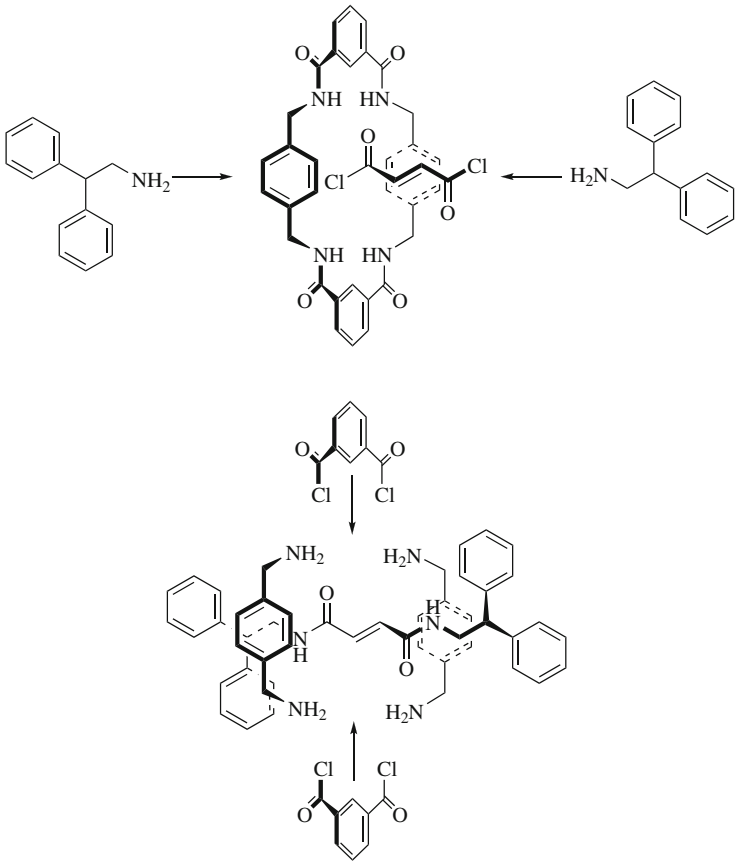



**Fig. 1.22** Two ways to assemble a rotaxane

Sauvage used his catenane method to thread a 1,10-phenanthroline derivative through a macrocycle formed from another, modified, 1,10-phenanthroline and a polyether. Again, copper(I) was used to hold the two components in place while two porphyrin rings were used to block the macrocycle's escape. The Stoddart group has exploited the rotaxanes' ability to respond mechanically to external stimuli in its work on molecular computers. Here they are used as bistable electronic components that perform as molecular switches. When the macrocyclic shuttle is energized it moves between two binding sites leading to two conductance levels. Programming high or low states into an array of rotaxanes allows them be used as computer memory devices [78]. Similarly, by linking shuttles together it is possible to construct a 'nano-elevator' [79]. Other rotaxanes have been prepared using a number of macrocyclic shuttles, including cyclodextrins and crown ethers, but most are based on molecules that incorporate rigid aromatic groups which can be manipulated by electronic means.

#### 1.6.1.1 Maxwell's Demon

One spectacular use of a rotaxane was to illustrate James Clerk Maxwell's 19th century thought experiment known as 'Maxwell's demon' as shown in Fig. 1.23. Maxwell proposed several experiments that would violate the Second Law of Thermodynamics to show how the entropy of an isolated system could be reduced without expending energy. To do so he invoked the idea of a demon who effortlessly operates a frictionless door between two compartments which contain particles at different temperatures. Whenever a particle approaches the door the demon decides whether to open it and allow the particle through. In this way the particles can be sorted so that one compartment contains only hot particles and the other only cold. A similar example can be envisaged in which particles at an equilibrium pressure are moved from one compartment to the other to increase its pressure without any work apparently being done.

As an example of a system related to the latter, the Leigh group prepared and tested a rotaxane with a thread containing two amine groups [80]. Between these sites there is a stilbene group that acts as a gate. The shuttle is a crown ether with an antenna group that can turn energy from photons into chemical energy that causes the stilbene to change conformation. The shuttle is threaded onto the rotaxane which is then stoppered at both ends. Both amine groups can be protonated to make them attractive stations for the crown ether shuttle. Initially the stilbene is in the *cis* form which, for steric reasons, stops the shuttle moving from one station to the other. When light shines on the shuttle it can transfer the energy to the stilbene gate which changes to the *trans* isomer and allows the shuttle to pass. Addition of benzil, an energy transfer agent, allows the gate to revert to the more stable *cis* form, trapping the shuttle on the other side. The station on this side is separated from the gate by a polyether spacer to make it harder for the shuttle to reverse its path. Consequently when the rotaxane is repeatedly irradiated the shuttles move in one direction only and are trapped following benzil addition. The compound therefore behaves as in

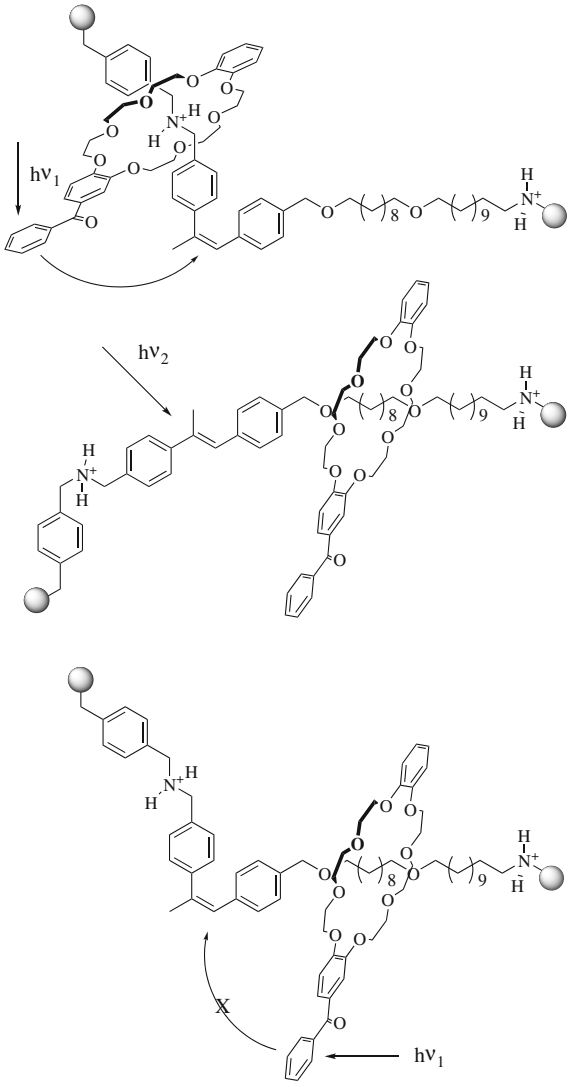



**Fig. 1.23** Maxwell's demon demonstrated by an irreversible rotaxane interconversion [80]

Maxwell's thought experiment, taking an equilibrium state and allowing it to spontaneously adopt a non-equilibrium state, though of course energy, in the form of light, and a demon, in the form of benzil molecules, have to be added. Thus it is not energy neutral and does not violate the Second Law.

### 1.6.1.2 Borromean Rings

As well as giving substance to a 19th Century thought experiment, this branch of chemistry can also produce entanglements that mirror some of humanity's earliest

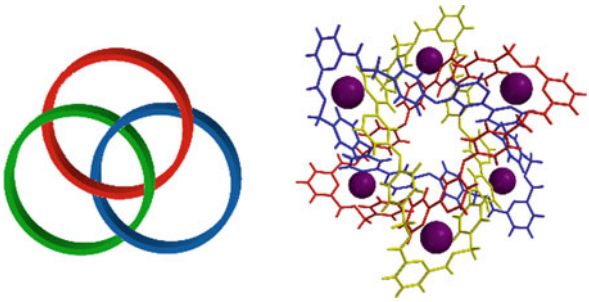


**Fig. 1.24** The Borromean motif (*left*) and a crystalline Borromean supramolecule (*right*) [83]

artistic motifs and religious symbols. The design of three interlocking rings, connected in such a way that should one be severed all three fall apart, is found in Norse, Christian and Hindu symbolism. The motif, shown in Fig. 1.24, was named the 'Borromean ring' following its prominent use as part of the Italian Borromeo family's coat of arms [81]. In essence the structure is a triply interlocked catenane and it can be constructed on the molecular scale. It has been observed in DNA but was thought to be too complex to synthesize by conventional means until the Seeman group's attempt [82].

Using a carefully considered, and distinctly unnatural, combination of transition metals and ligands, specifically 2,4-diformylpyridines and diamines containing bipyridyl binding sites, the Stoddart group was able to spontaneously form totally synthetic nanoscale Borromean rings. The interpenetrating structure was confirmed following crystallographic analysis by the Atwood group [83].

### 1.6.1.3 Natural Rotaxanes

Rotaxanes may appear to have no equivalents in biology but some of the most important proteins possess the same structural motif. For example, exo- and endonucleases have a toroidal structure through which DNA passes in a sequential manner and can also be considered as rotaxanes, or at least pseudorotaxanes.

The ribosome structure is perhaps the most impressive natural rotaxane. Messenger RNA, as the rotaxane thread, is clamped by the ribosomal protein subunits which read each codon and transcribe it into a sequence of amino acids that are introduced by transfer RNAs to build up the desired protein. The importance of the ribosomes has been reflected in the number of Nobel Prize recipients associated with their discovery (Palade in 1974) and elucidation of their structures and functions (Ramakrishnan, Steitz and Yonath in 2009).

α-Hemolysin is a transmembrane protein designed to disrupt cell membranes and initiate lysis. Ghadiri has shown that single strand DNA can be threaded through a heptameric α-hemolysin protein, derived from *Staphylococcus aureus*, that has been embedded in a lipid bilayer [84]. The DNA strands move through the central pore in response to an external voltage raising the fascinating possibility that the protein can be used to read the sequence of bases as the DNA rotaxane thread. Initial experiments show that polyadenine and polycytosine give significantly different

responses. Although much more work needs to be done to fine tune the response so that each individual base can be detected the rotaxane-based detection system has great potential. One of the main issues is the problem of differentiating the four bases. Fortunately many copies of the protein can be generated easily and placed in a bilayer so that the average response to identical DNA sequences should be amplified compared to the noise of incorrect detection or random base mutation. If this is achieved it will have a direct impact on the sequencing of entire genomes and cheap personal DNA sequencing.

### *1.6.2 Grids*

While supramolecular chemistry is often associated with organic synthesis, there are often occasions where a knowledge of simple inorganic coordination chemistry can be invaluable. One such case is the synthesis of supramolecular grids. Using a metal's preferred coordination environment to organize the relative position of ligands, as demonstrated in Sauvage's catenanes, can bring a powerful structural element to supramolecular chemistry. Catenane synthesis made use of the tetrahedral geometry generally adopted by copper(I) centres and this same preference can be exploited to prepare supramolecular grids. If a linear ligand containing numerous, evenly spaced metal binding sites is prepared then the addition of metals with a preference for a tetrahedral coordination environment should allow the spontaneous formation of a grid structure with metals binding ligands in two arrays orthogonal to each other. An excellent example can be found the work of the Lehn group which prepared a $3 \times 3$ grid composed of six linear ligands, each containing three convergent binding sites, and nine silver(I) cations [85]. Many transition metals are found in square planar and octahedral coordination geometries which also enables ligands to be positioned around the metals in parallel or perpendicular orientations. A good example of this is in the use of terpyridine derivatives to give meridional complexes. The motif can be extended to give polyfunctional ligands that can form discrete grid structures as demonstrated by the Lehn group, again, using a polyfunctional bis(terpyridine). In combination with several transition metals the ligand formed discrete 4:4 complexes as shown in Fig. 1.25. The complexes can be deposited as monolayers on a graphitic surface and, when analyzed by scanning tunnelling microscopy, reveal an extremely regular two dimensional packing arrangement. In a variation to this approach it is possible to prepare extended structures from two different ligands: one with multiple, convergent binding sites and one with two divergent sites. The second component can bridge between two metal-ligand strands so that the resulting complexes take the form of 'nano-ladders' [86].

### *1.6.3 Dynamic Combinatorial Libraries*

Many enzymes are known to fold into particular geometries in order to perform their catalytic functions. In some cases several proteins have to aggregate to form a very specific binding pocket. As there are many ways that proteins can fold they must be

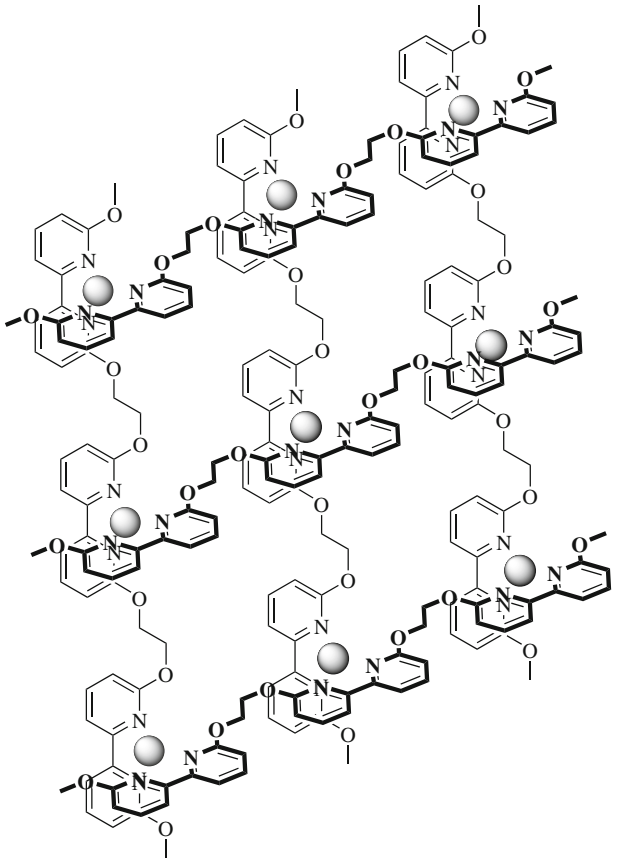


**Fig. 1.25** A supramolecular grid [85]

in a state of constant dynamic flux until the correct geometry is achieved. This often requires the template effects of small molecules involved in the catalytic process. Similar dynamic effects must also occur in processes such as DNA replication and protein synthesis by the ribosome. In supramolecular chemistry an analogue has been identified in dynamic combinatorial libraries. Here several molecules interact in a dynamic fashion until they form a stable supramolecular complex around a template [87].

The dynamic process is akin to the error checking mechanisms employed in protein synthesis: each reaction is reversible until the correct product has formed. In any evolutionary chemical system it is important to ensure copying fidelity and the success of dynamic combinatorial libraries indicates that concepts associate with supramolecular chemistry can be valuable in advancing chemical evolution.

## 1.7 Observing Supramolecules

How do we know when supramolecular interactions occur? Many of the standard analytical techniques used in synthetic chemistry can also be applied to supramolecular systems. Just as the formation of a new molecule can be followed by analysis of the bond-forming reactions between its precursors, so supramolecular products can be determined through analysis of the species formed as the components interact.

### *1.7.1 Isolation*

As with covalently linked products it may be possible to isolate supramolecules because they separate from the reaction mixture as precipitates, crystals or oils. Some chemical work up procedures, such as liquid-liquid extraction or chromatography, may be required to achieve separation but the methods are usually akin to those widely used to isolate reaction products. Here there is little difference between conventional and supramolecular chemistry, however, the larger size of the supramolecules may make them less amenable to some conventional methods, crystallization in particular.

### *1.7.2 Detection*

Detection of supramolecule formation relies on the same methods and instrumentation as is used for conventional molecules. Two main differences are encountered. By definition supramolecules are higher order aggregates of single molecules and are held together by non-covalent forces of variable strengths. If the interactions are very strong they often result in the formation of insoluble precipitates particularly if the supramolecule has a high molecular mass. Conversely, if the interactions are weak, the supramolecule may be quite fragile and easily destroyed. In either of these extreme cases analysis is difficult. Fortunately most supramolecular systems fall in between the two and can be analyzed by conventional techniques, albeit with some special considerations.

#### 1.7.2.1 Mass Spectrometry

Mass spectrometry is used to find the molecular mass of chemical samples that pass through the instrument. Using a range of instruments the technique can be used, at one extreme, to detect different isotopic ratios in low molecular mass gases such as $CO_2$, and therefore has found applications in archaeological radiocarbon dating, and, at the other end of the scale, can be used to accurately determine the masses of proteins. It is of great value in supramolecular chemistry as soft ionization methods can be used to detect weakly associated assemblies and their fragments. Classic

examples include the analysis of supramolecular capsules in which dimeric capsules containing guests can be identified as well as empty capsules and monomeric components.

#### 1.7.2.2 Infrared Spectroscopy

Infrared (IR) spectroscopy is a commonly used analytical technique in which a chemical sample is irradiated over a range of wavelengths. Different bonds absorb energy that can be recorded as a series of spectroscopic peaks. This technique is traditionally used to determine the presence of functional groups that have well known spectroscopic signatures, however, it can also be used to identify hydrogen bonds. The existence of strong hydrogen bonds is often evidence for supramolecular complex formation between molecules with complementary hydrogen bonding motifs. For example, Wolf's acetic acid *übermolecül* will be signalled by a much sharper peak due to the formation of the very strong self-complementary hydrogen bond network that arises when two molecules meet.

#### 1.7.2.3 Ultraviolet/Visible Spectroscopy

Ligands that contain conjugated organic groups usually absorb strongly in the ultraviolet or visible range of the electromagnetic spectrum. If these groups are involved in π-stacking or binding to a guest then the wavelength at which they absorb will be shifted in response to aggregation or host-guest binding. Excitation of the ligands and their complexes by an ultraviolet/visible (UV/vis) spectrophotometer and analysis of the absorption spectrum can give valuable information about supramolecular complex formation. The stoichiometry of host-guest complexes can be determined through a Job plot. Here a series of solutions are made up with varying ratios of host and guest from 0:1 through 0.5:0.5 to 1:0. A plot of relative absorbence at the wavelength of the peak associated with the complex reveals its stoichiometry. If the Job plot reaches a maximum at 0.33:0.67 the complex has a 1:2 stoichiometry, if the maximum is reached at 0.5:0.5 the stoichiometry will be 1:1 and so on. This method is most commonly used to identify transition metal complexes as these often have d–d electronic transitions at visible wavelengths that are strongly influenced by ligand interactions.

#### 1.7.2.4 Nuclear Magnetic Resonance Spectroscopy

Nuclear magnetic resonance (NMR) spectroscopy applies a changing magnetic field to a sample and records the absorption spectrum. While any nucleus that has one or more unpaired electrons is susceptible to this technique it is most commonly used to interrogate the hydrogen and carbon environments of a molecule. From NMR data it is possible to assign structural details through analysis of these environments. The method is complementary to IR spectroscopy, which reveals the functional groups, and mass spectrometry, which gives the molecular mass. Some aspects of supramolecular assembly, such as the ability to observe strong hydrogen bonds,

are obvious using basic experiments. Other applications allow the determination of host-guest binding by analyzing 'through space' interactions. Most NMR signals come from atoms that are related through bonds, such as hydrogen atoms on adjacent carbons, but through space interactions are indicated by shifts in the positions of spectral peaks that correspond to atoms of the host and guest that interact. The strength of the shifts can also reveal the orientation of the host-guest interaction and other useful information.

#### 1.7.2.5 X-Ray Crystallography

When compounds crystallize they do so because they can form stable repeating patterns in three dimensions. The minimal three dimensional repeating pattern is known as the unit cell and contains information about the relative positions of all the atoms within the crystalline structure. A unit cell may contain a single molecule or several that are related through symmetry. Furthermore, each molecule may have an internal plane of symmetry or similar geometric relationship. For example a crystal of [18]crown-6 could have a mirror plane so if the relative positions of half its atoms were known the structure of the entire molecule could be deduced. If the packing between adjacent molecules was known the unit cell could also be determined. This is precisely the information that can be found from a single crystal X-ray diffraction experiment. The method relies upon the fact that the wavelengths of X-rays are of a similar dimension to interatomic bonds and that these rays are diffracted by areas of electron density. In theory X-ray data can be collected from any regular crystal, thought the data may be of variable quality, so that unit cell sizes, molecular packing and relative atomic coordinates can be determined. Once a structure has been 'solved' it gives a three dimensional picture of the crystalline molecule or complex. This is particularly valuable in supramolecular chemistry where complexes may involve several components and solvent molecules. The complexity can also generate problems as large supramolecular systems with little or no internal symmetry require a large amount of data collection. If the molecules or complexes are flexible they may move over the course of the experiment so that some atomic positions are known with little certainty. This has been shown to be true of the contents of many supramolecular capsules particularly if they include volatile solvents. Despite these problems X-ray structures are sometimes the only way to characterize some complexes because of the difficulty in interpreting NMR or other data.

#### 1.7.2.6 Solution Techniques

In addition to spectroscopic techniques used to identify supramolecular behaviour in solution it is possible to observe molecular aggregation in solution and undertake analyses to give the mass distribution of the species present. Although not widely used in supramolecular chemistry, the technique of osmometry allows researchers to determine the masses of assemblies in solution and relate those to the masses of the single molecules to give aggregation numbers. It is of particular value in polymer science where it is used to determine the average mass, and mass distribution, of

polymers prepared by catalytic processes. Osmotic pressure is a colligative property of solutions and is therefore related to the number of particles present rather than any of their intrinsic properties. It can be measured with reference to boiling point elevation, freezing point depression or vapour pressure depression as all of these properties respond to changes in the number of species in solution. The osmotic pressure (Π) is related to molarity (M) through the expression:

$$\Pi = \mathrm{iMRT}$$

where i is the van t'Hoff factor, R is the gas constant and T the temperature in kelvin. From changes in the apparent molarity with changes in another parameter, such as temperature, the molar mass of the species in solution can be determined.

An alternative to osmometry is to determine particle size from the light scattering properties of the particles in solution. The technique of light scattering, again routinely used to characterize proteins and polymers, allows particle masses to be calculated based on their size and shape. Solutions are irradiated by coherent laser light which is scattered by the particles present: the diffraction angle and intensity of the scattered light is measured by a movable detector. The molecular masses of the particles are proportional to the intensity of the scattered light and can be calculated following a series of readings over a range of diffraction angles.

#### 1.7.2.7 Electrochemical Techniques

When macrocyclic compounds bind guests, particularly metals, there is usually a change in the ease with which redox chemistry occurs. This can be measured if a redox active function is coupled to the macrocycle and this is one way of turning a simple host-guest system into a valuable sensor. The classic redox reporting group ferrocene was used in this respect by the Beer group when it was linked through a conjugated spacer to a crown ether [88]. When a metal bound to the crown it shifted the ferrocene redox potential thus signalling the binding event. The electrochemical method used was cyclic voltammetry. In this technique a potential difference is applied to a solution containing the molecule of interest by a system comprising working, reference and auxiliary electrodes. The potential is varied smoothly between the system's resting potential, a predetermined limiting value, and back again. While this occurs the change in current is recorded to give a forward and reverse voltammogram. The process is repeated in the presence of guests to see how they affect the voltammogram. The simplest response would be a straightforward displacement of any existing potential peaks although it is not uncommon for new features to appear. One of the main attractions of this method is that it allows a direct calculation of the binding enhancement of the host for the guest through the Nernst equation. This can be expressed as:

$$\Delta \mathrm{E} = -\mathrm{RT/nF}\ [\ln K_1/K_2]$$

where $\Delta E$ is the difference between the redox potentials of the free host and its complex with the guest, $K_1/K_2$ is the binding enhancement ($K_1$ is the binding constant of the neutral complex, $K_2$ the binding constant of the reduced species), R is the gas constant, T the temperature in kelvin, and F is the Faraday constant.

#### 1.7.2.8 Molecular Modelling

Molecular modelling is not strictly an analytical tool that can be used directly. It is, however, a valuable way of visualizing supramolecular systems and predicting structures. The most sophisticated methods are able to predict properties associated with the model that can usefully be compared to data gathered on the real system. This is useful when several different interpretations of an experiment arise as one model may be shown to fit the data best and so be the most probable explanation. The main limitations of molecular modelling, and computational techniques in general, are the accuracy of the output and the the size of simulation that can usefully be attempted without recourse to a supercomputer or massively parallel facility.

Molecular mechanics (MM) methods are the simplest to use and can be used to model molecules up to the size of large proteins within a reasonable timeframe. The technique treats atoms as hard spheres that are joined by bonds that behave like mechanical springs. Using Hooke's law, or similar relationships between ideal and actual interatomic distances, a series of bond types are generated based on an optimum length between atoms. The values are determined by extensive analysis of spectroscopic and solid state information with the result that different parameters may be necessary to describe $sp^2$ carbon-nitrogen bonds and $sp^3$ carbon-nitrogen bonds. The exercise is repeated for all known interatomic interactions and is extended to include parameters for ideal bond angles, torsion angles, non-bonded interactions, charge-charge interactions and so on. The resulting 'forcefield' is used to generate an optimal geometry for the molecule that is being modelled. This is achieved when every interaction has been balanced against all others to give the lowest global energy structure. For various reasons, mostly dependent upon the algorithms used to navigate conformational space, the apparent lowest energy geometry may be at a local energy minimum and not represent the true energy minimum, and hence give an inaccurate structure. Other methods, notably Monte Carlo and annealed molecular dynamics, are available to search for conformers in other ways and can sometimes generate new geometries. Indeed, these methods can predict the existence of conformers that have yet to be observed experimentally [89].

Molecular mechanics simulations are useful methods when dealing with large molecules or when limited information is required. When more sophisticated analysis is desired, such as thermodynamic data, it is usually necessary to switch to ab initio methods that seek to solve, or approximate a solution to, the Schrödinger wave equation for the entire molecule. Programs are rapidly improving both in terms of time taken to generate a solution to the molecular orbital and in the size of molecule that can be analyzed by these methods. Despite these advances only the simplest of supramolecular systems can usefully be investigated at this level of

theory. Intermediate approaches, such as semiempirical methods that only consider valence electrons using high level theory, are fast but the quality of the resulting data often has to be questioned.

## 1.8 Summary

Supramolecular chemistry appears to have its origins in synthetic chemistry but its inspiration has often been biological. Common goals include mimicry of enzymes, where the selectivity and specificity of the host binding site must be optimized for the molecular guest, and the synthesis of molecules designed to bind to each other through complementary binding motifs, by analogy to protein-protein self-assembly and nucleic acid interactions. The molecular building blocks used to construct supramolecules can be likened to those found in biochemistry. There are macrocyclic hosts (crown ethers, calixarenes, phthalocyanines and others) that closely resemble the pyrrole-containing macrocycles at the active sites of enzymes, such as the cobalamins, and metal-containing biomolecules such as the porphyrins found in haemoglobin. The key interactions within supramolecular complexes are predominantly reversibly non-covalent in nature (hydrogen bonding, $\pi$–$\pi$ stacking, cation-$\pi$ interactions, hydrophoblic effects, etc). The same forces are seen extensively in biological macromolecules where they dictate the way proteins fold, cell membranes form, or DNA is copied and transcribed, via RNA, into protein sequences. The tools used to observe and interpret supramolecular behaviour are also common with many used in molecular biology. Whether mass spectrometry is used to determine the number of calixarenes in a cluster or the mass of a small protein, the technique is the same. Likewise, computer modelling, crystallography and NMR spectroscopy can be applied equally to biomolecules and artificial supramolecules. Subsequent chapters will show how key themes of complementarity, molecular recognition and host-guest complexation can bridge the disciplines of biology and chemistry. They will also demonstrate how an understanding of these phenomena enables chemists to design multifunctional molecules that can act as biological sensors, artificial cells and targeted therapeutics. They may even point towards the chemical origins of living systems.

## References

1. Lehn J-M (1978) Cryptates: inclusion complexes of macropolycyclic receptor molecules. Pure Appl Chem 50:871–892
2. Wolf KL, Frahm H, Harms H (1937) The state of arrangement of molecules in liquids. Z Phys Abt B 36:237–287
3. Pynchon T (1973) Gravity's Rainbow. The Viking Press, New York, NY, p 345
4. Century Dictionary, Supplement Vol XII (1909) The Century Company, New York, NY, p 1301
5. Baas-Becking LGM, Galliher EW (1931) Wall structure and mineralization in corraline algae. J Phys Chem 35:467–479

6. Palade GE (1955) A small particulate component of the cytoplasm. J Biophys Biochem Cytol 1:59–68
7. Mitchell P (1961) Coupling of phosphorylation to electron and hydrogen transfer by a chemi-osmotic type of mechanism. Nature 191:144–148
8. Luria SE (1970) Molecular biology: past, present and future. BioScience 20:1289–1296
9. Lehn J-M (2005) Quoted from a lecture given at the 30th international symposium on Macrocyclic Chemistry, Dresden
10. Fischer E (1894) Einfluss der configuration auf die wirkung der enzyme Ber Dtsch. Chem Ges 27:2985–2993
11. Fieser LE, Fieser M (1959) Steroids. Chapman & Hall, London, p 58
12. Lehn J-M (1988) Supramolecular chemistry – scope and perspectives. Molecules, supermolecules and molecular devices. Angew Chem Int Ed Engl 27:89–112
13. Drexler KE (1986) Engines of creation. Anchor Books New York, NY
14. Whitesides GM (2001) The once and future nanomachine. Sci Am 285:32–41
15. Wouters D, Schubert US (2004) Nanolithography and nanochemistry: probe-related patterning techniques and chemical modification for nanometer-sized devices. Angew Chem Int Ed Engl. 43:2480–2495
16. Saggimo V, Lüning U (2009) On the formation of imines in water – a comparison. Tetrahedron Lett 50:4663–4665
17. Morgan SG, Burnstall FH (1932) Dehydrogenation of pyridine by anhydrous ferric chloride. J Chem Soc 20–30
18. Sauvage J-P et al (1991) Photoinduced processes in dyads and triads containing a ruthenium(II) bis(terpyridine) photosensitizer covalently linked to electron-donor and acceptor groups. Inorg Chem 30:4230–4238
19. Barigelletti F et al (1993) Luminescence properties of rigid rod-like binuclear ruthenium(II)-osmium(II) terpyridine complexes – electronic interaction through phenyl bridges. JCS Chem Commun 942–944
20. Limburg J et al (2001) Characterization of the $O_2$-evolving reaction catalyzed by $[(terpy)(H_2O)Mn^{III}(O)_2Mn^{IV}(OH_2)(terpy)](NO_3)_3$ (terpy = 2,2′:6,2″-terpyridine). J Am Chem Soc 123:423–430
21. Fujita M et al (1994) Preparation, clathration ability, and catalysis of a 2-dimensional square network material composed of cadmium(II) and 4,4′-bipyridine. J Am Chem Soc 116: 1151–1152
22. Caulder et al (1998) The self-assembly of a predesigned tetrahedral $M_4L_6$ supramolecular cluster. Angew Chem Int Ed Engl 37:1840–1843
23. Leininger S, Olenyuk B, Stang PJ (2000) Self-assembly of discrete cyclic nanostructures mediated by transition metals. Chem Rev 100:853–907
24. Roche S et al (1998) Self-assembly of a supramolecular cube. Chem Commun 1681–1682
25. Hannon MJ et al (1997) An inexpensive approach to supramolecular architecture. Chem Commun 1807–1808
26. Pearson RG (1963) Hard and soft acids and bases. J Am Chem Soc 85:3533–3539
27. Constable EC (1992) Oligopyridines as helicating ligands. Tetrahedron 48:10013–10059
28. Bell TW, Jousselin H (1994) Self-assembly of a double-helical complex of sodium. Nature 367:441–444
29. Bell TW et al (1992) A supramolecular assembly of two torands, two lithium ions and three water molecules. Angew Chem Int Ed Engl 31:348–350
30. Schiff H (1864) Mittheilungen aus dem Universitätslaboratorium in Pisa: Eine neue Reihe organischer Basen Justus Liebigs. Ann Chem 131:118–119
31. Tummler B et al (1979) Open-chain polyethers – influence of aromatic donor end groups on thermodynamics and kinetics of alkali-metal ion complex-formation. J Am Chem Soc 101:2588–2598
32. Weber E, Vögtle F (1975) Crystal 1.1-alkali metal-complexes of non-cyclic neutral ligands. Tetrahedron Lett 29:2415–2418

33. Schmaltz et al (2001) Double stranded interwound infinite silver coordination network. Chem Commun 1242–1243
34. Abouderbala LO et al (2002) Cooperative anion binding and electrochemical sensing by modular podands. Proc Nat Acad Sci USA 99:5001–5006
35. Liu S et al (1992) Synthesis and characterization of lanthanide $[Ln(L)]_2$ complexes of $N_4O_3$ amine phenol ligands with phenolate oxygen bridges – evidence for very weak magnetic exchange between lanthanide ions. J Am Chem Soc 114:6081–6087
36. Hossain MA et al (2004) Anion binding with a tripodal amine. Inorg Chem 43:3751–3755
37. von Baeyer A (1872) Uber die Verbindung der Aldehyde mit den Phenolen. Ber Dtsch Chem Ges 5:25–26
38. Zinke A, Ziegler E (1944) Zur Kenntnis des Härtungsprozesses von Phenol-Formaldehyd-Harzen X. Ber 77B:264–272
39. Hayes BT, Hunter RF (1958) Phenol formaldehyde and allied resins VI: rational synthesis of a 'cyclic' tetranuclear *p*-cresol novolak. J Appl Chem 8:743–748.
40. Gutsche CD et al (1981) Calixarenes 4. The synthesis, characterization and properties of calixarenes from para-tert-butylphenol. J Am Chem Soc 103:3782–3799
41. D'Arcy Hart P et al (1955) Antituberculous effects of certain surface-active polyoxoethylene ethers. Br J Pharmacol 10:73–86
42. Colston MJ et al (2004) Antimycobacterial calixarenes enhance innate defense mechanisms in murine macrophages and induce control of *Mycobacterium tuberculosis* infection in mice. Infect Immun 72:6318–6323
43. Sessler JL, Camiolo S, Gale PA (2003) Pyrrolic and polypyrrolic anion binding agents. Coord Chem Rev 240:17–55
44. Niederl JB, Vogel H (1940) Aldehyde-resorcinol condensations. J Am Chem Soc 62: 2512–2514
45. Högberg AGS (1980) Two stereoisometric macrocyclic resorcinol-aldehyde condensation products. J Org Chem 45:4498–4500
46. Tunstad LM et al (1989) Host-guest complexation 48. Octol building blocks for cavitands and carcerands. J Org Chem 54:1305–1312
47. Robinson GM (1915) A reaction of homopiperonyl and of homoveratryl alcohols. J Chem Soc Trans 107:267–276
48. Miller B Gesner BD (1965) Conformation of cyclotriveratrylene by nuclear magnetic resonance measurements. Tetrahedron Lett 6:3351–3354
49. Cerrini S et al (1979) The crystal structure of the inclusion compound between cycloveratril, benzene and water. Acta Cryst B 35:2605–2609
50. Atwood et al (1996) Cyclotriveratrylene polarisation assisted aggregation of $C_{60.}$ Chem Commun 1449–1450
51. Pellegrin MM (1899) Contribution à l'études de la réaction de Fittig Recl Trav. Chim Pays-Bas 18:457–465
52. Brown CJ, Farthing AC (1949) Preparation and structure of di-p-xylene. Nature 164: 915–916
53. Cram DJ, Steinberg H (1951) Macro rings I. Preparation and spectra of the paracyclophanes, J Am Chem Soc 73:5691–5704
54. Braun A, Tcherniac J (1907) Über die Produkte der Einwirkung von Acetanhydrid auf Phthalamid. Ber Dtsch Chem Ges 40:2709–2714
55. Dandridge AG, Drescher HAE, Scottish Dyes Ltd (1928) Improvements in and relating to the manufacture and use of colouring materials. GB patent 322169
56. Tedesco AC, Rotta JCG, Lunardi CN (2003) Synthesis, photophysical and photochemical aspects of phthalocyanines for photodynamic therapy. Curr Org Chem 7:187–196
57. Chemical crowns and crypts. In: Roberts RM (ed) Serendipity: accidental discoveries in science, 1989, 1st edn. Wiley, New York, NY, pp 238–243

58. Pedersen CJ (1967) Cyclic polyethers and their complexes with metal salts. J Am Chem Soc 89:7017–7036
59. Greene RN (1972) 18-Crown-6:a strong complexing agent for alkali metal cations. Tetrahedron Lett 13:1793–1796
60. Gokel GW, Garcia BJ (1977) Crown-cation effects III. Chemistry and complexes of monoaza-18-crown-6. Tetrahedron Lett 18:317–320
61. Shultz et al (1985) 12-, 15-, and 18-Membered-ring nitrogen-pivot lariat ethers: syntheses, properties, and sodium and ammonium cation binding properties. J Am Chem Soc 107: 6659–6668
62. Simmons HE, Park CH (1968) Macrobicyclic amines I. Out-in isomerism of (l,k+2)diazabicyclo[k.l.m]alkanes. J Am Chem Soc 90:2428–2429
63. Dietrich B, Lehn J-M, Sauvage J-P (1969) Les cryptates. Tetrahedron Lett 10:2889–2892
64. Creaser II et al (1977) Sepulchrate: a macrobicyclic nitrogen cage for metal ions. J Am Chem Soc 99:3181–3182
65. Geue RJ et al (1984) Metal-ion encapsulation: cobalt cages derived from polyamines, formaldehyde, and nitromethane. J Am Chem Soc 106:5478–5488
66. Saenger W (1980) Cyclodextrin inclusion compounds in research and industry. Angew Chem Int Ed Engl 19:344–362
67. Terao J et al (2004) Synthesis of poly(para-phenylenevinylene) rotaxanes by aqueous Suzuki coupling. Chem Commun 56–57.
68. Caira MR, de Vries EJC, Nassimbeni LR (2003) Crystallization of two forms of a cyclodextrin inclusion complex containing a common organic guest. Chem Commun 2058–2059
69. Harata K (1987) The structure of the cyclodextrin complex 20. Crystal-structure of uncomplexed hydrated gamma-cyclodextrin. Bull Chem Soc Jpn 60:2763–2767.
70. Frisch HL, Wasserman E (1961) Chemical topology. J Am Chem Soc 83:3789–3795
71. Lüttringhaus A et al (1958) Cyclisationen von Langkettigen Dithiolen. Versuche zur Darstellung sich umfassender Ringe mit Hilfe von Einschlußverbindungen. Ann 613:185–198
72. Amabilino DB et al (1994) Olympiadane. Angew Chem Int Ed Engl 33:1286–1290
73. van Gulick N (1993) Theoretical aspects of the linked ring problem. New J Chem 17: 619–625
74. Dietrich-Buchecker CO, Sauvage J-P (1989) A synthetic molecular trefoil knot. Angew Chem Int Ed Engl 28:189–192
75. Ashton PR et al (1991) Molecular trains – the self-assembly and dynamic properties of 2 new catenanes. Angew Chem Int Ed Engl 30:1042–1045
76. Ono T et al (2007) Optodynamers: expression of color and fluorescence at the interface between two films of different dynamic polymers. Chem Commun 4360–4362
77. Harrison IT, Harrison S (1967) Synthesis of a stable complex of a macrocycle and a threaded chain. J Am Chem Soc 89:5723–5724
78. Bissell RA et al (1994) A chemically and electrochemically switchable molecular shuttle. Nature 369:133–137
79. Badjić JD et al (2004) A molecular elevator. Science 303:1845–1849
80. Serreli V et al (2007) A molecular information ratchet. Nature 445:523–527
81. Cromwell PR, Beltrami E, Rampichini M (1998) The Borromean rings. Math Intelligencer 20:53–62
82. Mao CD, Sun WQ, Seeman NC (1997) Assembly of Borromean rings from DNA. Nature 386:137–138
83. Chichak et al (2004) Molecular Borromean rings. Science 304:1308–1312
84. Cockroft SL et al (2008) A single-molecule nanopore device detects DNA polymerase activity with single-nucleotide resolution. J Am Chem Soc 130:818–820
85. Baxter PNW et al (1994) Self-assembly and structure of a 3 × 3 inorganic grid from 9 silver ions and 6 ligand components. Angew Chem Int Ed Engl 33:2284–2287

86. Reuben M, Lehn JM, Muller P (2006) Addressing metal centres in supramolecular assemblies. Chem Soc Rev 35:1056–1067
87. Lehn JM (1999) Dynamic combinatorial chemistry and virtual combinatorial libraries. Chem Eur J 5:2455–2463
88. Beer PD et al (1990) Redox-responsive crown ethers containing a conjugated link between the ferrocene moiety and a benzo crown ether. Inorg Chem 29:378–381
89. Beech J, Cragg PJ, Drew MGB (1994) Conformational study of the macrocycle 1,4,7-trithiacyclononane in metal complexes. J Chem Soc Dalton Trans 719–729

# Chapter 2
# Supramolecular Chemistry and the Life Sciences

## 2.1 Life as a Supramolecular Phenomenon

To understand the significance of supramolecular phenomena in the chemistry of life it is first necessary to consider what constitutes a living system. At its most basic, 'life' requires that the system in question is able to exist for a time by regenerating its essential components. It should also be able to produce copies of itself through replication and be able to evolve so that the copies can respond to, and survive, changes in external conditions. Associated with this is the concept of 'negative entropy'. Negative entropy was proposed by Schrödinger in his 1944 book 'What is Life?' [1]. Ordinarily, matter and energy tend to degrade over time: heat dissipates, molecules fall apart. In the process of life this is reversed as resources are stored and value is added to self-regulating systems. Consequently local entropy is reduced at the expense of its surroundings as the system exudes waste in the form of unwanted by-products. Nevertheless, in accordance with the Second Law of Thermodynamics, the overall entropy of the Universe will eventually increase when the system ceases to live and is eventually reduced to its constituent parts, releasing the energy it has stored. The above definition, and expansions of it given below, is at best highly simplistic. Defining life is analogous to holding a handful of water and using it to explain the flow of the river from which it came.

Observation of living systems shows them to be complex mixtures of different chemicals each of which functions to support one or more of the necessary tasks required to keep the organism 'alive'. At one end of the spectrum of complexity lie the viruses and related simple proteins, such as prions, which cannot replicate by themselves but reproduce with the aid of other, higher, life forms. They therefore fulfill the definition of 'life' to a very limited extent. At the other end of the spectrum there are the multicellular organisms such as humans with a vast diversity of processes that relate directly, and often more subtly, to the basic requirements of life. Even here, though, there is often a dependence on another living system. For example, humans cannot survive without vitamin $B_{12}$ but are unable to synthesize it and must rely on external agents for its production.

What does supramolecular chemistry have to do with life? This chapter explores the diversity of chemical species that are seen in living systems and those that

P.J. Cragg, *Supramolecular Chemistry*, DOI 10.1007/978-90-481-2582-1_2,

© Springer Science+Business Media B.V. 2010

may have been involved in the origin of life. A surprising number of these species rely on non-covalent, or reversible covalent, interactions now associated with supramolecular chemistry. As will be shown, this leads to an overarching concept of 'supramolecular biology' which links the almost infinite diversity of biological systems to the basic concepts of supramolecular chemistry.

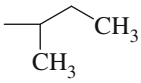
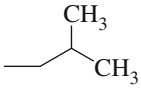
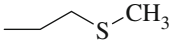
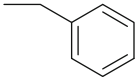
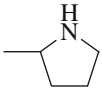
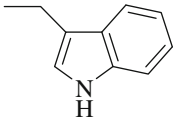
## 2.2 Supramolecular Interactions in Biological Systems

To understand the relationship between supramolecular chemical principles and their biological influences, and the reciprocal effect this has had on the application of supramolecular chemistry to challenges in biology, it is necessary to give a brief overview of some basic biological systems. There are many basic biological building blocks that combine to form larger structures and, ultimately, living organisms. Those with the greatest similarities to components in supramolecular chemical systems are discussed below.

### *2.2.1 Amino Acids*

Amino acids are the basic units that join together to form proteins. Each amino acid has an amine group and a carboxylic acid connected by a carbon atom with a side chain giving a generalized structure of $H_2NCHRCO_2H$. The central carbon atom confers the property of chirality, or 'handedness', to the amino acid. Each molecule exists in one of two geometric arrangements, mirror images of each other, that cannot be superimposed. For reasons which are not understood a vast majority of amino acids in Nature are in the so called L-form, a notation derived from laevorotatory meaning 'to rotate to the left'. In the context of amino acids this means that they rotate plane polarized light in an anticlockwise manner. Chirality is not actually related to the amino acid but to its geometrical similarity to the analogous *S*- or *R*-glyceraldehyde derivative; the former is laevorotatory, the latter, dextrorotatory. Most amino acids follow the rule though there are a few exceptions. Glycine, the simplest amino acid, has no side chain as the link is a simple $–CH_2-$, or methylene, group. By definition a methylene group has no chirality. The other exceptions are those amino acids containing sulfur, cysteine and methionine, together with their rare selenium analogues. The convention for naming chiral compounds depends on the relative masses of the groups attached to the chiral carbons and, because of the high mass of sulfur relative to carbon, nitrogen and oxygen, this results in the reversal of S and R descriptions for sulfur containing amino acids. Naturally occurring cysteine has the same 'handedness' as the L-amino acids but in this case naming conventions confusingly relate to the R form rather than the S. Although there are many known amino acids only 20, shown in Tables 2.1 and 2.2, are common. These can be divided into non-essential and essential groups. Mammals are able to biosynthesize the former but must rely on other sources for the rest. As a consequence,

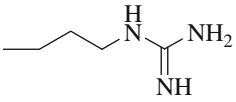
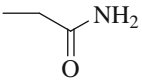
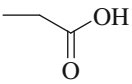
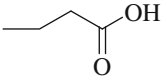
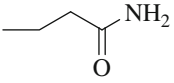
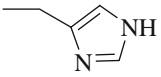
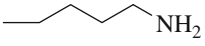
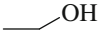
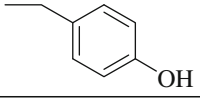
**Table 2.1** The common non-polar amino acids

| Name | Sidechain structure | Three letter code | One letter code |
|---|---|---|---|
| Alanine |  | Ala | A |
| Cysteine |  | Cys | C |
| Glycine |  | Gly | G |
| Isoleucine |  | Ile | I |
| Leucine |  | Leu | L |
| Methionine |  | Met | M |
| Phenylalanine |  | Phe | F |
| Proline |  | Pro | P |
| Tryptophan |  | Trp | W |
| Valine |  | Val | V |

humans can only synthesize ten of the common amino acids though others are simple derivatives of these ten. Essential amino acids can be accumulated from dietary sources or through symbiotic relationships with other organisms. For higher species this could be certain bacteria or microflora in the gut.

Amino acid side chains play a vital role in the shapes of proteins. They vary in their hydrophobicities and charges at physiological pH. Thus, when incorporated into proteins, hydrophobic amino acids that form part of membrane-bound proteins

**Table 2.2** The common polar amino acids

| Name | Sidechain structure | Three letter code | One letter code | Charge (pH 7.4) |
|---|---|---|---|---|
| Arginine |  | Arg | R | Positive |
| Asparagine |  | Asn | N | Neutral |
| Aspartic acid |  | Asp | D | Negative |
| Glutamic acid |  | Glu | E | Negative |
| Glutamine |  | Gln | Q | Neutral |
| Histidine |  | His | H | Neutral |
| Lysine |  | Lys | K | Positive |
| Serine |  | Ser | S | Neutral |
| Threonine |  | Thr | T | Neutral |
| Tyrosine |  | Tyr | Y | Neutral |

tend to bind to lipid membranes or in pockets within proteins that are protected from the external aqueous environment. Conversely, side chains incorporating amines or carboxylic acids that may be ionized or have water-soluble functional groups are likely to be found on the exterior of water-soluble proteins. A computer program that gives a measure of hydrophobicity was devised by Kyte and Doolittle [2] and has been used to determine the hydropathy index not only for individual amino acids

but, more importantly, for entire proteins. These weak, non-covalent interactions are familiar motifs in supramolecular chemistry.

Biosynthesis of amino acids is by one of several broad pathways as shown in Fig. 2.1. They are formed by enzyme-assisted reactions of common metabolites. By these routes, oxaloacetate yields aspartate and asparagine, and α-ketoglutarate

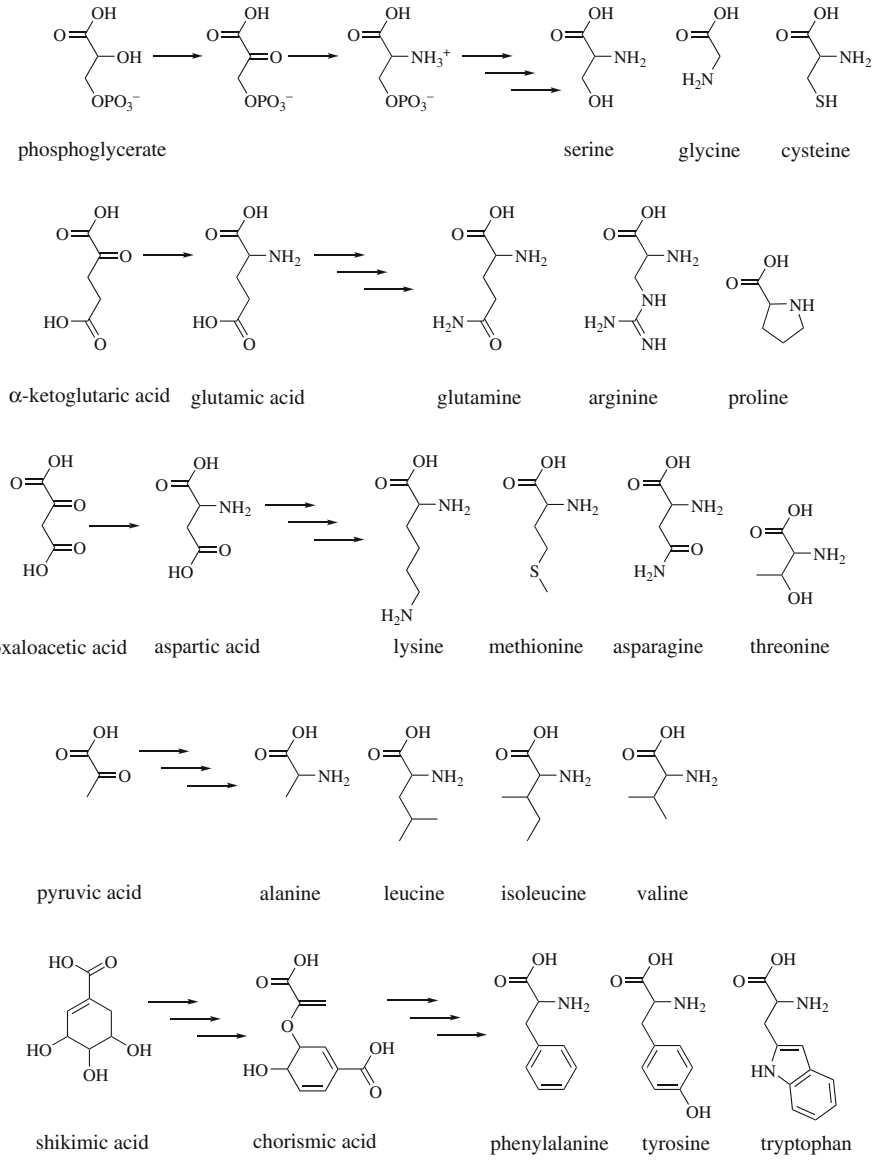



**Fig. 2.1** Some biosynthetic routes to amino acids

yields glutamine. The intermediate in glutamine synthesis, glutamate, is also used in the biosynthesis of proline and arginine. 3-Phosphoglycerate is biotransformed into serine, cysteine and glycine. Other essential amino acids fall into those derived from aspartate (lysine, methionine and threonine), pyruvate (alanine, leucine, isoleucine and valine), and chorismate (tryptophan, tyrosine and phenylalanine). Histidine is an oddity in that it is synthesized by a similar pathway to nucleotides, its imidazole ring motif is found in the purines, adenine and guanine.

### *2.2.2 Proteins*

If amino acids are the biological equivalents of words then proteins are biological sentences. The carboxylic acid groups and amines of amino acids can react to form a peptide bond, a single bond between the carboxylate carbon and nitrogen, but do not do so spontaneously. Bond formation, and elimination of water, can only occur by activating the carboxylic acid. In Nature this is facilitated in the ribosome by amino acids appended to RNA, an example of which is shown in Fig. 2.2. Amino acids fit into pockets to be manipulated by changes in the hydrogen bond environment and, in the case of amino acids with aromatic side chains, π–π interactions. Sequential reactions of amino acids results in protein formation as a linear polymer emerging from the ribosome. Synthesis is initiated when the ribosome is given a 'start' signal from mRNA. This 'start' signal is the adenine-uracil-guanine (AUG) triplet that codes for methionine. Subsequent amino acids are added to the carboxylic acid of the previous one such that the process ends with a free carboxylic acid, and each

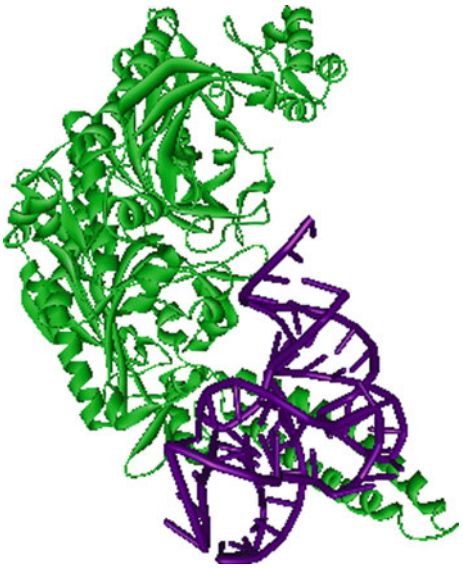


**Fig. 2.2** Protein biosynthesis: serine tRNA (in *purple*) coupled to tRNA serine synthase [3]

**Table 2.3** Codons for amino acids

| Codon | Anino acid | Codon | Amino acid |
|---|---|---|---|
| AUG | None (start) | UAA, UAG, UGA | None (stop) |
| GCA, GCC, GCG, GCU | Ala | AGA, AGG, CGA, CGC, CGG, CGU | Arg |
| UGC, UGC | Cys | AAC, AAU | Asn |
| GGA, GGC, GGG, GGU | Gly | GAC, GAU | Asp |
| AUA, AUC, AUU | Ile | GAA, GAG | Glu |
| CUA, CUC, CUG, CUU, UUA, UUG | Leu | CAA, CAG | Gln |
| AUG | Met | CAC, CAU | His |
| UUC, UUU | Phe | AAA, AAG | Lys |
| CCA, CCC, CCG, CCU | Pro | AGC, AGU, UCA, UCC, UCG, UCU | Ser |
| UGG | Trp | ACA, ACC, ACG, ACU | Thr |
| GUA, GUC, GUG, GUU | Val | UAC, UAU | Tyr |

one is chosen by reference to a three nucleotide sequence, or codon, as shown in Table 2.3. The amine end of the protein is called the N-terminus; the carboxylic acid end is the C-terminus. The order, or sequence, in which the amino acids are added to form a protein is its primary structure.

The amide bond formed between amino acids is rigid and planar. This constrains the hydrogen attached to the nitrogen and the adjacent carbonyl oxygen to point in opposite directions. However, the backbone is able to rotate freely around each chiral carbon atom which allows proteins to adopt a multitude of shapes. These are largely dictated by two factors: hydrogen bond formation between amide groups and the nature of the amino acid side chains. The former gives rise to two common structural motifs responsible for a protein's secondary structure; α-helices and β-sheets, illustrated in Fig. 2.3. Interactions between side chains, through electrostatic effects, hydrogen bonding, hydrophobic effects or disulfide bridge formation, are responsible for tertiary structures. These could relate, for example, to the packing of several α-helical domains together to form a cylindrical region of the protein. Finally there is the quaternary structure. This is the term used to refer to the structure formed when several proteins aggregate to generate a discrete functional unit. Examples of this are haemoglobin, an $\alpha_2\beta_2$ tetramer composed of myoglobin-like proteins, and the homotrimeric or tetrameric structures of some transmembrane proteins that function as ion channels. One such channel-forming protein complex that is responsible for $K^+$ crossing cell membranes is illustrated in Fig. 2.4.

Protein-protein interactions are of immense importance and yet are not well understood. It is probable that these involve the same principles as appear to regulate enzyme activity. The first attempt to understand how protein enzymes can have such great specificity for their targets was proposed by Fischer in 1894 in his conceptually groundbreaking 'lock and key' hypothesis [5]. Over half a century later Koshland proposed a modification to this theory arguing that it did not adequately explain the conformational rearrangements that must occur as the enzyme binds the target

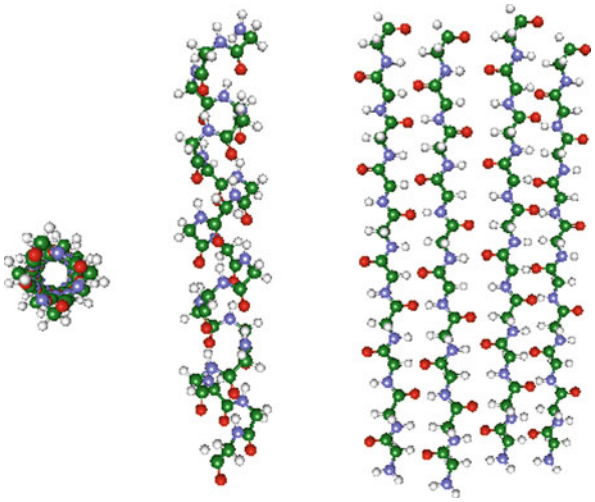


**Fig. 2.3** The main hydrogen bonding motifs in proteins: end- and side-on views of an α-helix (*left*) and the β-sheet (*right*)

molecules, induces a more favourable reaction environment (by reducing the energy required to reach the transition state), and finally releases the product. This 'induced fit' mechanism [6] has been a valuable extension of Fischer's original theory and certainly works well as a model. Improvements in analytical methods, particularly protein structure determination by NMR and prediction by computational methods, have allowed for far more extensive examinations of protein conformations over different timescales [7]. Studies of calmodulin and ubiquitin have shown that all the conformers adopted by the proteins when bound in protein-protein ensembles are

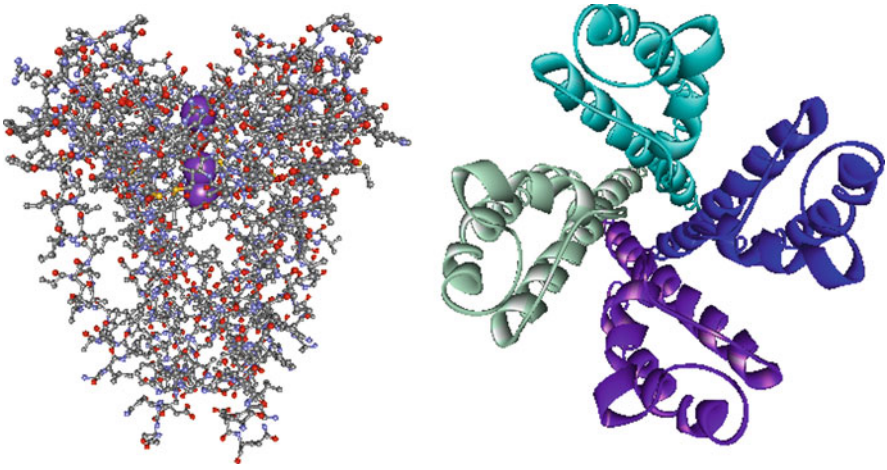


**Fig. 2.4** An example of protein aggregation: the $K^+$-specific transmembrane channel formed by KcsA [4]

also found in their free states [8]. This suggests that protein conformers that are initially complementary to their partners are bound preferentially, even if the numbers of those conformers are low. Once the conformer with the optimum geometry has been bound the remaining conformers will eventually repopulate the preferred state, only to be bound to further partner proteins. As will be seen in a later chapter, a similar complementarity drives self-replication in artificial systems.

### *2.2.3 Sugars*

Sugars are members of a group of carbohydrate molecules that adopt cyclic structures containing an ether bridge and appended alcohol groups, or are found in linear forms that contain aldehydes or ketones predisposed to cyclization. Both linear and cyclic forms exist in either the D- or L-form. Simple sugars, or monosaccharides, can link together through glycosidic bonds to generate disaccharides and higher order oligosaccarides. The basic monosaccharides start with the trioses, which cannot adopt cyclic structures as they would be too strained to be stable, followed by the larger tetroses, pentoses and hexoses. Of these it is the last two that are of greatest importance. The pentoses include fructose, mannose, and an essential component of RNA, ribose. The ribonucleic acids can be transformed enzymatically into the corresponding deoxyribonucleic acids. The hexoses include glucose and galactose. The former plays a key part in biochemical energy cycles. Plants produce it through photosynthesis and store it for future use in a polymeric form as the oligosaccaride, starch. Animals also produce glucose as its glycolysis to pyruvate followed by oxidation to carbon dioxide and water generates energy. Disaccharides are dimers composed of identical monomers, such as maltose from two glucose units, or different monomers, as in sucrose which is composed of fructose and glucose. Some examples of sugar chemistry are shown in Fig. 2.5.

Sugar chemistry has an added complexity. The carbon backbone of a linear monosaccaride is made up of linked chiral centres. For example, hexoses have four chiral carbons resulting in eight D- and L-stereoisomers of which seven pairs are known in biology. Upon cyclization a further stereocentre is formed which leads to α and β anomeric forms as shown below. Although they are distinct stereochemically, if the hexose is cleaved and reformed the new ring may have the same or opposite stereochemistry. As this process occurs naturally and quite rapidly in water the chirality is therefore not a permanent feature of the molecule.

Given the variation in size, isomeric form, conformation and degree of polymerization available to sugars, it is clear that their molecular recognition properties will be extremely complex. Unsurprisingly they perform a number of functions from simple energy sources to highly specific interactions with biological receptors. A good example is the incorporation of oligosaccharides into lipopolysaccharides (LPS). These are multicomponent molecules found in the exterior cell membranes of many bacteria. When LPS is detected an inflammatory response is elicited from higher organisms due to specific sugar-receptor interactions with the result that LPS acts as an endotoxin.

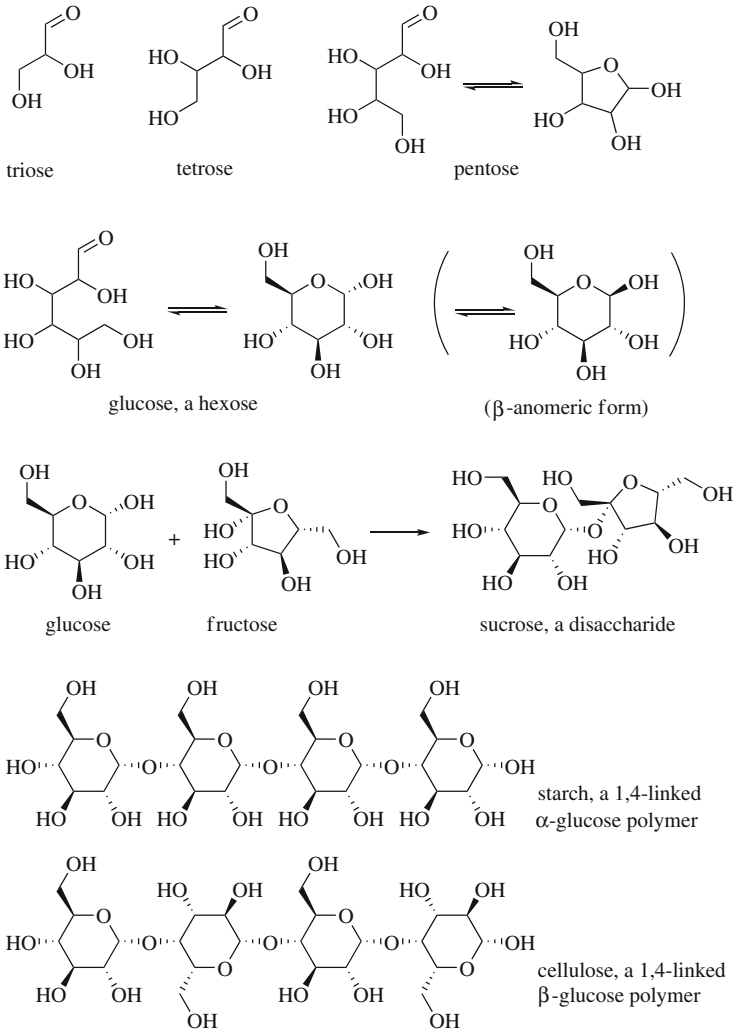



**Fig. 2.5** Structures and chemistry of sugars

## *2.2.4 Glycoproteins*

Glycoproteins are formed when a saccharide, or oligosaccharide, is attached to the side chain of an amino acid within a protein. Incorporation is mainly through serine and threonine (O-linked) or asparagine (N-linked) and results in simple ether or amide bonds being formed between the sugar and protein. The function of the saccharide may be to increase the specificity of molecular recognition for the protein, or influence protein folding by generating novel tertiary structures, or to increase

the level of hydration in the protein. This latter aspect has an impact on protein stability and is most evident in antifreeze glycoproteins found in fish, insects and plants that can withstand subzero temperatures. Other glycoproteins are found on cell surfaces, notably those involved in mammalian antinflammatory responses. The complex structures of many glycoproteins, together with difficulties in purification, have limited research into these biomolecules though their importance is recognized.

### *2.2.5 Lipids*

Lipids are molecules containing simple hydrophobic hydrocarbon chains as illustrated in Fig. 2.6. The hydrocarbons are insoluble in water and aggregate, or are attracted to other lipophilic molecules, through the hydrophobic effect. The most well-known of the lipids are those that form cell membranes. The simplest of these, the phospholipids, contain long chains attached to two termini of a triglyceride with a phosphate link to a water soluble head group attached to the third terminus. Similar compounds are formed when triglycerides link directly to one or more hydrocarbon chains. Other lipids include steroids, arachidonic acid and compounds with extensive conjugation such as retinol and carotene.

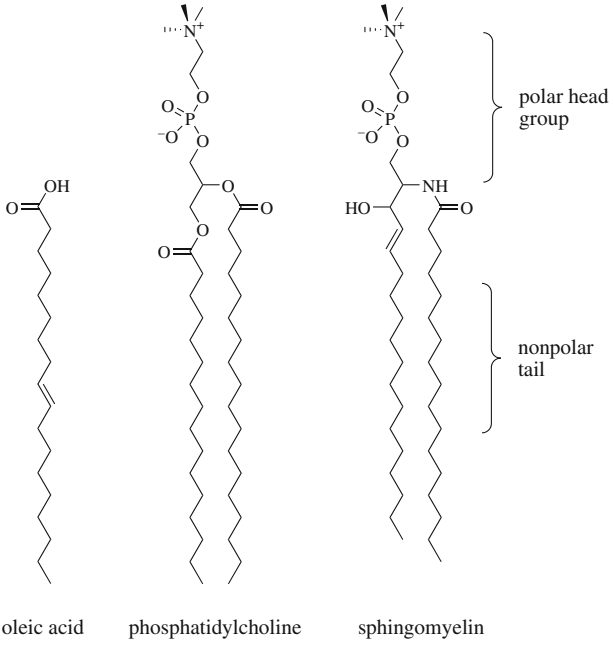



**Fig. 2.6** Some lipid structures

### 2.2.6 RNA and DNA

Ribonucleic acid (RNA) and deoxyribonucleic acid (DNA) are polymeric materials made up of nucleosides joined by phosphate esters. The nucleosides are composed of a monosaccharide (ribose or deoxyribose) joined to one of four planar molecules, or nucleobases: adenine, guanine, cytosine, or thymine. The first two are termed purines and the second two, pyrimidines. In RNA thymine is replaced by another pyrimidine, uracil. Each polymer naturally forms a helix because of the enforced twist in the phosphate ester backbone which exposes the nucleobases. This is fortunate as it allows two strands of DNA or RNA to form hydrogen bonded assemblies resulting in the iconic double helix structure necessary for these polymers to store, transcribe and translate genetic information. The nucleobases have hydrogen bonding motifs in which donor (**D**) amines bind to acceptor (**A**) imines or carbonyls of complementary bases. Thus adenine, with a **D**–**A** array, complements thymine or uracil, with an **A**–**D** array, and guanine (**A**–**D**–**A**) is complementary with cytosine (**D**–**A**–**D**). This match-up of bases is referred to as the Crick-Watson base pairing. When mismatch of bases or their orientation occurs other binding motifs such as Hoogsteen pairing are seen.

How do we know so much about DNA and RNA? Ever since the concept of heredity had been considered the question of its mechanism had been paramount. In the 1850s a monk, Gregor Mendel, noted that the peas he grew had different characteristics, such as flower colour, that may be the same through generations or may differ. By cross fertilizing plants with different characteristics he could control the resulting traits. This was nothing new: particular traits had been deliberately bred into, or out of, plants and animals for centuries. What Mendel was able to do was to analyze the result of many cross pollination experiments using flower colour and position, stem length, seed shape and colour, and pod shape and colour. He showed how the cross between a plant with purple flowers and one with white flowers results in a second generation that only has purple flowers, however, a third generation is composed of plants with approximately one quarter white and three quarters purple flowers.

Mendel used the term ‘heritable factor’ in 1866 [9] to describe the biological agent responsible for transmitting traits down through the generations; this factor is now termed the gene. Variations in a particular trait exist within the genetic material of an individual, such as flower colour in the first generation pea hybrids, are referred to as alleles. The traits described by Mendel can be dominant or recessive so that the second generation should display a higher proportion of the dominant trait. Although Mendel could show the outcomes of heredity he did not seek to identify the ‘messenger’ that carried the genetic code.

Later researchers and theorists presumed that genes were protein based molecules but in 1944 Avery, McCarty and MacLeod demonstrated that DNA, first extracted from cell nuclei by Miescher in 1869 [10] and consequently named nuclein, could be transferred between pneumococcal bacteria [11]. The traits of the DNA’s originator were then expressed by the infected strain suggesting that DNA could be the medium by which heredity occurred. The view was controversial as it

was known that DNA was comprised of only four different bases and it was unclear how such a limited chemical vocabulary could give rise to the complex instructions required to impart hereditary traits. The tetranucleotide composition had been proposed as early as 1919 by Levene [12] and the relationship between complementary bases reported by Chargaff in 1949 [13]. The bases have one of two basic chemical structures either the single six-membered pyrimidine ring or the linked five and six-membered rings based on purine. It was realized that the ratio of guanine, a purine, to cytosine, a pyrimidine, and adenine, a purine, to thymine, a pyrimidine, always approached unity, but that the ratio of G–C to A–T could vary between species. This observation led directly to the justly famous 1953 model designed by Crick and Watson. They reasoned that the preferred pairing must be due to optimized hydrogen bonding patterns as discussed above and, in one of the best known scientific quotations, noted that:

> [i]t has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material [14]

While the complementary double helical structure explained how particular sequences of bases could be used to store a genetic instruction it was not immediately clear how replication occurred or, indeed, how these instructions were used. Later work by Gamow linked DNA base pair sequences to protein synthesis [15] but it was not until 1961, when Nirenberg and Matthaei demonstrated that cell-free protein synthesis relied upon synthetic or natural polynucleotides [16], that the final link was made. The information held within the linear DNA sequence is replicated every time a cell divides. Replication is possible because of the unique double helical structure of DNA as shown in Fig. 2.7.

The two complementary strands of DNA are untwisted by a protein called topoisomerase then separated by helicase. As the two strands separate they form a ‘replication fork’ and a molecule of DNA polymerase attaches to each. The complementary sequence to that of the ‘leading strand’ is extended from the 5′ to 3′ direction as sequential bases are added. These bases interact with those of the original strand but rather than having a single phosphate group attached to deoxyribose they exist as the nucleoside triphosphates. DNA polymerase cleaves a phosphorus-oxygen bond to remove pyrophosphate: subsequent hydrolysis of pyrophosphate into two phosphate ions releases energy that in turn powers DNA helicase.

The second, ‘lagging’, strand of DNA cannot add in a 3′ to 5′ direction so DNA primase initiates the positioning of complementary bases. These are then joined by DNA polymerase in short sequences, called Okazaki fragments, starting at the highly constrained point where DNA forks and works from the 5′ end. The Okazaki fragments are connected by DNA ligase so that the combined effect of the DNA polymerase molecules is to make complete copies both original strands, hydrogen-bonded to their complementary sequences. The direction in which bases are added mirrors the direction in which amino acids are added to a protein.

A complete sequence of DNA bases, or chromosome, encodes instructions including those required by a complex supramolecular protein assembly, the ribosome, to form the extensive proteins that control much of biological structure and

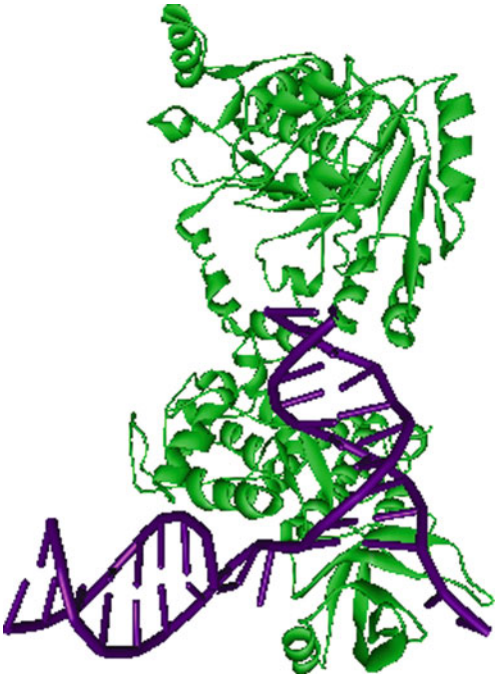


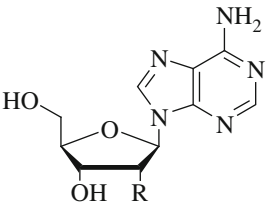
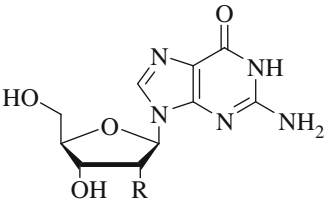
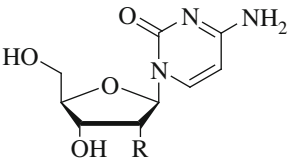
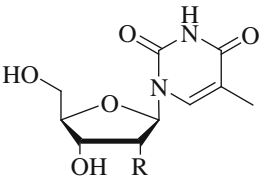
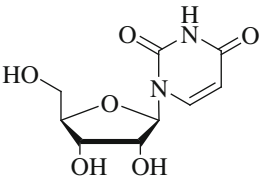
**Fig. 2.7** DNA replication [17]

function. This sequence, known as the 'sense' sequence, is paired with its non-coding ('antisense') complementary strand. The sequence of nucleobases is broken down into sets of three, or codons, each of which corresponds, indirectly, to an instruction for a particular amino acid to be added to a growing sequence. There are 64 possible three base sequences that can be generated from the four nucleobases leaving some redundancy as there are only 20 common amino acids.

As a consequence some amino acids can be coded in more than one way and there are also codons that regulate translation by acting as 'start' and 'stop' signals. This all appears extremely straightforward and allows for perfect transmission of a chosen DNA sequence through generations of dividing cells, however, as well as occasional mistakes when bases are introduced there are also chemical variants of DNA bases. These include purine variations xanthine, hypoxanthine and 7-methylguanine and pyrimidine variations 5,6-dihydrouracil and 5-methylcytosine. Rather than produce proteins directly through interaction directly with the ribosome the DNA sequence is transcribed into messenger RNA (mRNA).

In the transcription process the two DNA strands are separated and the antisense DNA strand paired with its complementary RNA bases by enzymes called RNA polymerases to produce mRNA that encodes the same sequence of bases as the sense DNA strand. The only difference between the DNA and RNA sequences is that the saccharide section of the nucleoside is ribose rather than deoxyribose and uracil takes the place of thymine. The effects of these changes are that the hydrophobic 5-methyl group of thymine has been removed to generate uracil and a 2′-hydroxy group is present in the linking saccharide (Table 2.4).

**Table 2.4** The common DNA and RNA nucleosides

| Base | Structure R═H (DNA) or OH (RNA) | Classification | Abbreviation |
|---|---|---|---|
| Adenine |  | Purine | A |
| Guanine |  | Purine | G |
| Cytosine |  | Pyrimidine | C |
| Thymine |  | Pyrimidine | T |
| Uracil |  | Pyrimidine | U |

The presence of the hydroxyl group makes RNA less stable in the long term, and more susceptible to attack under alkaline conditions, than DNA. It also inhibits the formation of the helical B-form of the nucleic acid. Thus RNA is usually single stranded and in the A-form which can fold into a much greater range of tertiary structures. This enables RNA to do far more than simply store a sequence of instructions but does, however, leave the RNA molecule more susceptible to attack by enzymes.

The replacement of thymine by uracil has no significant effect on the hydrogen bonding, as RNA does not use base pairing to form complementary dimers it is of less importance than it would be for DNA, but the removal of the methyl group may have an influence on the tertiary structures that RNA can adopt. From this it is clear that DNA is a better method of storing information whereas RNA is more suited to turn that information into a protein sequence. This is done by the ribosome, composed of ribosomal RNA (rRNA), which translates the codons of the mRNA sequence into a protein by matching three base sequences to those of tRNA that have the appropriate amino acids attached.

The process of transcription can also occur in reverse, from RNA to DNA, when the sequence coded in RNA is transcribed into antisense DNA by reverse transcriptase enzymes first discovered by Temin and Mizutani [18] and Baltimore [19]. Further integrase enzymes insert the DNA sequence into the native DNA. This is the mechanism by which viruses infect their host organisms and can be stopped by antiviral drugs, such as azidothymine (AZT) that inhibits HIV reverse transcriptase. Other processes such as the extension of telomeres at the ends of chromosomes by telomerase, to control programmed cell death, use the same mechanism.

## *2.2.7 Unusual Structural Forms of DNA*

Usually pairs of DNA molecules form a right-handed spiral with a regular pitch, the so called B-form, however other helical arrangements are also possible.

### 2.2.7.1 A-, B- and Z-DNA

Both the A- and B-forms of double stranded DNA have a right handed twist but they vary in how tightly they are wound: the B-form has a greater helical pitch and the helix consequently has a narrower diameter (Table 2.5).

By way of contrast, Z-DNA duplexes have a left-handed twist and a far greater helical pitch. The dramatic difference is due to the conformation of guanine relative to the deoxyribose ring. In the other forms of DNA all the bases are in the *anti* conformer so that the phosphate backbone is on the exterior of the helix with the base pointing inwards. In Z-DNA the connectivity through deoxyribose is the mirror image of that found for the A- and B-forms. Guanine is forced to fold over the

**Table 2.5** A-, B-, and Z-DNA

| | A-DNA | B-DNA | Z-DNA |
|---|---|---|---|
| Helix | Right handed | Right handed | Left handed |
| Average number of base pairs/repeat unit | 10.7 | 10 | 12 |
| Rotation per base pair | 33.6° | 35.9° | 30° |
| Base pair angle to DNA axis | 19° | −1.2° | −9° |
| Rise along DNA axis per base pair | 2.3 Å | 3.3 Å | 3.8 Å |
| Pitch of helix to DNA axis | 24.6º | 33.2º | 45.6º |

deoxyribose in order to bind to the complementary cytosine. The effect is to reverse the sense of the duplex whenever that particular base pair is found. A secondary consequence is that the major groove no longer features in the structure.

#### 2.2.7.2 Holliday Junctions and Other Features

Although it is usual for DNA to rely on helical twists to coil it around there are other mechanisms by which the direction of the duplex can be radically altered. The simplest occurs when one or more bases insert into an otherwise complementary pair of strands. Assuming that all the other bases pair up, this leaves a small loop, or bulge, where the unpaired bases reside. A more extensive internal loop forms when both strands are mismatched in the same region. Hairpins are formed when a self-complementary single strand of DNA folds back on itself to form a duplex. A loop due to unpaired bases remains at the point where the strand reverses.

Even more complex structures can be formed when DNA strands interact as seen in the G-quadruplex and Holliday junction (Fig. 2.8). The G-quadruplex is a complex structure formed by the end of a single strand of DNA that is rich in guanine. In humans this is manifested as a series of d(TTAGGG) sequences. The strand snakes in such a pattern that a series of stacked rosettes, composed of four guanines held together by a combination of hydrogen-bonding and electrostatic attraction to a centrally held cation, are formed. The apparent age of a cell can be determined by the length of these specific guanine-rich regions, or telomeres, attached to the ends of chromosomal DNA. In many species, including humans, telomeres become shorter with each successive replication cycle: once the telomere sequence is lost, replication ceases and cell death is initiated. Where this mechanism fails cells can proliferate and become cancerous.

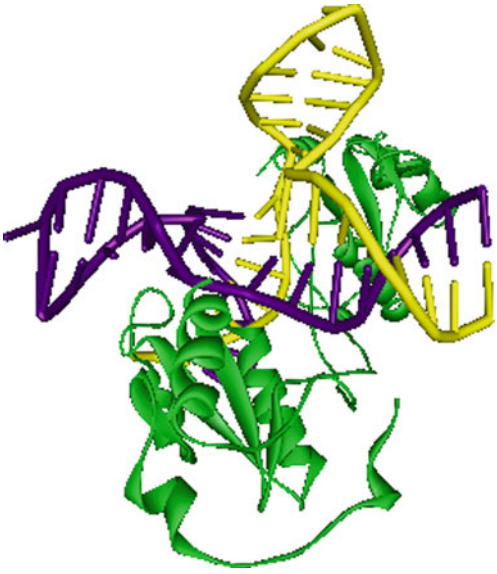


**Fig. 2.8** The Holliday junction, an unusual structure formed by DNA [20]

The Holliday junction is a cruciform structure composed of two unbound strands of DNA that have crossed over, that is, a single strand of one sequence is not paired with its complementary strand but with one from another sequence. Its existence has led to the use of DNA as a structural nanomaterial as will be seen in a later chapter.

## 2.3 Self-Replication as the Key to Life

A reductionist approach to single celled organisms shows them to have several basic characteristics. They require a mechanism for generating energy that fuels growth. Growth is facilitated by catalysts, enzymes in this context, that assemble molecular components such as amino acids or lipids from a readily available feedstock into more complex molecules capable of higher functions. To ensure that growth results in the promotion of the same organism, either through an increase in size or via cell division, through an increase in numbers some form of programmed self-replication is necessary. Replication must follow a rigorous code if it is to generate copies of the original instructions and must also be capable of self-correction through a reversible dynamic process. Several authors have viewed this phenomenon as the key to a self-sustaining molecular factory, from Dawkins' concept of a *Replicator* molecule [21] to Drexler's description of assemblers [22]. These processes must take place in concert. This is best achieved through compartmentalization which, in our predominantly aqueous world, makes use of lipids with water-soluble head groups. The lipids spontaneously form spheres corresponding to their lowest energy arrangement in water. A simple sphere where all the hydrophobic ends meet in the middle is called a micelle and cannot encapsulate cellular components but larger entities, with a double layer of lipids, can exist and are known as vesicles. A vesicle is bounded by two layers of interdigitated lipids which are hydrophobic where they meet but present hydrophilic surfaces. This feature allows vesicles to form in, and to contain, water but not to let water pass across the bilayer. The contents are therefore insulated from the external media and, more importantly, can be held in close proximity to energy sources and chemical building blocks within the vesicles. Single-celled organisms are essentially complex vesicles. Obviously there are secondary processes involved including transmembrane transport, manipulation of molecules, etc, but the major requirements remain: energy, macromolecular construction, compartmentalization and self-replication.

### *2.3.1 Replicators*

Dawkins' concept of a gene as a piece of information passed down the generations is worth examining in somewhat greater detail as it has implications for both natural and unnatural self-replicating systems. A successful replicator requires several characteristics that allow it to function better than its competitors. It must be able to make copies of itself from available raw materials. One example of this is the passive

effect of crystal seeding in saturated solution of a salt. The surface composition and geometry of the seed crystal dictates where solvated ions can attach themselves. The crystal then grows through reducing the entropy of the system while optimizing its own surface energy. Certain life forms use this principle directly: *Acantharia* are single celled marine organisms that use strontium sulfate as their skeletons. Each organism contains exactly 20 single sulfate crystals with a common origin that radiate out according to crystallographic rules. A cytoplasmic membrane forms between the radial crystals. So important is the crystallographic symmetry that every cell retains the relative dimensions of the original skeleton as it grows. However, in most cases the effects of crystal formation are of little importance in terms of information propagation through time. What is required is a more flexible and functional structure. Organic chemistry can supply this through the wealth of molecular structures available, many of which do not require biological processes in their synthesis.

Why is flexibility important? If a replicator is to act as a template it must bind adjoining subunits before bringing them close enough to undergo chemical reactions that form permanent covalent bonds between them. A rigid template would be able to bind one complementary subunit but, when another approaches, the steric effects of the reactive sites would prohibit reaction. Taking these considerations further, it is necessary to construct a template that is linear if the shape of the replica is to be identical. A linear molecule does not have to be flat but could, as in the case of DNA, form a helix. For the template effect to work the recognition sites would have to be on the interior of the spiral otherwise the replicator could not make copies. The external recognition sites would, as shown in Fig. 2.9, make fragments of the replicator but would not be able to reproduce it. Next we must consider the nature of the forces between replicator recognition sites and the subunits it needs to attract, and the forces that form links between subunits. The former need to be reversible, so that replicator and replica can separate once a complete copy has been made; the latter need to be strong and permanent. Nature uses hydrogen bonds extensively

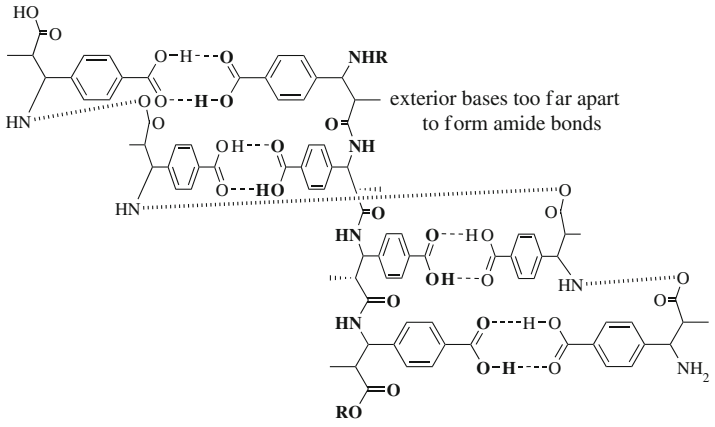



**Fig. 2.9** A non-functional replicator

for their strength in numbers and individual weakness though other weak forces could be used instead. The hydrophobic effect promotes binding of hydrocarbon-rich molecules in water though this is only easily reversed if a non-polar solvent is introduced to the system and solvates the hydrophilic regions of the molecules. Similarly the π-stacking motif, increasingly found to be important in biological systems, could also function as a weak attractant. The aromatic rings required for this type of interaction could have varied functional groups to give a range of solubilities though the chemical investment necessary (six carbons and five hydrogens for each interacting group) is far less efficient that the requirements of hydrogen bonding systems.

As Dawkins has observed, the greater the copying fidelity, the more successful the replicator will be. The act of copying must be perfect or the information passed on to each subsequent generation will become more and more corrupt and ultimately, as in the game of Chinese whispers, generate a completely different message which may not even make sense. To minimize the mistakes, or mutations, the replication process must have an error checking mechanism. It could be based on an optimal energy for each extension of the replica which is tested by a 'molecular checker' to determine if sequential interactions are favourable before committing to permanent bond formation between the penultimate and terminal subunits.

The replicator must also be more stable to decomposition and faster at forming than its competitors if it is to succeed. Two chemical processes are involved here. Stability relates to the energy gain made as a consequence of replica formation and implies that the most thermodynamically favoured product of a reaction will ultimately win out, even if other less stable products form faster. Greater chemical stability leads to a longer 'lifespan' for the molecule and therefore more chances to replicate. In turn this leads to more copies of the replicator which can broadcast its information more widely. The fastest forming compound is termed the kinetic product. Faster replication leads to what Dawkins calls fecundity. If a mutant replicator is faster at forming copies than its ancestor then it will take over, particularly if there is a limited pool of subunits available, so thermodynamic stability coupled to a rapid speed of replication must be the goal for any replicator.

### *2.3.2 Replicator Evolution*

The paradox implicit in the above description is that a successful replicator which rapidly produces long-lived, highly stable, perfect copies of itself cannot, by definition, evolve. Fortunately for us, evolution at the replicator scale occurs. If the replicator is a small molecule with perhaps only four or five subunits then one or two mistakes in the replication will generate a radically different product. It may be an even better replicator or completely inactive. If, however, the replicator has many hundreds or thousands of subunits then, while one key error might destroy its function, it is more likely that a few errors will have a marginal effect on its activity. Gradual evolution of this type would lead to a family of similar replicators

all with subtly different structures and functions. The end result is a diversification which could find each replicator in its own particular niche. Further changes over time may lead to competition or symbiosis. Indeed, the most successful replicators would be those that joined together to template the formation of multifunctional copies. To use a biological concept, individual replicator motifs are genes, with information copied during each replication, and the complete replicator sequence is a chromosome.

### *2.3.3 Orthogonal Translation*

Copying is only part of the story. The process of replication is linear; it merely produces more replicators. There needs to be another process that takes a functional property of the replicator, which may be derived from its self-replication motif, and translates it into other types of molecules. These other molecules should have some value to the replicator, for example if they either form a protective coat or help to collect and organize the subunits for subsequent incorporation.

Whatever the benefit it is clear that the type of molecule formed by translation must be quite different from the replicator or it would be in direct competition. It is precisely such a relationship that exists between DNA and protein synthesis, albeit with RNA acting as the go-between. Furthermore there has to be a mechanism for this 'orthogonal' translation. In mechanical systems the rotation of a motor can be translated into a variety of other motions through a gearing mechanism. In biological systems information encoded in an RNA sequence is translated into a protein by the complex concerted action of a ribosome. In one sense it is actually the translation device, whether it is a set of gears or a ribosome, that is the most important factor in the replicator spawning a long line of descendents.

Returning to the key issue of transcription, how simple could a replicator be? It needs a minimum of two binding sites to hold two subunit monomers in place while they react. Each replicator binding site must be specific to a complementary site on the target subunit. Several reversible chemical interactions could be employed. Simple $\pi$-stacking could be used, and would only require one type of self-complementary molecule to be generated, but for reasons given above it would require a large atomic investment. Imine bonds are stronger and reversible, although the process is dependent on changes in pH. Hydrogen bonds are much more attractive. If a subunit of the replicator is to be self-complementary then it must contain both a hydrogen bond donor and hydrogen bond acceptor. As illustrated in Fig. 2.10 this represents the simplest solution to replication as it requires only one type of molecule but, as a consequence, it will be information poor.

Assuming the translation process uses the same principles as RNA, where sequences of three subunits code for 'start', 'stop' or the addition of a particular amino acid, a replicator composed of a single donor-acceptor motif could only generate a replica with the same repeat pattern. This could succeed if translation was based on a different chemical function to transcription as shown in Fig. 2.11. Subunits incorporating a range of secondary chemical functions would

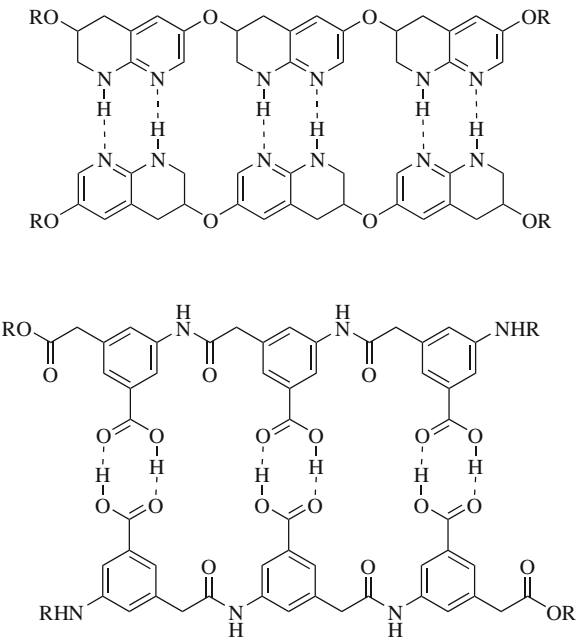


**Fig. 2.10** Possible self-complementary hydrogen bonding motifs for replication

lead to replicators with identical transcription patterns but vastly different translation motifs. Unfortunately such a simple transcription system would generate replicas with random secondary functions leading to a loss of fidelity during the replicator's translation phase. The only way to regulate the sequence of secondary functional groups would be to invoke a complex error checking mechanism.

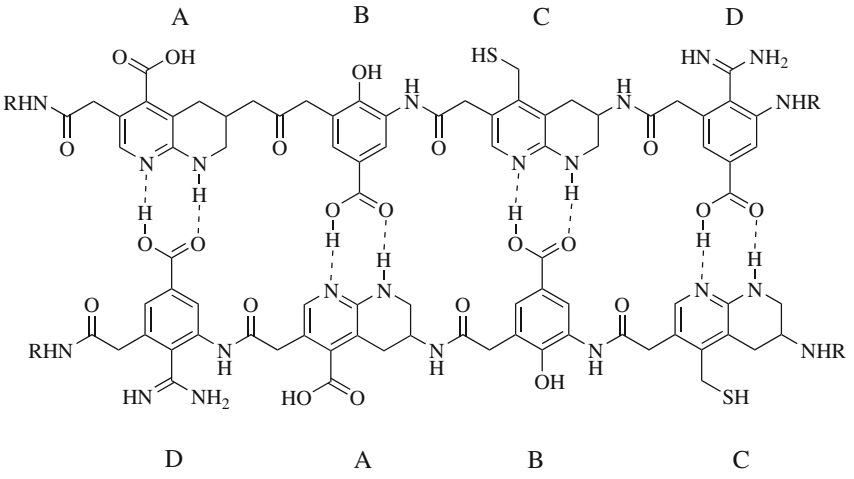



**Fig. 2.11** Low fidelity orthogonal replicator transcription

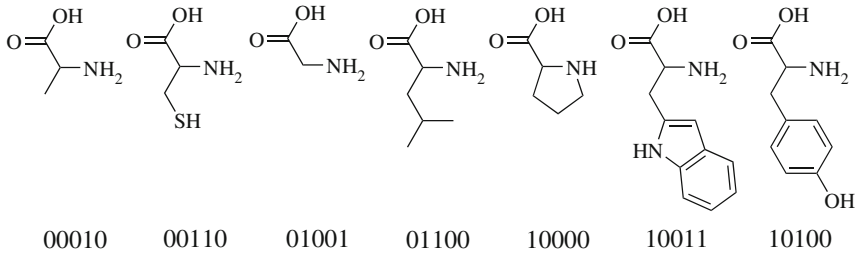



**Fig. 2.12** Some binary coded amino acids

In Nature, each subunit has multiple hydrogen bond donors and acceptors to promote accurate transcription. However, the replicators are flexible enough to allow different orientations of the subunits so a small number of mismatches can occur that are as stable as the intended interactions. The effect of this is that small replicator mutations can occur but most of the sequence is transcribed accurately.

Assuming that in an alternative system of replication the transcription hydrogen bonding motif was also to be used in translation then the simplest solution would be for replicators to be composed of complementary hydrogen bond donors and acceptors. If one subunit contained a donor group and the second had an acceptor group the resulting replicator would have a binary function.

For this system to code for our 20 standard amino acids together with start and stop instructions the translation device would need to use a five figure binary code as shown in Fig. 2.12. For example, a 'start' instruction could be given by 00001 and a 'stop' by 00000. Alanine could be coded alphabetically as 00010 with other amino acids having sequential codes (00011 for arginine, 00100 for asparagine, etc) up to valine coded as 10110. The method that Nature has adopted uses this mechanism but with four different subunits, though only groups of three are necessary to encode information for translation. The level of complexity is far higher than necessary; a base three system with only three different subunits would generate 27 possible outcomes, enough for all the major amino acids plus the 'start' and 'stop' instructions. A set of four bases has evolved because one pair has complementary hydrogen bonding motifs to the other pair. In a set of three bases two could be complementary to each other but one would have to be self-complementary and interact poorly with the other two to avoid mismatches occurring during replication.

## 2.4 Supramolecular Self-Replication

So, how close have supramolecular chemists come to realizing a self-replicating system? The answer is not too close. There are few chemists currently working on the problem from a perspective of a system based on supramolecular chemistry. One requirement of any system designed to perpetuate information is that it must encode a large amount of information. In the DNA/RNA system it is in the form of

codons, triplet motifs of nucleobases, that, when 'read' within the ribosome, trigger the attachment of a specific amino acid to a growing peptide chain. To do this we have established that at least three distinct subunits are required. In the DNA/RNA case each is coupled with its complementary base as the binding interactions are asymmetric. To date the most complex supramolecular systems can be classified either as compounds that form antiparallel double helices but do not self-replicate, or that self-replicate but have structures restricted to their dimeric components.

There are many examples of compounds, usually short polymers, that spontaneously form double helical structures. A classic example involves a number of bipyridyl ligands joined by short, flexible spacers. In the presence of Cu(I), which has both an affinity for nitrogen-containing ligands and a preference for a tetrahedral disposition of donor atoms, discrete double helical complexes are formed [23]. A similar phenomenon gives rise to a triple helical motif when a metal that preferentially adopts an octahedral coordination geometry is used [24].

Few examples exist of synthetic self-replicating compounds. Those that successfully act as replicators and produce copies that also replicate are formed from two subunits with complementary hydrogen-bonding motifs. Only one new bond forms when two subunits are held in close proximity by the replicating template which severely restricts the ability of the replicator to carry information. The molecules themselves have not been designed with the intention of forming self-replicating polymers but to prove that simple self-replication can be achieved. Unfortunately the chemical mechanisms used are not amenable to forming polymeric systems and, even if an 'original replicator' composed of multiple subunits was available, the reaction employed to bind two subunits together would not be able to generate anything other than a large number of dimers, as illustrated in Fig. 2.13. If the subunits could form templated polymers the likelihood is that they would be in a random order. Other, external, forces would be necessary to place the subunits in the correct sequence prior to coupling. Without this influence the chemical reactions would proceed without any reference to the replicator's sequence.

A cursory glance at the structure of DNA shows that it is composed of hydrogen-bonded units, the purine and pyrimidine bases, attached to sugars that are linked by phosphate groups. There is no chemical reason why the perfectly symmetric phosphates should bind in the orientation that they do. The same problem arises in when a synthetic analogue of a DNA-type replicator is considered. The most useful linkages are imine and peptide bonds. Both require a terminal amine; the former results from a reaction with an aldehyde and the latter with an activated carboxylic acid. The problem that occurs is that if these functional groups are present within the same molecule self-polymerization may occur unless a substantial effort is made to avoid this.

The only way to ensure that such a potentially random sequence is avoided is if an overarching scaffold exists to prevent it happening. This invokes a 'chicken and egg' scenario: for the original strand of DNA (or an artificial analogue) to polymerize in the correct sequence there must be some molecular construction present that regulates its formation. A similar scenario was envisaged by von Neumann when he considered the self-reproduction of artificial automata [26]. Von Neumann theorized

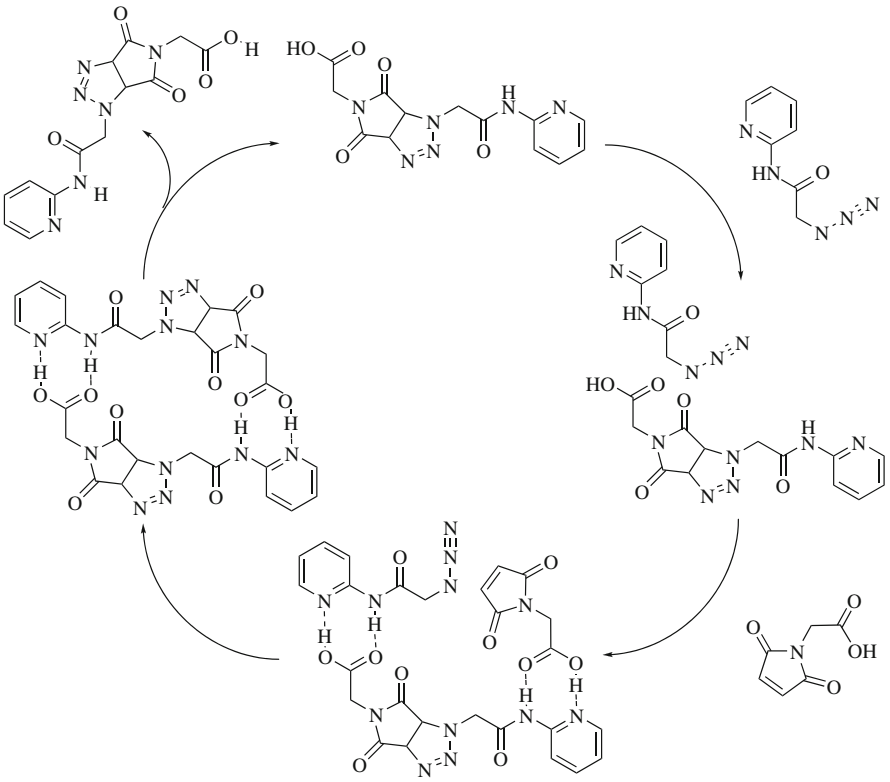


**Fig. 2.13** An example of supramolecular self-replication [25]

that a 'concept of complication' arose when one automaton had the ability to construct another automaton in its entirety. Intuitively the former must be more complex than the latter, however, von Neumann followed Turing's persuasive argument that if one automaton was capable of following a sequence of instructions it could construct copies of itself [27]. Where a complementary hydrogen-bonding motif lies at the centre of self-replication it is not too hard to see how isolated molecules, chosen from an available pool, could be positioned in the correct sequence ready to be bonded together.

## *2.4.1 Self-Assembling and Self-Replicating Motifs*

Supramolecular self-assembly can be as simple as acetic acid dimerization through mutually complementary hydrogen bonds or as complex as supramolecule formation from multiple components as demonstrated by numerous examples of resorcin[4]arene and calix[4]arene capsules. One of the more interesting directions to have been taken has been the design and synthesis of self-complementary

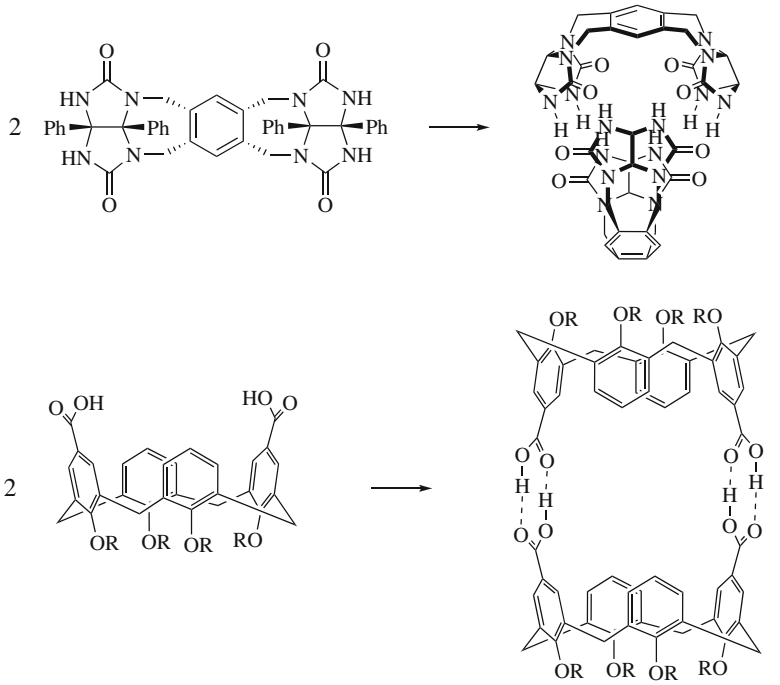


**Fig. 2.14** Supramolecular capsule formation: (*top*) the tennis ball [28] and (*bottom*) hemisphere approach [29]

nanocapsules. On the smallest scale two synthetic approaches predominate in which capsules form through homodimerization as shown in Fig. 2.14.

In the first a hemispherical molecule is prepared and functionalized on its open rim so that two molecules are mutually attractive. The second method, championed by the Rebek group, is to synthesize a linear molecule that curves such that two interact, like the halves of a tennis ball or baseball, with the complementary binding motifs forming the seam [28]. The interior space of these capsules is only capable of housing one or two small molecules though research on gas storage has shown that the capsules could be very specific in their choice of guests [30]. In general around half the internal space is occupied by guests which allows them some freedom to move.

Capsules composed of hemispherical compounds include calix[4]arenes with urea substituents on the upper rim that encapsulate tetraethylammonium cations, prepared by Böhmer's group [31], Cram's velcraplexes [32], and the cavitand dimers prepared by the Gibb group [33]. Dimerization can be solvent controlled with capsules driven by hydrogen bonding forming in non-polar solvents and those that form through hydrophobic effects being promoted by polar solvents. Hydrogen bonding motifs also appear in self-assembly systems designed by the Gokel group from diazacrown ethers with complementary adenine and thymine DNA bases that form

dimers [34]. The A-crown-A and T-crown-T derivatives formed both homodimers in aprotic solvents but, following stoichiometric mixing, formed heterodimeric boxes.

Self-assembly does not have to result in the formation of capsules. One of the best known non-biological self-assembly processes, the formation of melamine cyanurate utilizes similar hydrogen bonding motifs to those seen in DNA. Complementarity between hydrogen bond donors and acceptors can be seen in the secondary structures adopted by melamine and cyanuric acid [35]. Both are based on a hexagonal structure: the former has three nitrogen atoms alternating with three amine groups, the latter is a cyclotriamide with alternating carbonyl and amine groups. Melamine can attract three hydrogen atoms and form bonds with six amine hydrogen atoms whereas the cyanuric acid can form three bonds with its hydrogen atoms and interact with up to two hydrogens for every oxygen atom. Each component can thus take part in up to nine hydrogen bonds making any extended structures very stable. The melamine donor-acceptor-donor (**D**-**A**-**D**) arrangement is also an almost perfect geometric match for the cyanuric acid acceptor-donor-acceptor (**A**-**D**-**A**) sequence. As a consequence the two chemical species aggregate to form sheets, tapes and isolated hexamers as shown in Fig. 2.15. The Reinhoudt group has used melamine substituents on calixarenes to generate intricately interlocking

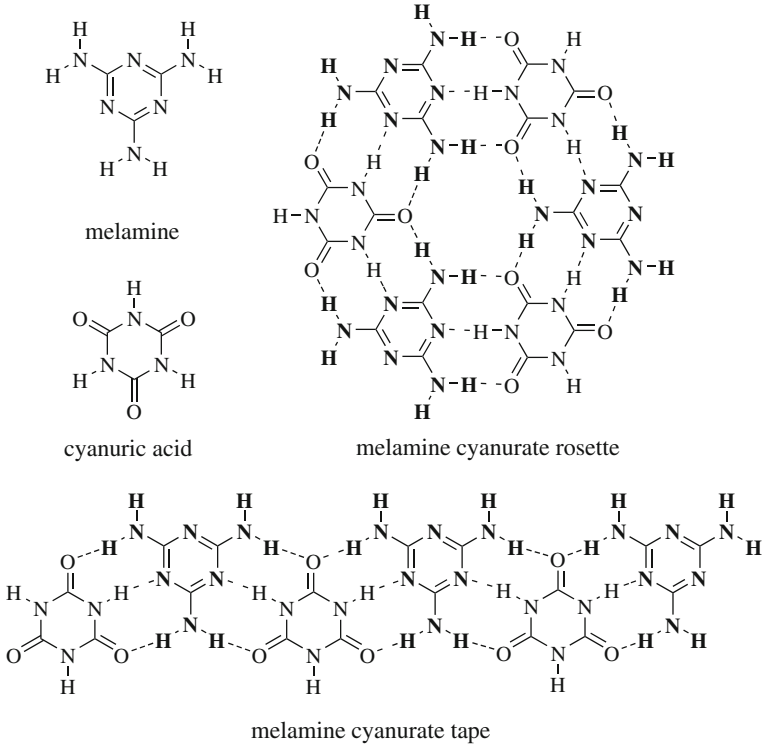



**Fig. 2.15** Planar melamine-cyanuric acid assemblies

supramolecular assemblies held together by the substituents' self-assembly [36]. While these motifs are valuable from a structural or materials chemistry perspective they give little insight into potential self-assembly mechanisms that could be analogous to those encountered in Nature.

Chemical self-replicating systems fall into two categories: those that are predicated on biological building blocks, such as amino acids or nucleotides, and those that avoid biological associations. While approaches based on the former have been quite successful, those based on the latter are usually simpler in design and execution.

The Rebek group undertook some of the earliest non-biological investigations into self-replication. Rebek's systems were derived from Kemp's triacid, an axially substituted cyclohexane in the chair conformation as shown in Fig. 2.16 [37]. The compounds incorporated the **A**-**D**-**A** motif and a side chain with an acceptor-donor (**A**-**D**) terminus. The sidechain contains an amide bond which, when considered retrosynthetically, suggests that two fragments, containing an activated carboxylic acid (***A***) and an amine (***B***), could be held in place by the template (***T***) while they react. Overall the self-replication can be described as the formation of a ternary complex between the template, ***T***, and fragments ***A*** and ***B*** (2.1), reaction of ***A*** and ***B*** to form self-complementary template ***T*** (2.2) and separation of the original template from that generated by the reaction of ***A*** with ***B*** (2.3).

$$\boldsymbol{A} + \boldsymbol{B} + \boldsymbol{T} \quad \rightarrow \quad \boldsymbol{A}:\boldsymbol{B}:\boldsymbol{T} \tag{2.1}$$

$$\boldsymbol{A}:\boldsymbol{B}:\boldsymbol{T} \quad \rightarrow \quad \boldsymbol{T}:\boldsymbol{T} \tag{2.2}$$

$$\boldsymbol{T}:\boldsymbol{T} \quad \rightarrow \quad 2\,\boldsymbol{T} \tag{2.3}$$

Von Kiedrowski [38] reported a similar system that operates using the same logic but with different fragments; a carboxylate group linked to an aromatic aldehyde and an amidium group linked to an aromatic amine. The strong affinity of the carboxylate for the amidium group ensures that aldehyde and amine are in close enough proximity for imine formation to be favourable. The reaction results in a copy of the template being formed.

Philp and his group also made an important contribution to self-replication with the design of a similar minimalist system which also catalyses its own template [25]. The template (***T***) contains complementary hydrogen bonding regions at its termini so that two reactants can be held in place while their reactive centres, an aldehyde (***A***) and amine (***B***) again, can be brought close together. Once the imine bond forms between the two, the new molecule separates from its template to catalyse further reactions leading to exponential growth in the template population.

In 1986 von Kiedrowski reported the autocatalysis of a palindromic nucleotide hexamer. Unlike simple (self-)complementary hydrogen bonding sequences, DNA can have both template and complementary strands [39]. In this example the components were CCG (***A***), protected at the 5′-terminus and activated at the 3′-phosphate terminus, and CGG (***B***), protected at the 3′-terminus. A doubly protected template,

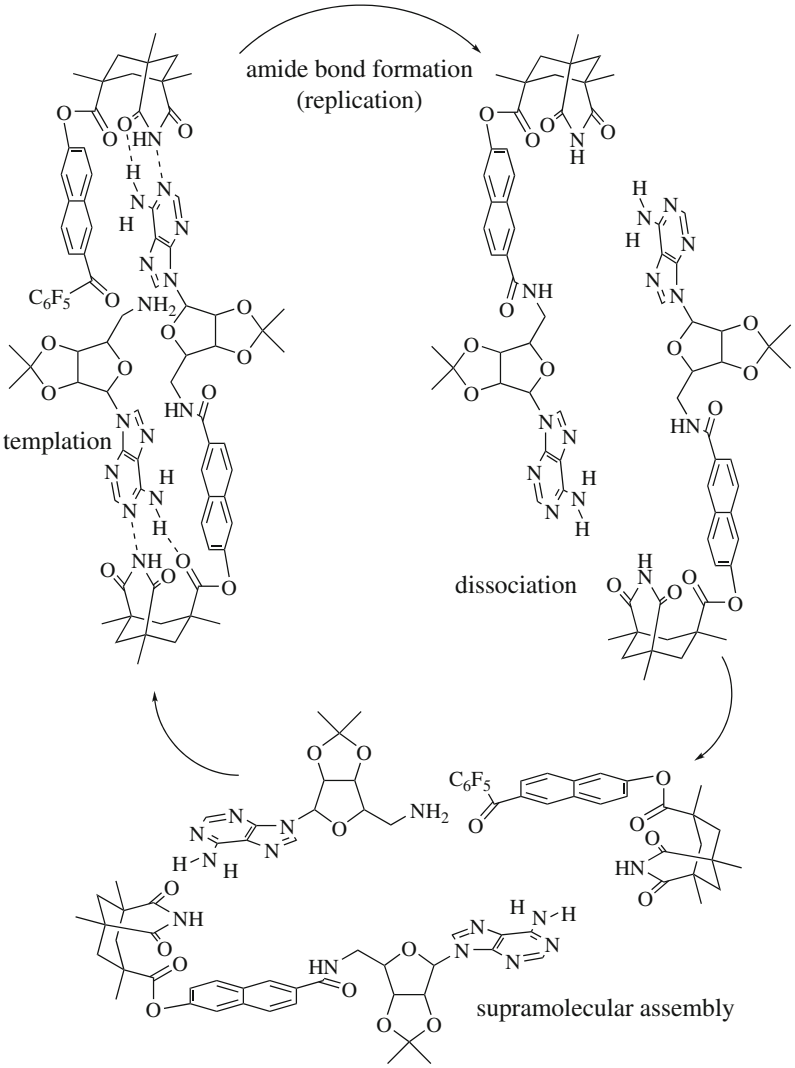



**Fig. 2.16** Self-replication based on Kemp's triacid [37]

CCGCGG (***T***), was prepared and a coupling reagent added to facilitate reaction between the trimeric components. Although some incorrect coupling was observed, the predominant chemistry was a hydrogen bonded complex formed by ***A*** and ***B*** in the presence of ***T*** ($\boldsymbol{A} + \boldsymbol{B} + \boldsymbol{T} \rightarrow \boldsymbol{A} : \boldsymbol{B} : \boldsymbol{T}$, using the same formalism for non-biological systems, as above) followed by reaction between ***A*** and ***B*** that formed a copy of ***T*** ($\boldsymbol{A} : \boldsymbol{B} : \boldsymbol{T} \rightarrow \boldsymbol{T} : \boldsymbol{T}$, as above). The cycle repeated following the dissociation of the template dimer ($\boldsymbol{T} : \boldsymbol{T} \rightarrow 2\,\boldsymbol{T}$, as above) and binding to further copies of ***A*** and ***B***. The main achievement of this research was to show that DNA replication,

albeit of a self-complementary motif, could occur without enzyme assistance. The von Kiederowski group later showed that more complex cross-catalytic methods between termolecular complexes of ***A*** and ***B*** could also occur by this method even if the template sequence was not self-complementary [40].

Taking a cue from those who propose that protobiotic systems were most likely based on amino acids, the Ghadiri group designed a self-replicating 32-residue α-helical peptide [41]. The sequence was based on the leucine zipper domain found in the protein GCN4, a yeast transcription factor. Unlike most other examples of self-assembly and autocatalysis, complementarity is not achieved by hydrogen bonding but hydrophobic interactions between the lysine and valine amino acids' side chains.

The sequence ArCONH-RMKQLEEKVYELLSKVACLEYEVARLKKVGE-$CONH_2$ was chosen as a template because valines align in an α-helix and are adjacent to lysines which also line up. The overall effect is to give a continuous hydrophobic face to one side of the helix. When two helices interact they are expected to do so in an antiparallel manner so that valines interact with lysines and *vice versa*. The reactants were simply fragments of the template though, unusually, the coupling method chosen involved thioester condensation in an effort to minimize side reactions. Thus a fragment with an electrophilic terminus, ArCONH-RMKQLEEKVYELLSKVA-COSBn, could be aligned with the hydrophobic face of the template and reacted with the N-terminus of $H_2N$-CLEYEVARLKKVGE-$CONH_2$. Notably it was expected that neither fragment would be long enough to form a stable helix so the template not only brought the polypeptide reactive centres into close proximity to each other but also enforced the helical form of the protein fragments. As expected, formation of the 32-residue peptide occurred significantly faster than the calculated rate if a non-autocatalytic mechanism was invoked. Replacement of lysine or valine in the template with a charged amino acid also led to polypeptide formation, but at a considerably slower rate, and absence of a template gave the slowest rate of all.

Perhaps the most important findings in the experiments of the von Kiedrowski and Ghaderi groups were that parabolic increases of the designed product were observed over other possible products. The implication is that, even in the presence of competitor reactants, the products with the greatest complementarity to their templates would form at rates which would effectively annihilate any opposition. The most successful chemical species would thrive at the expense of all others. This effect is also observed in the growing field of dynamic supramolecular combinatorial chemistry. There is a clear parallel with the way different organisms compete with each other for resources: those that are most successful at using those resources to reproduce will survive.

In an interesting development the Philp group investigated the effect of competition in a library comprising four molecules [42]. Two of the molecules contain donor groups which can bond with the two other molecules that have acceptor groups. There is the potential therefore to form donor-acceptor conjugates that template the formation of similar molecules, in the manner discussed above, that is, $\boldsymbol{A} + \boldsymbol{B} + \boldsymbol{T} \rightarrow \boldsymbol{A}:\boldsymbol{B}:\boldsymbol{T}$ followed by covalent bond formation ($\boldsymbol{A}:\boldsymbol{B}:\boldsymbol{T} \rightarrow \boldsymbol{T}:\boldsymbol{T}$)

and separation to double the concentration of templates ($\boldsymbol{T} : \boldsymbol{T} \rightarrow 2\,\boldsymbol{T}$) which leads ultimately to an exponential increase of the template molecule while a pool of components remains available. Here the library is dynamic, that is, the covalent bonds can form and break until the thermodynamically most stable product is formed.

In Philp's recent example the donors and acceptors were linked by spacers of different lengths but the longest template was shown to be the most successful replicator. From the available pool of precursors, **A**, **B**, **a** and **b** (where upper case denotes 'long' precursors and lower case, the 'short' precursors), four products are possible: **AB**, **Ab**, **aB** and **ab**. Equating **AB** to the template **T** we would expect, as before, to exhaust the supply of **A** and **B** and produce mainly **T** with perhaps some random distribution of other products. What actually occurs is, at first sight, quite unexpected. Initially all four products appear to form but, due to the reversibility of the coupling reaction the concentration of **T** increases. At the same time the concentrations of **Ab** and **aB** drop, as more **T** forms, leaving a pool of **a** and **b** molecules. These are then free to link and form **ab**, which could also be described as a 'short' template, **t**. As an unexpected consequence of removing **A** and **B** from the pool, the concentration of **t** increases. Because the formation of **t** is less efficient than **T**, the latter still comprises the majority of the molecules produced but the former is nevertheless a significant by-product of the reaction. On reflection the outcome is entirely reasonable: once the pool of two species has been exhausted the remaining two are free to combine, though perhaps not quite as successfully. Aspects of this observation may have implications for early molecular evolution and may go some way to explain how molecular diversity could have arisen in the presence of one very successful self-catalysing reaction.

## 2.5 Supramolecular Chemistry and the Origin of Life

It is inescapable that many of the features we now associate with a living process rely on classical supramolecular motifs: self-replication and random miscopying (to use Dawkins' phrase) through hydrogen bonding leading to evolution on the molecular scale, the formation and compartmentalization of cells through hydrophobic interactions, and the acquisition of energy from external sources to power the chemical processes involved through complex multicomponent interactions. Even the motion of simple bacteria is controlled by a molecular 'rotary engine' in which many proteins co-operate non-covalently to spin flagellae and provide forward movement. A similar level of complexity is seen in our existing DNA–RNA-protein system. Both DNA and RNA require that complex heterocyclic molecules are attached to sugar molecules and joined, in a highly specific order, through a backbone derived from phosphate anions. Furthermore, the information from one molecule, DNA, must be translated into functional proteins. The biochemistry needs to occur in an enclosed space where components are concentrated, as easy access to the precursor molecules required for synthesis is essential. All this also requires an energy source. Taken together all these interdependent processes are evidence of a complex and synergistic molecular evolution, however, one of these processes or its

primitive precedent must have been first. The chemistry involved in transcription and translation revolves around hydrogen-bonded species and their interactions with each other or their aqueous environment. Protein folding relies on hydrophobic and hydrophilic interactions as well as amide and reversible disulfide bond formation. Compartmentalization relies on hydrophobic effects and the alignment of amphiphilic molecules. It is therefore worth considering how several 'origins of life' theories can be analyzed in terms of the supramolecular chemistry involved.

There are many possible systems that may be envisaged as precursors to our DNA–RNA-protein driven biology and considered to be the first 'living' molecules. All appear to have been lost in the process of molecular evolution that has led us to our current system. It is highly likely that several systems evolved in isolation and died out through lack of resources, changes in conditions or were superseded by more successful forms. On the prebiotic Earth conditions for chemical synthesis would have been good. Water was available as a solvent, energy came from the sun, geothermal or atmospheric sources, and a ready supply of suitable elements, including carbon to supply a molecular backbone, was available. Despite all these factors, for atoms and molecules to meet and react in just the right way requires luck, or a great deal of time. As a result isolated 'living' systems may well have appeared and disappeared leaving no trace.

### *2.5.1 Compartmentalization: The Lipid World*

One way to improve the likelihood and frequency of chemical reactions between atoms or molecules in solution is to increase their concentrations. Today's cells function well because they encapsulate all the chemicals required for replication, protein synthesis and energy transfer. These chemical species are present at a far higher concentrations than in the extracellular environment and their levels are controlled by processes that transport molecules and ions across the cell membrane. To achieve compartmentalization the cells' contents must be isolated from their surroundings. In water this can be done in droplets of oil or similar substances when they form as an emulsion in water. These structures are often unstable; the hydrophobic compounds coagulate and rise to the water's surface or, if denser than water, adhere to the floor and walls of the structure in which the water is held. It would be advantageous if one end of the molecule could interact favourably with water as this would help to orient its 'tail' into the centre of an insoluble spherical molecular aggregate, or micelle. Molecules of this type with both hydrophilic and lipophilic character are classed as amphiphiles. A well-known example is the sodium salt of dodecylsulfonic acid, a linear 12-carbon chain ending in a water-soluble 'head group', that is found in many commercial detergents. Amphiphiles form micelles with their hydrophilic head groups found on the outer surface. Here they can interact favourably with water molecules while the 'tails' form the core due to their hydrophobicity. Unfortunately it is energetically disadvantageous to encapsulate water molecules at the hydrophobic micelle core and such a system would not be amenable to cell formation. Given

the right conditions, however, these molecules can form double layers, or bilayers, either 'head-to-head' or 'tail-to-tail'. If a large number of these polar hydrocarbons are present they may also form sheets in which the hydrophobic tails interdigitate and the head groups interact with water. At the edges of the sheet the molecules could form a curved edge with the head groups on the outside but this is energetically costly. A far better arrangement is for the sheet to fold up into a sphere with polar head groups covering the external and internal faces. This vesicle structure, also known as a liposome, and has the advantage of forming in water and being full of water yet with a membrane that is not water soluble. The membranes themselves thus have polar, hydrophilic internal and external faces but a hydrophobic region in between. It is therefore an ideal vessel to the house chemical components of a 'living' system.

The 'Lipid World' hypothesis states that polar hydrocarbons formed in a prebiotic Earth, or originated from extraterrestrial meteoric sources, and then went on to aggregate into vesicles. These vesicles then capture chemical species at random: in some cases the concentrating effect of the vesicle would facilitate chemical reactions and some of these would eventually lead to self sustaining chemical reactions. Eventually protein-based enzymes would emerge that could synthesize lipids and the entire system would then become symbiotic.

So, in our predominantly aqueous world, it makes sense to use a water-impermeable material to act as a boundary membrane. Where could the molecules from the membrane have originated? It is likely that the earliest cells were formed from lipid-like molecules that were found naturally in small pools of water. The reasoning behind this is that only in isolated, evaporating pools could potentially biotic molecules be concentrated enough to interact, a general requirement for biogenesis. If concentrations of reactants are very low, the chance of two molecules meeting, let alone reacting, is vanishingly small. To increase the concentration, liquid pools containing the reactants must evaporate. Alternatively some compounds could attach to a solid surface. This would reduce the degrees of freedom available to the molecules; instead of moving freely in solution some would be fixed and await a suitable reaction partners.

Much thought has been devoted to the 'chicken and egg' problem of the origins for biological building blocks and membranes are no exception. How could chemical reactions form amphiphilic lipids on a large enough scale unless the reactions themselves occurred within an enclosed space? There must be some non-biological processes capable of synthesizing membrane components. What types of non-biological reactions could produce the hydrocarbons necessary to form a membrane? Reduction of carbon dioxide by iron-nickel alloys can form methane but not higher hydrocarbons:

$$CO_2 + 4H_2 \rightarrow CH_4 + 2H_2O$$

In water containing large concentrations of dissolved carbon dioxide and hydrogen a variant of the inorganic Fischer-Tropsch reaction can produce hydrocarbons by the following reaction:

$$nCO_2 + (3n + 1)H_2 \rightarrow C_nH_{(2n+2)} + 2nH_2O$$

The Earth's early atmosphere provides the carbon dioxide. Several reactions could be the source of the hydrogen. Hydration of the mineral olivine to serpentine releases hydrogen as a by-product, as does the oxidation of methane and gasification of carbon.

$$6(Mg_{1.5}Fe_{0.5})SiO_4 + 7H_2O \rightarrow 3\,Mg_3Si_2O_5(OH)_4 + Fe_3O_4 + H_2$$

$$2CH_4 + O_2 \rightarrow 4H_2 + 2CO$$

$$C + H_2O \rightarrow H_2 + CO$$

All these reactions require high temperatures, high pressures, or catalysts. In a prebiotic atmosphere it is difficult to see how oxygen and silicates would have formed, though geochemical activity at great ocean depths or vulcanism would seem likely mechanisms. It has been shown that some low molecular weight hydrocarbons from deep sea vents can be formed by non-biological processes through analysis of the $^{13}C$ data and D/H ratios. Perhaps more surprising are the estimates of terrestrial carbonaceous material originating from comet and meteorite strikes which reach values of up to $10^{19}$ tonnes. Analysis of the Murchison meteorite, for which the carbon content is approximately 2%, showed that both polar and non-polar hydrocarbons were present in significant quantities. The conclusion drawn from these results is that chemistry occurring within, or on the surface of, comets and meteors is capable of synthesizing compounds that could self-assemble into micelles and vesicles. Whatever the exact process involved on early Earth, it may well have been in operation in prebiotic eras and instrumental in forming simple lipids in local concentrations high enough to form vesicles and micelles.

There are two interesting aspects of the Lipid World that merit further consideration. Firstly, until regulated cell division occurred it is likely that most of the new cells did not contain enough of the components to continue the cycle of replication. Once the 'nutrients' had been exhausted the cell would stop functioning. Alternatively, if nutrients were available, and could be transported through the membrane, the cell would presumably grow to the limit of its encapsulating membrane. Eventually this would burst and the contents would disperse. In a lipid-rich environment components of that first protocell could have been captured by other vesicles leading to a primitive mechanism of cellular proliferation. Prebiotic equivalents of cytokinesis and endocytosis may therefore have been responsible for the emergence of cellular-based self-replicating processes. Secondly, newly-formed 'second generation' cells may encapsulate new chemical species. If they interfered with the symbiotic workings inherited from the original cell these cells would fail to proliferate. Other cells may have been formed in the presence of new molecules that added beneficial functionality. These cells could be thought of as having evolved to include

improved properties. If several parallel examples of this process occurred it would lead either to competition between the differentiated cell types or, conceivably, the synergy associated with multicellular systems.

### 2.5.2 Catalysis: The Iron-Sulfur World

The Lipid World hypothesis is attractive because it offers a solution to the problem of widely dispersed nutrients required to sustain replication processes: localized concentration. Another way to increase the possibilities of molecular collisions, and thus chemical reactions, is to constrain one of the species. This is most easily achieved if one molecule adheres to a surface. There has been much discussion about lamellar clay-like minerals being good sites to host surface-bound molecules. Their preferred growth patterns have been suggested as reasons why particular chiral forms of organic compounds are more abundant than their enantiomers. What lamellar minerals do not obviously provide is any catalytic assistance that might enhance the production of new compounds from their surfaces. One mineral that could be a candidate is iron sulfide. Wächtershäuser has proposed that iron pyrites (iron(II) sulfide, $FeS_2$, or 'fool's gold') could have been responsible for organizing and catalysing chemical reactions in a prebiotic Earth [43]. Before the advent of photosynthesis it is unlikely that there was a significant concentration of oxygen in the atmosphere so complex minerals of the type existing today (sulfates, carbonates, phosphates, etc) would be rare, if they existed at all. While organic compounds could incorporate oxygen through reaction with carbon monoxide and dioxide, both of which are formed by geothermal activity, metal oxidation is far harder to accomplish. The formation of metal sulfides in a prebiotic sea is energetically possible under reducing conditions and could therefore have supplied a reactive surface for small molecules. Iron pyrites, in particular, has many fascinating aspects that may implicate it in a number of essential synthetic processes. Pyrites crystals often exhibit surface striations as a consequence of the interface between two crystalline forms and so fit the requirement of a lamellar structure. The crystalline surfaces act as Lewis acids and can accept anionic species through ionic interactions. Many of the functional groups associated with biologically important molecules are anionic – nucleic acids, coenzymes, several amino acids – and would be attracted to such a surface. Lastly, iron-sulfur clusters still play a major role in most organisms today. Is it not unreasonable to assume that their presence is a vestigial reminder of a time when all life depended on an iron sulfide surface?

### 2.5.3 Self-Replication: The RNA World

One of the major problems facing researchers looking for origins of 'living' processes is to make sense of how our current DNA-protein symbiosis evolved. DNA is an extremely efficient method of recording and replicating a vast amount of information, albeit with the occasional error, yet is essentially a linear molecule.

Alone, it lacks the ability to transfer its information into another form and so makes a wonderful repository but has no other function. Through the medium of RNA and the complex activity of ribosomes, information stored in DNA is translated into proteins that proceed to perform all of the many tasks required of the organism. Some of these involve catalytic activities in which the specific functionality of the proteins and their spatial arrangement allow them to make or break chemical bonds in other molecules. Crucially DNA replication requires several catalytic proteins, such as helicase to separate DNA strands and polymerase to create a new complementary strand, so the question arises, how can non-catalytic DNA catalyse protein formation while the proteins, unable to self-replicate, are essential to DNA synthesis? One possibility is that protein-like molecules emerged on Earth and were able to use their conformational flexibility and many weak interactions (hydrogen bonding, disulfide formation, π–π stacking and van der Waals forces) to catalyse self-replication, avoiding the need for DNA as a replication mechanism. Another possibility is for DNA to fold in a non-linear fashion, creating pockets in which metal-assisted autocatalysis could occur. Unfortunately neither of these scenarios is particularly appealing and they raise as many questions as they answer. If proteins can self-replicate why would DNA arise? If DNA was autocatalytic why would proteins be necessary?

A pleasing solution to this conundrum was discovered in 1986 by Westheimer who demonstrated that the mediator between DNA and proteins, RNA could have enzymic activity [44]. These ribonucleic acid enzymes, or ribozymes, are now known to catalyse the formation of RNA which can, in turn, catalyse protein formation. The term 'RNA World' to describe this phenomenon was coined by Gilbert writing in *Nature* [45] in the week following the publication of Westheimer's results. Gilbert's insight was that, as self-catalytic RNA could provide both a mechanism for information storage and replication, RNA alone would have been necessary for 'living' processes to start on Earth. In fact, Westheimer's experiments only confirmed similar ideas proposed by Woese and Rich who had also considered an RNA-only scenario in the 1960's [46].

As we have seen earlier in this chapter, DNA is more robust than its RNA equivalent due to the greater reactivity of the hydroxy group attached to the sugar that links the base to its phosphate backbone. For information to be encoded and transmitted down the generations DNA is a better medium, however, its relative inflexibility precludes it from forming the complex tertiary structures required of a catalyst. RNA can fulfill the latter function but does not to have long term stability as it is formed from the DNA template. Therefore, despite this elegant solution to the DNA-protein problem, there is still considerable scepticism surrounding synthetic routes to RNA in a prebiotic Earth.

Nevertheless, the RNA World hypothesis would seem to answer most of the questions raised above: how can an essentially linear molecule be autocatalytic, how can it synthesize proteins, and how can it replicate? Autocatalysis can occur because RNA can adopt a wide range of secondary and tertiary structures that position RNA monomers into a preorganized sequence and link them together, it can apply the same flexibility to bind other small molecules and catalyse their polymerization, and it could form weakly interacting sense and antisense duplexes. The main problem is

how such a complex molecule, comprising a sugar, a phosphate and a heterocyclic aromatic group, could possibly be synthesized. Sugars can form from formaldehyde under alkaline conditions and thermodynamics will dictate that the more stable hexoses and pentoses are available to couple with the other components. Phosphate derivatives occur naturally under a variety of conditions, though solubility of phosphate minerals is generally low under ambient conditions. Finally, the nucleobases, and many other related heterocycles, are known to form from cyanide and acetylene in water. So far, so good. Unfortunately, when it comes to joining the components together it seems that the necessary reaction pathways are unavailable.

A route to the pyrimidines, cytosine and uracil, has been found by Sutherland and colleagues that has as its intermediary a small heterocycle called 3-aminooxazole which forms through the reaction of glycolaldehyde and cyanamide [47]. This can react with glyceraldehyde to form a derivative of arabinose. Subsequent reaction with cyanoacetylene and dihydrogen phosphate yields a β-ribocytidine-2′,3′-cyclic phosphate. A final photochemical step leads to a rearrangement that generates the cytosine and uracil nucleotides, including the essential sugar and phosphate backbone, while destroying all other derivatives, as shown in Fig. 2.17.

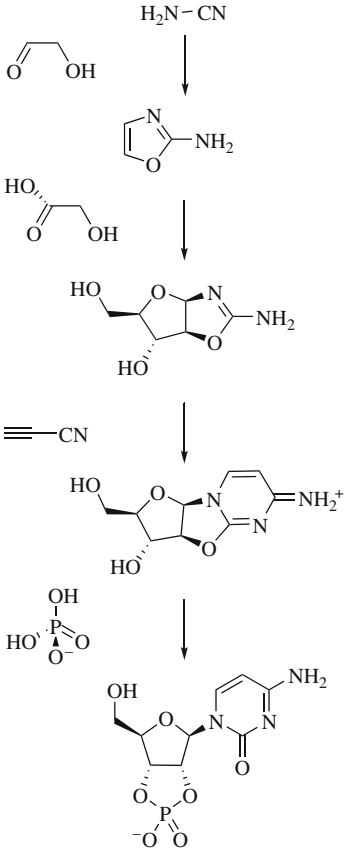



**Fig. 2.17** A simple pyrimidine synthesis [47]

An alternative to the terrestrial synthesis of the nucleobases is to invoke interstellar chemistry. Martins has shown, using an analysis of the isotopic abundance of $^{13}C$, that a sample of the 4.6 billion year old Murchison meteorite which fell in Australia in 1969 contains traces of uracil and a pyrimidine derivative, xanthine. Samples of soil that surrounded the meteor when it was retrieved were also analyzed. They gave completely different results for uracil, consistent with its expected terrestrial origin, and xanthine was undetectable [48]. The isotopic distributions of carbon clearly ruled out terrestrial contamination as a source of the organic compounds present in the meteorite. At 0°C and neutral pH cytosine slowly decomposes to uracil and guanine decomposes to xanthine so both compounds could be the decomposition products of DNA or RNA nucleobases. They must have either travelled with the meteorite from its extraterrestrial origin or been formed from components present in the meteorite and others encountered on its journey to Earth. Either way, delivery of nucleobases to a prebiotic Earth could plausibly have been undertaken by meteors. The conditions that formed the bases need not have been those of an early Earth at all but of a far more hostile environment elsewhere in the Solar System. That environment may have been conducive to the production of individual bases but they may never have been able to form stable DNA or RNA polymers: this development may have required the less extreme conditions prevalent on Earth.

## 2.6 Supramolecular Biology and Synthetic Biology

Supramolecular biology has been interpreted in many ways over the past half century being most closely associated with the structural effects of intracellular water. It seems much more holistic to define it as encompassing the same phenomena encountered in supramolecular chemistry but set in a biological context. Thus the construction of biological structures, whether they be simple phospholipid bilayers or complex motors powering bacterial flagellae, can be considered to be akin to the supramolecular assembly that occurs when carefully designed chemical components are brought together.

One principle behind supramolecular chemistry, the ability to design molecules that incorporate motifs intended to perform specific functions, has found its way into an emerging theme in the life sciences: synthetic biology. Synthetic biology, a term coined by Szybalski, refers to the introduction of ‘foreign’ genetic material into living systems to impart a new function [49]. One example of this would be the introduction of the gene that codes for green fluorescent protein (GFP), a small protein of 238 amino acids found naturally in some jellyfish, into other organisms. The discovery and subsequent use of GFP led to Chalfie, Shimomura and Tsien being jointly awarded the Nobel Prize in chemistry for 2008. As a direct result of their pioneering work it has become possible to introduce the protein through a variety of methods, including viral vectors, to anything from a bacterium to a pig. Consequently it is possible to label certain cells or whole animals so that they may easily be identified when mixed with controls. This has proved extremely important when visualizing

tissue or when the subject is being observed to determine the effects of other genetic differences. Despite the valuable uses to which GFP expressing organisms have been put, their only real claim to fame is that they glow in the dark. The power of synthetic biology is that it can identify many genes and introduce them to develop functional organisms. As a result it is possible to design yeast that produces diesel rather than alcohol as a waste product [50] or bacteria that respond to the presence of explosives [51]. In another application of synthetic biology Benner and co-workers have investigated methods of expanding the four DNA nucleobases to six by including the complementary bases isoguanine and isocytosine [52]. While sequences incorporating these new bases can be reproduced by polymerase chain reaction methods there is no natural ribosome that can use the new 'code' to generate entirely new proteins. It seems that synthetic biology is as diverse in its outlook as supramolecular chemistry and the two fields have much in common.

## 2.7 Summary

The key issue in this chapter is how small molecules can interact through both supramolecular interactions (such as vesicle formation and hydrogen bonding) and more conventional covalent bonds (such as protein formation from individual amino acids and DNA polymerization). The complex products of these reactions yield a microenvironment, the cell, in which chemical reagents are concentrated within a lipid membrane so that they can interact frequently enough, and with sufficient energy, to form polymeric species capable of catalysis and information storage. As all three of these aspects (lipids, proteins, DNA for information storage and RNA for translation) are present in cells it is reasonable to assume that they must have formed in the prebiotic Earth. The order in which these components emerged and, indeed, the mechanisms required for the synthesis of their constituent monomers (amino acids, nucleotides and lipids) are unknowable at present but many plausible theories, often backed up by careful experimentation, have been developed.

Of course there is more to 'being alive' than pure chemical self-replication. At the very least there must be an ability to evolve in response to non-catastrophic local environmental changes. As long as those changes occurred on a timescale that allowed several generations to pass then a viable population would be expected to have evolved to exist under the new conditions.

One aspect that must be central to any research in this direction is a consideration of ethics. The potential to create 'Life 2.0' would have an immense global impact even if its initial application was merely to understand how simplified biological systems interact.

## References

1. Schrödinger E (1944) What is life?. Cambridge University Press, Cambridge
2. Kyte J, Doolittle RF (1982) A simple method for displaying the hydropathic character of a protein. J Mo Biol 157:105–132

3. Biou V et al (1994) The 2.9 Å crystal structure of T. thermophilus seryl-tRNA synthetase complexed with tRNA(Ser). Science 263:1404–1410 PDBID:1SER
4. Doyle DA et al (1998) The structure of the potassium channel: molecular basis of $K^+$ conduction and selectivity. Science 280:69–77 PDBID:1BL8
5. Fischer E (1894) Einflus der Configuration auf die Wirkung der Enzyme. Ber Dtsch Chem Ges 27:2985–2993
6. Koshland DE (1958) Application of a theory of enzyme specificity to protein synthesis. Proc Natl Acad Sci USA 44:98–104
7. Pearson JG et al (1997) Predicting the chemical shifts in proteins: structure refinement of valine residues by using ab initio and empirical geometry observations. J Am Chem Soc 119:11941–11950
8. Lange OF et al (2008) Recognition dynamics up to microseconds revealed from an RDC-derived ubiquitin ensemble in solution. Science 320:1471–1475
9. Mendel G (1865) Versuche über Plflanzen-hybriden read to the Brünn Natural History Society on February 8th and March 8th
10. Meischer F (1869) Letter I to Wilhelm His Tübingen February 26th 1869. In: W. His et al (eds) Die Histochemischen und Physiologischen Arbeiten von Friedrich Miescher—Aus dem wissenschaftlichen Briefwechsel von F. Miescher vol 1 FCW Vogel Leipzig pp 33–38
11. Avery OT, MacLeod CM, McCarty M (1944) Studies on the chemical nature of the substance inducing transformation of pneumococcal types. J Exp Med 79:137–158
12. Levene PA (1919) The structure of yeast nucleic acid. J Biol Chem 40:415–424
13. Vischer E, Zamenhof S, Chargaff E (1949) Microbial nucleic acids: the desoxypentose nucleic acids of avian tubercle bacilli and yeast. J Biol Chem 177:429–438
14. Watson JD, Crick FHC (1953) Molecular structure of nucleic acids: A structure for deoxyribose nucleic acid. Nature 171:737–738
15. Gamow G (1954) Possible relation between deoxyribonucleic acid and protein structures. Nature 173:318
16. Nirenberg MW, Mattaei JH (1961) The dependence of cell-free protein synthesis in E coli upon naturally occurring or synthetic polyribonucleotides. Proc Natl Acad Sci USA 47: 1588–1602
17. Singleton MR, Scaife S, Wigley DB (2001) Structural analysis of DNA replication fork reversal by RecG. Cell. 107:79–89 PDBID:1GM5
18. Temin HM, Mizutani S (1970) RNA-dependent DNA polymerase in virions of Rous sarcoma virus. Nature 2261211–1213
19. Baltimore D (1970) RNA-dependent DNA polymerase in virions of RNA tumour viruses. Nature 226:1209–1211
20. Hadden JM et al (2007) The structural basis of Holliday junction resolution by T7 endonuclease I. Nature 449:621–624 PDBID:2PFJ
21. Dawkins R (2006) The selfish gene, pp 13–20. Oxford University Press, New York, NY
22. Drexler KE (2001) Machine-phase nanotechnology. Sci Am 285:66–67
23. Lehn JM et al (1987) Spontaneous assembly of double-stranded helicates from oligobipyridine ligands and copper(I) cations – structure of an inorganic double helix. Proc Natl Acad Sci USA 84:2565–2569
24. Kramer R et al (1993) Self-assembly, structure, and spontaneous resolution of a trinuclear triple-helix from an oligobipyridine ligand and Ni(II) ions. Angew Chem Int Ed Engl 32: 703–706
25. Quayle JM, Slawin AMZ, Philp D (2002) A structurally simple self-replicating system. Tetrahedron Letts 43:7229–7233
26. von Neumann J (1961) The general and logical theory of automata. In: Taub AH (ed) John von Neumann collected works Vol V. Pergamon Press, New York, NY
27. Ibid p 315
28. Wyler R, De Mendoza J, Rebek J (1993) A synthetic cavity assembles through self-complementary hydrogen-bonds. Angew Chem Int Ed Engl 32:1699–1701

29. Arduini A et al (1995) Calix[4]arenes blocked in a rigid cone conformation by selective functionalization at the lower rim. J Org Chem 60:1454–1457
30. Ajami D, Rebek J (2008) Gas behavior in self-assembled capsules. Angew Chem Int Ed Engl 47:6059–6061
31. Rudzevich V, Rudzevich Y, Böhmer V (2009) Dimerization and self-sorting of tetraurea calix[4]arenes. Synlett 1887–1904
32. Cram DJ et al (1992) Host-guest complexation 62. Solvophobic and entropic driving forces for forming velcraplexes, which are 4-fold, lock-key dimers in organic media. J Am Chem Soc 114:7748–7765
33. Liu S, Gibb BC (2008) High-definition self-assemblies driven by the hydrophobic effect: synthesis and properties of a supramolecular nanocapsule. Chem Commun 3709–3716
34. Schall O, Gokel GW (1994) Molecular boxes derived from crown ethers and nucleotide bases: probes for Hoogsteen vs Watson-Crick H-bonding and other base-base interactions in self-assembly processes. J Am Chem Soc 116:6089–6100
35. Conn MM, Rebek J (1997) Self-assembling capsules. Chem Rev 97:1647–1668
36. Timmerman P et al (1997) Noncovalent assembly of functional groups on calix[4]arene molecular boxes. Chem Eur J 3:1823–1832
37. Rebek J (1994) Synthetic self-replicating molecules. Sci Am 271:48–55
38. Siegel S et al (2000) Molecular recognition of a dissolved carboxylate by amidinium monolayers at the air-water interface. Progr Coll Poly Sci 115:233–237
39. von Kiedrowski G (1986) A self-replicating hexadeoxyribonucleotide. Angew Chem Int Ed Engl 25:932–935
40. Sievers D, von Kiedrowski G (1994) Self-replication of complementary nucleotide-based oligomers. Nature 369:221–224
41. Lee DH et al (1996) A self-replicating peptide. Nature 382:525–528
42. Sadownik J, Philp D (2008) A simple synthetic replicator amplifies itself from a dynamic reagent pool. Angew Chem Int Ed Engl 47:9965–9970
43. Wächtershäuser G (1992) Groundworks for an evolutionary biochemistry – the iron sulfur world. Progr Biophys Mol Biol 58:85–201
44. Westheimer FH (1986) Nature 319:534–536
45. Gilbert W (1986) The RNA world. Nature 319:618
46. Woese C (1968) The genetic code. Harper & Row, New York , NY
47. Powner MW, Gerland B, Sutherland JD (2009) Synthesis of activated pyramidine ribonucleotides in prebiotically plausible conditions. Nature 459:239–242
48. Martins Z et al (2008) Extraterrestrial nucleobases in the Murchison meteorite Earth Planet. Sci Lett 270:130–136
49. Szybalski W, Skalka A (1978) Nobel prizes and restriction enzymes. Gene 4:181–182
50. Kalscheuer R, Stölting T, Steinbüchel A (2006) Microdiesel: *Escherichia coli* engineered for fuel production. Microbiol 152:2529–2536
51. Smith RG, D'Souza N, Nicklin S (2008) A review of biosensors and biologically-inspired systems for explaosives detection. Analyst 133:571–584
52. Sismour AM, Benner SA (2005) The use of thymidine analogs to improve the replication of an extra DNA base pair: a synthetic biological system. Nucleic Acids Res 33:5640–5646

# Chapter 3
# Artificial Cells

## 3.1 Cells as Capsules

Nature is full of complex constructions designed to encapsulate molecules within a defined space. At one extreme there are cells, enclosed within phospholipid bilayers, and sea dwelling algae like *Emiliania huxleyi* that are surrounded by calcium carbonate coccoliths which overlap to form an exoskeleton. On a smaller scale are iron storage proteins such as ferritin that are amorphous in the absence of metal ions but form globular structures once metals are bound.

As will be shown, model systems for cells employing lipids or composed of polymers have been in existence for some time. Model systems for coccolith-type structures are well known on the nanoscale in inorganic and materials chemistry. Indeed, many complex metal oxides crystallize into approximations of spherical networks. Often, though, the spherical motif interpenetrates other spheres making the formation of discrete spheres rare. Inorganic clusters such as quantum dots may appear as microscopic spheres, particularly when visualized by scanning electron microscopy, but they are not hollow, nor do they contain voids that would be of value as sites for molecular recognition. All these examples have the outward appearance of cells but not all function as capsules for host molecules.

One of the driving forces behind the development of self-assembling supramolecular capsules is their potential as delivery systems for guest drug molecules as will be seen in a later chapter. Central to the formation of these three dimensional structures is the mechanism of aggregation and folding into the correct shape. While natural and artificial capsules may be composed of very different building blocks and be very different in scale it is worthwhile considering what is known about the formation of natural structures as the underlying principles may be very similar.

## 3.2 Natural Capsules

The largest natural capsule of relevance to this chapter is of course the cell itself and the method of encapsulation is by the membrane. Disregarding the myriad of proteins, steroids and other small molecule that inhabit the membrane, it is mainly

P.J. Cragg, *Supramolecular Chemistry*, DOI 10.1007/978-90-481-2582-1_3,

© Springer Science+Business Media B.V. 2010

composed of phospholipids with very similar chemical structures. The formation of cell membranes has been discussed previously but it is worth remembering that the driving force for lipid bilayer aggregation is essentially an effort to reduce energy. This is accomplished in two ways. On one scale the lipids align so that groups with similar hydrophobicities are adjacent and, to reduce unfavourable interactions between the lipophilic substituents, they form bilayers with more polar groups on the external faces. On a larger scale the entire bilayer folds to reduce surface energy and attempt to adopt the most energy efficient shape possible, that of a sphere. The same energetic considerations exist for other biological and non-biological capsules though other factors control the architectures that arise in each case. Few capsules have the flexibility and surface homogeneity that is seen in an idealized cell membrane so geometric alignment and subunit complementarity become more important.

### *3.2.1 Clathrins*

Clathrins, first observed by Pearse, are structures with pentagonal and hexagonal faces that can form three dimensional structures with the same truncated icosahedral morphology as a $C_{60}$ molecule or a soccer ball [1]. The vertices of the clathrin coating are formed by three intertwined proteins composed of heavy and light chains that emanate from a central hub in a structure known as a triskelion, shown in Fig. 3.1.

Strands from different triskelia intertwine to form the edges of the structure with the hubs at the vertices. The protein framework encapsulates a vesicle that started out by budding from the plasma membrane or Golgi apparatus. The bud attracts adaptor proteins that in turn interact with the clathrin triskelia. Once the budding is

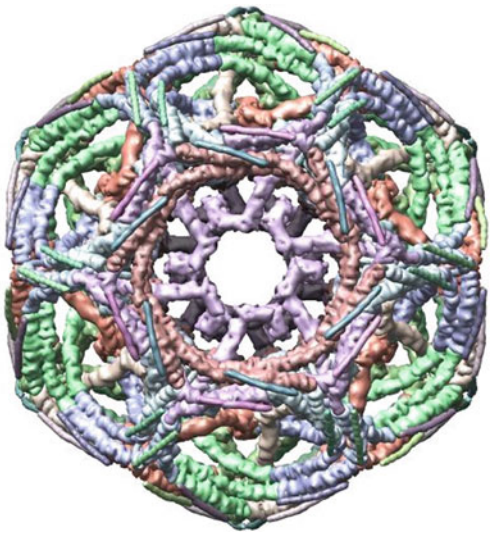


**Fig. 3.1** A clathrin [2]

complete the vesicle is removed from the surface by the clathrin and is employed to carry materials around the cell. The clathrin framework disassembles to release the vesicle but the triskelia can subsequently reform. This process is highly relevant to artificial capsule formation. In vitro the formation of clathrin capsules from solution is pH dependent with the rates fastest at low pH which would indicate, unsurprisingly, that hydrogen bonding is of importance. In vivo the assembler proteins seem to be responsible for bringing the intertwined triskelia strands together.

### *3.2.2 Viral Capsids*

One way to proceed when designing a cell de novo is to refer back to existing biological systems. Anyone interested in the design elements required for a supramolecular capsule need look no further than the capsid viruses. These, some of the smallest sub-cellular entities, are the capsules that transport and deliver viral DNA or RNA. These are generally composed of a number of identical protein subunits that link together in a predetermined manner to encapsulate the oligonucleotides. The capsule is composed of interlocking protein plates and makes great use of symmetry to reduce the number of building blocks required. An important facet of these protein spheres is that they can open their triangular doors in response to specific stimuli and deliver their payload.

Two particular architectures predominate: the icosahedral structures with triangular faces, described above, that approximate to the low energy geometry of a sphere, and helical structures. In the former the genetic material is folded up inside the capsule whereas in the latter it lies down the axis of the helix. As a potential model for cellular development the icosahedral architecture is clearly of greater interest. Here three proteins join to make a planar triangular structure. Twenty of these then aggregate and fold along protein hinges to adopt as close to a spherical structure as possible resulting in an icosahedron, illustrated in Fig. 3.2. The capsid proteins are

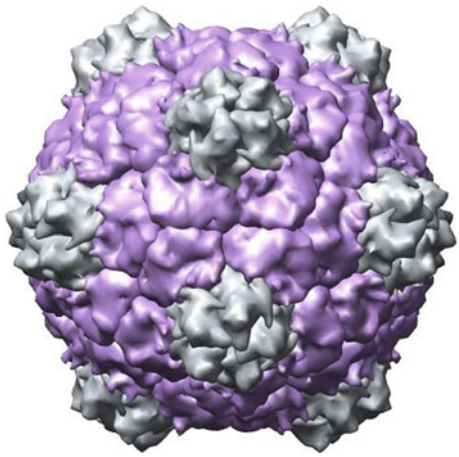


**Fig. 3.2** A capsid virus [3]

often covered in a further protein coat, decorated with glycoproteins, or by a lipid membrane sequestered from the host organism.

Some of the most interesting work on the dynamics of capsid formation has come from Zlotnick [4]. He has proposed that the process does not require a complex biological template but can proceed from an initial contact between two protein monomers. Following this hypothesized 'nucleation' event, which will occur readily at almost any protein concentration through random interactions between monomers in solution, it had been proposed that hexamers form which rapidly assembled into viral capsules. However, there is also evidence from experiments on cowpea chlorotic mottle virus that dimers polymerize under slightly acidic conditions to form pentamers. The disappearance of dimers, based on light scattering data, follows a sigmoidal trajectory consistent with the observed species being consumed through some form of self-association but only when the concentration data are fitted to the fifth power. Energetically pentamers are more strained than hexamers and, based on electron microscopy, curve slightly. Consequently, unlike planar hexamers, pentamers are likely to form non-planar aggregates with each other. Once curvature has been initiated it is likely to continue until a pseudospherical structure arises.

Similar attempts to minimize surface energy are also seen throughout the natural world, from water drops to cells, and in the chemical world where fullerenes and boranes, capsular compounds formed of boron hydride networks, both form hollow shells.

### *3.2.3 Coat Proteins*

Icosahedral capsid viruses and clathrins are examples of coat proteins of which there are many. Another example that has been extensively studied is coat protein II, or COPII, which is composed of an inner cage and outer coat [5]. The inner cage is a cuboctahedron approximately 60 nm across. It has square and triangular faces which can only be constructed if four protein strands emanate from the structure's hub, rather than the three seen in clathrins. It also transpires that the proteins interact with each other at the vertices without any of the extensive interdigitation seen in clathrin cages.

In contrast to the COPII cage, the outer coat is an icosidodecahedron composed of pentagons and triangles. With a diameter of 100 nm it is able to encapsulate the cage and its contents. COPII carries materials downstream from the endoplasmic reticulum to cell surfaces and other routes of excretion.

### *3.2.4 Vault Proteins*

One of the most unusual and fascinating groups of protein structures are those known as vaults, first identified in 1986 by Rome and Kedersha [6], which exist in almost all organisms. They are composed of copies of a single peptide that

**Fig. 3.3** Vault protein architecture [7]

self-assembles so that distinct regions, known as the shoulder, cap helix and cap ring domains, align to give a hollow cup shape. Two of these structures meet at the wider end to give a barrel with caps on both ends. Consequently, and in stark contrast to many other proteins, vaults are huge with masses around 13 MDa. Low level X-ray structures and electron microscopy indicate that some human vaults are composed of 96 proteins with 48 of the monomers forming each half of the overall structure.

Other vaults are subtly different. The rat liver vault is composed of 78 protein monomers in total, with 39 in each half, to give the highly unusual 39 fold symmetry, shown in Fig. 3.3. They have an overall length of 67 nm, of which the main body accounts for 35 nm and the shoulder and cap regions the remainder, and a greatest width of 40 nm. The cap ring region is about 13 nm across and the opening at the cap ring, 5 nm. The protein wall is 1.5–2.5 nm thick. Rat liver vaults are known to comprise vault RNA (vRNA) strands of 141 bases and three proteins [8]. Of these, one is the major vault protein (MVP) that forms the vault itself, and the other two a telomerase associated protein (TEP1) and a vault poly(ADP-ribose) polymerase (VPARP). Multiple copies of all the components are present with TEP1 and VPARP found inside the vault formed by MVP. The exact purpose of these vaults is unknown although it is speculated TEP1 and VPARP appear to be involved in the transport of genetic material across the nuclear envelope.

## 3.3 Unnatural Capsules

The earliest molecules designed specifically to encapsulate other species were of a type now known as cryptands, including sepulchrates and sarcophagenes, and have been described in Chapter 1. These compounds make use of two bridgehead atoms, often carbon, nitrogen or boron, which are linked by three molecular strands. These molecules are intentionally quite flexible to allow small guests to enter but, once inside, donor atoms within the strands surround the guests with numerous weak

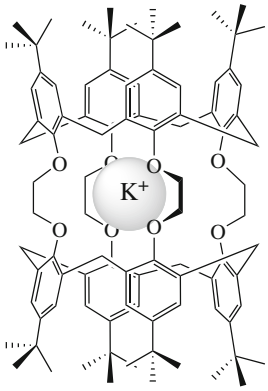



**Fig. 3.4** A calixtube [12]

interactions to keep them in place. As a consequence the guests are bound very strongly. Given the extremely high stability constants of these complexes it may be more appropriate to replace the term 'host-guest' with 'host-hostage'.

Similar encapsulating molecules have been prepared by the Schmidtchen [9] and Lehn [10] groups based on molecular tetrahedra with bridgehead atoms at the four vertices. A separate class of capsules can be made by linking two macrocycles, as exemplified by Cram's cavitands [11] and Beer's calixtubes [12], where a narrow annulus exists at both ends of the capsule that is usually too narrow to allow guests to pass through. These molecular capsules, for example the calixtube shown in Fig. 3.4, have two things in common: they are composed entirely of covalent bonds formed by normal synthetic methods and they have very small internal volumes. To make larger capsules an alternative, supramolecular strategy is required.

### *3.3.1 Self-Complementary Capsules*

One of the simplest methods to design capsules capable of opening and releasing their contents is to construct a hydrogen bonded supramolecule composed of self-complementary subunits. If this is to approach a spherical geometry two strategies can be adopted, one requiring a hemispherical component and another that assembles like a tennis ball. Few simple molecules are hemispherical so that examples based on this design are generally prepared from macrocycles that taper slightly and can be derivatized at the wider rim. The tennis ball model has been explored extensively by the Rebek group as discussed in Chapter 2. In this approach short, tape-like molecules that curve in a semicircle are synthesized. An array of complementary hydrogen bonding groups is incorporated at the ends of the molecules to encourage self assembly.

In reality the compounds prepared lack true curvature and lead to molecular tetrahedra but without points or, to use the Archimedean terminology, truncated tetrahedra. One aspect of these simple capsules is that it is possible to investigate the encapsulation of small molecules. Not only can the energetics of capsule

assembly and disassembly be monitored but the number and orientation of guest molecules can be examined. The number of guests influences the equilibrium between monomers and dimers, which can be controlled by external forces, and the guest orientation shows how guest properties can dictate where they reside within the capsule.

### *3.3.2 Boxes with Metal Hinges*

Using a knowledge of coordination chemistry is possible to design ligands with convergent or divergent donor groups that have an affinity for particular metals. These metals in turn will have a preference for specific coordination environments: tetrahedral, square planar or octahedral, illustrated in Fig. 3.5. The combination of the two will dictate the structures that it is possible to create. So for 2,2′-bipyridine the nitrogen donors converge at an acute 60° angle on a metal, for 3,3′-bipyridine the nitrogen atoms diverge at a 120° angle and for 4,4′-bipyridine they diverge at a 180° angle.

These simple motifs, or slight variations on them, can be combined in programmed ways to create supramolecular tetrahedra, cubes and more complex species with metal atoms forming the vertices and ligands the edges or, in some cases, the faces of the capsules. By increasing the distances between donor atoms, and keeping rigid ligands, a series of homologous capsules with increasing volumes can be prepared and tailored to guests that match the shape and electrostatic charges of the internal void. Furthermore, the numerous weak interactions that formed the capsules are reversible so that guest release can be triggered by simply changing external conditions such as pH. This strategy has been adopted by the research groups of Raymond, Thomas, Fujita, Nitschke and others.

A highly original aim, synthesizing a supramolecular cube, was realized by Thomas and co-workers [13] who linked metal corners with bifunctional, rigid, linear spacers. The complete the metals' coordination environment the corners were capped with [9]ane-$S_3$ macrocycles. There is a central cavity but it is too small to encapsulate molecules of the size of proteins or genetic material.

The $M_4L_6$ tetrahedral capsule motif was first exploited by Saalfrank and co-workers [14] who produced tetrahedral species from bis(acetylacetonate) derivatives, with binding sites that diverged by 60°, and $Mg^{2+}$ cations. Salts formed

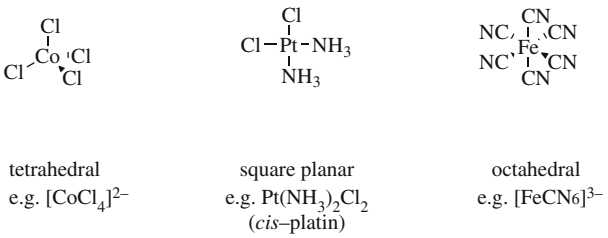



**Fig. 3.5** Common geometries in coordination chemistry

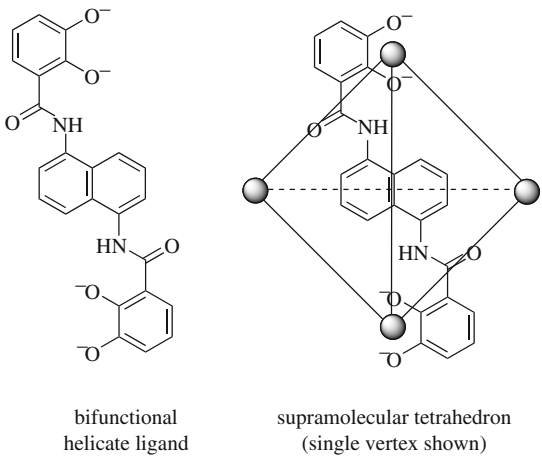
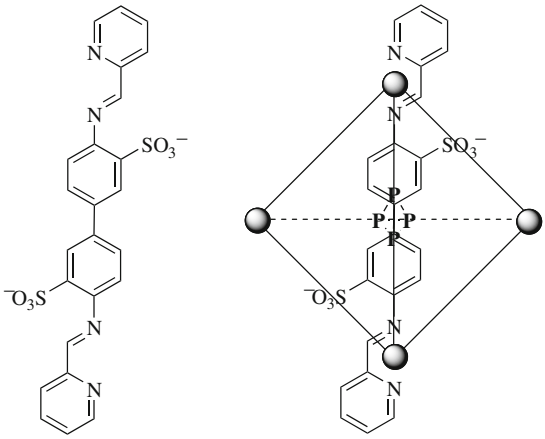



**Fig. 3.6** A cage formed from four metals and six ligands [15]

with ammonium cations but the anionic tetrahedra were too small to encapsulate any guests.

The Raymond group used a rigid linear bis(catechol) ligand to bind metals that adopt six fold coordination environments [15]. Iron(III) is well known to bind to the bidentate catechol ligand and was one of the first metals to be used to initiate cage formation. As the ligand's catechol termini face in opposite directions, as seen in Fig. 3.6, they can impart chirality to the capsule and, once the first ligand has bound, the overall chirality becomes fixed. All the ligands must bind in the same way if a capsule is to form: any deviation from this leads to open structures. Consequently when the capsules form the chirality around each metal is either $\Delta\Delta\Delta\Delta$ or $\Lambda\Lambda\Lambda\Lambda$. Each of the 12 catechols has a –2 charge giving an overall charge of –24. The four metals have a +3 charge each making the capsule anionic with an overall –12 charge. This also means that, with suitable counter ions, it is water soluble. The size of the ligands gives the capsule a considerable internal volume, in the order of 400 $Å^3$, so that guests such as the tetraethylammonium cation can be encapsulated.

In a clever application of Raymond's approach, the Nitschke group has prepared a capsule using bidentate Schiff base ligands formed by reacting 4,4′-diaminobiphenyl-2,2′-disulfonic acid with two molecules of 2-formylpyridine [16]. A tetrahedral capsule forms when the ligands react with iron(II), though not zinc(II) or cobalt(II), but does not bind tetrahedral cations as the internal void is only in the region of 140 $Å^3$. Small organic molecules including cyclopentane and cyclohexane can be included in the cavity and take up between 50 and 60% of the available space. The capsule, shown in Fig. 3.7 is pH sensitive and dissociates when acid is added. Usefully this is reversible and the capsule reforms when sodium bicarbonate is added making it mimic some of the 'capture and release' behaviour of natural capsules, though on a much smaller scale. Where this capsule has achieved something that natural capsules have not is in the capture of white phosphorus. This form of phosphorus comprises tetrahedra made up of four phosphorus atoms and




bifunctional helicate ligand

supramolecular tetrahedron (single edge shown) with included $P_4$

**Fig. 3.7** A cage to bind phosphorus [16]

is extremely air sensitive: it spontaneously combusts in air through rapid oxidation. $P_4$ tetrahedra are small enough to fit inside the capsule but the spaces between the ligands are too small to allow oxygen through thus preventing reaction. If oxides should form they would, in any case, be too large to fit inside the capsule. Despite the reactive nature of the guest the complex is even soluble in water and stable in air. If the complex is added to a two phase mixture of water and benzene, or cyclohexane, the organic solvent displaces $P_4$which then reacts with water to form phosphoric acid.

Fujita's group has produced a number of molecular capsules through careful consideration of transition metal binding requirements [17]. Threefold coordination around a metal can lead to ligands diverging at 90° or 109.5° to generate tetrahedral or cubic capsules, respectively.

Larger capsules prepared by the same group have been used extensively as nanoreactors in which exotic chemical species can be formed. The group has also used a molecular 'panel' approach as shown in Fig. 3.8. These compounds are, however, too small for any biomimetic activity.

In an exciting combination of spontaneous supramolecular aggregation followed by programmed covalent bond formation, Hiraoka, Shionoya and colleagues have prepared much larger hexasubstituted benzene panels in which metal coordinating pyridyl groups alternate with alkyl chains that terminate in allyl groups [18]. Eight of these panels spontaneously form an octahedron in the presence of six $Pd^{2+}$cations which coordinate at the corners. The allyl groups are then covalently linked through the action of Grubbs' catalyst. The metal corners are removed by competitive ligands, such as ethylenediamine, to leave all eight panels covalently bound to each other at the edges but not the vertices. Reduction of the double bonds to more flexible single bonds allows the panels greater movement.

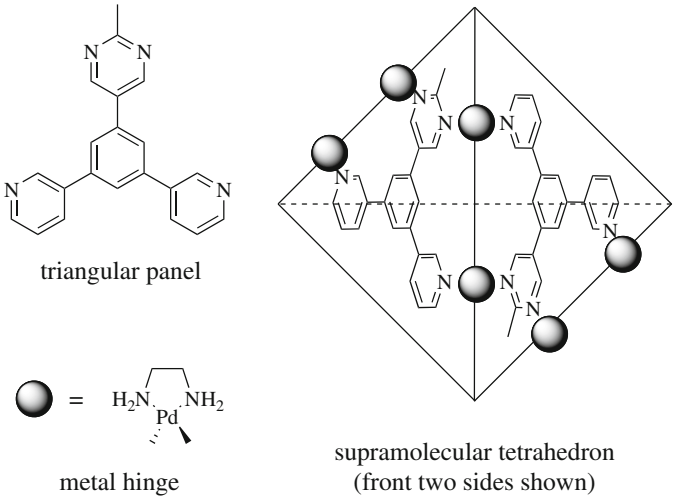



**Fig. 3.8** A cage formed from panels [17]

### *3.3.3 Capsules as Reaction Flasks*

The importance of encapsulation is to isolate molecules from environments in which they would be unstable and to store chemical species safely. Capsules also act as an arena for chemical reactions by concentrating the reagents needed for synthesis and aiding the formation of the desired products. On a large scale this can be represented by a cell. In the case of eukaryotic cells there are capsules within the cell each with its own function. Organelles exist to house and transcribe DNA, to remove waste, to generate energy and a host of other functions. Do artificial capsules have the same capabilities? The answer depends on the complexity of the capsule's function. Most synthetic capsules are only large enough to contain volatile solvent molecules though the Cram group was able to trap the highly reactive cyclobutadiene inside a hemicarcerand to illustrate the concept of a molecular reaction flask, as shown in Fig. 3.9 [19].

Larger capsules, such of those of the Rebek and Gibb groups, can hold larger molecules and have them react with each other in the confined space available. For example, Rebek has demonstrated that coencapsulation of a benzene molecule with a cyclohexadiene molecule accelerates the formation of the Diels-Alder cycloaddition product [20]. Often the shape of the capsule can promote the formation of one product over others due to the constraints of size and charge imposed by the internal surface. This can lead to a single isomer, or particular chirality, being favoured over others. In that sense the capsules do replicate a function of natural compartmentalization: the effect of encapsulating the chemical species controls the production of one product from many possibilities. However, the small size of the capsules limits the range of molecules that can be produced.

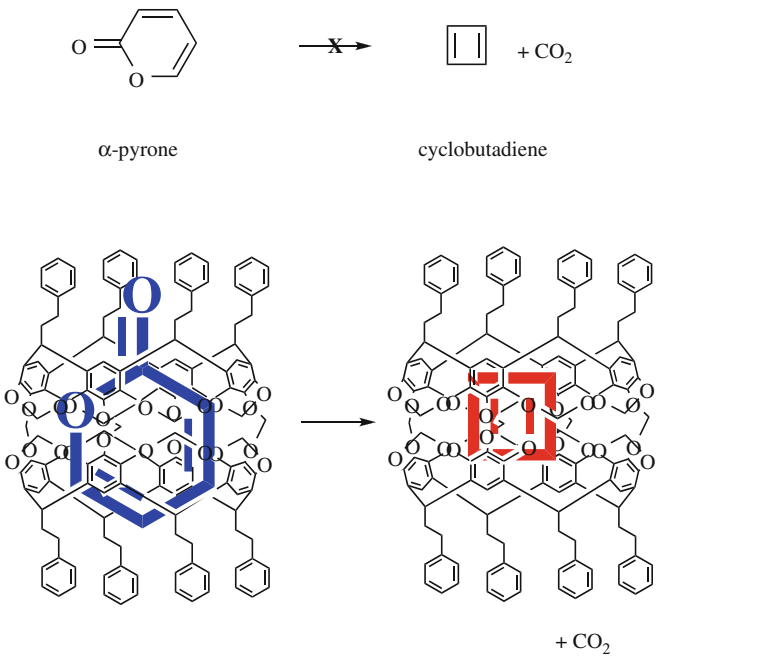



**Fig. 3.9** A cage to trap the reactive species cyclobutadiene [19]

### *3.3.4 More Complex Geometries*

Of course an icosahedron is not the only three dimensional design that can form a capsule. MacGillivray and Atwood proposed a structural classification for supramolecular assemblies based on the five Platonic and 13 Archimedean solids [21]. The Platonic solids, illustrated in Fig. 3.10, are the tetrahedron, cube, octahedron, dodecahedron and icosahedron.

Each face of a particular solid is the same shape: tetrahedra, octahedra and icosahedra are composed of equilateral triangles; cubes have square faces; dodecahedra are composed of regular pentagons. The Archimedean solids are composed of at

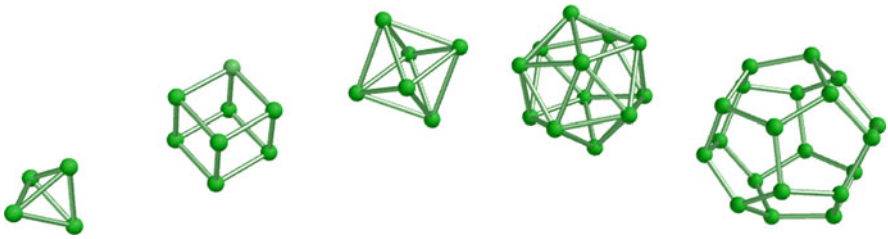


**Fig. 3.10** The Platonic solids: (*left* to *right*) tetrahedron, cube, octahedron, dodecahedron and icosahedron

least two different types of faces such as squares and triangles, pentagons and hexagons, or squares and pentagons. Further complexity is possible if prisms, such as a trigonal prism composed of three squares or rectangles and two triangles, and their related antiprisms are considered. The ultimate goal is to reduce the surface energy of the resultant polygon by approximating a sphere.

## 3.4 Synthetic Cells

In a review of the biotechnological potential of minimal cells Pohorille and Deamer describe seven properties that an artificial cell must have to function [22]. The entire process must be encapsulated within a boundary membrane composed of material gathered from the environment or manufactured within the cell. Assuming that the cells are to proliferate they must contain a self-replicating, information-carrying polymer which can be constructed within the cell. The materials from which the polymer is to be constructed, and any essential chemicals, must be sourced externally and transported across the cell membrane. An energy source is required to sustain the chemical reactions necessary to synthesize the diverse products. The replication process must be linked to the production of other chemical species by a catalytic mechanism. It must be possible for the cell to grow and split if cells, rather than just the information-carrying polymer, are to replicate. Finally, all these processes must be integrated within a particular time frame so that no process occurs too quickly or too slowly. Any, or all, of these requirements could be based on naturally-derived systems. In practice these are the properties and mechanisms that most researchers have targeted in their attempts to prepare artificial cells.

By definition the first requirement of any cell, natural or artificial, is to be compartmentalized. Once a ‘closed system’ has been created it is possible to develop, or evolve, mechanisms that control the concentrations of essential nutrients and waste materials within that system. This will inevitably involve a route for chemical species of different sizes and properties to move between the cell’s interior and its external surroundings. There are several candidates for the cell boundary material.

### *3.4.1 Capsules with Mineral Walls*

The first candidate is an inorganic mineral formed either by chemical processes or through some form of biomineralization. The latter development may have been an evolutionary response by the cellular material to improve on its coincidental use of an existing mineral structure. The previous examples all presume a cell, or cell model, with a flexible outer membrane composed of a water impermeable material such as a lipid or phospholipid. However, as mentioned at the start of this chapter it is possible to form capsules from materials that are rigid solids. In Nature the best examples of these are algae that form plates of metal carbonates, coccoliths, and

then exude them so that they act in the manner of an exoskeleton. While there is no consensus for the reasons behind the synthesis of coccoliths it is clear that they have not developed as a cell membrane substitute, as coccolith-forming cells also possess a cell membrane. Theories range from protection from strong sunlight, as calcium carbonate has a high refractive index, or a method for trapping water so that solution phase reactions can be controlled close to the cell, to a localized mechanism for storing and releasing carbon dioxide. The latter possibility relies on changes in the concentration of metal carbonate and pH of the water near the cells: at high concentrations of carbonate, solids will be deposited on crystal surfaces causing the coccolith to grow, whereas low pH will lead to dissolution of the coccolith in the acidic solution releasing bicarbonate and, ultimately, carbon dioxide. In an environment controlled by the cell this may be a valuable method to control the release of carbon for biosynthesis, however, it has its dangers. Most seawater is slightly above pH 7 but when carbon dioxide is absorbed it reacts slowly with water to form carbonic acid and the pH drops.

$$CO_{2(g)} + H_2O_{(1)} \rightarrow H_2CO_{3(aq)}$$

Consequently the coccolith-forming algae in carbon dioxide rich water may be unable to prevent their 'exoskeletons' from being dissolved.

Given the possible origins of cell membrane formation on solid supports, as envisaged in a Lipid World where lipids dispersed in pools become concentrated when the water evaporates and they adhere to the underlying rocks, it would not be too far-fetched to consider the process of coccolith formation as a relic of early life. If, instead of forming metal carbonate structures, the cell had originally required a metal carbonate support from which to grow then the two processes, cell growth and metal carbonate crystallization, would be an example of symbiosis. In other single celled organisms, such as *Acantharia*, cellular growth follows crystallographic growth down mathematically precise axes.

To what extent can the example of a solid exoskeleton be replicated in the laboratory? Going against most contemporary examples of flexible artificial cells, Müller and Rehder published an example of a complex molybdenum oxide that spontaneously forms discrete nanospheres [23]. The hollow spheres were porous and allowed lithium cations to pass through the exoskeleton. While this a perhaps an extreme example of what may be considered an artificial cell, the authors assert that the presence of ion selective channels through the encapsulating oxide is directly analogous to natural ion channels in organic cells.

The use of biomineralization to compartmentalize materials occurs on many scales from the coccolith-forming algae discussed below to the shells of birds and reptiles. It has also been demonstrated by Martel and Young that combinations of calcium salts and proteins generate hollow calcium rich shells which bear very close similarities to purported remains of nanobacteria [24].

In an interesting fusion of biology and materials chemistry, the group of Tang [25] demonstrated that individual cells of yeast, *Saccharomyces cerevisiae*, can be covered in successive layers of calcium phosphate with the assistance of

poly(diallyldimethylammonium) chloride and sodium poly(acrylate) electrolytes. Within the mineralized shell the cells remained viable for at least a month in water, and were in fact protected from attack by zymolase, an enzyme that can digest the yeast cell wall. Upon treatment with dilute hydrochloric acid the calcium phosphate shell dissolved and the intact cells were recovered.

### *3.4.2 Polymer Based Capsules*

An alternative to an extensive inorganic structure is an extensive organic one. Candidates for this could include any number of polymers. One approach is to make self-assembled capsules from complementary polymers that form layer by layer (LbL) vesicle-like structures [26]. This has been achieved by templating the co-assembly of polymers around a removable core. The alternation of polymers with opposite charges allows the composition and thickness of the artificial cell walls to be controlled. The size of the core determines if the resulting capsule is a model for a cell or a smaller capsule like an organelle within a cell. The porous nature of the polymer allows chemical species to enter and leave the capsule but the potential for capsule growth and division, even with the presence of polymers in the external solution is very limited.

While this notion may conjure up visions of plastic materials it is important to remember that proteins and nucleic acids are also polymers. Many proteins form globular structures and, indeed, may interlock to encapsulate a large volume of space as exemplified by the coatings of capsid viruses. In a prebiotic world, polypeptides could have formed in aqueous solution through the sequential reaction of amino acids. The individual amino acids' hydrogen bond donor and acceptor groups, amines, carbonyls and carboxylic acids, would all have helped to keep the molecules in solution. Once a polypeptide had formed, however, many of these would be unavailable as they became incorporated in the hydrogen bond network that formed the secondary and tertiary structure. This would result in a more hydrophobic surface for the protein capsule which would make an effective cell.

Capsule-like polymer structures, especially those involving biopolymers such as proteins or RNA, may also assemble in the presence of metal ions as can be seen in the case of the protein ferritin. Concentrations of iron at the levels held by ferritin would ordinarily be very toxic but, fortunately, the protein contains large numbers of sulfur-containing amino acids that effectively isolate the metal clusters and impedes their reactivity. Although the purpose of the protein is as a controlled release store for iron, the mechanism by which individual metal ions enter and leave the structure are debated as it does not appear to contain any pores large enough for them to pass through. Nevertheless it is possible to remove the native iron and replace it with another metal without destroying the protein's tertiary structure. Although this is an attractive method to form a cell-like structure the drawback is that proteins of this type lack a permanent central void and can only adopt a globular structure in the presence of appropriate metal co-factors.

### 3.4.3 Lipid Capsules

An alternative candidate for a cell membrane is one composed of mostly hydrophobic molecules that would spontaneously aggregate if suspended in water. It is important that the molecule is not entirely hydrophobic, like hexane or benzene, as some part of it must be solvated if it is to aggregate with others to form a structure capable of encapsulating other species. Candidates for this type of molecule are the lipids and related compounds. The formation of protocells by lipids has been discussed previously as part of the Lipid World hypothesis. In the context of compartmentalization it is important to consider how such an aggregate could evolve.

Luisi has shown that membrane material itself can be formed autocatalytically in an experiment to investigate the base catalysed ester hydrolysis of hydrophobic ethylcaprylate [27]. Hydrolysis occured initially at the aqueous-organic interface where the products were micelle-forming sodium caprylate and ethanol. Once the critical micelle concentration, or cmc, was reached an exponential increase in hydrolysis was observed. The rate of hydrolysis in this second phase was almost 1000 times greater than in the initial phase suggesting that a catalytic mechanism had been activated. Luisi and co-workers hypothesized that once the cmc had been reached hydrolysis occurred within the micelles and, as the reactants were then constrained within a more hydrophobic environment, the increased rate was due to autocatalysis. Below pH 7 the micelles reorganized into unstable vesicles, in the order of 150 nm across as verified by freeze-fracture electron microscopy.

The autocatalytic hypothesis was backed by the reaction's pH sensitivity. Addition of $CO_2$ to the micelles increased acidity resulting in vesicles as the pH dropped below 7. Overall the experiment showed that vesicles could form from a reaction that generates amphiphilic molecules that in turn form autocatalytic micelles. Upon addition of a gas, common in planetary atmospheres, the micelles form vesicles that could function as protocells. The protocells are self-replicating so this process can be considered to be autopoietic, from the Greek for self-forming, and therefore fulfils an essential step in the chemical evolution of life.

Research by Zhu and Szostack has shown that vesicles composed of oleic acid lipid bilayers, with a water filled core, grew when fed a diet of fatty acid micelles [28]. Furthermore the vesicles underwent several cycles of growth followed by division. In doing so they became thread-like before generating daughter vesicles rather than enlarging to the point where they collapsed. To demonstrate how this mechanism could carry genetic information Zhu and Szostack included short strands of fluorescently tagged RNA within the original vesicles. After division the RNA strands were distributed to the daughter vesicles as would be expected with normal cell proliferation. The RNA was not able to replicate under the experimental conditions used but had it been incorporated with an appropriate RNA polymerase and a source of nucleotides, either within the vesicles or in the surrounding media, this too may have been achieved within the daughter vesicles. Based on these and other

experiments it is clear that lipid vesicles are the most flexible candidate molecules from which cell membranes can form.

One other, more complex, approach has recently been devised to prepare nanospheres with organic membranes. Armaroli and Bonifazi reported [29] that when symmetric hydrogen bonding molecules with the **D**-**A**-**D** motif combine with molecules containing the **A**-**D**-**A** motif and a non-bonding anthracene group, a ternary membrane is formed comprising two **A**-**D**-**A** molecules sandwiching one **D**-**A**-**D** molecule as shown in Fig. 3.11. The membrane forms spheres between 10 nm and 1 μm depending upon the solvent used.

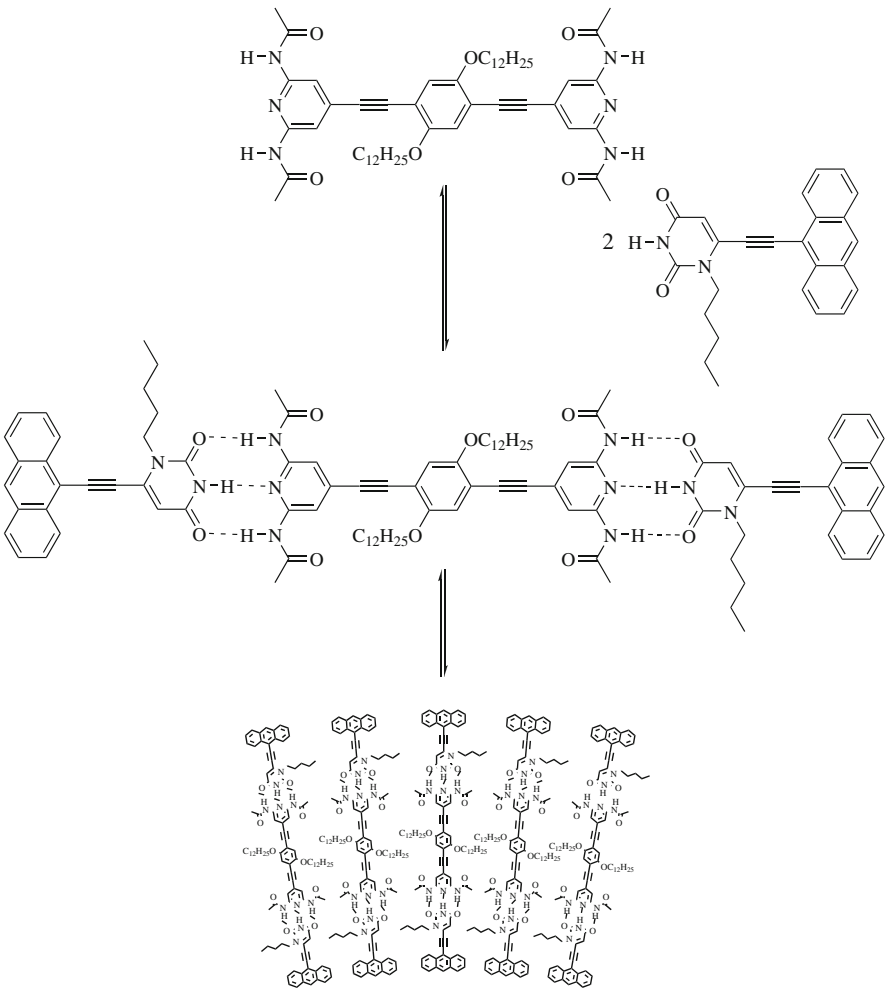


**Fig. 3.11** Ternary membranes formed by complementary hydrogen-bonding motifs [28]

### *3.4.4 Capsid Virus Mimetics*

An excellent example where a capsid virus has been given a new supramolecular application can be found in the work of Nolte who took an icosahedral capsid virus, cowpea chloritic mottle virus (CCMV) and used it as a nanoreactor for polymer synthesis [30]. Natural CCMV spontaneously assembles in acidic aqueous solution and disassembles in basic solution. The capsid contains pores open at pH 5 to release RNA into the host. Once the RNA leaves, the empty capsule is left. The Nolte group was able to assemble the subunits around polystyrene sulfonate with a mass of 9.9 kDa but the resulting structure had a different morphology to the natural system. Indeed, capsules formed around polymers with masses between 2 and 85 kDa but not around those with masses above 100 kDa. This raised the question of the potential for polymers to form within a capsid but to test the possibility a mixture of botanical, biological and chemical approaches was needed.

Cowpeas were grown, infected with CCMV, and the virus particles isolated. RNA was extracted from the particles and amplified by PCR to generate cloned plasmids. These were transfected into *Escherichia coli* that expressed the capsid proteins which were subsequently isolated and purified ready for self-assembly. Two types of intra-capsid synthesis were attempted. In the first Prussian blue was prepared photochemically from ammonium salts of hexacyanoferrate(III) and trisoxalatoferate(III); in the second conventional polymerization was undertaken in which it was possible to link ten monomers inside the capsid to give a polymer of 1 kDa mass. Larger polymers, up to 500 kDa, were also observed but only outside the capsids. This was consistent with capsids encapsulating monomers and catalysts, leading to the generation of polymers with a size limited by the capsid, while free monomers and catalysts in the remaining solution reacted but were unconstrained in the size of polymer that resulted. Considering the ease with which capsid proteins can be prepared through the *E. coli* transfection route this fascinating work at the interface between chemistry and biology could lead to significant advances in novel capsid-derived biotransformations with applications in materials chemistry.

## 3.5 Towards a Minimal Synthetic Cell

If we were to construct a cell from first principles what would be required? At its most basic a cell is a microenvironment in which desirable chemical processes, including replication and energy production, are favoured. To do so it must have a membrane, capable of encapsulating the reactants and products, through which small chemical species can pass. One approach that is gaining ground is to consider the minimum requirements that a functioning cell requires based on existing methods of encapsulation (phospholipid membranes), information storage (DNA), replication (tRNA) and transcription into structural and catalytic polymers (ribosomal production of proteins from mRNA). The structures that undertake these tasks

have been named artificial protocells in recognition of their unnatural origins and possible similarities to primordial cells.

To make protocells from a pool of precursor chemicals requires a genetic code that generates the instructions necessary to prepare all the components. Genes for phospholipid producing enzymes, RNA synthases, and all the other essential functions of a cell would have to be identified and the DNA sequenced. But, in a classic chicken and egg scenario, without pre-existing ribosomes and other essential protein forming tools the DNA would be inactive. To kick start the system it is necessary to add these from other sources but, once initiated, the protocell's own artificial DNA could take over.

Vaults, composed of proteins and discussed above, are able to encapsulate mRNA, RNA-binding proteins and some transferases, in short almost everything a primitive self-replicating system would require, so could represent an interesting model for a protocell. It is a particular attractive proposition as it only requires a single protein type that can self-assemble to form the entire structure.

Church, of Harvard University's Origins of Life Initiative, has identified a minimal set of proteins necessary for a functioning protocell [31]. It is possible to code for all of these using known genes and thereby create an entirely unnatural protocell genome. Based on the minimal set of 528 genes present in *Mycoplasma genitalium*, as determined by the Venter group [32], Church has designed a genome comprising 115 genes that covers all essential functions with the exception of those associated with the biosynthesis of membrane proteins, lipids and polysaccharides. It has been proposed that membrane encapsulation is unnecessary for an in vitro experiment, making lipid production and chemical transport across the lipid bilayer redundant, and that polysaccharides are not required in the proteins produced by the ribosomes. The functions of a small number of genes in *M. genitalium* are unknown but the effect of their absence should be readily apparent assuming the minimal set functions.

The genome could be constructed by excising the DNA sequences from *E. coli*, or simple viruses, but it would be more appropriate to undertake synthesis de novo without reference to an existing biological entity. The question then arises as to the type of genetic material to be employed. Entirely artificial methods that do not use a phosphate backbone or conventional nucleobases would not generate the same proteins so the choice is between DNA and RNA. While RNA would be an easier starting point, as it would remove the requirement for any genes associated with DNA, other complexities arise such as stable double strand sequences and reproducible folding to form the correct enzyme geometries. DNA therefore is the medium of choice.

To make replication and transcription simple a rolling circle of DNA is envisioned which would react with coilphage DNA polymerase to produce a single strand of DNA that can be transcribed into RNA. This must include the sequence of ribosomal RNA for protein synthesis, segments of which would also have to function as DNA primers. While this appears to be simple at the theoretical level the major challenge will be in the complex interplay between replication, transcription and eventual translation into functional proteins. The key step will be to generate the correct translation machinery. As Church points out, almost all the essential genes

in the minimal genome are associated with this task so successful synthesis of the ribosome will eventually dictate the success, or failure, of the venture.

## 3.6 Cellular Aggregation

Within the cell different systems may be kept separate through isolation within their own sub-compartments. These organelles are replicated during cell division ensuring that each new cell contains the same material. The single cell is the simplest level at which an organism can function but the size of the cell limits the diversity of tasks that it can undertake. While cells of this type can form colonies, and indeed are usually more successful when they do, they still embody the same simple functions. Multicellular organisms make use of different cell types that allow different, specialized functions to emerge. These could be for structure, vision, movement or any other beneficial modification. As a result they are greater than the sum of their parts but to succeed they must be able to interact. Much is known about the chemical signals that cells give out to their neighbours to ensure that they interact favourably: less is known about how artificial cells could be controlled.

In an interesting twist on the usual method for forming liposomes, vesicle-like structures prepared by recombining a lipid dispersion, several groups have inverted the technique. Rather than form emulsions of lipids in water they have created microdroplets of water in oil but, by incorporating phospholipids, are able to introduce a monolayer interface between the water and surrounding oil. The advantage of this is that each one can be mechanically placed next to another whereupon bilayers form between adjacent droplets. The properties of these linked systems can be probed by standard microelectrode techniques. In one instance droplets were filled with a mixture of salts, to mimic the contents of a cell, and a channel-forming heptameric protein, α-hemolysin (α-HL), added to the first. As the protein was localized in cell membranes the ionic current that flowed following insertion could be monitored as shown in Fig. 3.12. When a second droplet contained either a biomimetic

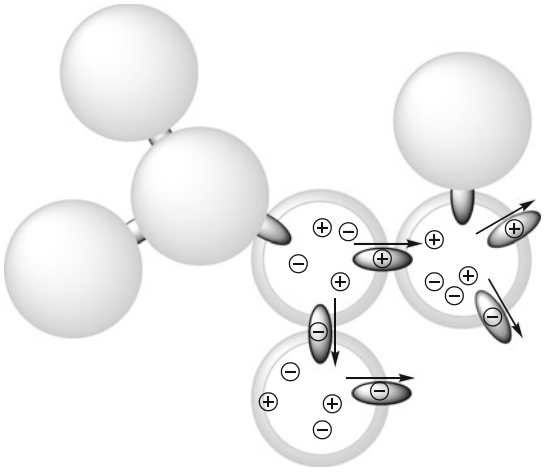


**Fig. 3.12** Transmembrane channel containing droplets acting as models for cellular aggregation [33]

salt solution or a salt solution with γ-cyclodextrin (γ-CD), which is known to block channels in α-HL, the ionic current increased upon α-HL addition but was intermittently blocked in the presence of γ-CD [33].

## 3.7 Summary

Compartmentalization is a key biological phenomenon and is achieved through the formation of capsular structures. These may be inorganic in origin, as is the case of coccoliths and protein-induced biomineralization, or make use of organic biomolecules. Proteins can form capsules through metal-induced folding leading to a globular structure. They can also aggregate as is seen in clathrins, coat proteins, capsid viruses and the intriguing vault proteins. A simpler solution is to use small molecules that will self-assemble to form membranes which curve round in order to adopt a low energy, spherical, capsule. This is achieved by lipids, phospholipids and other molecules with simple hydrophobic hydrocarbon groups that adopt lamellar structures to form an impenetrable barrier to water.

Artificial capsules come in many shapes and sizes but are generally too small to function as cell mimics. Where a supramolecular chemical approach can help is in designing self-complementary molecules that can aggregate to enclose large volumes of space. At the same time the design of a minimal genome reduces the size of an artificial cell required to surround it, bringing it within the realms of existing supramolecular capsules.

## References

1. Pearse BMF (1976) Clathrin – unique protein associated with intracellular transfer of membrane by coated vesicles. Proc Nat Acad Sci USA 73:1255–1259
2. Fotin A et al (2004) Molecular model for a complete clathrin lattice from electron cryomicroscopy. Nature 432:573–579 PDBID:1XI4
3. Ochoa WF et al (2006) Generation and structural analysis of reactive empty particles derived from an icosahedral virus. Chem Biol 13:771–778 PDBID:2BFU
4. Zlotnick A et al (2000) Mechanism of capsid assembly for an icosahedral plant virus. Virology 277:450–456
5. Stagg SM et al (2008) Structural basis for cargo regulation of COPII coat assembly. Cell 134:474–484
6. Kedersha NL, Rome LH (1986) Isolation and characterization of a novel ribonucleoprotein particle: large structures contain a single species of small RNA. J Cell Biol 103:699–709
8. Tanaka H et al (2009) The structure of rat liver vault at 3.5 angstrom resolution. Science 323:384–388
7. Tanaka H et al (2009) The structure of rat liver vault at 3.5 angstrom resolution. Science 323:384–388 PDBID:2ZUO
9. Schmidtchen FP (1977) Inclusion of anions in macrotricyclic quaternary ammonium-salts. Angew Chem Int Ed Engl 16:720–721
10. Graf E, Lehn JM (1975) Synthesis and cryptate complexes of a spheroidal macrotricyclic ligand with an octahedrotetrahedral coordination. J Am Chem Soc 97:5022–5024
11. Bryant JA et al (1991) Host guest complexation 55. Guest capture during shell closure. J Am Chem Soc 113:2167–2172

12. Matthews SE et al (2002) Calix[4]tubes: A new class of potassium-selective ionophore. J Am Chem Soc 124:1341–1353
13. Roache S et al (1998) Self-assembly of a supramolecular cube. Chem Commun 1681–1682
14. Saalfrank RW et al (1988) Adamantoid chelate complexes 1. The 1st adamantoid alkaline-earth metal chelate complex – synthesis, structure, and reactivity. Angew Chem Int Ed Engl 27:851–853
15. Caulder DL et al (1998) The self-assembly of a predesigned tetrahedral $M_4L_6$ supramolecular cluster. Angew Chem Int Ed Engl 37:1840–1843
16. Mal P et al (2009) White phosphorous is air-stable within a self-assembled tetrahedral capsule. Science 324:1697–1699
17. Fujita M et al (2001) Molecular paneling via coordination. Chem Commun 509–518
18. Hiraoka S et al (2009) Template-directed synthesis of a covalent organic capsule based on a 3 nm-sized metallocapsule. J Am Chem Soc 131:11646–11647
19. Cram DJ, Tanner ME, Thomas R (1991) The taming of cyclobutadiene. Angew Chem Int Ed Engl 30:1024–1027
20. Kang JM et al (1998) Self-assembled molecular capsule catalyzes a Diels-Alder reaction. J Am Chem Soc 120:7389–7390
21. MacGillivray LR, Atwood JL (1999) Structural classification and general principles for the design of spherical molecular hosts. Angew Chem Int Ed Engl 38:1018–1033
22. Pohorille A, Deamer D (2002) Artificial cells: prospects for biotechnology. Trends Biotechnol 23:123–128
23. Müller A et al (2004) Artificial cells: Temperature-dependent, reversible $Li^+$-ion uptake/release equilibrium at metal oxide nanocontainer pores. Angew Chem Int Ed Engl 43:4466–4470
24. Martel J, Young JD (2008) Purported nanobacteria in human blood as calcium carbonate nanoparticles. Proc Nat Acad Sci USA 106:5549–5554
25. Wang B et al (2008) Yeast cells with an artificial mineral shell: protection and modification of living cells by biomimetic mineralization. Angew Chem Int Ed Engl 47:3560–3564
26. Wilson A et al (2009) A reliable method for attaching biological molecules to layer-by-layer self-assemblies. Chem Commun 2487–2489
27. Bachman PA, Luisi PL, Lang J (1992) Autocatalytic self-replicating micelles as models for prebiotic structures. Nature 357:57–59
28. Zhu TF, Szostak JW (2009) Coupled growth and division of model protocell membranes. J Am Chem Soc 131:5705–5715
29. Yoosaf K et al (2009) Engineering spherical nanostructures through hydrogen bonds. Chem Commun 2830–2832
30. de la Escosura A, Nolte RJM, Cornelissen JJLM (2009) Viruses and protein cages as nanocontainers and nanoreactors. J Mater Chem 19:2274–2278
31. Forster AC, Church GM (2006) Towards synthesis of a minimal cell. Mol Syst Biol 2:ARTN 45
32. Gibson DG et al (2008) Complete chemical synthesis, assembly, and cloning of a Mycoplasma genitalium genome. Science 319:1215–1220
33. Holden MA, Needham D, Bayley H (2007) Functional bionetworks from nanolitre water droplets. J Am Chem Soc 129:8650–8655

# Chapter 4
# Supramolecular Enzyme Mimics

## 4.1 Enzymes

Enzymes are biological structures that speed up chemical reactions. They may be composed entirely of proteins or incorporate other chemical species such as metal ions. Enzymes work by acting catalytically on particular molecules, known as substrates, converting them into other molecules in accordance with the definition of a catalyst as a substance that increases the rate of a chemical reaction yet remains unchanged at the end of each catalytic cycle. The rate of reaction is increased by lowering the activation energy barrier to that reaction occurring and this is often accomplished by stabilizing the transition state which is usually where the high energy barrier occurs as shown in Fig. 4.1.

The effect of enzymes was first described by Pasteur in his work on yeast fermentation. The term enzyme (from the Greek, to leaven) was first used by Kühne [1] to describe reactions brought about by chemicals, rather than organisms, but it was Buchner who first used the suffix '-ase' to denote a biomolecule with enzymic activity when he named the chemical agent that promoted sugar fermentation 'zymase' [2]. The method by which enzymes could be so specific was coined the 'lock and key' mechanism by Fischer as discussed in Chapter 1. He proposed that the substrate

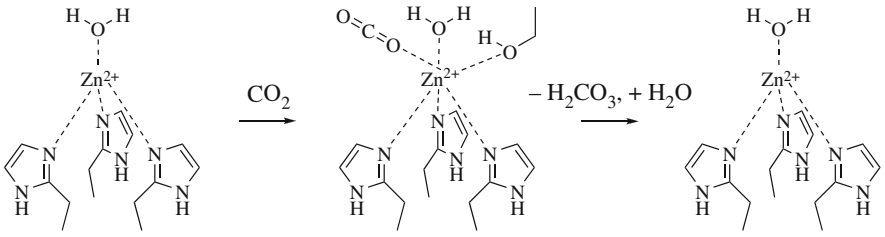



**Fig. 4.1** The catalytic effect of enzymes illustrated by a low energy route from a tetrahedral ground state to an octahedral transition state

P.J. Cragg, *Supramolecular Chemistry*, DOI 10.1007/978-90-481-2582-1_4,
© Springer Science+Business Media B.V. 2010

fitted into the enzyme like a key in a lock and that a good fit ensured specificity. This concept is also to be found at the core of supramolecular chemistry where molecular host molecules are designed to be complementary to the binding properties of the desired guest.

While Fischer's idea works well with relatively simple systems it does not explain how the transition state of a reaction may be stabilized as this involves one or more molecules being held at the point of reaction. How can an enzyme have a specific binding site for a substrate yet also be specific for its transition state? This problem became even more pressing when, in 1926 after a decade of research, Sumner demonstrated that enzymes were proteins with defined structures by growing crystals of urease isolated from the jack bean, *Canavalia ensiformis* [3]. The discovery was met with some scepticism at the time, particularly by the eminent chemist Willstätter, in part because its reproducibility was highly dependent on the source of the beans. It was only accepted when Northrop independently crystallized the enzyme pepsin [4]. In recognition for his work on enzymes Sumner shared the 1946 Nobel Prize for Chemistry with Northrop and Stanley, who had earlier crystallized the self-assembling tobacco mosaic virus [5]. Jack bean urease is now known to have an active site containing two nickel atoms, although its crystal structure, shown in Fig. 4.2, has yet to be determined in complete detail.

Prior to the crystallization of jack bean urease it was assumed by the biochemical community that enzymes had no ordered structure. In 1965 the first crystallographic evidence for the mechanism by which enzymes work when Phillips and his group solved the lysozyme structure [6]. Details of the structure indicated how the enzyme could bind the oligosaccharides present in its target, bacterial cell wall peptidoglycans, and could respond to the binding event by changing its structure.

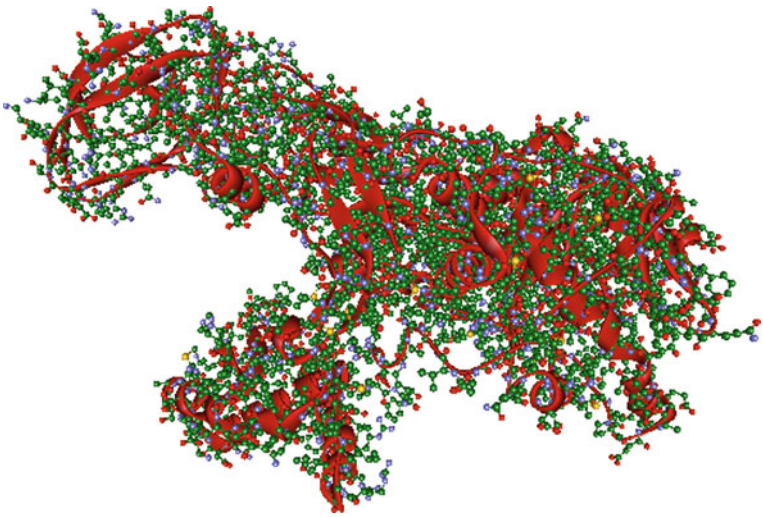


**Fig. 4.2** The first enzyme to be characterized: jack bean urease [7]

This development was important as it gave support to Koshand's induced fit model that had superseded Fischer's lock and key hypothesis. From a mechanistic point of view Koshland argued that the so called 'induced fit' model could account for enzymes' structural order which must exist if they can crystallize [8]. It also explained the conformational changes essential if reactants were to be bound, reactions occur, and products disgorged. These two iconic concepts are shown in Fig. 4.3.

The flexible nature of proteins allows their secondary and tertiary structures to alter in response to substrate binding. Thus an enzyme may be specific for a certain substrate, rearrange to stabilize the transition state and finally release the transformed substrate which is no longer complementary to either the initial protein conformation or the conformation that bound the transition state. It is now possible to determine which factors are vital to enzyme activity through two methods: structural determination and site-directed mutagenesis. The most obvious approach is to form crystals of the enzyme and its target, then solve the crystal structure. This is never an easy prospect as very few complexes crystallize well enough to generate high quality data. There is also the criticism that crystal structures do not reflect the hydrated, flexible structures of the proteins that exist in organisms. An alternative structural method to crystallography is to analyze the enzyme and its complex by NMR spectroscopy. Though this is also limited in its application, it can give a number of conformers and may generate a better representation of the complex. Both crystallographic and NMR methods can identify the important enzyme-target interactions. The alternative approach is to vary specific amino acids in the protein backbone, replace them with others, and assess the mutated enzyme's activity. For example, if a thiol group is suspected to be responsible for holding the target species in place, a mutant where cysteine had been replaced by methionine or serine should exhibit lower activity. This can either be achieved by synthesizing the protein directly or by altering the gene responsible through site-directed mutagenesis.

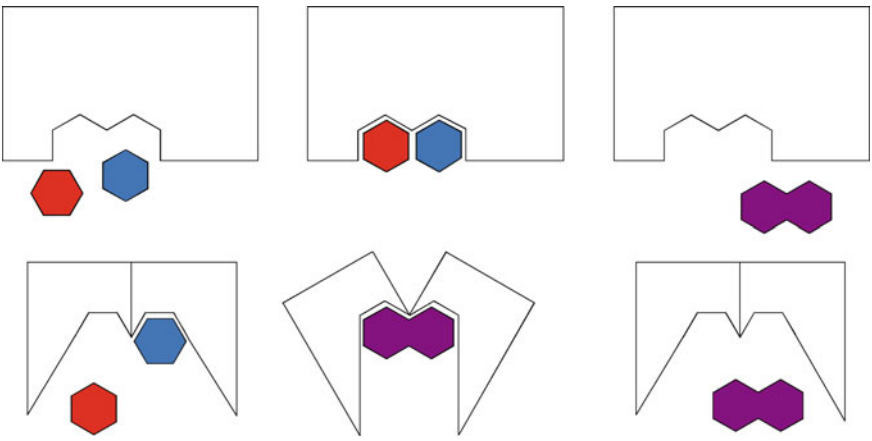


**Fig. 4.3** Models of enzyme specificity: lock and key (*top*) and induced fit (*bottom*)

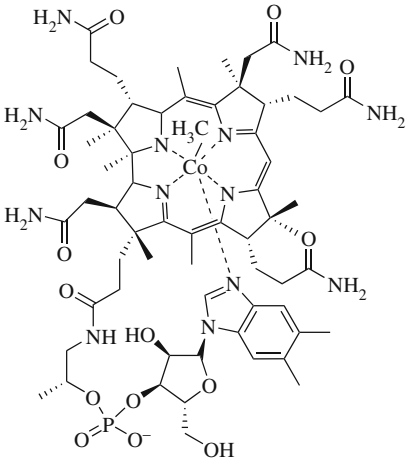



**Fig. 4.4** The active site of vitamin $B_{12}$, as the methylcobalamin complex

A single protein may act as an enzyme by having a binding site that is complementary to the target substrate. In a more complex system several proteins may interact to form a binding site; each co-enzyme with slightly different properties that are required to control separate aspects of the catalysis. Another group of enzymes exist that incorporate metals, particularly transition metals, either through interactions with peptide side chains or within non-peptidic structures such as protein bound porphyrin-type ring structures.

These essential biomolecules were the targets of both synthetic and structural research efforts by leading scientists in the mid-20th century. Notable amongst these were the crystallographic studies of the vitamin $B_{12}$ core by Hodgkin [9], shown in Fig. 4.4, and the synthesis of the same compound by the Woodward group [10]. The research was sparked by the realization that lack of vitamin $B_{12}$, which is synthesized by bacteria and biotransformed in humans, leads to pernicious anemia. The condition could be reversed through direct injection of the vitamin so a synthetic analogue that also performed a therapeutic function was an important target. Both Hodgkin and Woodward were rewarded with Nobel prizes in for their achievements in unravelling the complex behaviour of $B_{12}$ and, perhaps more significantly, a number of synthetic chemists who would later become leading proponents of supramolecular chemistry were amongst those involved in the synthetic work.

## 4.2 Metal Complexes as Enzyme Mimics

Once it was established that a number of enzymes contained non-peptidic structures, and that these features were present as binding sites for catalytic metals, researchers started to look for simple analogues that could be used as models for the more complex natural systems. The motivation behind these investigations was to gain a greater understanding into how the natural systems work by reducing the active

site to the simplest possible structure. Many metals are held by rigid ligands which often have extensive conjugation. This can be achieved in the laboratory through the use of imine bond formation. The resulting Schiff base ligands are easy to produce on a large scale and also lend themselves to modification. For example, ligands based on the porphyrin structure can be synthesized through the condensation of pyrrole and benzaldehyde which results in the formation of the purple, crystalline tetraphenylporphyrin. Introduction of metals is generally a trivial exercise so that numerous transition metal complexes can be prepared from the same ligand and their relative enzymic activities assessed. Similarly, Schiff base complexes of cobalt have been used as models for $B_{12}$ [11] with some success, despite the acyclic nature of the ligand, and complexes of manganese with another Schiff base, salen, have been shown to have promising superoxide dismutase activity [12]. Two of these compounds are illustrated in Fig. 4.5.

Although many metalloenzymes and related metallobiomolecules have planar four-coordinate ligating sites for metals, often extended into octahedral six-coordinate centres through the addition of axial groups, a large number have been found to have tetrahedral geometries. This is particularly true when non-transition metals, or those with a full *d*-shell of electrons, are concerned. Classically these include zinc enzymes such as carbonic anhydrase in which $Zn^{2+}$ is bound by three histidine groups and an apical water molecule. The metal is located in the bottom of a cleft in the protein, shown in Fig. 4.6, so that only the desired small molecules, carbon dioxide in the case of carbonic anhydrase, can approach. Model ligands for

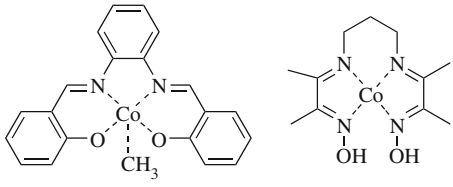



**Fig. 4.5** Schiff base complexes as models for enzyme active sites

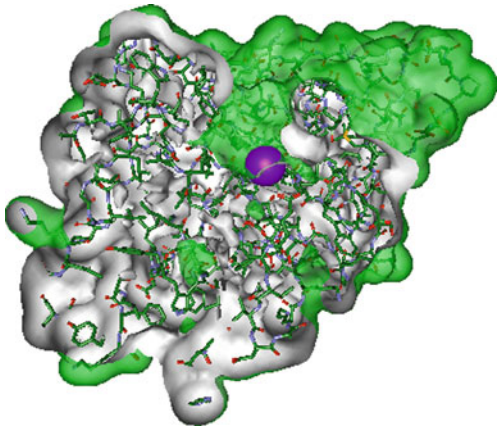


**Fig. 4.6** Carbonic anhydrase showing the zinc ion at its core (*purple*) accessible only to small molecules [13]

these binding sites are generally cyclic amines in which the nitrogen atoms are separated by two or three methylene groups to maximize the preferred ligand geometry while minimizing ligand strain energy.

In mimicking enzymes it is always important to determine which aspects of the natural system are to be modelled. It may be that a simplified complex which dispenses with the protein structure of the original is desired so that more light can be shed on the metal-centred catalytic mechanism. To be successful the model compound must still perform the catalytic function of the natural enzyme and do so in the same fashion otherwise any mechanistic details may only be applicable to the model. Such a model would be said to have both a structural similarity with the enzyme and a functional similarity. In practice it is usual to focus on one aspect or the other. Thus it may be possible to model the metal binding site quite closely but for the model to have no activity. Nevertheless, such a model would give valuable information on metal substrate interactions and point to potential transition states in the catalytic process. An alternative approach is to devise a system that copies the reaction which a particular enzyme catalyses, to a greater or lesser extent, but which has very little in common with the enzyme's structure. A good example of this is the use of organic dyes or ruthenium tris(bipyridyl) complexes as light harvesting molecules in solar cells. The cells mimic an aspect of the natural photosynthetic process undertaken by enzymes in the photosystem II complex of proteins but have a much simpler structure.

## 4.3 Enzymes and Their Supramolecular Analogues

Enzymes are complex, multicomponent systems and hard to use in isolation from their natural biological microenvironment. This has led to three distinct methods of exploitation. Enzymes may be extracted from cells and concentrated for use as catalytic promoters in other systems or comparators with similar systems. Common examples of this include the use of DNA polymerase to amplify DNA in Mullis' polymerase chain reaction (PCR) [14] and assays that compare artificial antioxidant activities of potential therapeutic compounds with natural catalytic antioxidants. Enzymes may be used in situ through the process of biotransformation. Here the entire organism is used to bring about a desired reaction without recourse to traditional chemical methods except in the final extraction of the product. Often the biological agent is 'fed' an unusual precursor molecule and then carries out its normal enzymic action on the precursor to generate a novel and useful compound. The potential to genetically modify simple organisms so that their enzymes function in a different way, or to introduce different enzymes to existing organisms, renders this technique extremely powerful. One application, discussed by Ventor, is to modify bacteria so that they can perform biotransformations on atmospheric carbon dioxide to reduce the impact of its industrial release [15]. A third way that enzymes can be exploited is through analysis of their mechanisms. Once the catalytic process has been understood it may be possible to replicate the activity using

an entirely synthetic molecule. This is highly advantageous if the enzyme is required to operate in unnatural conditions. One such application, discussed later, could be an artificial analogue of photosystem II, a complex biological pathway involving several enzymic transformations that turn photonic energy into electrical energy. The analogue could be incorporated into the external surfaces of buildings to generate electricity from sunlight. Given the complex nature of all the weak interactions involved it is not surprising that attempts to mimic enzymes are often encountered in supramolecular chemistry.

### *4.3.1 Haemoglobin, Myoglobin and Their Models*

Haemoglobin and myoglobin are not generally thought of as catalysts. Between them they bind gaseous oxygen from the lungs, transport it through the bloodstream, and deliver it to the muscles where it is released for use in cells. In catalytic terms they facilitate the transfer of oxygen from one environment where reactions do not occur to another where they do. It is an example of phase transfer catalysis in which the oxygen binding species acts as the transfer agent. Usually this type of catalysis involves one type of molecule forming a complex with another to move it from one liquid phase, in which it is soluble, to one in which it is not, but where its activity is required. Ordinarily the two liquid phases are immiscible and the phase transfer catalyst operates at the interface by shuttling the target species from one phase to another. In a non-biological context crown ethers are often used as phase transfer catalysts as they can bind water soluble metals in the aqueous phase and transport them to an organic phase where they initiate reactions with hydrophobic organic molecules. By analogy the movement of oxygen in the respiration process clearly fits as an example of biological phase transfer catalysis, albeit a complex one.

Haemoglobin is composed of four similar subunits to give an $\alpha_2\beta_2$ structure. The $\alpha$ and $\beta$ structures are closely related to that of myoglobin. As illustrated in Fig. 4.7,

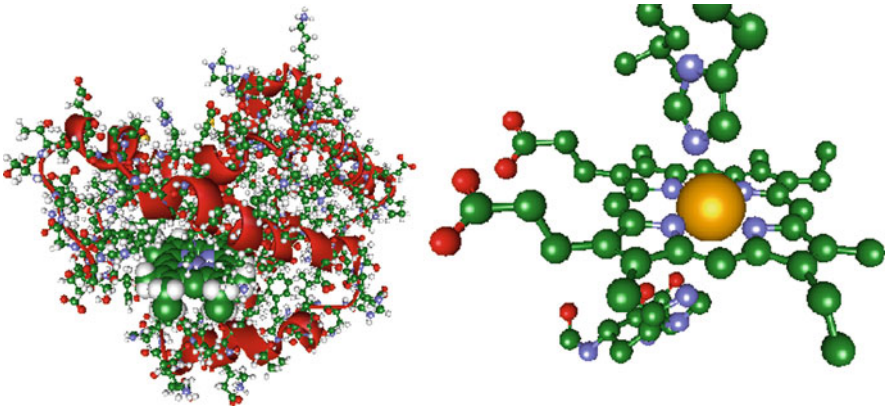


**Fig. 4.7** A haemoglobin subunit [16] and its porphyrin core (*left*, hydrogen atoms removed for clarity)

at the core of each haemoglobin subunit is a single iron(II) ion held within a planar macrocycle comprised of four linked pyrrole groups: the porphyrin ring. The pyrroles are in turn decorated with a number of residues that link the ring to the parent protein. Iron coordinates to the four pyrrole nitrogen atoms in a square planar geometry and a further imidazole group in an axial position. The sixth site is available to bind an oxygen molecule. Iron(II) is initially in the high spin electronic state which is slightly too large to fit within the binding pocket of the porphyrin so it sits slightly above the coplanar nitrogen atoms. Upon binding to oxygen the low spin state is adopted by iron(II) which has the effect of reducing the metal's ionic radius and allows it to sink comfortably into the porphyrin binding pocket. The movement is reinforced by a stronger interaction between iron(II) and the axial imidazole nitrogen donor atom that has the effect of pulling the metal cation into the pocket. This process is reversible allowing oxygen to be released under the appropriate conditions. It is a good example of an allosteric effect as binding in one part of the molecule causes movement in a remote region.

Both haemoglobin and myoglobin can be irreversibly poisoned if certain chemical species bind more strongly to iron than dioxygen. Typical examples include cyanide, carbon monoxide and hydrogen sulfide. The toxicity of these molecules can be estimated by a concept known as the '*trans* effect'. This states that the rate of substitution in a particular position around a transition metal, such as iron, is greatly influenced by the nature of the group bound in the opposite, or *trans*, position. In this case the *trans* position to the oxygen binding site is an imidazole group. If the species binding in place of oxygen has a greater *trans* effect than imidizole then it is the dative bond between imidazole and iron which will be preferentially broken. This is the case for chemical species such as cyanide. A further problem can arise if the porphyrin ring is exposed, through protein degradation or similar damage, and allowed to come into contact with a second ring. A bound oxygen molecule can then bridge two iron(II) centres and, through the action of forming a peroxide, oxidize both centres to iron(III). Further rearrangement results in a strong iron-oxygen-iron bond being formed which destroys any oxygen transporting ability as shown in Fig. 4.8.

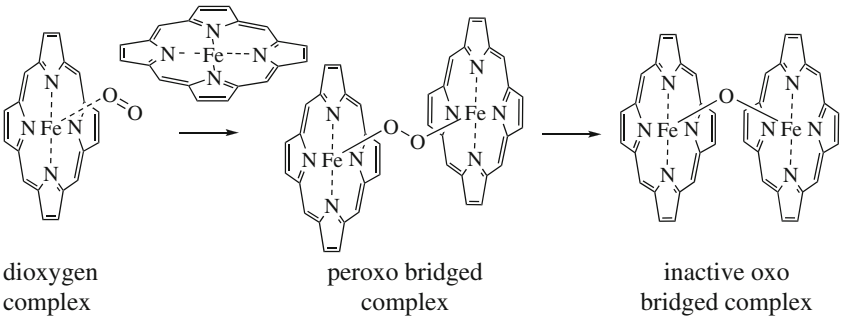



**Fig. 4.8** Deactivation through oxidation of face-to-face iron porphyrins

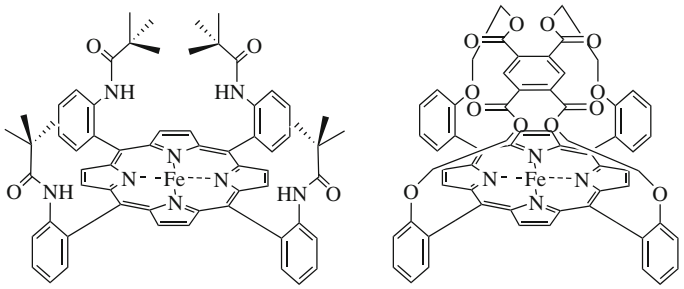


**Fig. 4.9** Picket fence [18] (*left*) and roofed [19] (*right*) models for porphyrin containing enzymes

Iron(II) porphyrins are simple to synthesize in the laboratory. A porphyrin analogue, tetraphenylporphyrin, can easily be prepared by boiling equimolar amounts of benzaldehyde and freshly distilled pyrrole in an organic acid. Despite the fact that the reaction involves eight molecules being brought together the yield is usually about 15% which would correspond with a linear eight step reaction sequence in which each step is 80% efficient. Once the product has been prepared it can be treated with an iron(II) salt to form the iron-porphyrin mimic. Early studies on iron-porphyrin complexes showed that in the presence of dioxygen the metals were oxidized from iron(II) to iron(III), as occurs in haemoglobin, but that pairs of complexes rapidly stacked together with the metals connected by a peroxide bridge.

Several methods have been used to prevent dimerization starting with Collman's 'picket fence' porphyrin [17]. This tetraphenylporphyrin derivative has short organic chains extending from the phenyl rings at right angles to the porphyrin plane. The architecture allows dioxygen to be bound by the complex but keeps adjacent binding sites apart through the steric bulk of the side chains. Other model systems exist that further protect the dioxygen binding site as shown in Fig. 4.9.

### *4.3.2 Cytochromes*

Cytochromes encompass a large group of iron-haem complexes that are involved in electron transport and are usually membrane bound. Originally described by MacMunn [20] in the 1880's, cytochromes were named and given alphabetical designations by Keilin [21] to denote the different wavelengths at which they absorb. For example, cytochrome *a* absorbs at 605 nm and cytochrome *c* at 550 nm. Other members of the family, such as cytochrome $P_{450}$, include the wavelength at which the pigment absorbs. The latter enzyme is a monooxygenase which catalyses the reaction of dioxygen and an organic substrate to insert one oxygen atom while the other oxygen is reduced to water as in:

$$\text{R–H} + \text{O}_2 + 2\bar{\text{e}} + 2\text{H} \quad \rightarrow \quad \text{R–O–H} + \text{H}_2\text{O}$$

Other reactions include epoxidation of double bonds and alcohol oxidation. These reactions have an important role in detoxification in the liver and, for that reason, the enzyme has attracted much interest. Alongside investigations of the natural enzymes have come model complexes that seek to catalyse the same reactions. The active site of the enzyme is an iron-containing haem ring but crystal structures show the sulfur atom of a cysteine side chain coordinated to the porphyrin-bound iron in an axial position. This key interaction holds the iron, as $Fe^{3+}$, in a low spin configuration. The substrate binding site is initially filled by water which is displaced when the organic target is introduced. Catalysis is initiated by the transfer of an electron from cytochrome $P_{450}$ reductase to give $Fe^{2+}$ that can then bind dioxygen which accepts the electron to generate a superoxide species. A second electron is delivered to form the peroxide which reacts with a proton to give an iron(III) hydroperoxo complex. A second proton then reacts to form water and leave an iron(IV) oxo species that reacts with the target organic substrate and the approach of another water molecule displaces the oxidized product.

The reactivity of the iron centre is linked to the thiolate ligand which means that any enzyme model must contain both an iron(II)-porphyrin centre and a sulfur donor. This combination is hard to achieve under aerobic conditions as both the iron and the sulfur are readily oxidized. Furthermore both are light sensitive to a degree.

Two measures to reproduce the key aspects of the enzyme have had some success: to use an aromatic thiol and to block attack on the iron by building organic frameworks over the metal. Other approaches, illustrated in Fig. 4.10, include substituting iron with manganese, which is less sensitive to aerial oxidation, and to tether a cyclodextrin to the porphyrin so that the organic substrate could be positioned correctly for oxidation.

One of the first examples of this can be seen in the work of the Breslow group [24] which synthesized nickel and copper containing functional esterase mimics. The concept was taken further by Groves and Neumann [22] who appended four cholesterol derivatives to a manganese porphyrin so that the compound would be held within a lipid bilayer. When introduced to vesicles the complex buried itself in the membrane and, once flavoprotein pyruvate oxidase and ethylbenzene were added, generated acetophenone. The use of cyclodextrins as hydrophobic pockets to preorganize artificial enzyme substrates remains a lure to many. The Breslow group later attached a cyclodextrin to a manganese porphyrin and found that it was able to catalyse the regiospecific hydroxylation of steroids [23].

### *4.3.3 Protection from Radicals: Catalytic Pro- and Antioxidants*

It is ironic that while most organisms rely on oxygen for respiration, its mere presence can have damaging effects. Oxidation occurs slowly in an oxygen rich atmosphere but is accelerated by many transition metals such as those regularly encountered in biological systems. When oxidation of lipids, DNA, RNA, proteins and other biomolecules occurs a major problem is encountered as it affects their functions. There are, of course, many necessary enzymes that catalyse oxidation

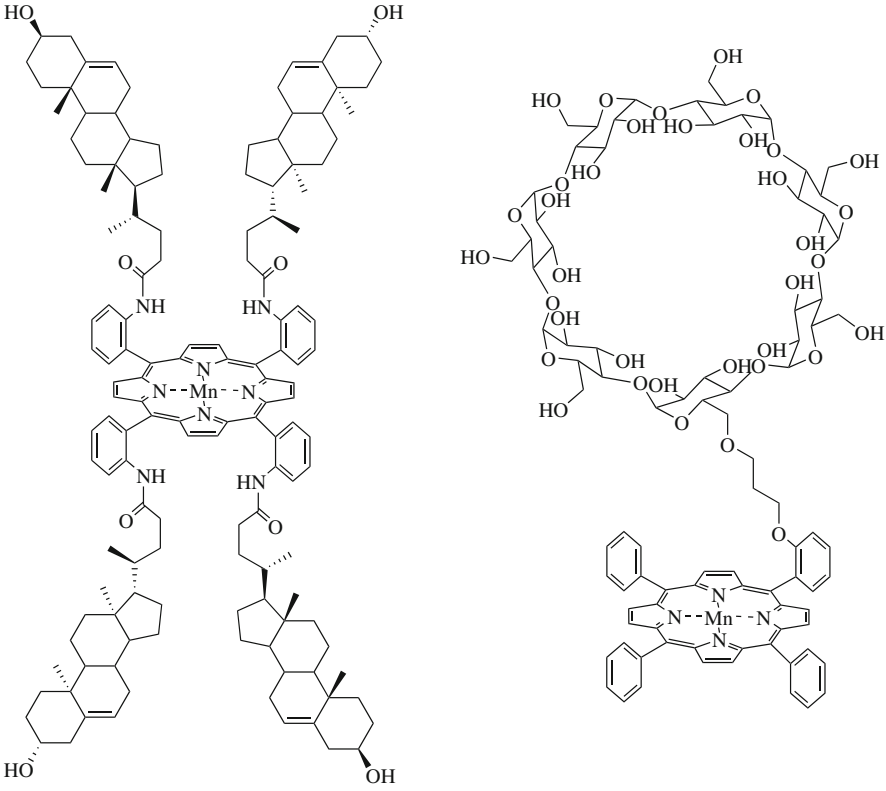



**Fig. 4.10** Cytochrome model complexes [22, 23]

of specific molecules as can be seen in the biosynthetic pathway that makes the stress hormone and neurotransmitter noradrenaline from tyrosine. Key steps such as the conversion of tyrosine into L-DOPA and later conversion of dopamine into noradrenaline involve a hydroxylase and β-monooxygenase, respectively, to introduce hydroxy groups. These examples of oxidation are necessary but also carefully focused so that only specific products result. The widespread and non-specific oxidation of organic biomolecules is undesirable and requires defence mechanisms. Many of the dangerous oxidizing species are free radicals, which have an unpaired electron making them highly reactive, and often incorporate nitrogen atoms. Consequently these molecules are collectively known as reactive nitrogen and oxygen species, or RONS. Examples include the hydroxyl ($OH^{\cdot}$), superoxide ($O_2^{\cdot -}$) and nitric oxide ($NO^{\cdot}$) radicals, peroxide ($O_2^{2-}$) and many combinations of nitrogen and oxygen. They arise naturally through the interaction of energetic photons in sunlight, particularly those in the ultraviolet spectral region, or high temperature reactions including those occurring at the tips of cigars and cigarettes. Many of the resulting radical species are mopped up by small organic antioxidant

molecules like vitamin C (ascorbic acid), the vitamin E group, polyphenolic compounds such as resveratrol found in red grape juice, and glutathione. While these molecules react with radical species they can do so only once. Other molecules, the catalytic antioxidants, are able to undertake these reactions repeatedly and are therefore much more effective protecting agents. There is a catch to this as, in order for the antioxidants to be recycled into their reducing forms, they have to oxidize another substrate. So, depending upon where in the cycle the molecules are, they can act as either pro- or antioxidants.

Superoxide dismutase (SOD) is an enzyme that reacts with the toxic radical superoxide species and transforms it into a mixture of dioxygen and peroxide. The latter is also damaging but is targeted by catalase which turns it back to dioxygen. Natural superoxide dismutases fall into two general classes: those with bimetallic copper and zinc centres, and those with a single metal, usually manganese, at their cores. While both classes have been modelled, the manganese systems are easiest to prepare.

Superoxide dismutase turns out to be the key mammalian antioxidant enzyme. SOD is essential: genetically modified mice lacking the enzyme die shortly after birth. It exists in several forms but the most closely studied version is SOD1, a protein containing zinc and copper. The catalytic site, shown in Fig. 4.11, contains a copper cation held by three neutral histidines and one histidine anion which in turn is bound to a zinc atom.

The enzyme removes the toxic superoxide free radical by the following pathway:

$$[CuZn]^{n+} + O_2^{\cdot -} \rightarrow [CuZn]^{(n-1)+} + O_2$$

$$[CuZn]^{(n-1)+} + O_2^{\cdot -} + 2H^+ \rightarrow [CuZn]^{n+} + H_2O_2$$

Overall this gives:

$$2O_2^{\cdot -} + 2H^+ \rightarrow O_2 + H_2O_2$$

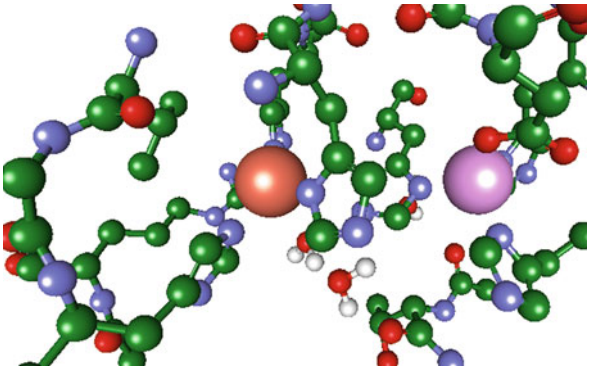


**Fig. 4.11** The catalytic site in copper-zinc superoxide dismutase (SOD1) [25]

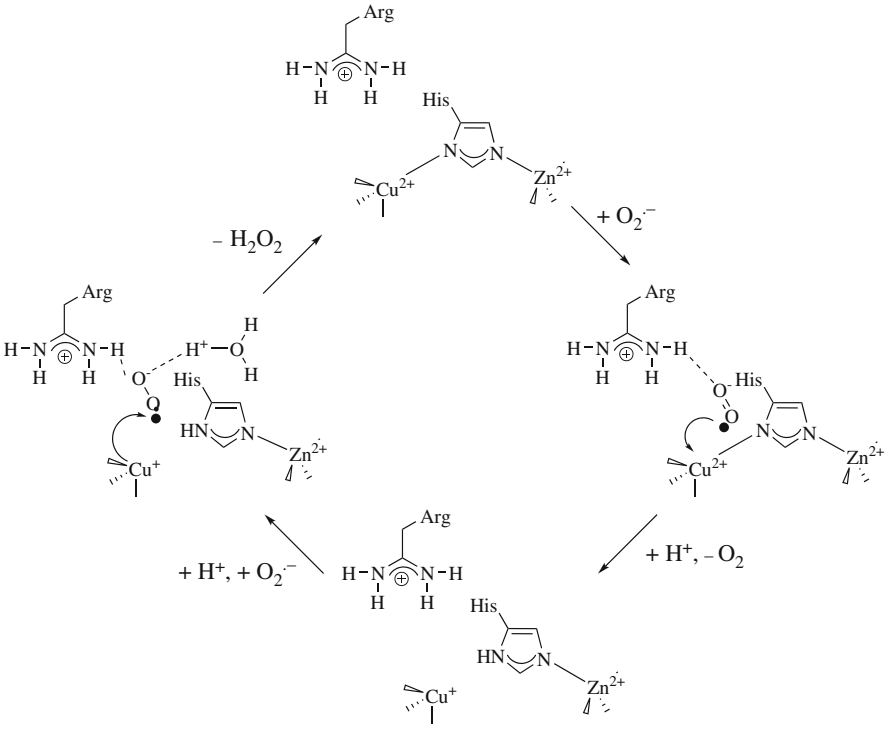



**Fig. 4.12** The superoxide dismutase catalytic cycle

A closer look at the catalytic cycle, as shown in Fig. 4.12, shows how zinc plays a largely structural role while copper continually changes oxidation state and coordination environment. It should be immediately apparent that one of the by-products is the equally toxic reactive species, hydrogen peroxide. Fortunately a second enzyme, catalase, is present to scavenge the peroxide and convert it to oxygen and water.

Catalase is a haem-type enzyme containing an iron atom in a porphyrin ring. A two step reaction removes the dangerous reactive oxygen species through the oxidation and subsequent reduction of the central iron atom:

$$[\mathrm{Fe}]^{3+} + \mathrm{H_2O_2} \quad \rightarrow \quad [\mathrm{Fe{=}O}]^{4+} + \mathrm{H_2O}$$

$$[\mathrm{Fe{=}O}]^{4+} + \mathrm{H_2O_2} \quad \rightarrow \quad [\mathrm{Fe}]^{3+} + \mathrm{O_2} + \mathrm{H_2O}$$

The overall reaction is:

$$2\mathrm{H_2O_2} \quad \rightarrow \quad \mathrm{O_2} + 2\mathrm{H_2O}$$

Given the obvious health benefits of these enzymes there has been extensive research into mimetic compounds. Three main hurdles exist to replicating the

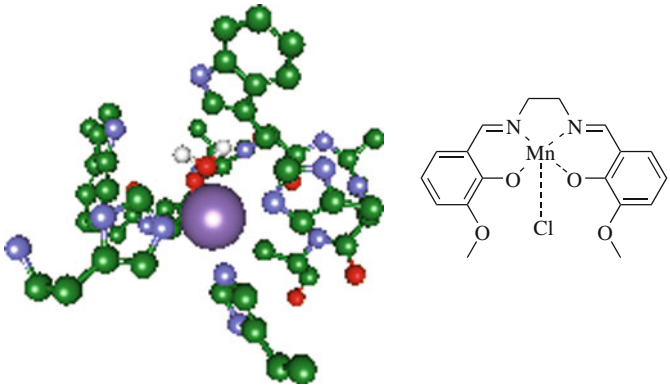



**Fig. 4.13** The catalytic site in manganese superoxide dismutase (SOD2) [26] and a model compound

enzyme's function. Firstly, model systems do not incorporate the protective protein that surrounds the enzyme's active site and effectively blocks intrusion by other small molecules. Secondly, the structural variations that occur during the catalytic cycle, specifically bonds breaking and forming between the histidine bridge and metal ions, are hard to replicate. Finally, the complexity of synthesis necessary to introduce a zinc and copper into two subtly different binding sites far outweighs the benefits of making a SOD1 mimic.

As noted above, other forms of SOD, and SOD-like enzymes, exist that contain a single metal site. One such is the manganese-centred mitochondrial matrix protein SOD2. In this enzyme the metal is held not in a planar haem ring but in a tetrahedral geometry through coordination to three histidine side chains set at 120° to each other with coordinating nitrogen atoms pointing towards the metal ion. The fourth coordination site is occupied by water that can be replaced by $O_2^{\cdot -}$ during the catalytic cycle.

An early model for the active site was found in the manganese salen class of compounds. The active sites of the enzyme and model are shown in Fig. 4.13. The Schiff base model complexes are unusual in that they are able to maintain manganese in the highly reactive +3 oxidation state, ready to catalyse the oxidation of superoxide, yet still be flexible enough to bind the larger manganese(II) ion.

Strangely, these small molecules are also used industrially as oxidation catalysts, however, small changes in their structures lead to dramatically different behaviour. In these model compounds, unlike SOD2, the metal is bound by two oxygen and two nitrogen atoms in a square planar orientation with axial positions occupied by water or reactive oxygen species. The superoxide dismutase catalytic cycle has been proposed to occur by a one electron mechanism:

$$[Mn]^{3+} + O_2^{\cdot -} \rightarrow [Mn]^{2+} + O_2$$

$$[Mn]^{2+} + O_2^{\cdot -} + 2H^+ \rightarrow [Mn]^{3+} + H_2O_2$$

As with SOD this gives an overall reaction of:

$$2O_2^{\cdot -} + 2H^+ \quad \rightarrow \quad O_2 + H_2O_2$$

An additional catalase-type cycle may occur by a two electron mechanism:

$$[Mn]^{3+} + H_2O_2 \quad \rightarrow \quad [Mn{=}O]^{5+} + H_2O$$

$$[Mn{=}O]^{5+} + H_2O_2 \quad \rightarrow \quad [Mn]^{3+} + O_2 + H_2O$$

As with catalase, the overall reaction is:

$$2H_2O_2 \quad \rightarrow \quad O_2 + 2H_2O$$

Many claims have been made for manganese salen derivatives, from the hotly debated report that they extend the lifespan of nematode worms [27, 28] to curative properties in cases of lung damage [29]. Unfortunately the compounds suffer from the same problems as other simple enzyme mimics in that they are not shielded from deactivation mechanisms by a protein shell. Furthermore they can also act as prooxidants through mechanisms such as the following:

$$[Mn]^{2+} + O_2 \quad \rightarrow \quad [Mn]^{3+} + O_2^{\cdot -}$$

$$[Mn]^{2+} + H_2O_2 \quad \rightarrow \quad [Mn]^{3+} + HO^- + HO^{\cdot}$$

$$[Mn{=}O]^{5+} + \text{alkene} \quad \rightarrow \quad [Mn]^{3+} + \text{epoxide}$$

If this was not bad enough the manganese(III)salen cation is attracted electrostatically to DNA where it binds strongly, presumably through a combination of electrostatics and π–π stacking, and cleaves the DNA strand [30]. Here some fundamental supramolecular interactions are responsible for contributing considerable cytotoxicity. This would be desirable if the compound was intended as an anticancer agent but not if it was intended to supplement the body's existing antioxidant mechanisms. Fortunately a simple non-binding derivative has been made which retains beneficial antioxidant properties and appears to have no deleterious effects. It is also possible to significantly reduce the molecule's affinity for DNA by introducing bulky antioxidant molecules such as thioctic acid that bind to the manganese salen complex to keep it clear of major and minor grooves in DNA [31]. Similar modifications have been so successful that a derivative has even been added to a skincare product to enhance its antioxidant properties.

### *4.3.4 Copper-Containing Enzymes*

Historically there is a correlation between the use of copper in enzymes, the reactions of small non-metal compounds and the advent of a dioxygen atmosphere.

Copper is a recent addition to the list of bioavailable elements because of the high potential necessary to oxidize $Cu^0$ to $Cu^{2+}$. Mineralized $Cu^+$ is insoluble at pH 7 so that the soluble $Cu^{2+}$ was unlikely to be in existence before the onset of photosynthesizing plants. Once an oxidizing atmosphere emerged on Earth the soluble $Fe^{2+}$ was oxidized to insoluble $Fe^{3+}$ leaving copper as the best metal to take over iron's catalytic function in enzymes. For this reason copper containing enzymes are so common today. Most enzymes that incorporate copper bind the metal in a distorted trigonal or tetrahedral pocket through interactions with three nitrogen-containing amino acid side chains and a fourth small molecule such as water. The amino acid is usually histidine and the fourth site is where catalytic activity generally occurs.

Copper-containing biomolecules have been divided into types 1, 2 and 3. The simplest systems contain a single copper atom with clustering increasing to the six to eight atoms in yeast metallothioneins. Type 1 centres include the 'copper blues', so called because of their colours, such as azurin (with a mass of 16.6 kDa), plastocyanin (10.5 kDa) and stellacyanin (16.8 kDa). Single copper atoms are held by three or four ligating amino acid side chains: two histidines, a cysteine anion, and methionine (glutamine in the case of stellacyanin). Crystallography shows a strained geometry for $Cu^{2+}$ but a better fit for the less stable $Cu^+$ species, as seen in Fig. 4.14. This induces a fast rate of electron transport mediated by cooperative behaviour of protein partners to catalyse:

$$O_2 + 4H^+ + 4\bar{e} \rightarrow 2H_2O$$

In many cases the metal is unusually close to the surface of the protein, unlike most enzymes, so that several enzymes can cooperate through a synergy, that perhaps exemplifies supramolecular biology, and to catalyse reactions that would be impossible with a single metal.

Type 2 centres include superoxide dismutase, discussed above, and cytochrome c oxidase (COX). COX consists of two haem iron centres, cytochrome a and cytochrome $a_3$, and two electronically distinct copper centres, $Cu_A$ and $Cu_B$, which catalyse:

$$O_2 + 4\text{cyt c}(Fe^{2+}) + 8H^+ \rightarrow 2H_2O + 4\text{cyt c}(Fe^{3+}) + 4H^+$$

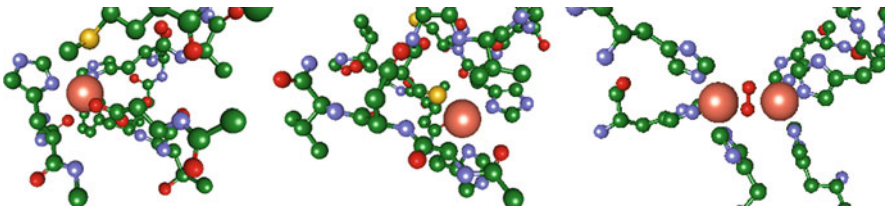


**Fig. 4.14** Metal binding sites in copper-containing enzymes: (*left* to *right*) azurin [32], stellacyanin [33] and haemocyanin [34]

COX is an extensive membrane-bound ensemble in which cytochrome $a_3$ and $Cu_B$ cooperate to form an oxygen reduction site and reduced cytochrome c containing iron(II) is oxidized through proximity to $Cu_A$. As a consequence of this reaction four protons are pumped across the membrane to set up a potential that helps to power ATP synthase. In plants, and some bacteria, the latter enzyme is serviced by another complex, membrane-bound protein ensemble, photosystem II.

Type 3 centres include haemocyanin which is used by arthropods and molluscs to transport dioxygen in the haemolymph. Generally hamocyanin contains two copper ions each coordinated to three histidines. In the colourless, $Cu^+$, deoxy form these are separated by 3.6 Å. Once dioxygen binds between the two coppers they both adopt a tetrahedral geometry and are oxidized to $Cu^{2+}$ turning the complex blue. Binding has not been fully characterized and may be 'side-on' or 'end-on'.

One way to replicate a binding environment for copper in an artificial enzyme is to start with a molecule that has threefold symmetry and incorporate three imidazole groups. While there are many ways to achieve this, one of the most interesting has been pioneered by the group of Reinaud [35]. The group has used a calix[6]arene scaffold and generated a trimethyltris(imidazole) derivative that binds $Cu^+$ in a tetrahedral pocket. The metal is bound in its preferred geometry and is protected from random reactions by the bulk of the calixarene framework. The synthesis, given in Fig. 4.15, starts from 4-t-butylcalix[6]arene with methoxy groups on alternating phenolic rings which reacts with 2-chloromethyl-1-methyl-*1H*-imidazole hydrochloride. The water-soluble sulfonated and nitrated derivatives allow the compounds' catalytic effects to be followed in aqueous solution.

Another way is to hold copper in a small macrocycle for which it has a high affinity and append several complexes to a secondary scaffold. This has been achieved by Reinhoudt, Ungaro and co-workers who introduced azacrown groups to the upper rims of calix[4]arenes [36]. The calix[4]arenes had been made water soluble through the incorporation of lower rim ether substituents. Compounds containing one, two and three azacrowns were prepared, with both 1,2- and 1,3-disubstituted regioisomers being synthesized. The copper complexes, in which each centre was initially copper(II), were tested as artificial ribonucleases (RNAses) as these RNA cleaving enzymes are known to rely on divalent metals as cofactors. Crystallographic studies have indicated that RNAse A binds one nickel(II) ion and two copper(II) ions suggesting that the latter model is likely to be an active analogue of the enzyme.

The model compounds containing two or three copper ions, the former illustrated in Fig. 4.16, were shown to be the most active at RNA cleaving at pH 7.4 in water with the three centre compound the best. Specificity for preferential cleaving of the cytosine-adenosine phosphate ester was similar to natural RNAse A. It was postulated that the multi-metal site was necessary to bind RNA in an appropriate conformation for cleavage. The cooperativity of the metal ions, coordinating binding site and molecular scaffold successfully mimicked the ability of proteins to bind metal ions in their preferred environments through particular side chains whilst ensuring that ideal spatial constraints are present through secondary and tertiary structures.

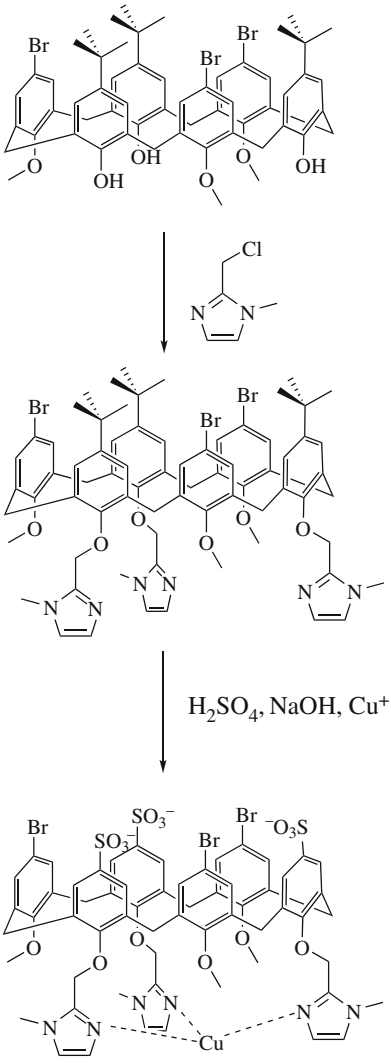



**Fig. 4.15** The synthesis of a copper enzyme model [35]

Cytochrome c oxidase mimics have long been a goal because they catalyse the reduction of dioxygen to water without the generation of other reactive oxygen species. Mimetic compounds are a necessity for any researcher interested in the native enzyme system because of its intrinsic complexity. X-Ray crystallography revealed that the active site consisted of an iron-haem complex with an adjacent copper cation bound by three histidines, in the expected motif, and a nearby tyrosine group. The importance of tyrosine had been overlooked in many model systems but was introduced as an imidazole substituent by the Collman group [37]. The enzyme active site and its model can be seen in Fig. 4.17.

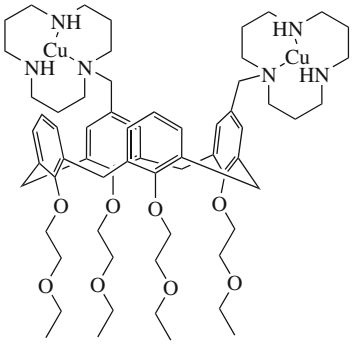



**Fig. 4.16** A calixarene-based artificial ribonuclease [36]

The Collman model has an iron-copper separation of 5 Å and catalyses the reduction of dioxygen without generating any toxic species. The mimic was incorporated into a self assembled monolayer (SAM) on a gold electrode. It was attached by either a non-conducting alkyl chain, to slow the rate of electron transfer from the electrode, or a short conjugated linker, to allow fast electron transfer. In the former SAM the rate limiting step for the reduction is the electron transfer, in the latter it is the rate of oxygen diffusion. When electron transfer was limiting it was found that removal of copper or methylation of the tyrosine resulted in significant generation of reactive oxygen species. The catalytic environment formed by copper and iron coordination sites, a tyrosine group and a dioxygen molecule is an excellent example of supramolecular cooperativity in which each component is essential but together they can perform an extremely complex function.

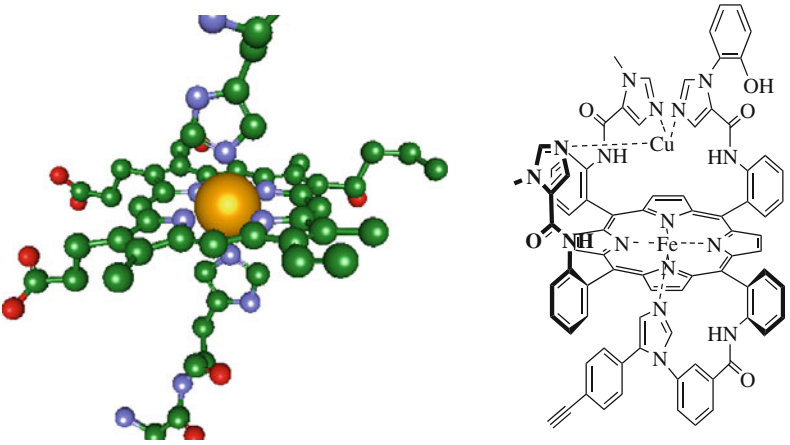



**Fig. 4.17** The active site of cytochrome c oxidase [38] and a model compound [37]

### 4.3.5 Zinc-Containing Enzymes

Zinc was one of the first metals to be identified as an essential element and the history of zinc enzymes is extensive. Lack of zinc was found to inhibit the growth of the mould *Apergillus niger* as early as the mid nineteenth century and the first zinc enzyme was identified in 1940. Since then zinc has been found to be present in well over a hundred biomolecules. Why such ubiquity? Zinc is a good Lewis acid and more accessible as the divalent cation than copper, cobalt, cadmium or even iron. As $Zn^{2+}$ it is very soluble and resistant to oxidation and reduction. It is found in seawater and has probably been the most bioavailable transition metal for most of Earth's history. Any biochemical process requiring a non-redox active metal would be most likely to encounter zinc first. The cation binds strongly to amino acid residues and preferentially forms tetrahedral complexes. This is due to zinc's electronic configuration: as $Zn^{2+}$ it has an empty *s* shell and full *d* shell. The geometric preferences observed for transition metals usually arise from an overlap of *d*-orbitals that result in six coordinate octahedral complexes or four coordinate square planar and tetrahedral complexes. When the *d* shell is empty or full no geometric preferences are imposed and the orientation of ligands is dictated by steric constraints. Usually this results in an arrangement of four ligands in a tetrahedral arrangement though, if ligands are small, six coordinate species may be formed as is believed to occur during some examples of zinc centred catalysis. With the exception of superoxide dismutase and carbonic anhydrase, zinc enzymes are generally associated with protein degradation as exemplified by carboxypeptidases and numerous matrix metalloproteinases. In these and many other examples the zinc is found in a tetrahedral pocket often comprised of three histidines with a water molecule in the fourth site.

In mammals the bicarbonate ion, $HCO_3^-$, is used as a buffer in blood and is catalysed by the zinc-containing enzyme carbonic anhydrase:

$$CO_2 + H_2O \quad \rightarrow \quad H^+ + HCO_3$$

The uncatalysed reaction is slow ($k = 9.5 \times 10^{-2}$ $Lmol^{-1}s^{-1}$ at 25°C), however, in the presence of carbonic anhydrase the rate increases to $5 \times 10^7$ $Lmol^{-1}s^{-1}$ which represents 500,000 turnovers per second for each enzyme molecule. Carbonic anhydrase has a globular structure and has a mass of about 29 kDa. The single zinc ion is bound to three nitrogens (from histidine residues) and a water molecule or, as in Fig. 4.18, nearby amino acid occupies the fourth site.

Experiments have shown that a metal is required for catalysis to occur but activity is seen for metals other than zinc. Its activity is inhibited by other small molecules that can bind to zinc in place of water and carbon dioxide, in particular cyanide, hydrogen sulfide and chloride that all bind tenaciously to transition metals. As well as its buffering ability, this enzyme represents a valuable method of converting carbon dioxide into carbonate so any advances in mimicking the behaviour of this enzyme may have implications for carbon storage. Conceivably carbon dioxide could be passed through a vat containing an aqueous solution of a carbonic

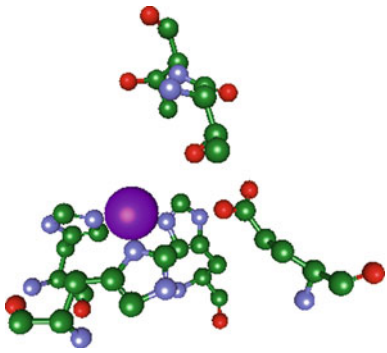


**Fig. 4.18** Three zinc-binding histidines at the active site of carbonic anhydrase [13]

anhydrase mimic where bicarbonate would be formed then the solution treated with a metal salt to precipitate an insoluble metal carbonate.

Carboxypeptidases catalyse protein degradation into their constituent amino acids by hydration and target amino acids with aromatic, hydrophobic or positively charged side chains depending on the type of carboxypeptidase. The enzymes are egg-shaped and with masses in the region of 35 kDa. Here the binding site is slightly different as a single zinc cation is held by two histidines, glutamic acid and water, as shown in Fig. 4.19.

As with carbonic anhydrase the metal is in a cleft that exposes the active site. Interestingly the metal-free enzyme is inactive but the cobalt and nickel analogues are more active. It appears that the transition state complex, where the terminal amino acid side chain is held in place while the peptide bond is hydrolysed, requires six-fold co-ordination. The activation energy required to change from the tetrahedral to octahedral geometries is higher for zinc than the other metals.

The final exemplars of a tetrahedral zinc complex are the transcription factors known as zinc fingers. These are proteins about 30 amino acids in length that recognize regions of DNA to be transcribed into RNA sequences. Zinc binds in particular

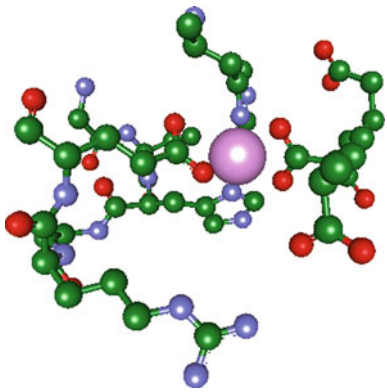


**Fig. 4.19** The zinc binding site in carboxypeptidase [39]

sites in order to form a protein loop. The final structure contains an α-helix that interacts with specific sequences of DNA bases. Typical binding is between two histidines (H) and two cysteines (C) in a sequence below where X is an amino acid other than histidine or cysteine.

$$\text{-F-X-C-X}_{(2\text{-}4)}\text{-C-X}_{(12)}\text{-H-X}_{(3\text{-}5)}\text{-H-}$$

Given the range of zinc enzymes it is not surprising that model systems have been developed to mimic the active site. There are parallels with copper enzymes which are also often in tetrahedral pockets so it is no surprise that groups with an interest in copper enzymes have also turned their attention to zinc. Indeed, the zinc(II) binding sites in enzymes are very similar to those used to bind copper(I); three histidines and a fourth small ligating molecule. The Reinaud group has successfully bound zinc in calix[6]arenes with three pendent imidazole groups, as shown in Fig. 4.20, and X-ray structures show that the zinc cation binds all three pendent groups while being drawn into the macrocyclic cavity [40]. Crystallographic evidence for propionitrile and heptylamine coordination is suggestive of the proposed structure of the carboxypeptidase active site while $^1$H NMR analysis was indicative of bound water when the zinc complex was crystallized from organic solvents.

The potential for macrocycle complexes that bind zinc in a tetrahedral pocket to mimic natural enzymes such as carbonic anhydrase has been investigated for many years, most notably by the Kimura group. Models for carbonic anhydrase have been based on tri- and tetraazamacrocycles with pendent sidearms that coordinate to the apical position of zinc. X-Ray data for related zinc enzymes, the liver alcohol dehydrogenases, reveal a zinc pocket that comprises two cysteines, one histidine and a water molecule in the expected tetrahedral geometry [41]. The acidity of the water molecule is an important factor in activating the complex and is controlled through the interaction with zinc. The Kimura group investigated the reactions between simple hydrated zinc complexed by either [12]aneN$_3$ (1,5,9-triazacyclododecane) or [14]aneN$_4$ (1,4,8,11-tetraazacyclotetradecane, also called cyclam) and found that $\{Zn{\cdot}H_2O{\cdot}[12]aneN_3\}^{2+}$ was highly effective [42]. When the catalyst was added to propan-2-ol in the presence of 4-nitrobenzaldehyde the former was dehydrogenated and the latter transformed into 4-nitrobenzyl alcohol in a yield of over 7000% based on the concentration of the zinc complex.

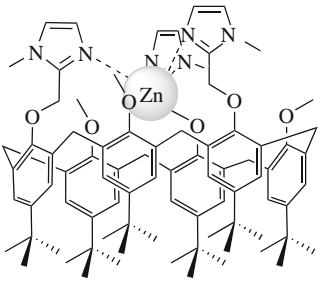



**Fig. 4.20**
A carboxypeptidase model [40]

The Walton group has also made models for zinc enzymes but attempted to incorporate another feature common to many of them, positioning the metal ion at the bottom of a constrictive cleft. The approach has the advantage that the only competing substrates ought to be small molecules that can fit inside the cleft. In the case of a compound based on cyclohexane the zinc is bound to acetate, as a model for bicarbonate, in a manner reminiscent of the enzyme. The compound, *cis*,*cis*-1,3,5-tris[*E*,*E*-3-(2-furyl)acrylideneamino]cyclohexane, has a pocket at its base, as shown in Fig. 4.21, for those metals that prefer a tetrahedral coordination geometry [43].

Ordinarily the ligand would be expected to adopt an open, almost planar, shape but when zinc(II) is added, as shown in Fig. 4.21, it pulls the three amine groups together which creates an extended hydrophobic pocket with the zinc cation buried at the bottom. Exchange between a zinc-bound hydroxide group and methanol, to form the methoxide, would not ordinarily be expected to occur but is observed. It has been hypothesized that displacement of hydroxide by the much less acidic methoxide anion can only proceed because of the hydrophilic nature of the cavity and the focused hydrogen bonding array formed by acidic alkene hydrogen atoms around the active site. By a pleasing coincidence ethanol displaces water in a similar manner at the active site of liver alcohol dehydrogenase.

Phosphate ester hydrolysis mimicry has also been demonstrated by several groups including those of Kimura [44] and Reinhoudt [45]. Here the zinc takes an analogous role to that which it has in alkaline phosphatase (AP) where two zinc ions work with a magnesium ion to dephosphorylate biopolymers, including RNA and DNA, that contain phosphate ester bonds. Two phosphate oxygens bridge between the zinc ions, which are only 2.9 Å apart, so that it can be attacked by, and then attached to, a serine residue that is in close proximity to the zinc ions. In Kimura's compound, illustrated in Fig. 4.22, a tetrahedral zinc is additionally bound by a pendent alcohol, analogous to the position of serine in AP, within an azacrown complex. In Reinhoudt's compounds two zinc ions are bound by pyridinyl substituents on the upper rim of a calix[4]arene. The pyridines have pendent alcohol groups and are attached to the calix[4]arene by ether links so that they can bind the zinc ions strongly while leaving space for further reactions with phosphate groups over the central cavity of the calix[4]arene.

Poijärvi-Virta and co-workers synthesized di- and trinuclear zinc complexes from benzene substituted in the 1,3 or 1,3,5 positions by 1,5,9-triazacyclododecane ligands via methyloxy linkers [46] as shown in Fig. 4.23.

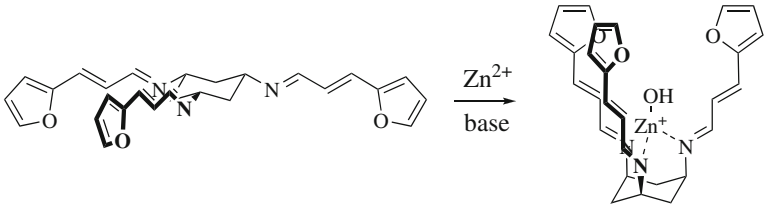



**Fig. 4.21** A model for the carbonic anhydrase active site [43]

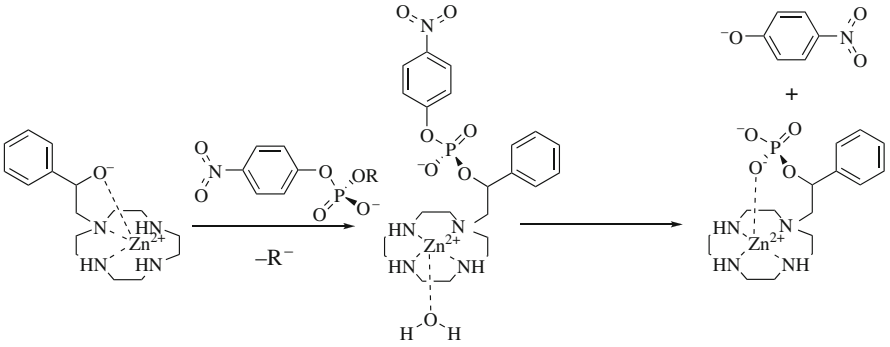
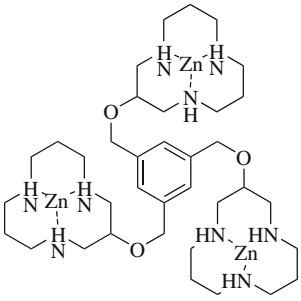


**Fig. 4.22** A model for phosphate ester hydrolysis [44]

**Fig. 4.23** A sequence-specific artificial ribonuclease [46]



Both compounds appear to behave as zinc RNAses. The disubstituted complex was effective at cleaving short oligoribonucleotides into fragments at the points where uracil is linked to one other base by phosphate, such as UpG or CpU, but did not affect UpU. Conversely, the trisubstituted complex favoured cleavage at UpU. The complementarity between binding motifs available to the zinc-macrocycle complex, together with the geometric disposition, suggests a supramolecular solution to the mechanism of the artificial RNAse. The mechanism proposed for the latter is that two zinc-macrocycle moieties bind uracil through complementary hydrogen bonding and a nitrogen-zinc interaction while the third macrocycle binds to, and activates, the phosphate bridge so that an adventitious water molecule can initiate phosphate cleavage.

### *4.3.6 Photosynthesis and Artificial Leaves*

Of the many complex natural mechanisms for generating useful biochemical energy perhaps the most interesting is photosynthesis. The ability of plants to turn sunlight into energy has fascinated humans for as long as they have watched plants grow. More recently we have been able to unravel the biochemical pathways involved

and even the structures of the active sites within the photosynthetic molecules. An interest in photosynthesis would seem to be limited to botanists and biochemists but it is becoming clear that isolated aspects of the process can be copied by unnatural molecules. One simple observation is that the energy liberated by photosynthesis could be harnessed for use in domestic and industrial applications but of greater relevance, given the current reliance on fossil fuels for energy, is another step in which water is cleaved.

In the photosynthetic cycle oxygen is liberated as $O_2$ gas but the hydrogen evolved is in the form of protons that are transported to one side of the lipid bilayer. A potential gradient is set up, with the charge balanced by electrons that reduce quinones. If we could construct a synthetic version of this portion of the enzyme, but design it to produce both hydrogen and oxygen gases, it would have immense potential as a source of clean energy through the storage and later recombination of the gases. They react exothermically, and so generate energy, yet the only other product of the reaction is pure water. To see how numerous supramolecular interactions are central to the photosynthetic process, and how unnatural supramolecular systems can mimic the natural processes, it is important to look at the chemical pathways involved and the chemical structures responsible for each stage.

Plants, and some bacteria, generate dioxygen and carbohydrates from carbon dioxide through two stage dark and light reactions. In plants the light independent Calvin cycle uses energy collected by chlorophyll containing light harvesting antennae to convert carbon dioxide into triose phosphate glyceraldehyde-3-phosphate that later reacts in the mitochondria to form the familiar hexose carbohydrates, or sugars.

Calvin cycle:

$$3CO_2 + 6NADPH + 5H_2O + 9ATP \rightarrow C_3H_5O_3\text{-}PO_3^{2-} + 2H^+ + 6NADP^+ + 9ADP + 8P_i$$

Light dependent stage:

$$2H_2O + 2NAD^+ + 2ADP + 2P_i + h\nu \rightarrow 2NADPH + 2H^+ + 2ATP + O_2$$

Overall reaction:

$$6CO_2 + 6H_2O \rightarrow C_6H_{12}O_6 + 6O_2$$

Similar reactions occur in red algae and cyanobacteria but because the full spectrum of sunlight does not always penetrate very far below the surface of the sea they make use of different pigments. These may be structurally similar to those used by plants, and are therefore classed as bacteriochlorophylls, or belong to the structurally distinct phycobilin class of compounds. Unlike chlorophyll, which has a cyclic tetrapyrrole structure containing magnesium, the phycobilins are acyclic tetrapyrroles with similarities to the breakdown product of the haem ring, bilirubin.

As the only light available may be from a weak source emanating from geothermal vents the antenna efficiency must be high. This is achieved through the unusual step of attaching the chromophores to a protein which serves to amplify the energy received before transmitting it to more conventional chlorophyll antennae.

Focusing on the more conventional behaviour of plants, the first process in the photosynthetic pathway is the collection of photons, or light harvesting, through which the energy of sunlight is transferred to the leaf. This is achieved by exciting molecular antennae. Although there are many examples of excitation in small molecules it is hard for the excited species to retain the energy involved or to channel it into a useful direction. In larger organic molecules, and indeed some metal complexes, that have numerous low lying molecular orbitals the energy can be redistributed into other forms. In certain examples the incoming photon can initiate a change in molecular conformation, as in retinal, but in the photosystems the photons excite the antenna molecules, pigments such as chlorophyll, and that excitation energy is transmitted firstly to a local pheophytin and then on to quinones, specifically plastohydroquinone. The quinone molecules then cross the cell membrane, accompanied by charge neutralizing protons, and are transferred through a multicomponent membrane-bound protein complex, that includes a cytochrome and plastocyanin, to photosystem I (PSI). The energetic quinones are processed by PSI, a ferredoxin centre and ferredoxin-NADP reductase to generate NADPH. The protons are all generated in the interior of the cell and go on to power the propeller-like ATP synthase that adds a phosphate group to adenosine diphosphate (ADP). ATP acts an energy store: when it is hydrolysed back to ADP and phosphate this energy is released to be used where required.

As a consequence of the transfer of an electron the chlorophyll centre becomes positively charged and attracts an electron from a tyrosine side chain which in turn requires an electron from another reactive site. In photosystem II (PSII) this is a cluster formed of four manganese ions and one calcium ion generally known as the oxygen evolving complex. Three of the manganese cations and the calcium cation are linked by bridging oxygen atoms, to form a cube, while the fourth manganese cation is linked to the cube by a further oxygen atom. The cluster is held in place by numerous amino acid side chains but the strongest interactions appear to be with one histidine, two aspartate and three glutamate groups. Anionic cofactors are involved though whether chloride or carbonate is critical remains speculation. Regardless of the exact nature of the reaction at the oxygen evolving centre, electrons are generated as a result of water cleavage though the exact mechanism has yet to be observed. What is known is that there are four chlorophyll harvesting centres and the water-splitting reaction generates four electrons as shown:

$$2H_2O + 4\,h\nu \rightarrow O_2 + 4H^+ + 4\bar{e}$$

Each of the four manganese ions cycles between Mn(II) and Mn(III) to generate the four electrons necessary. Had each pathway consisted of a single antenna and manganese centre the overall water cleaving reaction would take far longer and require a complex stepwise pathway: consequently the entire photosynthetic

process has evolved to produce four electrons in parallel so that it works at optimal efficiency.

Two goals are evident in mimics for the complicated chemistry of photosynthesis. Firstly there is the attempt to generate energy from sunlight which requires an antenna that can be linked to a molecule which can translate the energy from photons into some more useful form. The second goal is to devise a model for water splitting to generate molecular hydrogen for storage and later use as a fuel. While both goals would benefit from better crystallographic and mechanistic data this has not prevented researchers from designing and synthesizing model compounds.

One key issue associated with natural photosynthesis is the energy of the photons involved. Initially light interacts with the $P_{680}$ species in PSII, a pigment (chlorophyll *a*) that absorbs at 680 nm. The energy involved is enough to power the water splitting reaction but not enough for the reactions of PSI. A second pigment, $P_{700}$, is involved in energy transfer within PSI. In both cases the amount of incoming full spectrum sunlight that can be utilized is only between 30 and 60%. It is not beyond reasonable speculation that an artificial light harvesting system could be devised with a greater efficiency and, given the increasing need for more 'clean' energy, an artificial form of photosynthesis would seem to be a very worthy goal.

Photosystem mimics are modelled along the lines of many supramolecular sensors in that there is a recognition, or receptor, site linked to a transduction site. The molecules either link antennae to a water splitting centre, for the production of hydrogen, or to a complex that can turn the energy from photons into a more useful form such as electrical energy. In either event the first requirement is to identify a suitable antenna. Nature uses chlorophyll or a similar conjugated molecule so that the energy gathered at one end can be directly transferred for a reasonable distance. In practice chlorophyll is too complicated a molecule to use in model systems so most are based on a transition metal complex.

One such example involves a ruthenium tris(bipyridine) complex, illustrated in Fig. 4.24, in which the metal can accept energy from photons and transfer it to the bipyridine ligands. The process of metal to ligand charge transfer is a well known phenomenon in coordination chemistry and the experimental conditions needed to form the complexes are fairly well understood. Added complexity is encountered when the bipyridine units have been modified as in the work of Hammarström [47].

Early attempts to link ruthenium to manganese involved a single ion of each rather than the four chlorophyll antennae and tetramanganese cluster seen in natural examples. One important observation was that the light harvesting ruthenium site had to be remote from the redox active manganese site to avoid the latter quenching the excited state of the former. At 10 Å it was a major problem, with quenching occurring in less than 10 ns, however when the separation was increased to 14 Å quenching was reduced by a factor of 1000. It is perhaps no coincidence that the manganese clusters in PSII are separated from the chlorophyll harvesting site by 15–20 Å.

Having demonstrated that single electron transfer could be modelled successfully the next stage was to put several of the systems together so that enough manganese

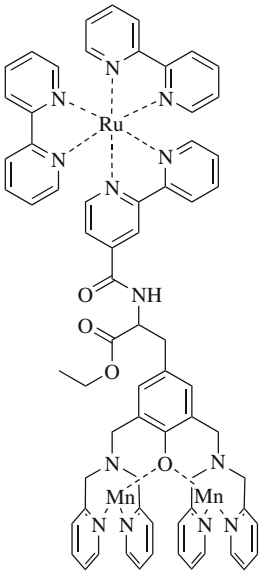



**Fig. 4.24** A multifunctional model for photosystem II [47]

centres are focused together to mimic the single water cleaving site of PSII. Hammarström linked a dimanganese cluster to a single ruthenium tris(bipyridyl) centre that has two light harvesting napthylene bis(amide) antennae attached by aromatic spacers. So far it has not been possible to oxidize the cluster four times to initiate water cleavage but it is undoubtedly the best biomimetic approach to date.

Other approaches for translating light energy into chemical energy have been investigated. These range from analogues of PSII that make use of alternative antennae and metal centres to materials with intrinsic electron transfer properties. A good example from the first camp is a system devised by Albinsson that links fluorescein to the ends of a short strand of DNA [48]. A complementary strand is attached to a hydrophobic zinc porphyrin at its mid point that embeds itself in a phospholipid membrane. Light is harvested by the fluorescein antennae and electrons transferred through the DNA strand to the porphyrin. Once the electrons arrive at the membrane bound porphyrin they are transferred to membrane bound quinones in an identical manner to the PSII mechanism. The novelty of this model system is that it is membrane bound and mimics the PSII quinone excitation pathway by receiving light on one side of the membrane and transmitting the effects to species that also only exist in the membrane. What the model does not do is to power the water splitting reaction. Nor does it take light energy and turn it into a useful form that could be stored for future use in a fuel cell or similar device.

A supramolecular approach to light harvesting has also been taken by Hamilton's group. In earlier work they had designed compounds with hydrogen bond donor-acceptor motifs to complement barbituric acid [49]. They used the affinity between these host-guest systems to prepare a self-assembling complex between a porphyrin with an appended barbiturate group and a receptor with a fluorescent dansyl

group attached. The complex was modelled and distance between the chromophore centres found to be about 23 Å. Fluorescence experiments demonstrated that the two chromophores were able to communicate as occurs in natural photosystem architecture [50].

One of the difficulties in designing PSII models is that although X-ray structures for the manganese cluster have been determined they may not accurately reflect the biologically active unit. The X-ray crystallographic experiment, by definition, relies on high energy electrons to determine the positions of atoms within the structure. Ordinarily this is not a problem, however, the electrons can interact with the metal centres and affect their oxidation states. Consequently the oxidation states determined may not be those of the active catalytic site but be artefacts of the experiment. The oxidation states alone would not matter if the metals were in isolation but they will influence the preferred interactions between the metals and their surroundings. As this includes bond lengths and ligand geometry, the potential for Mn(IV) and Mn(III) species to be reduced to Mn(II) may impart a significant level of uncertainty in the validity of the final crystal structure. Fortunately the structure solved by Barber and his group that forms the basis of most mimetic attempts has been compared to computational models and is in very close agreement [51]. Nevertheless it is important to bear in mind that artefacts in crystal structures of proteins are possible and secondary methods of analysis are advisable before proposing hypotheses based on structural data.

Alternatives to biomimicry include materials based approaches as shown in Fig. 4.25. Mallouk has used commercial solar cell technology in which a dye is

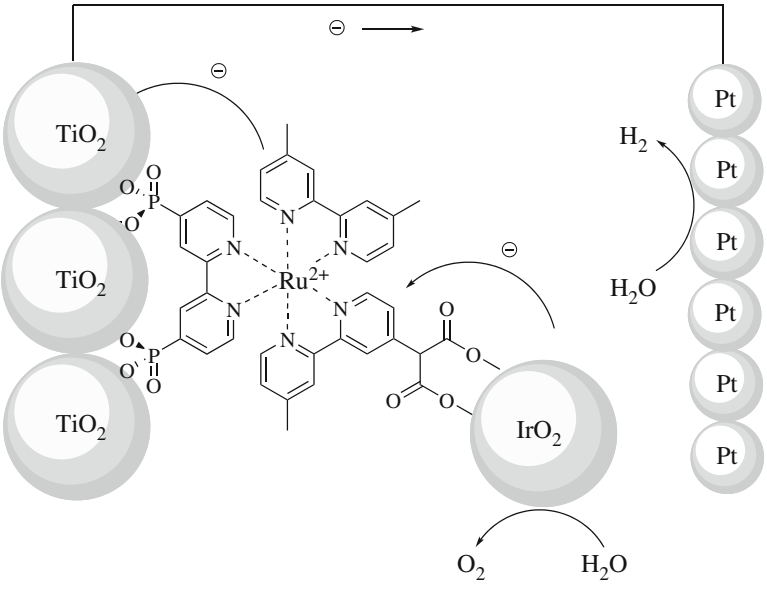



**Fig. 4.25** A solar cell design inspired by photosystem II [53]

incorporated within a thin film of titanium dioxide. The dye is excited by oncoming photons and rapidly transmits the energy to the solar cell. The Mallouk group has incorporated a surface bound complex that links a tris(bipyridyl) ruthenium centre to iridium oxide. The transfer of electrons from iridium through the ruthenium centre to the cell sets up active iridium sites for water cleavage.

The only problem is that the system requires electrical energy, rather than sunlight, to operate. Despite this, it demonstrates that the water splitting aspect of PSII can be reproduced [52]. An improved system has been designed by Nocera whose group modified an indium-tin oxide electrode by incorporating cobalt phosphate to fabricate an oxygen evolving catalyst [53]. The catalyst is exceptional in that it works in water at pH 7 and room temperature. Furthermore it is able to repair itself. The natural photosynthetic process is extremely demanding and plants are constantly removing damaged material and replacing it. As the same pressures exist for model systems many are unable to undergo more than a few catalytic cycles before breaking down. In Nocera's system the material decomposes to lose cobalt but the local environment rapidly reoxidizes the metal which becomes active once more.

### *4.3.7 Cyclodextrins as Artificial Enzyme Supports*

Cyclodextrins, described in Chapter 1, are naturally occurring macrocycles that exist in a number of different sizes. Externally they are decorated with hydroxyl groups but have hydrophobic central cavities that can bind appropriately sized guests. In fact they appear to be ideal molecules to use as a basis for an enzyme mimic. Furthermore the hydroxyl groups can be regioselectively functionalized.

An inspired example of cyclodextrin enzyme mimicry came from Breslow's group in 1970 [24]. One equivalent of pyridine-2-carboxylic acid was attached to α-cyclodextrin through esterification of one hydroxyl group. The nickel chelate was prepared by the addition of nickel(II) chloride and the complex treated with pyridine carboxaldoxime which bound to the nickel. The result, illustrated in Fig. 4.26, was a complex nickel hydrolase mimic. It was known that 4-nitrophenyl derivatives bound within the α-cyclodextrin cavity. Upon hydrolysis these compounds generate brightly coloured nitrophenolate anions, the formation of which was followed spectroscopically by the increase in the intensity of a signal at 370 nm. Breslow studied the hydrolysis of 4-nitrophenyl acetate and found an uncatalysed rate of $7 \times 10^{-5}\text{min}^{-1}$ at ambient temperatures and under alkaline conditions used to slow down the kinetics to an observable rate. The rate increased almost one thousand fold, to $5 \times 10^{-2}\text{min}^{-1}$, with the addition of 10 mM of the cyclodextrin catalyst.

Later work by the Bresow group dispensed with metals entirely. Disubstitution of β-cyclodextrin with imidazole groups, analogous to the histidine side chains of amino acids in ribonuclease A, led to a catalyst that increases the rate of a small cyclic phosphate by a factor of 120. The selectivity for one product over the other was 99:1 but the rate increase and selectivity were only observed for cyclodextrins

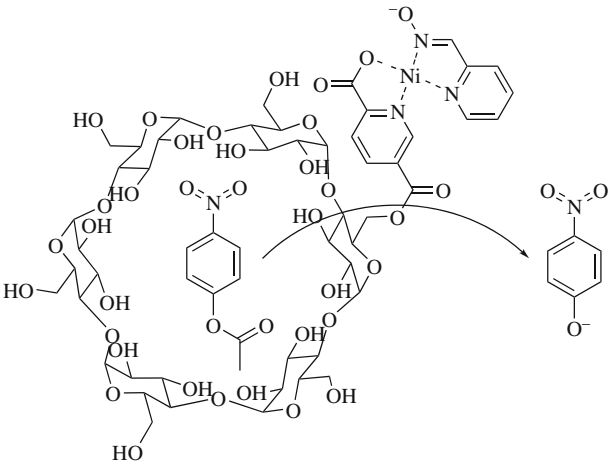


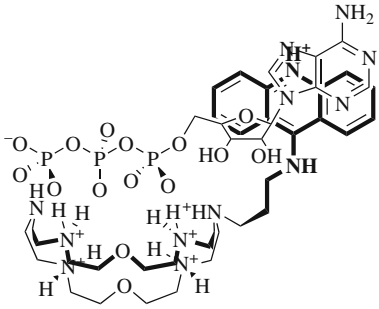
**Fig. 4.26** A cyclodextrin scaffold for an artificial esterase [24]

where the two substituents were on adjacent rings. Since the success of Breslow's approach many other cyclodextrin based enzyme mimics have been synthesized. In particular they have been incorporated to mimic hydrophobic binding sites that are observed in many enzymes. A good example comes from Kim and Lee who appended β-cyclodextrin to tri- and tetraazamacrocycles which were then used to bind zinc and perform as carboxypeptidases. The mimics enhanced the hydrolysis rates of 4-nitrophenyl acetate by up to 300 times by binding the aromatic ring of the substrate whilst exposing the terminal ester to a nearby zinc-bound hydroxide [54].

### *4.3.8 Model Enzymes that do not Require Metals*

Not all enzymes are metal centred. Many rely on carefully positioned amino acid side chains to bind the reactant, or reactants, and promote the formation of the transition state of the reaction before expelling the product, or products. The Lehn group synthesized a hexaaza[24]crown-8 macrocycle that bound ATP very effectively and was able to catalyse its hydrolysis to ADP while simultaneously phosphorylating the macrocycle [55]. Rates were 100 times that of the uncatalysed reaction and were found to be insensitive to ionic strength but modified macrocycles octaaza[24]crown-8 and diaza[24]crown-8 enhanced hydrolysis rates by 300 times under alkaline conditions. The same group later prepared derivatives of hexaza[24]crown-8 with an appended acridine group which acted both to enhance binding to the nucleotide by π–π stacking and as a self-indicator as binding could be followed by shifts in the acridine $^{1}$H NMR signals [56]. A crown ether-ATP complexes is shown in Fig. 4.27 to illustrate the complementary binding sites.

**Fig. 4.27** A crown ether ATPase [55]



## 4.3.9 Molecularly Imprinted Polymers

Molecularly imprinted polymers (MIPs) form when template molecules are surrounded by monomers which then react with cross-linkers to produce a porous polymer around the templating species. Once the template has been removed a porous polymer remains in which the pores are ideal binding sites for a target with the same molecular shape as the template as shown in Fig. 4.28.

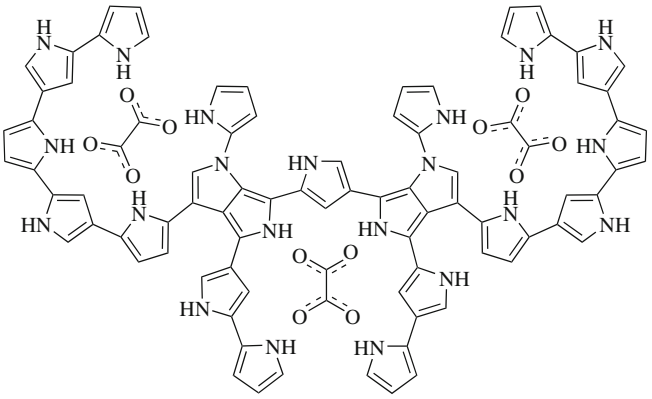


**Fig. 4.28** A molecular imprinted polymer for the oxalate anion

If the template was a transition state analogue for a particular enzyme reaction then the resulting polymer should contain myriads of sites capable to catalysing that reaction. So far this method has been used successfully to prepare catalysts for hydrolysis reactions [57] and Diels-Alder cyclizations [58] but it could equally easily be applied to mimic more traditional enzymes.

## 4.3.10 Combinatorial Polymers

Combinatorial chemistry was developed as a method to generate large numbers of compounds that could be screened for numerous applications, primarily as drug

candidates. The background to the approach comes from solid phase peptide synthesis and, latterly, from dendrimer formation. In both cases an initial building block or functionalized support material is activated and reacted with a 'first generation' substituent. The first generation material is itself activated, usually through deprotection, and the process repeated to synthesize successive generations. The reactant added at each stage could be different from the previous one and the order in which the different reactants added can be varied to generate libraries of compounds that can then be screened for desirable properties which will depend upon the application to which they will be put. Combinatorial polymers are prepared by varying the side chains on a functionalized polymer and then test its catalytic activity. As an extension to this the groups of Lehn [59] and Sanders [60] have pioneered the concept of a dynamic combinatorial library based on reversible reactions between small molecules. Candidates emerge from this process that are able to act as catalysts.

### *4.3.11 Dynamic Combinatorial Libraries*

Dynamic combinatorial libraries employ reversible supramolecular reactions to generate the most stable species as shown in Fig. 4.29. In many examples the supramolecular product forms in response to a template: in the case of self-replication this is autotemplation. According to Sanders, one of the field's pioneers, the inspiration behind the approach comes from mimicking the immune system's response to pathogens and antigens that leads to adaptive immunity. From this perspective the self-replication of supramolecules from a dynamic library could be viewed as an autoimmune response. As with much of supramolecular chemistry it is a vast oversimplification of the natural phenomenon, however, a better understanding of the fundamental mechanisms involved in this model of the immune process may have profound consequences for the way future therapies are designed.

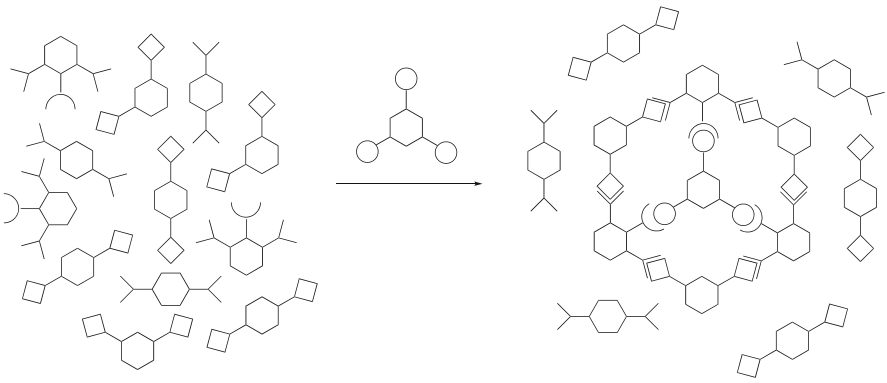


**Fig. 4.29** A dynamic combinatorial library resolving in the presence of a template

## 4.4 De novo Design and Evolutionary Development of Enzymes

Many enzyme cores are based on natural macrocycles, including porphyrin and corrin rings found in haemoglobin and $B_{12}$, so their synthetic analogues are commonly encountered in biochemistry. While these synthetic enzymes, or synzymes, are often quite effective in mimicking natural enzyme activity there is a secondary feature which is often overlooked. The metal-macrocycle structures are never encountered in isolation; they are always incorporated within a protein. This allows the protein to fold around the active site and act as a support for substrates and the products of catalysis. Furthermore, the protein's tertiary structure can fold in such a way as to form channels, grooves or pockets that allow only the target molecules to approach the metal centre. Selectivity is therefore based on a combination of weak forces, a classic mechanism in supramolecular chemistry. The substrate's journey to the active site will be subject to exclusion based on size, conformation, electrostatic charge distribution, hydrogen bond strength, $\pi-\pi$ interactions and a wealth of other influences. The absence of a complex scaffold around the active site greatly affects the selectivity and activity of a synzyme. In particular it increases the likelihood of poisoning which occurs when a molecule binds to irreversibly to the active site, for example when cyanide rather than carbon dioxide and water bind to zinc in carbonic anhydrase. Nevertheless, a large number of enzyme mimics are based on macrocyclic or acyclic ligand complexes of transition metals.

One problem with enzyme mimetics is that they are usually designed to catalyse the same reactions as a biological analogue. As none approach the catalytic activity of the original there is a sense that biotransformations should be left to the natural enzymes and that organisms that can express the catalytic activity should be cultured specifically for that purpose. Certainly this approach has been applied widely in the past: the use of yeast in baking and, more particularly, brewing is a good example of using the biotransformation of sugars into carbon dioxide (raising dough) and ethanol (brewing). It should not be forgotten that penicillin is produced enzymatically by *Penicillium* fungi in response to stress.

If the targeted chemical transformation is unknown in Nature then the catalyst designed for the purpose usually bears little relationship to a natural enzyme. Often the catalysts are relatively simple molecules such as the metallocenes used in Ziegler-Natta polymerization. Important structural factors such as steric inhibition of particular pathways for incoming substrates or chirality of the product are usually enforced by quite crude methods. Thus a symmetric catalyst may be synthesized if a syndiotactic polymer is required as this can best be achieved if sequential substrate molecules alternate in their approach to the active site. Similarly a bulky organic substituent may be attached to the active site in order to block a particular route for an approaching substrate molecule. Enzymes, as has been shown, are much more sophisticated. The difficulties in reproducing thier level of specificity and catalytic activity in an artificial system are, conceptually and practically, immense for a synthetic chemist or biochemist. Firstly, the active site has to be designed and demonstrated to have the desired catalytic activity. Secondly,

the functional catalyst would have to be incorporated within a protective environment that reduces poisoning by substrates that bind irreversibly to the active centre. Finally, the complete ensemble must be an improvement over other catalysts that perform the same function. Therefore, to really be of use, an artificial enzyme must have enhanced catalytic activity over existing, natural enzymes that could be exploited in the laboratory. Alternatively the artificial enzyme must catalyse an important reaction that is unknown in Nature and do so in a very efficient manner. To do either necessitates a multidisciplinary approach to the design, synthesis and exploitation of catalysts based upon careful control of weak, reversible interactions that are at the core of supramolecular chemistry. The challenge has been taken up by many groups but the first real success came from Baker and colleagues.

In 2008 the Baker group reported not one but two de novo enzymes. The group's approach started with a search for suitable unnatural reactions that could be catalysed within an active site decorated with amino acids. Two were chosen: the retro-aldol reaction [61] and the Kemp elimination [62]. The former was initiated by nucleophilic attack by lysine on a ketone group in the substrate to form a carbinolamine intermediate from which a water molecule was eliminated to leave an imine. Next the alcohol group was deprotonated leading to cleavage of a carbon-carbon bond. The resultant enamine rearranged to an imine and was hydrolysed to the desired product which was ejected from the active site leaving it vacant for another cycle. The Kemp elimination is simpler in that it requires a basic amino acid sidechain, such as glutamate or histidine, to abstract a hydrogen atom in order to stabilize the transition state which quickly rearranges into the product.

Using computational techniques an active site was designed that would catalyse both of the reactions. This was followed by an exhaustive search of known protein structures that contained, or could be mutated to contain, the active sites chosen by the design process. To undertake the search a global network that used idling computers to evaluate candidate proteins was accessed. In both cases a number of potential targets were discovered and encoded within the genes of *Escherichia coli*. The catalytic activity was then expressed by these bacteria. The catalytic activities of the bacteria were determined and the most active subjects were then subjected to random site-directed mutagenesis in an elegant example of in vitro directed evolution. The outcome of this method was that synthetic enzymes had been created for two unnatural chemical reactions. These enzymes were expressed by a bacterium that was well-suited to being 'farmed' for its novel abilities thus giving access to the new catalysts on a potentially industrial scale. Further, natural, evolution could be used to improve the existing catalytic activity or to generate strains that are specific for other substrates.

Some caution should be exercised however when working with synthetic enzymes. A similarly designed and expressed de novo enzyme to mimic the naturally occurring triose phosphate isomerase was reported in 2004 only for the papers to be retracted when it appeared that contamination by unmodified *Escherichia coli* had been responsible for the observed enzymic activity. For this reason it seems preferential to target reactions that have no known natural catalysts.

## 4.5 Summary

Enzymes are remarkably sophisticated molecules that catalyse reactions, essential for life, which could otherwise not occur on Earth under ambient conditions. Only after structural elucidation, often decades after the first detection of catalytic activity, has it been possible to understand how some of them operate at an atomic level. Most have not been characterized to this level and their chemical mechanisms remain an enigma. What all have in common is a complex binding site that changes to accommodate molecular guests and transform them into valuable chemical products. The interactions required to entice and then bind the guests are those so often seen in supramolecular chemistry so it is no surprise that most of the early work undertaken by supramolecular chemists was aimed at producing artificial enzymes. Where some structural information is known it is possible to mimic the enzyme's activity through careful design of an artificial analogue. Rarely do these work with anything like the activity of their natural analogues, however, they are often easier to analyse and so give valuable mechanistic information.

Given the advances in molecular biology and computer modelling it is now possible to design biological systems, based on known biomolecules, that catalyse reactions of industrial importance, currently unknown in Nature. The design of de novo enzymes owes a lot to the principles of supramolecular chemistry though the synthetic methods are entirely within the realm of the biologist. While this field is very much in its infancy it has the potential to deliver huge rewards in terms of biologically engineered catalysts that can produce valuable chemicals, pharmaceuticals in particular, or that can enzymatically degrade harmful molecules.

## References

1. Kühne WF (1877) Erfahrungen und Bemerkungen über Enzyme und Fermente Untersuchungen a d physiol. Institut der Universität Heidelberg 1:291–324
2. Buchner E (1907) Nobel prize lecture
3. Sumner JB (1926) The isolation of the enzyme urease: preliminary paper. J Biol Chem 69:435–441
4. Northrop JH (1930) Crystalline pepsin. J Gen Physiol 13:739–766
5. Stanley WM (1935) Isolation of a crystalline protein possessing the properties of tobacco-mosaic virus. Science 81:644–645
6. Blake CCF et al (1965) Structure of hen egg-white lysozyme: a three-dimensional fourier synthesis at 2 Å resolution. Nature 206:757–761
7. Jabri E et al (1995) The crystal structure of urease from Klebsiella aerogenes. Science 268:998–1004 PDBID:2KAU
8. Koshland DE (1994) The key-lock theory and the induced fit theory. Angew Chem Int Ed Engl 33:2375–2378
9. Brink C et al (1955) Structure of vitamin $B_{12}$: X-ray crystallographic evidence on the structure of vitamin $B_{12}$. Nature 174:1169–1171
10. Woodward RB (1973) The total synthesis of vitamin $B_{12.}$ Pure Appl Chem 33:145–177
11. Dreos R, Siega P (2006) Kinetics and mechanism of metallacyclization in a chloromethylcobalt complex with a salen-type ligand. Organomet 25:5180–5183

12. Baudry M et al (1993) Salen-manganese complexes are superoxide dismutase-mimics. Biochem Biophys Res Com 192:964–968
13. Avvaru BS et al (2010) A short, strong hydrogen bond in the active site of human carbonic anhydrase II. Biochem 49:249–251 PDBID:3KS3
14. Mullis K (1999) Dancing naked in the mind field Bloomsbury, New York, NY, pp 3–14
15. Ventor JC (2007) A DNA-driven world (Richard Dimbleby Lecture). http://www.bbc.co.uk/print/pressoffice/pressreleases/stories/2007/12_december/05/dimbleby.shtml
16. Xu Y et al (2006) A new strategy for structure determination of large proteins in solution without deuteration. Nat Methods 3:931–937 PDBID:2H35
17. Collman JP et al (1973) Reversible oxygen adduct formation in ferrous complexes derived from a picket fence porphyrin – model for oxymyoglobin. J Am Chem Soc 95:7868–7870
18. Collman JP et al (1975) 'Picket fence porphyrins'. Synthetic models for oxygen binding hemoproteins. J Am Chem Soc 97:1427–1439
19. Almog et al (1975) Condensation of tetraaldehydes with pyrrole. Direct synthesis of 'capped' porphyrins. J Am Chem Soc 97:226–227
20. MacMunn CA (1886) Researches on myohaematin and the histohaematins. Phil Trans R Soc 177:276–298
21. Keilin D (1925) On cytochrome, a respiratory pigment, common to animals, yeast, and higher plants. Proc R Soc Lond B 98:312–339
22. Groves JT, Neumann R (1989) Regioselective oxidation catalysis in synthetic phospholipid-vesicles – membrane-spanning steroidal metalloporphyrins. J A Chem Soc 111:2900–2909
23. Yang J, Breslow R (2000) Selective hydroxylation of a steroid at C-9 by an artificial cytochrome P-450. Angew Chem Int Ed 39:2692–2694
24. Breslaw R, Overman LE (1970) An 'artificial enzyme' combining a metal catalytic group and a hydrophobic binding cavity. J Am Chem Soc 94:1075–1077
25. Strange RW et al (2007) Molecular dynamics using atomic-resolution structure reveal structural fluctuations that may lead to polymerization of human Cu-Zn superoxide dismutase. Proc Natl Acad Sci USA 104:10040 PDBID:2V0A
26. Quint P et al (2006) Crystal structure of nitrated human manganese superoxide dismutase: mechanism of inactivation. Free Radic Biol Med 40:453–458 PDBID:2ADQ
27. Melov S et al (2000) Extension of life-span with superoxide dismutase/catalase mimetics. Science 289:1567–1569
28. Keaney M, Gems D (2003) No increase in lifespan in Caenorhabditis elegans upon treatment with the superoxide dismutase mimetic EUK-8. Free Rad Biol Med 34:277–282
29. Gonzalez PK et al (1996) Role of oxidant stress in the adult respiratory distress syndrome: Evaluation of a novel antioxidant strategy in a porcine model of endotoxin-induced acute lung injury. Shock 6:S23–S26
30. Fucassi F et al (2001) Characterisation of small molecule binding to DNA using a quartz crystal resonant sensor. Chem Commun 841–842
31. Fucassi F et al (2007) α-Lipoic acid and glutathione protect against the prooxidant activity of SOD/catalase mimetic salen Mn(III) derivatives. J Inorg Biochem 101:225–232
32. Lancaster KM et al (2009) Type-zero copper proteins. Nat Chem 1:711–715 PDBID:3FPY
33. Hart PD et al (1996) A missing link in cupredoxins: crystal structure of cucumber stellacyanin at 1.6 Å resolution. Protein Sci 5:2175–2183 PDBID:1JER
34. Hazes B et al (1993) Crystal structure of deoxygenated Limulus polyphemus subunit II hemocyanin at 2.18 Å resolution: clues for a mechanism for allosteric regulation. Protein Sci 2:597–619 PDBID:1NOL
35. Sénèque O et al (2000) Calix[6]arenes and zinc: biomimetic receptors for neutral molecules. J Am Chem Soc 122:6183–6189
36. Cacciapaglia R et al (2007) Efficient and selective cleavage of RNA oligonucleotides by calix[4]arene-based synthetic metallonucleases. J Am Chem Soc 129:12512–12520
37. Collman JP et al (2007) A cytochrome c oxidase model catalyzes oxygen to water reduction under rate-limiting flux. Science 315:1565–1568

38. Qin L et al (2009) Redox dependent conformational changes in cytochrome c oxidase suggest a gating mechanism for proton uptake. Biochem 48:5121–5130 PDBID:3FYI
39. Pallares I et al (2008) Direct interaction between a human digestive protease and the mucoadhesive poly(acrylic acid). Acta Cryst Sect D 64:784–791 PDBID:2V77
40. Rondelez Y et al (2002) The first water-soluble copper(II) calix[6]arene complex presenting a hydrophobic ligand binding pocket: a remarkable model for active sites in metalloenzymes. Angew Chem Int Ed Engl 41:1044–1046
41. Rubach JK, Plapp BV (2002) Mobility of fluorobenzyl alcohols bound to liver alcohol dehydrogenases as determined by NMR and X-ray crystallographic studies. Biochem 41:15770–15779
42. Kimura E et al (1992) A model for catalytically active zinc(II) ion in liver alcohol dehydrogenase: a novel 'hydride transfer' reaction catalyzed by zinc (II)-macrocyclic polyamine complexes. J Am Chem Soc 114:10134–10137
43. Cronin L, Walton PH (2003) Synthesis and structure of [Zn(OMe)(L)]·[Zn(OH)(L)]·2($BPh_4$), L = *cis*,*cis*-1,3,5-tris[*E*,*E*)-3-(2-furyl)acrylideneamino]cyclohexane: structural models of carbonic anhydrase and liver alcohol dehydrogenase. Chem Commun 1572–1573
44. Koike T, Kimura E (1991) Roles of zinc(II) ion in phosphatases – a model study with zinc(II) macrocyclic polyamine complexes. J Am Chem Soc 113:8935–8941
45. Molenveld P et al (1999) Dinuclear and trinuclear Zn(II) calix[4]arene complexes as models for hydrolytic metallo-enzymes. Synthesis and catalytic activity in phosphate diester transesterification. J Org Chem 64:3896–3906
46. Laine M et al (2009) Base moiety selectivity in cleavage of short oligoribonucleotides by di- and tri-nuclear Zn(II) complexes of azacrown-derived ligands. Org Biomol Chem 7: 2780–2787
47. Hammarström L (2003) Towards artificial photosynthesis: ruthenium-manganese chemistry mimicking photosystem II reactions. Curr Opin Chem Biol 7:666–673
48. Börjesson K et al (2009) Membrane-anchored DNA assembly for energy and electron transfer. J Am Chem Soc 131:2831–2839
49. Chang SK et al (1991) Hydrogen-bonding and molecular recognition – synthetic, complexation, and structural studies on barbiturate binding to an artificial receptor. J Am Chem Soc 113:7640–7645
50. Tecilla P et al (1990) Hydrogen-bonding self-assembly of multichromophore structures. J Am Chem Soc 112:9408–9410
51. Ferreira KN et al (2004) Architecture of the photosynthetic oxygen-evolving center. Science 303:1831–1838
52. Youngblood WJ et al (2009) Photoassisted overall water splitting in a visible light-absorbing dye-sensitized photoelectrochemical cell. J Am Chem Soc 131:926–927
53. Kanan MW, Nocera DG (2008) In situ formation of an oxygen-evolving catalyst in neutral water containing phosphate and $Co^{2+}$. Science 321:1072–1075
54. Kim DH, Lee SS (2000) Origin of rate-acceleration in ester hydrolysis with metalloprotease mimics. Bioorg Med Chem 8:647–652
55. Hosseini MW et al (1987) Supramolecular catalysis in the hydrolysis of ATP facilitated by macrocyclic polyamines: mechanistic studies. J Am Chem Soc 109:537–544
56. Hosseini MW, Blacker AJ, Lehn JM (1988) Multiple molecular recognition and catalysis. Nucleotide binding and ATP hydrolysis by a receptor molecule bearing an anion binding site, an intercalator group, and a catalytic site. J Chem Soc Chem Commun 596–598
57. Liu J, Wulff G (2008) Functional mimicry of carboxypeptidase A by a combination of transition state stabilization and a defined orientation of catalytic moieties in molecularly imprinted polymers. J Am Chem Soc 130:8044–8054
58. Visnjevski A et al (2005) Catalysis of a Diels-Alder cycloaddition with differently fabricated molecularly imprinted polymers. Cat Commun 6:601–606

59. Lehn JM (1999) Dynamic combinatorial chemistry and virtual combinatorial libraries. Chem Eur J 5:2455–2463
60. Rowan SJ et al (2002) Dynamic covalent chemistry. Angew Chem Int Ed Engl 41:898–952
61. Jiang L et al (2008) De novo computational design of retro-aldol enzymes. Science 319: 1387–1391
62. Röthlisberger D et al (2008) Kemp elimination catalysts by computational enzyme design. Nature 453:190–195

# Chapter 5
# Natural and Synthetic Transmembrane Channels

## 5.1 Cells and Their Membranes

The history of life on Earth is bound up with the symbiotic relationship between replication processes and the microenvironment in which they can occur. It is likely that the combination of cell formation and its encapsulation of replicators occurred three to four billion years ago. Today the cell represents the fundamental biological unit containing genetic material together with a host of enzymes and proteins that allow it to function. As discussed in previous chapters, concepts central to supramolecular chemistry can be used to inform research in both replication mechanisms and capsule formation. The focus of this chapter is on an essential aspect of cell-based life; the transport of materials from an external source across a membrane so that they can fuel the chemical processes found there. Central to this is the ability to recognize targets and transport them to the exclusion of others. Unsurprisingly this has caught the attention of supramolecular chemists who have applied their knowledge of molecular recognition to synthesize molecules that can also act as selective transmembrane channels. Given the complexity of natural transmembrane channels, model compounds have been largely restricted to the simplest transporters, those for ions. To appreciate the biological inspiration behind the artificial compounds this chapter starts with an overview of natural transmembrane transport systems before describing how supramolecular chemists are attempting to mimic them.

### *5.1.1 Cell Membranes*

Current theories propose that the development of cell membranes accelerated early life. Enclosing chemicals responsible for replication, protein synthesis and energy generation within a confined space allows these essential life processes to occur far more frequently than they would do otherwise. Biological membranes comprise two phospholipid layers held together by interdigitation of their hydrophobic alkyl tails as shown in Fig. 5.1. They form a lamellar structure with their polar internal and external surfaces separated by a non-polar region. In their extended form each

P.J. Cragg, *Supramolecular Chemistry*, DOI 10.1007/978-90-481-2582-1_5,

© Springer Science+Business Media B.V. 2010

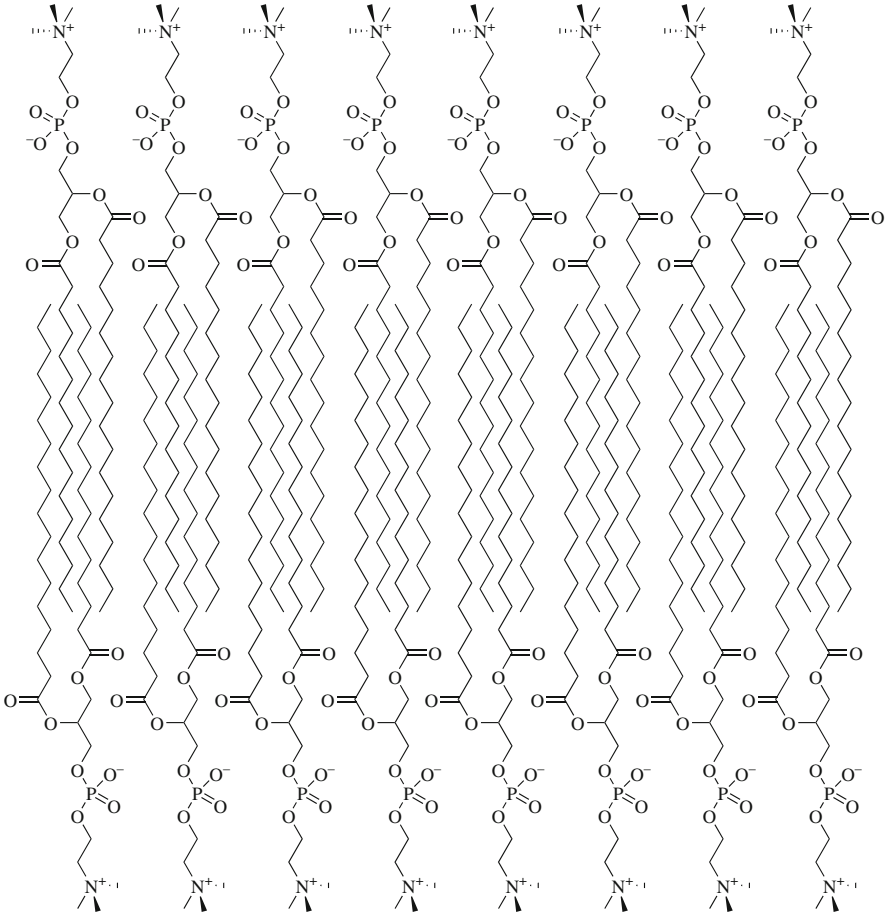


**Fig. 5.1** An idealized lipid bilayer

phospholipid is about 3 nm in length, however, when the hydrophobic tails of two lipid monolayers intertwine a bilayer 4–5 nm thick results. The membrane is also host to large biomolecules with particular functions that can bind to, or in some cases penetrate, its surfaces. The bilayer's 'fluid mosaic' structure allows it to reorganize when incorporating the macromolecules but remains intact unless seriously compromised [1].

Bilayers vary in thickness depending on external influences including the insertion of transmembrane proteins. Both extra- and intracellular environments are essentially aqueous but need to be kept separate. This is achieved by having polar head groups attached through mid-polar ester links to hydrophobic alkyl chains. Thus the internal and external surfaces of the bilayer are hydrophilic and polar enough to allow proteins to bind, the mid-polar region allows specific chemical groups to insert slightly deeper in the membrane, and the interlocking hydrophobic

region acts to insulate the internal and external surfaces from each other. Although it forms an effective barrier between the extra- and intracellular environments, the hydrophobic core can still be penetrated by a number of biological molecules, such as proteins, and synthetic compounds designed to insert through the polar exterior to the non-polar interior of the bilayer. The challenge is to use this phenomenon to transport specific ions through this apparently impermeable barrier. Active transmembrane transport occurs in two ways: migration of facilitated by species-specific complexing agents that shuttle through the bilayer, and channels formed by transmembrane proteins and similar molecules.

## *5.1.2 Transmembrane Migration: Molecular Shuttles*

Simple migration occurs when a carrier, or host, molecule traps a guest ion, ion pair or molecule, and transports it across the cell membrane. To function effectively the host must be selective for only one guest species. The host needs to be water soluble to bind the guest in solution but the complex has to enter the polar region of the bilayer then cross the non-polar core before the process is reversed for the guest to be released on the other side of the membrane. The host must have polarizable regions to bind a guest, or coordinate to water, and the ability to change conformation, inverting these regions to encapsulate the guest and present a lipophilic exterior, when passing through the hydrophobic region of the bilayer. There are many naturally occurring highly selective carriers and most are able to transport up to $10^4$ guests per second.

### 5.1.2.1 Ionophore Mediated Transport

Organic ligands that bind to alkali metal and alkaline earth cations are known as ionophores, literally 'ion carrying' compounds. Valinomycin, a cyclic natural antibiotic consisting of alternating D- and L-valine separated by D-hydroxyisovalerate and L-lactate groups shown in Fig. 5.2, binds $K^+$ one thousand times better than

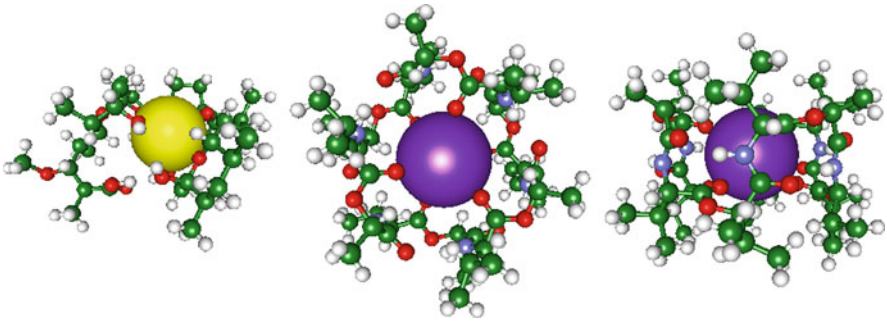
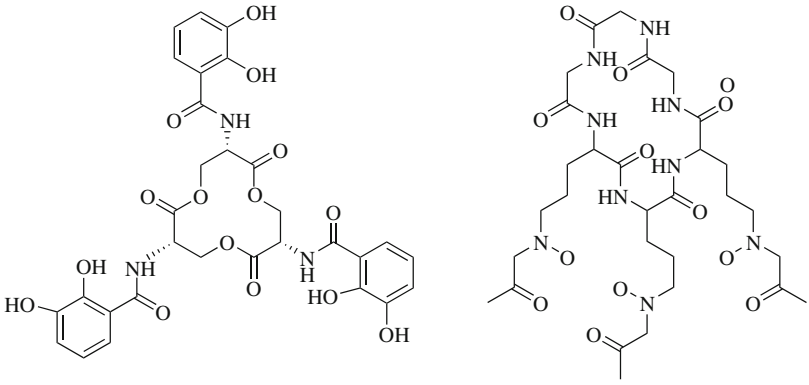


**Fig. 5.2** Ionophores: the $Na^+$-monensin complex [2] (*left*) and $K^+$-valinomycin complex [3] (*right*, *top* and side views)

$Na^+$ and exemplifies carrier facilitated transport. The free ionophore crystallizes as a rectangular box, roughly 1.7 nm long, 0.9 nm wide and 1.0 nm high, held in place by intramolecular interactions. The complex has a tubular shape, 1.0 nm high with an internal cavity of 0.4 nm, containing the metal ion, and an overall diameter of 1.5 nm. Six carbonyl oxygen atoms bind the metal in an octahedral environment forcing valine and isovalerate sidechains to form a hydrophobic exterior to the tubular complex. It is the inversion that occurs upon metal binding that turns the hydrophilic ligand into a hydrophobic, lipid-soluble complex.

### 5.1.2.2 Siderophore Mediated Transport

Yeast, fungi and bacteria, such as the *Salmonella* genus and *Escherichia coli*, need to sequester iron from aqueous solution. Unfortunately water soluble $Fe^{2+}$ is easily oxidized to $Fe^{3+}$ in air which makes it much less soluble. At concentrations of around $10^{-17}$ M for free $Fe^{3+}$ in aqueous solution it is essential that every available ion is bound effectively and transported across the cell membrane. This is achieved by siderophores, iron-specific chelating agents that bind $Fe^{3+}$ in an octahedral pocket usually comprising three bidentate catechol or hydroxamate groups, illustrated in Fig. 5.3. Binding is specific and very strong with binding constants ranging from $10^{20}$ to $10^{50}$. The resulting complex is targeted by membrane-bound receptors to return it to the bacterium. Once inside the protection of the cell localized conditions reduce $Fe^{3+}$ back to $Fe^{2+}$ which is then released as the siderophores' affinities for the reduced form are much lower than for the oxidized form.



enterobactin ferrichrome

**Fig. 5.3** Siderophores

### 5.1.2.3 Anion Transport

Anions can also be transported across membranes. In Nature, $Cl^-$ is transported by prodigiosins which comprise three conjugated pyrrole groups and a lipophilic alkyl

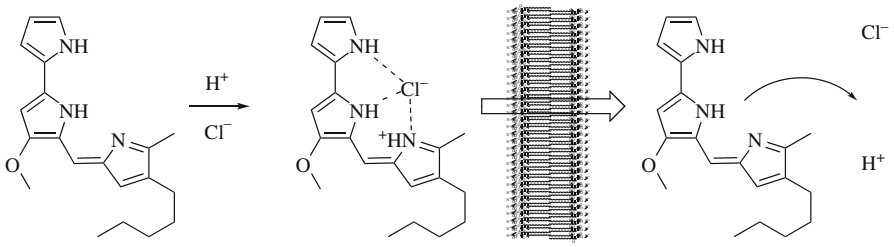



**Fig. 5.4** Proton ancd chloride co-transport by prodigiosin

tail as shown in Fig. 5.4 [4]. These simple compounds simultaneously transport $Cl^-$ and $H^+$ by what is termed a symport mechanism. As it is neutral it can traverse the phospholipid bilayer with much greater ease than a charged complex.

## 5.2 Transmembrane Channels: Selectivity and Gating Mechanisms

Rather than ferry ions or small molecules across a membrane one at a time it is much more economic to move them, with great selectivity and efficiency, through specially designed transmembrane channels. Channel formation occurs when a macromolecule, or group of macromolecules, pierce the cell membrane to open a pore through the lipid bilayer. The pore dimensions dictate the maximum size of guests that may enter but more subtle interactions determine how and when they are transported through the membrane. The pore is often able to discriminate in favour of one particular species and be able to switch the current, caused by the movement of the chosen species, on and off in response to an external stimulus. Transport is several orders of magnitude faster than is observed for a shuttle mechanism with up to $10^8$ ions passing through some channels each second. Ions can be 'pumped' across membranes against the concentration gradient, particularly by transporter proteins. This is an energy intensive process often powered by ATP hydrolysis.

Most natural channel forming proteins assemble from identical, or nearly identical, complex polypeptides that aggregate to form binding regions for selectivity. Natural channels are often gated, that is they open only upon a certain electrical, physical or chemical signal, to control the flow of chemical species through the central pore, and may also contain a constricted region, or filter, to bar all but one chemical species. Simple channels form when two membrane-spanning α-helices meet across the membrane as seen for the gramicidin group of polypeptides. Another type of channel architecture is seen when several molecules come together to form a barrel in which each molecule forms a stave, as is observed for the antibiotic amphotericin. These channels are often non-specific and allow many species to diffuse through the central pore, attracted to the polarized hydrophilic regions within the cavity.

### 5.2.1 Voltage Gating

Voltage gating occurs when the conformation of the transmembrane protein assembly alters upon a change in membrane potential. Restructuring brings about an open state in the protein which allows species, usually ions, to pass through. In the case of a $K^+$ channel, KcsA, depolarization of the membrane forces positively charged, arginine-rich voltage sensor regions of the protein through the bilayer [5]. This in turn opens the channel. After a channel has opened it enters a non-conducting rest stage until repolarization returns the channel to its original closed state. Evidence from the ClC $Cl^-$ channel indicates that conducting states may be induced by the voltage controlled unblocking of the channel by the movement of a charged amino acid side chain that initially resides within the channel [6].

### 5.2.2 Ligand Gating

Inactivation of transmembrane ion channels by small molecules is a well known phenomenon. For example the puffer fish poison, tetrodotoxin, blocks $Na^+$ channels but not those for $K^+$ [7], and sulfonatocalix[4]arene blocks $Cl^-$ channels [8]. It is also possible that metals or small molecules binding in a region remote from the channel to induce a conformational change. This type of ligand gated behavior, known as an allosteric effect, can be positive or negative. $Ca^{2+}$-activated $K^+$ channels, $K_{Ca}$, regulate neuronal function and are activated by increasing concentration of intracellular $Ca^{2+}$ ions which bind to regions of the protein that extend out from the cell membrane [9]. Similarly, the nicotinic acetylcholine receptor opens in response to acetylcholine which is proposed to bind to extramembrane subunits [10]. These rotate as a result and the movement is transmitted to the gating region through helical peptide domains that link the two.

### 5.2.3 Gating by Aggregation

There are many examples of peptides that cannot form transmembrane channels on their own but can do so through aggregation. The gramicidin antibiotics, produced by bacteria as part of their chemical defence system, are only 1.6 nm or so in length [11]. Specific placement of side chains, such as four tryptophan residues towards the C-terminus, ensures that the helix penetrates cell membranes to a particular depth but does not pass through the membrane.

To span a typical phospholipid bilayer gramicidin molecules dimerize through one of two mechanisms: end-to-end hydrogen bonding through N-terminus aldehydes or antiparallel double helix formation. These two binding modes are illustrated in Fig. 5.5. When two molecules interact through their N-termini a network of hydrogen bonds forms to give the dimer an overall length of approximately 2.8 nm. Conversely, if a double helix forms through intermolecular β-sheet interactions this

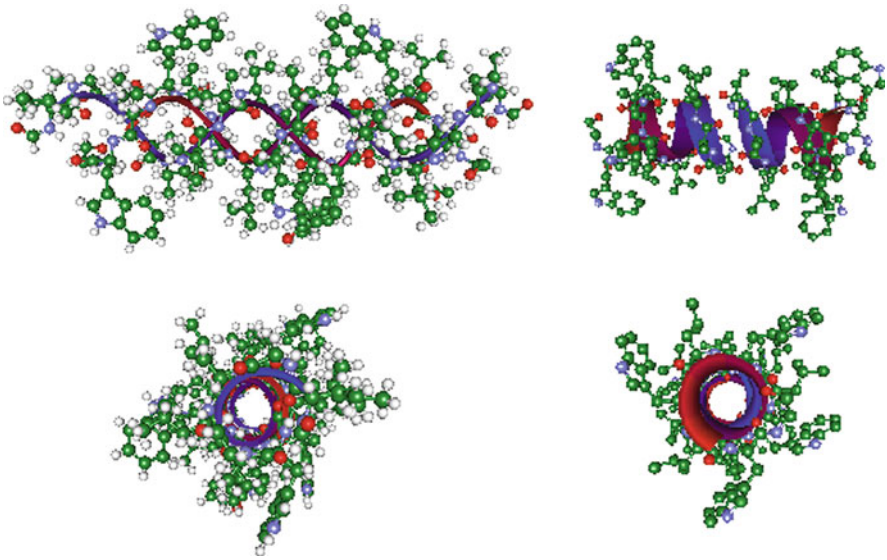


**Fig. 5.5** Gramicidin dimers: double helix form [12] (*left*) and end-to-end form [13] (*right*)

distance is up to 3.5 nm. Small molecules and cations can pass through either dimer: the channel diameter of the end-to-end form estimated to be between 0.11 and 0.22 nm and that of the double helix between 0.04 and 0.14 nm. This is in itself a potential gating mechanism based on simple size exclusion with one form unable to transport large cations. Simple cations (alkali metals, $Tl^+$ and $Ag^+$, $NH_4^+$ and $H_3O^+$) can pass through the electronegative channel along a spiral formed by carbonyls that run along the gramicidin backbone.

Many natural channels, particularly those for simple ions, form at the confluence of complex proteins. This mechanism has been simplified by the Gokel group which has prepared alkyl-terminated hexapeptides that contain a short proline containing peptide sequence, GGGPGGG, similar to that found in natural $Cl^-$-selective channels [14]. These compounds give $Cl^-$ selectivity when anchored in phospholipid vesicles. It is assumed that channels form by supramolecular aggregation of the hexapeptides around the proline motif, shown in Fig. 5.6. Such a simple transport mechanism has yet to be seen in Nature but it would not be too surprising if one were to be found.

### *5.2.4 Gating by pH and Membrane Tension*

Channels' internal walls and external mouths contain groups that are easily ionized, such as tyrosine, or protonated, such as tryptophan, and are likely to be affected by changes in pH. The $M_2$ transmembrane protein, encoded by influenza viruses, forms a tetrameric transmembrane channel for protons when it infects cells [15].

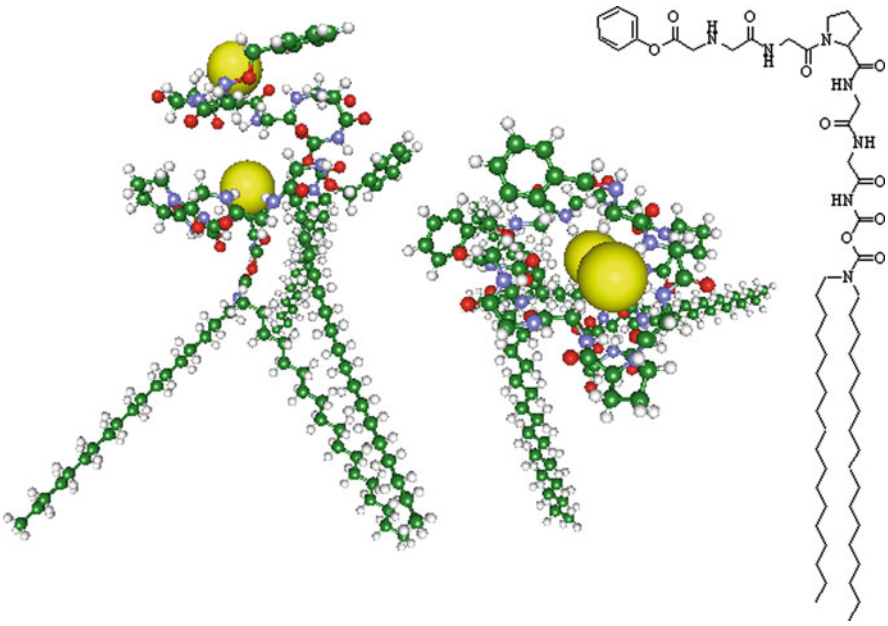


**Fig. 5.6** Chloride transport by peptide aggregation in an artificial system [14]

Tryptophan residues near the entrance to the channel block the central pore. The residues are mechanically responsive to changing concentrations of $H^+$: at about pH 5.5 they start to interact with increasingly protonated histidine residues further up the channel walls, opening up the pore, and at pH 4 the process is complete allowing protons to traverse the membrane. When tryptophan is replaced with the smaller phenylalanine the pore remains open regardless of pH [16]. By way of contrast, $K^+$ movement through some $K^+$ channels is inhibited at low pH by the protonation of lysine and histidine residues that lie outside the central pore [17]. These so called 'ball and chain' mechanisms result in the protonated residues physically blocking the channel thus stopping transmembrane ion transport.

Mechanosensitive channels respond to changes in membrane tension. A prokaryotic large-conductance mechanosensitive channel, MscL, opens in response to osmotic stress to form a water filled channel between 3 and 4 nm across [18]. The change in pressure on the bilayer imparts a small movement in a transmembrane helix that is then followed by a dramatic rearrangement of the transmembrane domain to a fully open state.

### *5.2.5 Light Gating*

Employing light as a gating mechanism may appear highly unusual, however, it is in constant use in the eyes where the light-induced *cis-trans* isomerization of retinal dictates how we see. Incoming photons have enough energy to turn 11-*cis*

retinal into the all *trans* isomer. Retinal is coupled to proteins known as opsins but is released upon photoisomerization. This initiates a sequence of reactions that close $Na^+$-channels in photoreceptor rod and cone cells but leave $K^+$-channels unaffected. Ultimately there is a reduction in the release of glutamate triggering the on/off visual response. The influence of light gating is indirect but recent research has identified algal opsins that respond directly to changes in light frequency.

*Chlamydomonas reinhardtii* has two types of rhodopsin in its eyespot that have been shown by Nagel and co-workers to form non-specific transmembrane channels for mono- and divalent cations with protons being the most permeable [19]. The proteins have been named channelrhodopsin-1 and 2 (ChR1 and ChR2) and are still poorly understood. Heberle and co-workers have proposed that ChR2 is normally non-conductive [20]. Following the absorption of a photon there is a change in the protein backbone as a consequence of the light induced *trans-cis* isomerization though the protein remains in the closed state. Further transitions lead to a conducting state which closes over a matter of seconds to a desensitised state. This in turn slowly returns to the original state. Since the original discovery of ChR1 and 2 by Nakanishi and co-workers the protein has been expressed in other species and cell types, including mammalian neurons, leading to the possibility that light activated ion channels are more prevalent in Nature than at first thought [21].

## 5.3 Channel Architecture

Channel-forming proteins exhibit a number of structural motifs: the influenza virus $M_2$ proton channel and voltage-gated channels for $K^+$, $Na^+$ and $Ca^{2+}$are all composed of four identical subunits that aggregate to form a central pore as shown in Fig. 5.7; an acid sensing $Na^+$ ion channel has a similar structure but with threefold symmetry; $Ca^{2+}$ release channels, the divalent metal ion transporter CorA, the

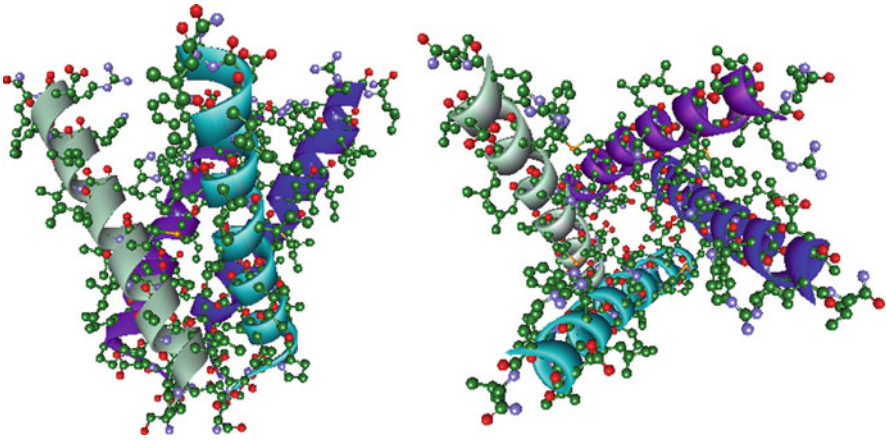


**Fig. 5.7** The $M_2$ influenza channel [22]

nicotinic acetylcholine receptor pore, and MscL have fivefold symmetry; aquaporins form a β-pore as the protein folds in an antiparallel β-sheet structure; alamethicin aggregates to form a barrel-stave channel; the bee sting protein mellitin forms an α-helix to penetrate membranes; the ClC $Cl^-$ channel comprises two interlocking subunits each with a central pore and, as discussed above, the gramicidins dimerize to form transient channels. Ion pairs are often ferried across membranes together as neutral complexes pass through the hydrophobic core easier than those that are charged. Transport of several species in the same direction across a membrane is known as a symport mechanism; when species are moved in opposite directions at the same time an antiport mechanism results.

### *5.3.1 Channels for Neutral Molecules*

Aquaporins, denoted by the acronym AQP, are found across species, from plants and bacteria to humans, where they facilitate the movement of water and some small molecules across lipid bilayers [23]. In AQP1, where the channels form from a bundle of tilted α-helices, selectivity occurs through a 0.28 nm constriction in the transmembrane region of the protein where water encounters an arginine residue. Transport is further assisted by an asparagine/proline/alanine (NPA) dipole half way through the channel. The protein subunits are attracted to each other through complementary hydrophilic residues to give the tetramer a lipophilic exterior that can penetrate the cell membrane. The presence of the arginine residue, in particular, prevents charged species such as the hydronium ion, $H_3O^+$, from passing through. Bacterial porins also allow siderophore complexes to pass through the outer membrane. These larger channel-forming proteins consist of three or four linked monomeric transmembrane pores formed by 16-stranded antiparallel β-barrels and can have cavities in the region of 0.9 nm. An example of aquaporin architecture can be seen in Fig. 5.8.

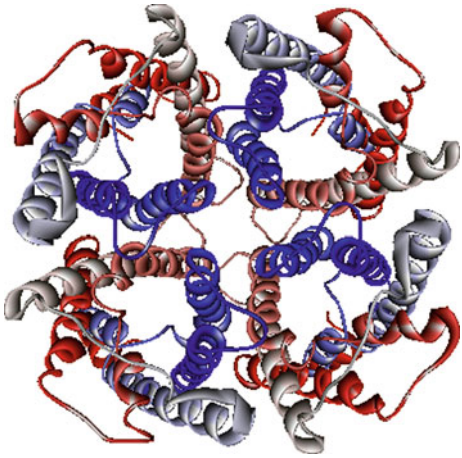


**Fig. 5.8** The architecture of an aquaporin channel [24]

### 5.3.2 Anion Channels

As with neutral and cation channels, anion channel-forming proteins are found throughout Nature [25]. In 2002 the first high resolution (0.3 nm) X-ray structures of anion channels were published by the MacKinnon group. Although they are from the ClC family these proteins conduct $Br^-$ as well as $Cl^-$ though the former is less common in a physiological environment. The ClC channels from *Escherichia coli* (EcClC) and *Salmonella typhimurium* (StClC), illustrated in Fig. 5.9, comprise 18 α-helices linked in an antiparallel manner [26].

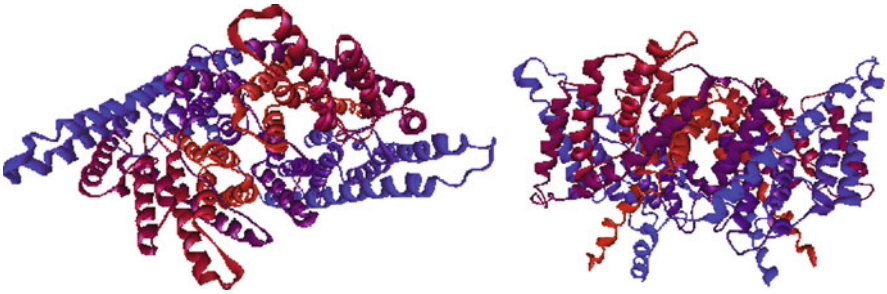


**Fig. 5.9** Bacterial chloride channel-forming proteins [26]

The helices interact with each other to leave a central pore and two of these channels align, also in an antiparallel fashion, to give a structure with two channels. Anions pass through a selectivity filter comprising isoleucine, phenylalanine, serine and tyrosine though, interestingly, there appear to be no strong interactions with positively charged residues: it is the hydrophobic regions that induce anions into the filter. A further feature of the channel is the presence of a conserved proline-containing sequence of amino acids at the N-termini of the α-helices that give the channel a twist. As noted above the proline-twist motif has been exploited in the synthesis of artificial anion channels.

### 5.3.3 Cation Channels

Cations come in many shapes and sizes. The simplest is the lone proton which may jump from base to base along a small channel. Then there are inorganic ions with no directional preferences for bonding, such as the alkali or alkaline metals, and $NH_4^+$ which is tetrahedral but appears spherical when hydrated. At the other end of the spectrum of structural complexity we have organic cations and hydrated transition metal complexes with non-uniform charge densities.

Four metal cations, $Na^+$, $K^+$, $Mg^{2+}$ and $Ca^{2+}$, are of primary importance. $Na^+$ and $K^+$ are ubiquitous in biological systems where they are employed to control membrane potential as well as intracellular levels of hydration. Several types of $Na^+$ and $K^+$ transporting channels are known to exist and some, notably the inwardly

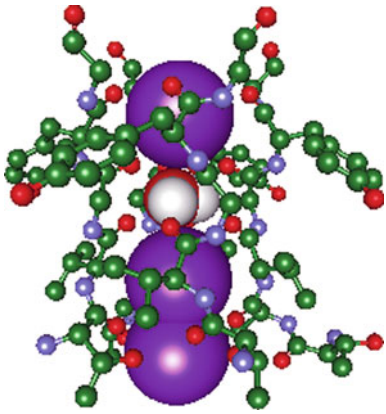


**Fig. 5.10** The $K^+$ selectivity filter region in the transmembrane region of the KcsA protein showing partially hydrated $K^+$ ions (*purple*) passing through the pore [5]

rectifying-$K^+$ channel KcsA isolated from *Streptomyces lividans*, the filter region of which is illustrated in Fig. 5.10, have been the subject of intense investigation by multiple techniques.

#### 5.3.3.1 $Na^+$ Channels and Transporters

Voltage-gated $Na^+$ channels usually remain in the closed state but open in response to minute voltage changes. The techniques of channel cloning and site-directed mutagenesis have allowed the $Na^+$ permeability of several mutant channels to be probed. This is done by analysing the responses of the channels to site-specific anaesthetic molecules and toxins. Tetrodotoxin, isolated from puffer fish, binds to particular domains found near the outer mouths of $Na^+$ channels and causes channel failure. The response to tetrodotoxin can therefore be used to determine the presence of specific protein sequences in the channels. Certain scorpion toxins can also be used for this purpose. Voltage-gated channels, activated by changes in membrane potential leading to conformational changes, allow the channel mouth to open to accommodate appropriately sized ions but, once they enter the channel protein, these ions face a selectivity filter.

At present very few filter regions have been analysed at the level of detail necessary to determine the exact selectivity mechanism. A voltage-gated $Na^+$ channel isolated from the electric eel, *Electrophorus electricus*, has been determined by electron microscopy and shows a square-based, bell-shaped molecule [27]. There is a large central cavity 4.0 nm high and 3.5 nm wide with four chambers (1.5 nm high and wide) which are connected to the central cavity by four small openings in the extracellular region of the channel protein. Each chamber appears to connect to a helical region which presumably contains a selectivity filter but the structure of the filter has yet to be imaged at atomic level.

In 2007 the high resolution structure of an acid sensitive ion channel was obtained in the 'closed' position at low pH [28]. As shown in Fig. 5.11 the molecule is composed from three identical interlocking proteins and has a transmembrane region

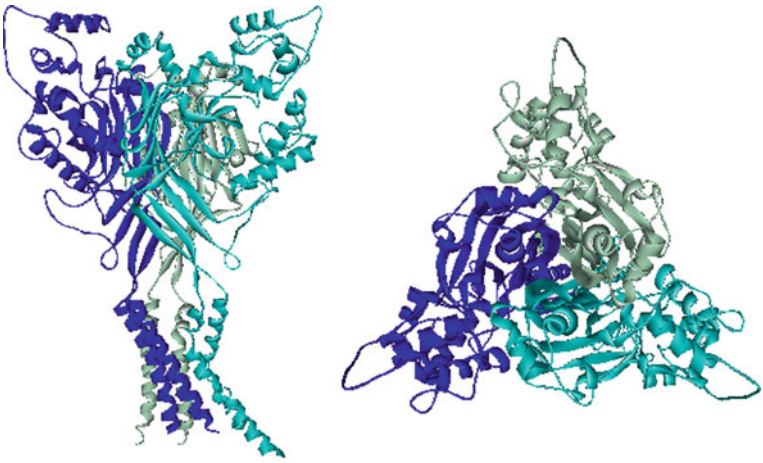


**Fig. 5.11** An acid sensitive transmembrane ion channel-forming protein [28]

formed by six α-helices, three of which are believed to open at high pH and allow $Na^+$ to pass through. The channel's threefold symmetry implies that sodium ions coordinate to six carbonyl oxygen atoms as they traverse the channel. This is consistent with sodium's affinity for oxygen and preferred octahedral geometry.

The structure of a protein that transports $Na^+$ and $Li^+$, but not $K^+$, from the cytoplasm to the periplasm of *Escherichia coli* by an antiport mechanism reveals that one or two protons are required to assist the process [29]. *E. coli* is a Gram-negative bacterium and so has an internal and external cell membrane with the region between the two known as the periplasmic space. The antiport mechanism, in which two different chemical species move in opposing directions across a cell membrane, was elucidated using a combination of crystallography and molecular dynamics computational simulations. Models indicate that a single cytoplasmic $Na^+$ ion approaches an aspartate residue through a narrow channel and displaces the proton which returns to the cell's cytoplasm. A second, periplasmic, proton travels down a channel, binds to a neighbouring aspartate, and triggers a conformational change that moves the bound $Na^+$ closer to the periplasmic space. The periplasmic proton displaces $Na^+$, which finally enters the periplasm. The overall mechanism moves two protons into the cytoplasm as one $Na^+$ enters the periplasm. This is consistent with observations that the channel-forming protein's activity is much lower below pH 8 than above it or when the bacterium is exposed to high $Na^+$ or $Li^+$ concentrations.

#### 5.3.3.2 $K^+$ Channels

There are two main groups of transmembrane $K^+$ channels: the first includes proteins with six transmembrane domains such as the voltage-gated-$K^+$ ($K_v$) channels and the $Ca^{2+}$ activated-$K^+$ ($K_{Ca}$) channel; the second includes the inwardly

rectifying-$K^+$ ($K_{ir}$) channels with two transmembrane domains. $K_v$ channels are hetero- or homotetramers according to electron microscopy and mass determinations implying that ions are channelled through a pore formed at the confluence of the proteins. Scorpion venom inactivates the outer mouths of the $K_v$ channels by binding in an analogous manner to tetrodotoxin in voltage-gated $Na^+$ channels [30]. The selectivity filter appears to be about 0.5 nm from the extracellular face of the protein and mutagenesis experiments have shown that removing a tyrosine-glycine dipeptide sequence allows other alkali metals to pass at rates similar to $K^+$. $K_{Ca}$ channels require a particular $Ca^{2+}$concentration to initiate $K^+$ conductance though little is known about their structures [31].

Much more is known about the structure of $K_{ir}$ channels following the outstanding crystallographic work of MacKinnon's group to determine the structure of KcsA [5], which led to him sharing the 2003 Nobel Prize in chemistry. A theoretical model was devised by Bernèche and Roux to analyse $K^+$ movement through the KcsA selectivity filter in greater detail [32]. KcsA also has much in common with $K_v$ and $K_{Ca}$ channels so aspects of the selectivity mechanism used by KcsA may well be found in other ion channels.

The transmembrane region of KcsA is formed from four subunits with extracelluar turrets and arginine-rich paddles that respond mechanically to changes in membrane polarization. Each subunit has an inner and outer α-helix that interlock, constricting the channel as it enters the intercellular region. One $K^+$-binding site is found in the internal cavity of the channel protein with four further sites inside the selectivity filter, and two other binding sites on the extracellular side of the protein. As a result $K^+$ ions are forced through a funnel and dehydrated as they pass through. A combination of computational simulations and crystal structure determinations at different concentrations has shown that not all sites are occupied simultaneously but that $K^+$ ions are linked by water molecules in an alternating pattern through the filter. As each $K^+$ ion approaches the internal opening to the filter it is bound to four water molecules from the hydrated internal cavity and to four carbonyl groups at the mouth of the filter thus conserving the preferred cubic geometry of the $K^+$ primary coordination sphere. The resulting low energy pathway for transmembrane transport explains the high efficiency of the channel.

The selectivity filter comprises four carbonyl groups from each of the four constituent protein monomers which are oriented into the central pore. At high concentrations of $K^+$ each cation interacts with four carbonyl groups above and four below to give the cubic binding domain. Additional water molecules may bridge between successive cations. The resulting binding pocket corresponds to the preferred disposition of oxygen-containing ligands around $K^+$ in many of its non-biological complexes. $Na^+$, by way of contrast, is more usually found in a six-coordinate, octahedral binding site and is not selected by KcsA. As proof of this, MacKinnon's group solved the crystal structure of a variant of KcsA in which one glycine residue in the selectivity filter had been replaced by D-alanine [33]. The resulting binding site was now more amenable to binding $Na^+$ and conductance measurements confirmed that the channel allowed $Na^+$ ions to traverse a phospholipid

bilayer as long as $K^+$, which competes for the same binding site more efficiently, was absent.

The skeletal muscle ryanodine receptor, RyR1, controls the release of calcium, through a central channel, from the intracellular sarcoplasmic reticulum [34]. The channel forms at the convergence of four identical protein subunits each of which contains two α-helices. One helix from each subunit is approximately 0.45 nm long and has a central kink. The overall effect is to give the channel a funnel-like entrance about 0.3 nm across leading to a central pore with a 0.15 nm diameter. The pore is defined by four further α-helices approximately 0.22 nm long, one from each subunit. Unfortunately the low resolution of the structure does not allow for detailed study of the filtering mechanism.

#### 5.3.3.3 Cation Selectivity

Sodium and potassium cations are often encountered in the same biological environment and the transmembrane movements of both are required as part of an enzymatic pathway as in $Na^+$, $K^+$-ATPase. Under these circumstances it is essential that cation-specific channels are formed. What features of the channels contribute to the selectivity? Earlier the preferred geometries of $Na^+$and $K^+$, sixfold octahedral and eightfold cubic respectively, were proposed as the main discriminatory factors. A computational analysis by Dudev and Lim [35] has considered the effect of coordinated water, number of available coordination sites in the channel walls, and the dipoles of the coordinating groups. The researchers investigated cation complexes with valinomycin and the protein KcsA, both $K^+$-selective, and compared these with a non-selective NaK channel.

Although the coordination geometry around the cation is a major factor, with its preferred coordination number as the hydrate being reflected in the environment available within the pore of the protein, the solvent accessibility within the pore and its flexibility were also influential. Thus when $K^+(H_2O)_8$ enters KcsA it is stripped of its almost all of its solvent and the coordination sites replaced by eight oxygen atoms from peptide carbonyl groups. The same environment is not attractive to $Na^+$ because it prefers six coordination sites and the distances to those carbonyl groups are longer than the ideal value for $Na^+$ binding. The KcsA channel is also too rigid to respond to the preferred $Na^+$ environment. By way of contrast the NaK channel is much larger and more flexible allowing both cations to pass through in a reasonably close approximation of their preferred geometries.

## 5.4 Structural Determination

Primary sequences of proteins known, or likely, to incorporate membrane-spanning regions are a useful place to start when investigating the structural aspects of transmembrane ion channels. Once a sequence has been identified it is possible to generate a secondary structure by matching regions of the protein with known

sequences that have been determined previously by crystallography. This widely used technique is known as homology mapping. Computational models can also be invoked to give a 'best guess' secondary structure. An analysis of the hydrophilic and hydrophobic character of the protein, its hydropathy index, will indicate which regions are likely to be membrane bound and which would be expected to be in inter- or extracellular positions. The combined information will give a crude picture of the tertiary structure but, crucially, it will allow the transmembrane regions to be identified. It is these regions that need to be the focus of further study to determine if they contain protein sequences with the potential to form channels. Often the regions are similar to others known to form channels though on occasion they may reveal a novel channel-forming motif.

It is unnecessary to try to determine the structure of the entire protein if only a small section is of interest. Where a particular region has been identified as containing a channel-forming structure it is possible to excise that specific part of the protein thus simplifying the crystallization process somewhat. Despite the intense interest surrounding natural ion channels' methods of activation and transport mechanisms little is known about them at the atomic level. Structures of some natural channels have been determined by protein crystallography at varying resolutions. The sheer effort required to grow single crystals suitable for crystallographic study makes the publication of every new structure a major event. At high resolution (below 0.2 nm) the relative positions of individual atoms become unambiguous. Unfortunately most structures cannot be resolved to this level so some uncertainty will always exist regarding the exact orientations of peptide side chains. Computer simulations can be used to refine the picture, using X-ray data as a starting point to generate optimized geometries for protein sequences, but both crystallographic and computational models suffer from shortcomings.

Crystallography offers a snapshot of the protein structure but gives no information about dynamic activity such as protein unfolding and refolding in response to external factors. An inevitable consequence of crystallization is that the level of hydration is reduced. As a consequence, experimental artifacts are introduced into the resultant structure, in particular a greater degree of hydrogen bonding and unnaturally strong ion binding, leading to misleading interpretations. Heavy atoms, like $Hg^{2+}$, are routinely introduced as part of the crystallization protocol to identify particular residue positions but this too may lead to unnatural structures. Computational models suffer because of the sheer size of the proteins. Prediction of polypeptide geometries is accurate for small sequences but large numbers of residues can only be modelled at present by simple molecular mechanics methods to give structural information. Even this requires considerable computational resources and time as simulations need to consider the effects of the phospholipid bilayer, water and ions that would be expected to be present. Once high resolution X-ray data provide initial atomic coordinates then simulations can be conducted at levels of theory which enable accurate dynamic behaviour to be investigated. Given the complexity of the model this remains computationally expensive: Roux's simulation of the KcsA channel contained over 40,000 atoms and required supercomputers to carry out the calculations.

Advances in nuclear magnetic resonance (NMR) techniques mean that it is now possible to undertake structural studies of proteins, including transmembrane channel-forming proteins, in solution. This avoids the complex processes required to grow good quality crystals that diffract well enough to submit to structural solution, and the large element of luck that is associated with this endeavour. Artificial systems usually employ molecules that are smaller than their natural analogues and so are more amenable to study by computational or single crystal X-ray diffraction experiments.

## 5.5 Measuring Channel Activity

Channel activity is best studied electrochemically as charged species cross a cell membrane or artificial lipid bilayer. There is a difference in electrical potential between the interior and exterior of a cell leading to the membrane itself having a 'resting' potential between –50 and –100 mV. This can be determined by placing a microelectrode inside the cell and measuring the potential difference between it and a reference electrode placed in the extracellular solution. Subsequent changes in electrical current or capacitance are indicative of a transmembrane flux of ions.

### *5.5.1 Voltage Clamping*

The voltage clamp technique uses this principle to keep a constant transmembrane potential. If the membrane is intact and contains no channels there will be no detectable current. The membrane's reduced resistance at the 'clamped' potential indicates that ion transport is occurring through opening of endogenous channels or activity due to ion channel mimics. Unfortunately there are some drawbacks to this method. The main problem is that the experiment measures the whole cell current, that is, the changes in membrane resistance due to the movement of affected ions. The method cannot discriminate between different ion channels or ions and therefore gives an overall current. It is possible to use compounds known to block particular channels to give some measure of discrimination but in practice this is quite unsatisfactory as the same compounds will have unknown effects on other channels. To investigate single channel events the patch clamp method is more useful.

### *5.5.2 Patch Clamping*

Patch clamping requires that an electrode, housed within a micropipette, is attached to the cell to make an almost perfect seal with the cell membrane. This generates a very high resistance between the cell and the pipette wall, typically 10 GΩ. The resulting transmembrane currents, measured between microelectrodes inside and outside the cell, generate extremely low noise so single channel events can be

detected. As with voltage clamping, the holding potential can be changed to investigate voltage sensitive effects. In addition to the 'cell attached' experiment, the ion flux can be measured through an isolated 'patch' of cell membrane, and whole cell currents can also be determined by bursting the membrane with the pipette. The currents recorded are due to endogenous single channel activity which can complicate matters if the effects of artificial channels, added to the extracellular solution, are being investigated.

### *5.5.3 Bilayer Methods*

The planar bilayer technique is by far the most useful by which to test artificial channel mimics and is also used to investigate the activity of natural channel forming proteins. A purified lipid is painted across a small hole in a dividing wall between two compartments and the compound to be studied incorporated into the bilayer. The compartments are filled with an aqueous solution that includes the ion of interest and the transmembrane resistance measured across the bilayer. As with other methods the bilayer may also be voltage clamped at set potentials. By systematically changing the potential, from −100 mV to +100 mV for example, it is possible to find out if the channel conducts in both directions or only in one. A similar experiment may be performed in which the ions are changed, such as substituting $Na^+$ for the larger $K^+$ or the more highly charged $Ca^{2+}$, to determine specificity. The influence of counterions can be determined by preparing solutions from different salts. NaCl, $NaNO_3$, $Na_2CO_3$ and $NaOCH_3$ could be used to find out if the transport of $Na^+$ was affected by $Cl^-$, $NO_3^-$, $CO_3^{2-}$ or the organic methoxide anion, $OCH_3^-$.

### *5.5.4 Dye Release Methods*

A technique that employs liposomes (spherical water-filled structures formed by phospholipid bilayers in aqueous media) filled with the fluorescent carboxyfluorescein dye has been used successfully to investigate $Cl^-$ transport. At high concentrations, as are found within the liposome, the dye self-quenches so no fluorescence is observed. When the liposome is treated with a channel-forming compound, particularly one designed to transport hydrated anions, the dye is released into the extraliposomal solution and fluoresces. The increase in fluorescence intensity can be visualized to give an indication of channel or carrier activity. Although the dye is relatively large the method has been used successfully to investigate the behavior of artificial anion channel-forming compounds.

### *5.5.5 NMR Methods*

$^{23}Na$ NMR can be used to follow transmembrane $Na^+$ transport. $Na^+$ is added to a suspension of liposomes incorporating $Na^+$-transporting compounds, or $Na^+$-specific channels, which are subjected to NMR spectroscopy. It has been shown

that $Na^+$ held within a cell or liposome has a different chemical shift to that in the bulk solution. Thus the change in intensity between the 'internal' and 'external' peaks can be used over time to monitor the transport of $Na^+$. In an extension of this technique, so called 'magic angle spinning' $^{23}Na$ NMR spectroscopy has been used to determine the presence of $Na^+$ in three different environments within a G-quadruplex structure composed of guanosine derivatives [36]. The method spins a crystalline sample at 54.7° (the 'magic angle') to an applied magnetic field which simplifies the analysis of the resulting spectrum.

## 5.6 Transmembrane Transport by Artificial Systems

Ions and small molecules may be transported across cell membranes or lipid bilayers by artificial methods that employ either a carrier or channel mechanism. The former mechanism is worthy of brief investigation as it has several ramifications in the design of selectivity filters in artificial transmembrane channels. To date there are few examples where transmembrane studies have been carried out on artificial transporters. The channel mechanism is much more amenable to analysis by traditional biological techniques, such as planar bilayer and patch clamp methods, so perhaps it is not surprising that more work has been done to model transmembrane channels.

### *5.6.1 Transporters*

Aside from opening a pore, the easiest method to transport material from one side of a phospholipid bilayer to the other is to design small molecules that mimic naturally occurring siderophores or ionophores.

#### 5.6.1.1 Synthetic Ionophores

Many artificial complexants, in particular the crown ethers, are proposed to work by ionophore-facilitated ion transport. Crown ethers, such as those in Fig. 5.12, have

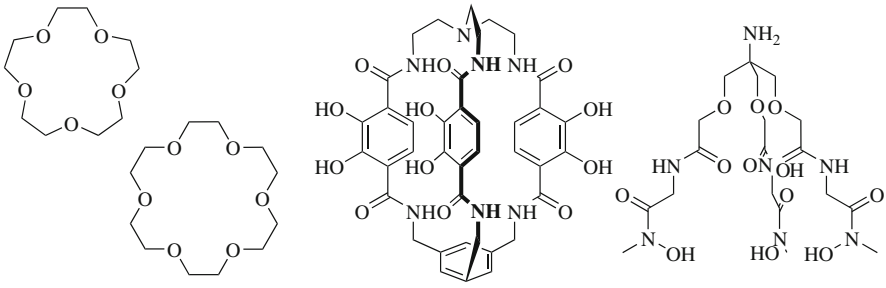



**Fig. 5.12** Artificial ionophores and siderophores: (*left* to *right*) $Na^+$-selective [15]crown-5, $K^+$-selective [18]crown-6, $Fe^{3+}$-selective cryptand and podand

a high affinity for alkali metals and can transport cations from aqueous solution to non-polar solvents [37]. This particular attribute has led to their use in phase transfer catalysis. Each member of the crown family has a preference for a particular alkali metal so they have also been incorporated into artificial transmembrane channels where they act as selectivity filters. Many channel models incorporate [18]crown-6 derivatives which bind $K^+$ (0.133 nm ionic radius) preferentially over the smaller $Na^+$ (0.097 nm). The combination of the flexible crown molecules and alkali metals is ideal as the latter, unlike transition metals, have no directional preference when it comes to bonding. The crown ethers can therefore wrap them up, with the electronegative oxygen atoms attracted to the metal cation, while simultaneously exposing a lipophilic exterior to allow the complex passage across a hydrophobic bilayer core.

#### 5.6.1.2 Artificial Siderophores

As noted above, siderophores are used by bacteria as a method of extracting iron from their surroundings. Enterobactin, a cyclotripeptide comprising three L-serines with pendent catechol groups, is produced naturally by Gram-negative bacteria and binds $Fe^{3+}$ with an association constant of $10^{52}$ M. This siderophore has been synthesized in the naturally occurring Δ-*cis* form and the unnatural Λ-*cis* form, enantioenterobactin. Both have been shown to be active iron transporters [38]. Examples of artificial analogues are also shown in Fig. 5.12.

#### 5.6.1.3 Anion Transport

Artificial analogues of the chloride transporter prodigiosin are effective symport HCl carriers as exemplified in the model systems developed by the groups of Gale [39] and Davis [40]. Biological inspiration is also behind another $Cl^-$ carrier. Cholic acid is a naturally occurring bile acid that functions as a surfactant in the intestine. Derivatives with three binding sites known as cholapods are able to transport isolated anions across lipid bilayers [41].

### *5.6.2 Channel-Forming Systems*

Successful synthetic transmembrane channels must have three characteristics if they are to replicate the behaviour of natural systems. They must span the cell membrane, implying a single molecule or stable self-assembled complex over 4 nm in length. Ideally they should also be able to discriminate in favour of one chemical species, if they are to mimic the highly selective channels, and transport that species at rates in the region of $10^4$ to $10^8$ ions per second to match the efficacy of natural channels.

Several channel architectures have been considered when designing artificial mimics. The most obvious biologically inspired method is to prepare extended helical molecules in an attempt to reproduce the channels either within the helix or where the molecules meet. A small number of systems have been designed using this

principle, primarily to generate membrane-spanning molecules with the potential for aggregation to form channels, in the hope that ion specificity follows. A similar concept, the 'barrel stave' model, proposes that several non-peptidic membrane-spanning molecules, such as amphotericin, aggregate to form a pore has inspired several designs based on natural analogues. A highly unnatural design has also been proposed based on the barrel stave. The active molecules are linear polymers, long enough to span the membrane, that have ion-selective groups attached that form a continuous channel through which the ions pass. An alternative to the insertion of staves is to design macrocyclic hoops that stack on top of each other until they reach the 5 nm or so required to form a pore.

Most of these methods are closely related to biological systems, however, supramolecular research into size-based selectivity has been used to design artificial ion channel mimetic compounds through careful consideration of their fundamental requirements. Macrocycles, known to be selective for particular species, are used to form selectivity filters and amphiphilic membrane-spanning molecules are attached to guide the ions across the membrane. This design-based approach has much to recommend itself. It can combine rigid ion-filtering molecules with lengthy substituents that allow the entire molecule to insert in a bilayer. Furthermore, the substituent can be tuned to give optimum solubility in both the aqueous phase, for ease of delivery, and the lipid phase, to ensure it remains bound within the membrane. Natural non-peptidic channel-forming compounds often contain amide or ester links that are highly susceptible to cleavage by a range of enzymes. The chemical links between the membrane-spanning and size selective components in the unnatural systems are not restricted to these groups so the resulting compounds are likely to be more stable in vivo and less susceptible to enzymatic degradation.

#### 5.6.2.1 Artificial Channel-Forming Peptides

In 1988 the DeGrado group reported a number membrane active oligopeptides prepared from serine and leucine [42]. Heptapeptides of varying sequences were prepared and linked to form dimers and trimers. The 21-residue sequences extended to between 3 and 4 nm so were capable of spanning a phospholipid membrane as shown in Fig. 5.13. The combination of serine and leucine gave the helicates a polar yet hydrophobic exterior. The internal cavity formed by the α-helix was too small to act as a channel so it is assumed that the helices aggregate due in part to their hydrophobicity. This mechanism is supported by computer simulations which showed that the sequence $H_2N$-LSSLSSLLSSLSSLLSSLSSL-$CONH_2$ could form interlocking hexamers with an internal channel diameter of approximately 0.8 nm. This is large enough to accommodate a string of water molecules and the peptide was shown in lipid bilayer experiments to be permeable to a range of alkali metals but without any significant selectivity. A second peptide, in which leucine substituted for the second serine, formed trimeric or tetrameric aggregates but the central cavity could transport $H^+$ only as the internal diameter was in the region of 0.1 nm. Peptides composed of 14 residues were unable to form stable channels at

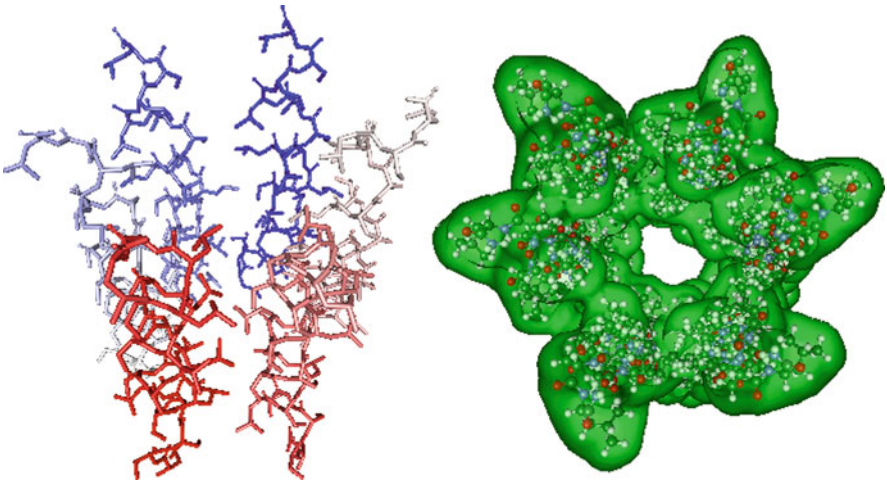


**Fig. 5.13** Hexameric transmembrane channels formed by synthetic peptides [42]

all. Presumably they aggregated in the same manner as the longer homologues but were unable to span a bilayer fully.

Using the knowledge that a helical sequence of 21 amino acids can generate a structure that induces transmembrane conductance Voyer's group prepared a crown ether-peptide hybrid containing 21 residues [43]. The sequence comprised alanine and a synthetic phenylalanine derivative of the ionophore [21]crown-7. The artificial peptide formed an α-helix with a series of crown ethers emanating from the external surface. Their insertion as every third or fourth residue resulted in a stack of ionophores running down one side of the peptide backbone forming a channel for alkali metal cations, as can be seen in Fig. 5.14. Planar lipid bilayer experiments

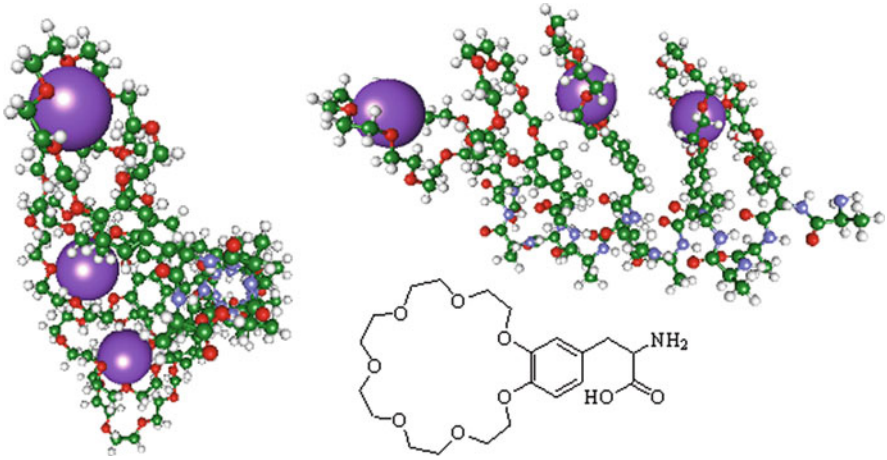



**Fig. 5.14** Helical proteins incorporating a crown ether amino acid analogue: a view down the α-helix (*left*) and along the channel (*top*) [43]

indicated $Na^+$ conductance which could be reduced by guanidinium or $Cs^+$. As both cations are known to fit well into the [21]crown-7 cavity the experiments support the interpretation that, unlike DeGrado's aggregating compounds, cations travel through a channel formed by a single molecule.

### 5.6.2.2 Crown Ether Derivatives

The first attempts to prepare artificial ion channels came as a result of work on [18]crown-6 derivatives, examples of which are shown in Fig. 5.15. The Lehn group's crystal structure of the potassium salt of their tetracarboamido[18]crown-6 showed that the macrocycles formed an extended one-dimensional stack with

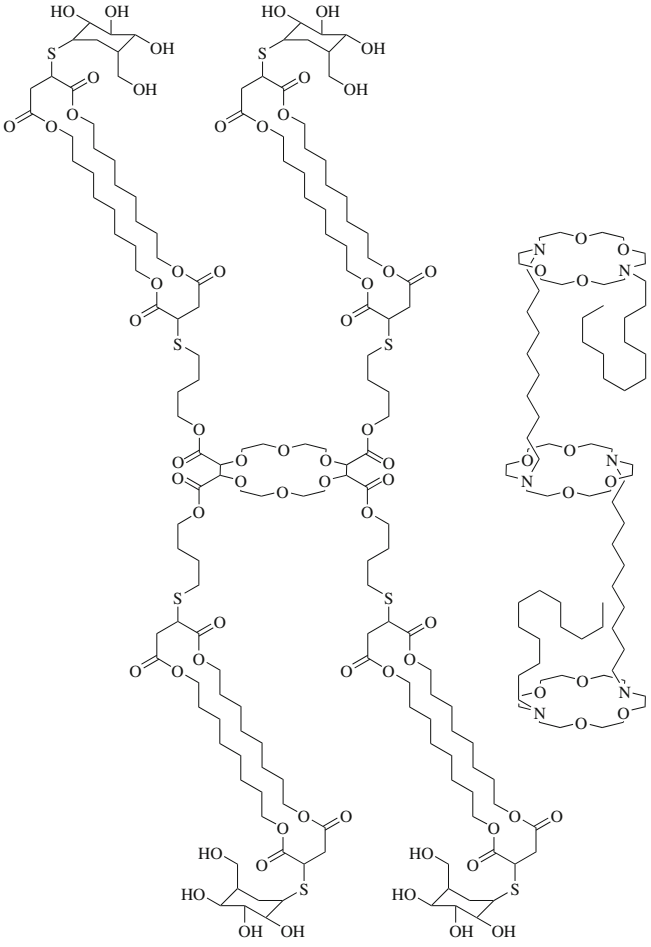


**Fig. 5.15** Transmembrane channels formed by crown ethers: a single filter chundle approach [45] (*left*) and multifilter hydraphile [46] (*right*)

hydrated cations in channels formed by the crowns [44]. The crown ether substituents were not long enough to span a lipid bilayer but other derivatives formed from polyether-polyamide conjugates, or 'chundles' as they were named, were considerably longer. An alternative, modular, approach was also explored by the Fyles group in which [18]crown-6 di-, tetra- or hexaesters were prepared [45]. The crown acted as the selectivity filter and the cyclic tetraesters, extending from alternating faces of the crown cavity, provided the membrane-spanning arms. The arms ended in hydrophilic acid, alcohol or glucose groups which would be pulled into the polar lipid bilayer exterior. Several of the compounds had transmembrane channel properties: all had low transport rates for $Li^+$, favouring $Na^+$ or $K^+$. It was assumed that the channel mimics inserted through the membrane with the crown ether in the hydrophobic central region. Here the ether oxygens would point inwards to form an ideal binding pocket for cations. The large hydrophobic arms would collapse into impermeable cables within the non-polar region of the membrane but would open temporarily to allow hydrated cations to pass through to the central crown ether filter and exit through the opposite end of the tubular molecule. The overall effect has been memorably likened to 'a snake swallowing an egg'.

Using a simpler approach, the Gokel group pioneered the development of 'hydraphiles' which use crown ethers as relays within membrane spanning structures [46]. The central diaza[18]crown-6 has alkyl substituents attached to each nitrogen. Each alkyl chain is linked to another diazacrown that has a substituent designed to reach into the polar region of a membrane. The entire molecule is about 3 nm in length though how extended it is remains unknown. Derivatives of [18]crown-6 are usually selective for $K^+$ but the hydraphile compounds exhibit $Na^+$ conductance. It transpires that both cations can be bound by [18]crown-6 in aqueous solution. The selectivity (based on relative binding constants) for $K^+$ is eighteen times that for $Na^+$ but this is based on two phenomena: the rate of binding and the rate of release. The rate of release is ten times greater for $Na^+$ than $K^+$ so the compounds transport $Na^+$ preferentially because $K^+$ is released more slowly.

Hall and Kirkovits demonstrated voltage-dependent channel opening using a hydraphile that incorporated a ferrocene group linked to two crown ether relays [47]. The electrochemically active ferrocene redox centre could influence the rate of transport through a conformational change. This resulted from a change in oxidation state and concomitant lipophilicity that was imparted to the molecule's termini. Studies using patch clamp techniques indicated channel opening at −60 mV in the presence of the hydraphile, no activity at 0 mV and further activity at +60 mV, consistent with a symmetric channel.

#### 5.6.2.3 Rigid Cavity Approaches

Crown ethers are attractive platforms from which membrane-spanning substituents can be appended but suffer from a lack of rigidity. It can be argued that cyclic compounds with well-defined central cavities should be more robust and selective. The cyclodextrins and calixarenes have both been employed in ion channel mimetic compounds. These compounds can have different functional groups at their

upper and lower rims and are therefore often encountered as 'half channel' models. Unlike crown ethers, which usually reside in the lipophilic region of a membrane, rigid macrocycles probably lie at the polar surface with hydrophobic substituents descending into the lipophilic region.

A cyclodextrin derivative developed by the Tabushi group exemplifies this strategy [48]. Cyclodextrins are formed biosynthetically from starch containing plants and sugar by *Bacillus macerans*. They are cyclic compounds, as their name implies, usually containing six (α), seven (β) or eight (γ) D-glucopyranoside units linked in the 1,4-positions. These three members of the family have internal diameters of 0.57, 0.78 and 0.95 nm, respectively. The Tabushi group's compound was functionalized with 14 hydroxyl groups on one rim, allowing the molecule to sit in the polar region of the bilayer, and four thiols on the other rim that were linked to amide containing alkyl chains as shown in Fig. 5.16.

Calixarenes, in particular calix[4]arene, have been seen as potential ion selective filters around which ionophore or channel frameworks can be constructed. Calix[4]arenes exist in different conformers, two of which are of interest as platforms for transmembrane ion transport: the *cone* conformer, in which all four

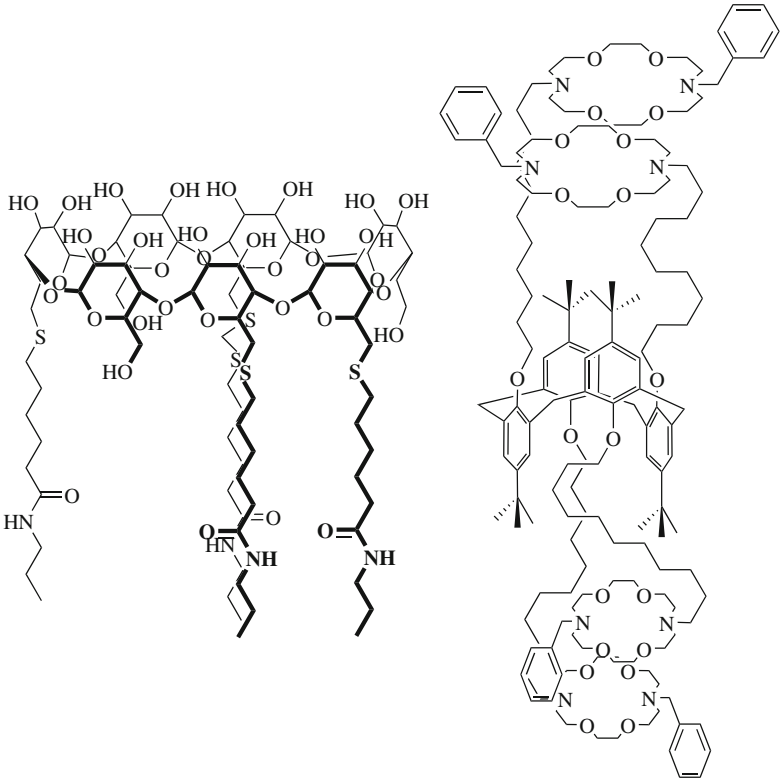


**Fig. 5.16** Channels formed by a cyclodextrin [48] (*left*) and calixarene [49] (*right*)

functionalizable phenols point in the same direction, and the 1,3-*alternate* (*1,3-alt*) conformer, in which alternating phenolic rings point in opposite directions. The aromatic rings of the *cone* conformer are separated by 0.7 nm to give a central pore about 0.3 nm in diameter through which species with $r \approx 0.15$ nm can pass.

*Cone*-4-*t*-butylcalix[4]arenetetra(diethylamide) binds alkali metals within an extended central cavity and extracts $Na^+$ from aqueous solution with a selectivity factor of 68 over $K^+$ [50]. Although this compound cannot span a phospholipid membrane it clearly displays the level of selectivity required for an artificial ion channel. Fortuitously the Raston group reported the crystal structure of *O*-alkyl substituted calix[4]arene with *n*-octadecyl substituents that completely interdigitate to form crystalline bilayers approximately 3.3 nm thick [51]. The calix[4]arene framework therefore lends itself to membrane-spanning derivatives. As an exemplification of this, the Gokel group appended four dodecyl ethers, terminating in *N*-benzyldiaza[18]crown-6, from calix[4]arene in both *cone* and 1,3-*alt* conformations [49]. The former was nonconductive, being unable to span a bilayer fully, but $Na^+$, with a hydrated radius of 0.18 nm, burst through a planar lipid bilayer incorporating the latter, illustrated in Fig. 5.16.

Noting that the resorcin[4]arenes are amphiphilic analogues of the calix[4]arenes, the Kobuke group designed a series of calix[4]resorcarenes, extended with undecyl and heptadecyl substituents to span a lipid monolayer [52]. The former conducted both $Na^+$ and $K^+$, with selectivity for $K^+$ over $Na^+$ by a factor of three, though $K^+$ conduction was blocked by addition of $Rb^+$. Compounds with the shorter undecyl substituents did not conduct at all suggesting that the compounds insert in inner and outer lipid layers with cation transport occurring when the alkyl substituents of two molecules met in a gramicidin-like manner.

Similar results were reported by the Beer group for resorcin[4]arenes containing phenoxyalkyl ether, phenoxypolyether or phenoxycrown ether substituents [53]. These displayed $K^+$ selectivity over other alkali metal cations with the acetylated derivatives able to transport $K^+$ across a chloroform bulk liquid membrane with varying efficacies. Compounds that extended to 2 nm were active in phospholipid bilayers but those with shorter alkyl chains (ca. 1.6 nm) were not. Gramicidin-like behaviour was again invoked as the channel forming mechanism.

Two examples of channel-forming calixarene derivatives help to support assertions, discussed above, that selectivity is enhanced when the preferred coordination geometry of the cation is matched by donor atoms in the transmembrane channel. Firstly, the compounds tested by the Beer group were all *cone* conformer calixarenes with four polyether groups extending from the lower rim. This presents an environment where a cation could be surrounded by eight oxygen atoms from four $–OCH_2CH_2O-$ motifs that repeat for the length of the polyethers. Such an arrangement favours $K^+$ over $Na^+$ as was shown by its transmembrane activity. Another derivative, a bis(4-t-butylcalix[4]arene) compound known as a calixtube, gave further evidence for the cation binding mode. This symmetric compound resembled a Chinese lantern in which calixarenes are at the top and bottom, separated by four quite inflexible $–OCH_2CH_2O-$ linkers. The crystal structure of the $K^+$ complex showed the cation to be held by all eight oxygen atoms in a cubic arrangement [54].

Secondly, a calix[4]arene with four polyether substituents prepared by the Cragg group demonstrated no $K^+$ transport but significant $Na^+$ transport [55]. Here, as with the Gokel group's more complex molecule described above, the calixarene was in the *1,3-alt* conformation so that at best the cation could only bind to two $–OCH_2CH_2O-$ chelation sites at any time. This would allow two water molecules to bind the cation giving a sixfold coordination centre ideal for $Na^+$. Consequently *1,3-alt* calix[4]arenes with simple polyether substituents appear to be selective for $Na^+$ over $K^+$.

A similar result has been observed for another of the Cragg group's modular systems employing a *cone*-4-t-butylcalix[6]arene filter and a membrane-disrupting surfactant, Triton-X100$^{TM}$, that can span a phospholipid membrane [56]. Three lower rim methoxy groups alternate with Triton-X100$^{TM}$ substituents so that a channel, lined with three strands of repeating $–OCH_2CH_2O-$ sites, was available to the cation. This gave the preferred sixfold environment for the cation and, with the surfactant behaving like a molecular harpoon, allowed $Na^+$ to cross a lipid bilayer at a rate of $7 \times 10^6$ ions per second consistent with channel-like behaviour.

### 5.6.2.4 Acyclic Models

Nature does not use a filter system in every case. Simple alkyl esters and ethers, such as alamethicin, monensin or amphotericin B, are able to aggregate and form

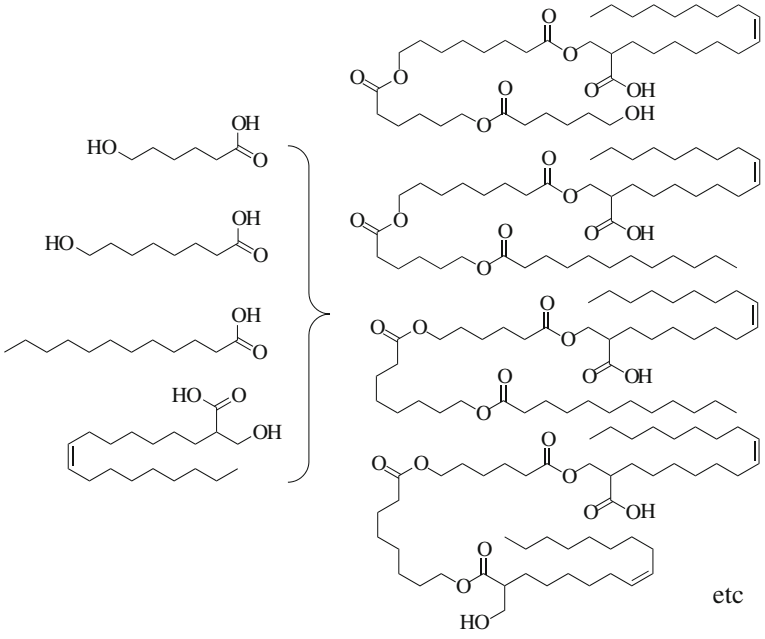



**Fig. 5.17** Systematic structural variation to form membrane piercing compounds [60]

barrel stave channels. Analogues of these have been prepared by the groups of Menger (alkyl esters of polyethylene glycols containing benzyl head groups) [57], Regan (bis(polyethyleneglycol)esters of alkanes and alkenes) [58] and Kobuke (alkylammonium cations linked to alkylpolyether carboxylates) [59]. Only the latter displayed any ion selectivity, with channels opening in the presence of both $Na^+$ and $K^+$.

A more systematic approach has been taken by the Fyles group. In an effort to understand the important structure-activity relationships underpinning simple ion transport, alkyl chains of variable lengths, polyethers and amino acids have been linked by esters to produce an extensive library, examples of which are to be seen in Fig. 5.17 [60]. The transport rates and selectivities of these compounds gave valuable information on the essential and desirable characteristics of these, the simplest, transmembrane ion channels.

## 5.7 Summary

The methods by which ions and small molecules are transported across cell membranes are varied. They range from highly selective siderophore shuttles through channels formed by complex proteins to non-specific aggregates of simple ethers and alkyl esters. A number of methods are available to study ion transport and range from those that determine the current resulting from the movement of ions to X-ray crystallographic structures which show the relative positions of atoms, including those being transported, in three dimensions. No single approach can give the entire picture though computer simulations may come close. Nevertheless, many features of ion transport are clear. The transport mechanism must be highly specific in favour of one water soluble chemical species, it must be able to move that species from one aqueous region to another through a bilayer of variable polarity with a hydrophobic core, and it must do this fast enough to respond to rapidly changing inter- and intracellular conditions. The complex nature of the process has been called ‘supramolecular biology’ by some to reflect the many non-covalent interactions at play between the cell membrane, the transmembrane transporter (whether ionophore, aggregate or channel-forming protein), and the transported species.

Chemists are able to approach the problem from a minimalist angle. In essence they are trying to answer the question: what is the simplest molecule that can accomplish the transport process? The principles behind supramolecular chemistry have been essential in constructing artificial analogues of natural transporters. Knowledge of host-guest preferences and the influence of the macrocyclic effect allows an informed choice of ionophore to be made. Similarly, matching the hydropathy index of the bilayer to potential membrane spanning compounds allows membrane piercing compounds to be made. A combination of the two concepts can result in a compound with channel-like behaviour and levels of activity.

# References


1. Singer SJ, Nicolson GL (1972) The fluid mosaic model of the structure of cell membranes. Science 175:720–731
2. Ward ML et al (1978) Crystal and molecular-structure of sodium bromide complex of monensin, $C_{36}H_{62}O_{11}{\cdot}Na^{+}Br^{-}$. Acta Cryst Sect B 34:110–115
3. Neupert-Laves K, Dobler M (1975) Crystal-structure of a $K^{+}$ complex of valinomycin. Helv Chim Acta 58:432–442
4. Sato T et al (1998) Prodigiosins as a new group of $H^{+}/Cl^{-}$symporters that uncouple proton translocators. J Biol Chem 273:21455–21462
5. Doyle DA et al The structure of the potassium channel: molecular basis of $K^{+}$ conduction and selectivity. Science 280:69–77
6. Niemeyer MI et al (2003) A conserved pore-lining glutamate as a voltage- and chloride-dependent gate in the ClC-2 chloride channel. J Physiol 553:873–879
7. Cestèle S, Catterall WA (2000) Molecular mechanisms of neurotoxin action on voltage-gated sodium channels. Biochimie 82:883–892
8. Atwood JL et al (1996) Calixarene chloride channel blockers. US Patent 5:489– 612
9. Minami K et al (1993) Mechanism of activation of the Ca(2+)-activated $K^{+}$ channel by cyclic AMP in cultured porcine coronary artery smooth muscle cells. Life Sci 53:1129–1135
10. Miyazawa A, Fujiyoshi Y, Unwin N (2003) Structure and gating mechanism of the acetylcholine receptor pore. Nature 423:949–955
11. Wallace BA (1990) Gramicidin channels and pores. Annu Rev Biophys Chem 19:127–157
12. Burkhart BM et al (1998) Heterodimer formation and crystal nucleation of gramicidin. D Biopys J 75:2135–2146 PDBID:1ALZ
13. Lomize AL, Orekhov VI, Arsenèv AS (1992) Refinement of the spatial structure of the gramicidin A ion channel. Biol Membr (USSR) 18:182–200 PDBID:1GRM
14. Schlesinger PH et al (2002) SCMTR: a chloride-selective, membrane-anchored peptide channel that exhibits voltage gating. J Am Chem Soc 124:1848–1849
15. Schnell JR, Chou JJ (2008) Structure and mechanism of the M2 proton channel of influenza A virus. Nature 451:591–595
16. Stouffer AL et al (2005) Sequence determinants of a transmembrane proton channel: an inverse relationship between stability and function. J Mol Biol 347:169–179
17. Yellen G (2002) The voltage-gated potassium channels and their relatives. Nature 419:35–42
18. Perozo E et al (2002) Open channel structure of MscL and the gating mechanism of mechanosensitive channels. Nature 418:942–948
19. Nagel G et al (2003) Channelrhodopsin-2, a directly light-gated cation-selective membrane channel. Proc Nat Acad Sci USA 100:13940–13945
20. Radu I et al (2009) Conformational changes of Channelrhodopsin-2. J Am Chem Soc 131:7313–7319
21. Takahashi T et al (1991) Photoisomerization of retinal at 13-ene is important for phototaxis of *Chlamydomonas reinhardtii* – simultaneous measurements of phototactic and photophobic responses. Biochem Biophys Res Com 178:1273–1279
22. Stouffer et al (2008) Structural basis for the function and inhibition of an influenza virus proton channel. Nature 451:596–599 PDBID:3BKD
23. Gonen T, Walz T (2006) The structure of aquaporins. Quart Rev Biophys 39:361–396
24. Horsefield R et al (2008) High-resolution x-ray structure of human aquaporin 5. Proc Natl Acad Sci USA 105:13327–13332 PDBID:3D9S
25. Nilius B, Droogman G (2003) Amazing chloride channels: an overview. Acta Phys Scand 177:119–147
26. Dutzler R et al (2002) X-ray structure of a CIC chloride channel at 3.0 angstrom reveals the molecular basis of anion selectivity. Nature 415:287–2940 PDBID:1KPL, 1KPK
27. Sato C et al (2001) The voltage-sensitive sodium channel is a bell-shaped molecule with several cavities. Nature 409:1047–1051

28. Jasti J et al (2007) Structure of acid-sensing ion channel 1 at 1.9 Å resolution and low pH. Nature 449: 316–323
29. Hunte C et al (2005) Structure of a $Na^+/H^+$ antiporter and insights into mechanism of action and regulation by pH. Nature 435:1197–1202
30. Legros C, Bougis PE, Martin-Eauclaire MF (1999) Molecular biology of scorpion toxins active on potassium channels. Persp Drug Disc Design 16:1–14
31. Wulff H et al (2007) Modulators of small- and intermediate-conductance calcium-activated potassium channels and their therapeutic indications. Curr Med Chem 14:1437–1457
32. Bernèche S, Roux B (2001) Energetics of ion conduction through the $K^+$ channel. Nature 414:73–77
33. Valiyaveetil FI et al (2004) Glycine as a D-amino acid surrogate in the $K^+$-selectivity filter. Proc Nat Acad Sci USA 101:17045–17049
34. Serysheva I et al (2005) Structure of $Ca^{2+}$ release channel at 14 angstrom resolution. J Mol Biol 345:427–431
35. Dudev T, Lim C (2009) Factors governing the $Na^+$ vs $K^+$ selectivity in potassium channels. J Am Chem Soc 131:8092–80101
36. Wu G, Wong A (2001) Direct detection of the bound sodium ions in self-assembled 5′-GMP gels: a solid-state Na-23 NMR approach. Chem Commun 2658–2659
37. Landini D, Montanar F, Pirisi FM (1974) Crown ethers as phase-transfer catalysts in 2-phase reactions. J Chem Soc Chem Commun 879–880
38. Marinez ER et al (1996) Enterobactin and enantioenterobactin. J Org Chem 61:3548–3550
39. Gale PA et al (2005) Co-transport of $H^+/Cl^-$ by a synthetic prodigiosin mimic. Chem Commun 3773–3775
40. Seganish JL, Davis JT, Prodigiosin is a chloride carrier that can function as an anion exchanger. Chem Commun 5781–5783
41. McNally BA et al (2008) Structure-activity relationships in cholapod anion carriers: enhanced transmembrane chloride transport through substituent tuning. Chem Eur J 14:9599–9606
42. Lear JD, Masserman ZR, DeGrado WF (1988) Synthetic amphiphilic peptide models for protein ion channels. Science 240:1177–1181
43. Voyer N, Potvin L, Rousseau E (1997) Electrical activity of artificial ion channels incorporated into planar lipid bilayers. J Chem Soc Perkin Trans 2:1469–1471
44. Jullien L, Lehn JM (1988) The chundle approach to molecular channels synthesis of a macrocycle based molecular bundle. Tetrahedron Lett 29:3803–3806
45. Fyles TM, James TD, Kaye KC (1993) Activities and modes of action of artificial ion-channel mimics. J Am Chem Soc 115:12315–12321
46. Gokel GW (2000) Hydraphiles: design, synthesis and analysis of a family of synthetic, cation-conducting channels. Chem Commun 1–9
47. Hall CD et al (2003) Cation transport by a redox-active synthetic ion channel. Org Biomol Chem 1:2973–2982
48. Tabushi I, Kuroda Y, Yokata K (1982) A,B,D,F-Tetrasubstituted beta-cyclodextrin as artificial channel compound. Tetrahedron Lett 23:4601–4604
49. de Mendoza J et al (1998) A synthetic cation-transporting calix[4]arene derivative active in phospholipid bilayers. Angew Chem Int Ed 37:1534–1537
50. Arduini A et al (1988) Molecular inclusion in functionalized macromolecules 15. para-tert-Butylcalix[4]arene tetra-acetamide – a new strong receptor for alkali cations. J Incl Phenom 6:119–134
51. Clark TE et al (2006) Supersized bilayers based on an O-alkyl substituted calix[4]arene. CrystEngComm 8:707–711
52. Kobuke Y, Ueda K, Sokabe M (1992) Artificial nonpeptide single ion channels. J Am Chem Soc 114:7618–7622
53. Wright AJ et al (2001) Novel resorcin[4]arenes as potassium-selective ion-channel and transporter mimics. Chem Eur J 7:3474–3481

54. Matthews SE et al (2002) Calix[4]tubes: A new class of potassium-selective ionophore. J Am Chem Soc 124:1341–1353
55. Lawal O et al (2009) An artificial sodium ion channel from calix[4]arene in the 1,3-alternate conformation. Supramol Chem 21:55–60
56. Iqbal KSJ et al (2007) Artificial transmembrane ion channels from commercial surfactants. Chem Commun 3951–3953
57. Menger FM et al (1990) Synthetic flux-promoting compounds – exceeding the ion-transporting ability of gramicidin. J Am Chem Soc 112: 2451–2453
58. Nagawa Y, Regen SL (1991) Membrane-disrupting surfactants that are highly selective toward lipid bilayers of varying cholesterol content. J Am Chem Soc 113: 7237–7240
59. Kobuke Y, Nagatani T (2001) Transmembrane ion channels constructed of cholic acid derivatives. J Org Chem 66:5094–5101
60. Fyles TM, Luong H (2009) Structure-activity relationships in linear oligoester ion-channels. Org Biomol Chem 7:733–738

# Chapter 6
# Diagnostic Applications

## 6.1 Applications of Supramolecular Chemistry in Medical Diagnostics

The basic premise of medical diagnostics, and indeed all chemical tests, is that an unequivocal signal is given by the diagnostic agent in response to the presence of a specific analyte. Simple chemical spot tests could fall into this category: addition of a colourless silver nitrate solution to a colourless solution of sodium chloride results in the precipitation of white solid silver chloride. At a chemically more complex level, the addition of protons to certain organic molecules, which changes their electronic conjugation resulting in a distinctive colour change, is the basis of pH sensitive dyes. Organic dyes are central to biomedical science and are widely used to distinguish different cell types in bioassays. As Table 6.1 shows, these dyes and stains can operate in a wide spectral range, however, they represent a very simplistic approach to imaging.

Applying the principles of supramolecular chemistry to imaging, particularly with respect to molecular recognition by a designed host molecule, it is possible to generate much more sophisticated diagnostic agents.

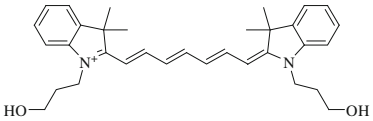
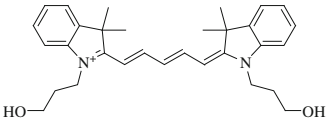
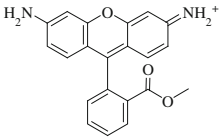
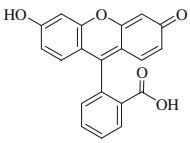
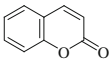
## 6.2 Design Principles

In common with other supramolecular systems the process of guest specific binding and signal transduction will involve a complex interaction of forces. In supramolecular chemistry a design principle has emerged that considers sensors to have three requirements. There must be a region of the sensor molecule that binds to the target with high specificity. This receptor can take many forms although most utilize a combination of weaker forces generally associated with supramolecular chemistry. At one end of the scale the probe could be a macrocycle, with particular affinity for cations or anions of particular sizes and geometries, whereas at the other it could be a very specific protein fragment or antibody complementary to that of the target molecule or biological feature. The sensor must also function as a reporter so that once the binding event has occurred it has to be signalled in some manner. Typically

P.J. Cragg, *Supramolecular Chemistry*, DOI 10.1007/978-90-481-2582-1_6,

© Springer Science+Business Media B.V. 2010

**Table 6.1** Dyes used in biological imaging

| Filter | Wavelength (nm) | Example | Structure |
|---|---|---|---|
| Near infrared (NIR) | 770 – 850 | Cyanine dye (Cy7) |  |
| Deep red | 650 – 740 | Cyanine dye (Cy5) |  |
| Red | 570 – 630 | Rhodamine |  |
| Green | 510 – 570 | Fluorescein |  |
| Blue | 425 – 480 | Coumarin |  |

this is achieved through a molecule that can respond to the binding event by a change colour, fluorescence, electrochemical potential or other readily detectable output. The signal transduction requirement separates sensors from simple host molecules that do not necessarily signal their binding to a target substrate. The last requirement is that the probe and reporter are linked in some way leading to the sensor being referred to as a receptor-spacer-reporter system, illustrated in Fig. 6.1.

Initially it may appear that an ideal sensor would be one where the same chemical group functioned as both probe and reporter. For many applications, however, the two need to be separated. This is most clearly apparent when the target is in a cleft in a large biomolecule: if the reporter is too close to the probe any change in signal may be masked by the surrounding protein chains, as in Fig. 6.2. Other applications may require a degree of conjugation between probe and reporter so that the electronic changes that occur upon binding are transferred directly to the reporter group.

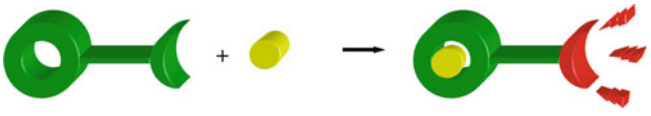


**Fig. 6.1** The receptor-spacer-reporter concept

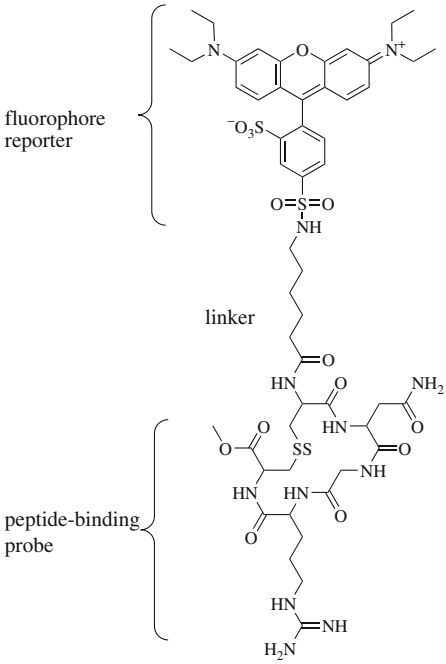



**Fig. 6.2** A remotely linked reporter

This is most clearly seen in the work of Beer whose group linked cation specific crown ethers to electrochemically responsive ferrocene reporter groups shown in Fig. 6.3 [1].

Reversibility is also an issue if the sensor is to be used more than once which, given the investment necessary to synthesize any complex molecule, will be highly desirable. The sensor's solubility in aqueous or organic solvents must be considered. Most targets are likely to be in water so fluorescent and optical sensors will need to be water soluble. Other sensors may be surface bound, or incorporated into other materials as signal modifiers, in which case aqueous solubility is not an issue.

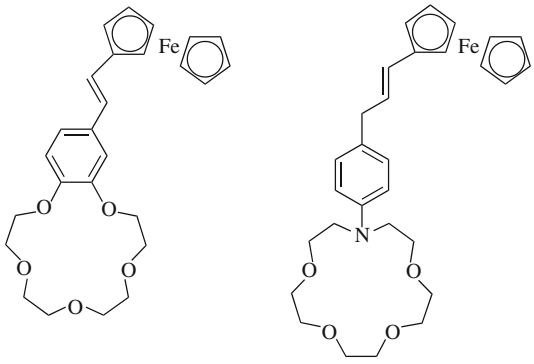



**Fig. 6.3** Crown-spacer-ferrocene reporter systems [1]

## 6.3 Supramolecular Sensors

Sensors based on molecules designed along supramolecular principles are widespread in the scientific literature though few have found their way into practical devices. In most cases the sensor molecules follow the receptor-spacer-reporter principles outlined above. The choice of reporter depends on how the signal is to be used. Where the sensor molecule is to be used in solution an optical, or fluorescent, response may be desirable. If this can be detected by eye so much the better, but in reality the concentrations of analytes are likely to be low and there may be interfering species, such as natural chromophores, present in a biological sample. Fortunately uv/visible and fluorescence spectrophotometers are common in analytical laboratories making detection relatively simple. An important aspect of the reporter is its ability to work at a wavelength which can be imaged easily and cheaply. It is also useful to develop reporting groups that have a well defined response so that detection is unequivocal. Furthermore it is valuable to have a number of chemical systems that respond in different spectral regions. These varied approaches to sensing can be illustrated by considering the main types of signal transduction: optical, fluorescent and electrochemical responses.

### *6.3.1 Optical and Fluorescent Biosensors*

The principle of fluorescent detection systems is that an organic molecule, or in some cases a metal complex, can absorb photons in a particular range and then reemit them at a longer wavelength, illustrated by a Jablonski diagram in Fig. 6.4. For this to be used as a detection method the fluorescent species must have an affinity for a specified target or be attached to a molecule that does.

Once the fluorescent species has been excited it responds by channelling the energy into a conformational change, intramolecular rearrangement, bond vibration or similar effect. This process usually lasts for no more than a few nanoseconds. The species can also lose energy to external systems such as solvent molecules which reduces the intensity of the response. Finally the energy is released through fluorescence emission at a longer wavelength than that of excitation which returns the molecule to its unexcited ground state. To avoid overlap between excitation and emission wavelengths should be separated as far as possible: the difference between the two wavelengths is known as the Stokes shift.

Many detectors have sets of filters that can be combined to show responses to different biological features. For example, two molecules, one incorporating a red dye attached to a receptor that interacts with cell membranes and the other having a blue dye that targets the cell nucleus, can be used independently but give more information if used together. Thus the same microscope image can show both the cell membranes and nuclei. The approach only works if the dye derivatives are specific to the target features, do not overlap in their emission wavelengths, and can be differentiated from the background colours, or autofluorescence in the case of a fluorescent dye.

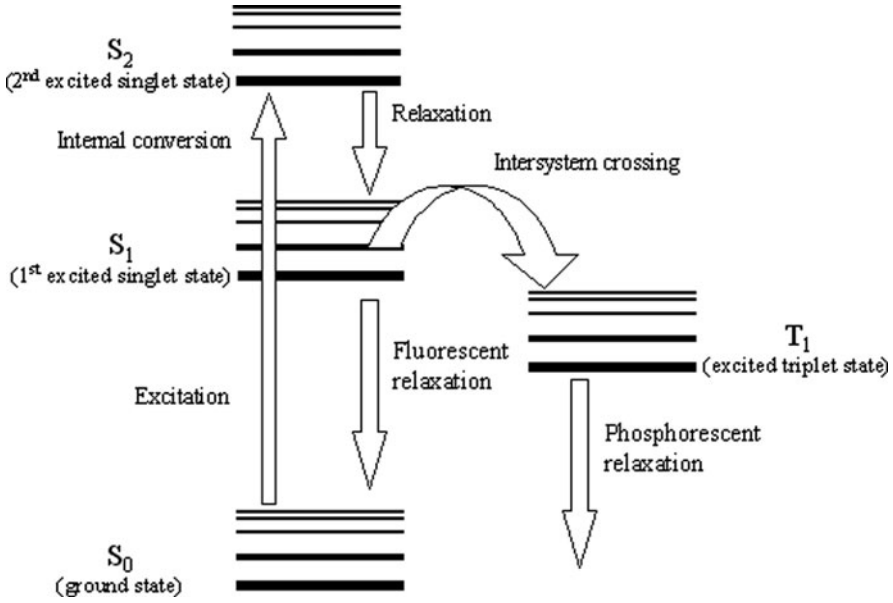
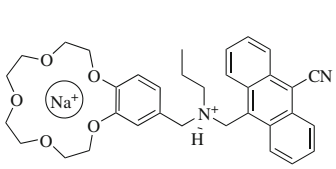


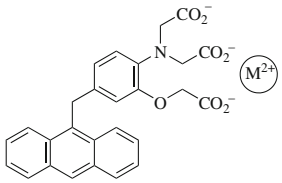

**Fig. 6.4** The energetics of reporter excitation and relaxation

Autofluorescence is a particular problem as it is observed naturally in the green region of the electromagnetic spectrum when imaging many cell types. Two methods to overcome the problem are either to use dyes that respond outside these regions, usually at wavelengths over 500 nm, or use a reporter that has a different type of response such as radioactive decay. As the latter is often undesirable, due to the inherent dangers of working with radioactive isotopes, a third method employs species with long lived fluorescence. Complexes of lanthanides such as europium and terbium, which give orange and pale blue responses respectively, are often encountered as reporter molecules. The advantage of these complexes is that the decay of background fluorescence is rapid so by delaying detection by a few milliseconds only the response due to the complexes, and their targets, remains.

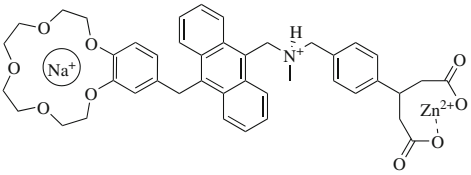
A number of metals give fluorescent or optical emissions in response to irradiation in various regions of the electromagnetic spectrum. Examples of these include transition metals that can be imaged in the visible spectrum through *d–d* orbital electronic transitions and lanthanides that have distinctive fluorescent signals from *f–f* orbital transitions. One of the best methods to achieve high resolution imaging is to use quantum dots, small aggregates of semiconducting metals that exhibit intense fluorescence emissions at wavelengths directly related to their sizes, which tend to be on the 2–10 nm scale. Unfortunately the toxicity of these assemblies is high and their application in biological and clinical contexts is currently quite limited [2]. Consequently the receptor-spacer-reporter design has been used for a variety of diagnostic purposes.



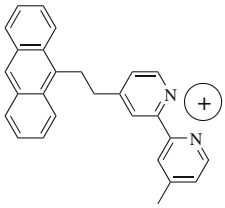

AND gate: $Na^+ + H^+$ = fluorescence




OR gate: any $M^{2+}$ = fluorescence




AND gate: $Na^+ + H^+ + Zn^{2+}$ = fluorescence




NOR gate: fluorescence quenched by any guest

**Fig. 6.5** Examples of logic gates

### 6.3.1.1 Logic Gates

The receptor-spacer-reporter motif may appear quite simple yet, as the group of de Silva has shown, can result in some extremely sophisticated responses [3]. Through a careful choice of fluorophore reporter, short spacer and nitrogen-containing probe it is possible to construct a series of molecules that can perform Boolean logic in response to external stimuli, illustrated in Fig. 6.5. These mechanisms are currently used in electronic logic components and form the basis of many computational functions.

In a simple form the chemical function can take the form of a low pH YES gate that has its fluorescence quenched when protonated such that total loss of fluorescence would constitute the 'YES' response. Other logical operators include NOT, AND, OR and NOR which can all be used in combinations to give unambiguous chemical signals. For example if two sensors were used, one that was quenched below pH 7 and one below pH 4, there would be no loss of fluorescence intensity above pH 7, a significant reduction between pH 4 and 7, and complete quenching below pH 4. If these pH sensors were used in combination with ion selective hosts a combination of responses could meet highly specific analytical challenges. A response could be given not only to the pH but also to the particular ions present at that pH. Further applications are also possible, including a three-state logic gate that is normally only encountered in electronic components.

### 6.3.1.2 Detection of Critical Analytes

The de Silva group has also pioneered the use of the classic receptor-spacer-reporter molecular recognition stratagem in supramolecular diagnostics. Receptors for key inorganic analytes in blood were linked to optical reporter groups. These receptors have since been built into commercial the OPTI Critical Care Analyzer [4] which measures a wide range of 'critical analytes' in blood (oxygen, carbon dioxide, pH, key cations and anions, and specific metabolites) using optical sensors [5].

The sensors for $Na^+$, $K^+$ and $Ca^{2+}$ are based on a crown ether, cryptand and podand, respectively as shown in Fig. 6.6. The sensors are surface bound to a polymer disc housed in a disposable cassette. Attachment to the polymer is through a 4-aminonaphthalimide linker to create a fluorionophore. The receptor group reports analyte binding and fluorescence increases proportionately to the concentration of the cations. The sensor gives an optical response and is therefore functions as an optode. In a similar approach, ionophore based sensors composed of cryptaspherands with chromogenic substituents have been developed to detect alkali metals in blood serum [6].

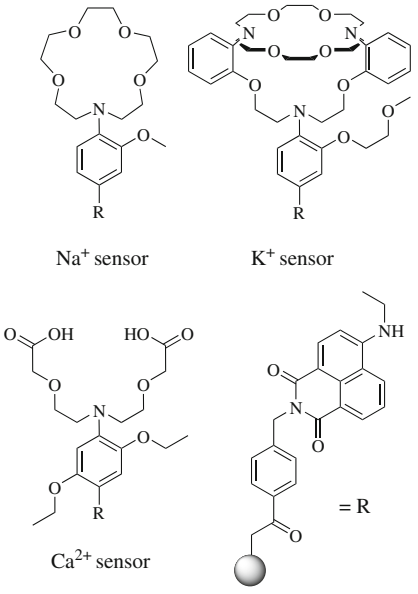



**Fig. 6.6** Polymer bound critical analyte sensors [6]

### 6.3.1.3 Displacement Assays

The Anslyn group has pioneered supramolecular displacement assays as a distinctive method of analysis. A typical example used a tripodal molecule, or podand, to determine the amount of citrate in drinks and the quality of wine made from Pinot Noir grapes [7]. The initial work on citrate utilized a complex between a tris(guanidinium) derivative of 1,3,5-triethylbenzene with a fluorescent dye, carboxyfluorescein. In the presence of citrate the dye is ejected and replaced with the

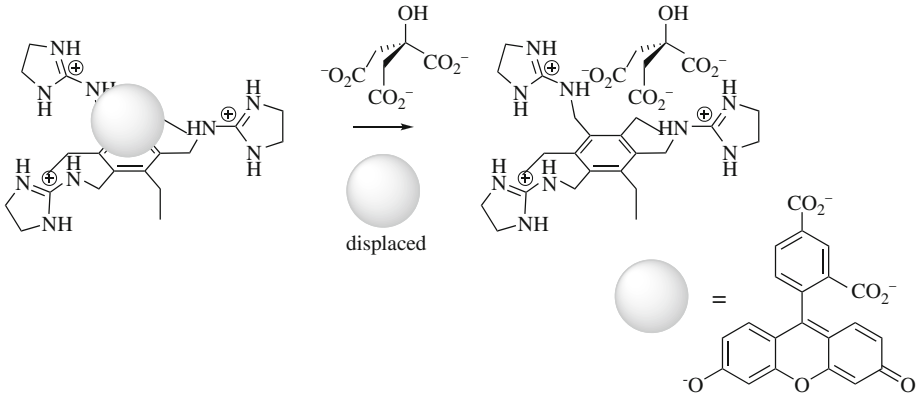



**Fig. 6.7** The principle of a displacement assay

more tightly binding anion. The dye's fluorescence is quenched when bound by the podand but once it is free it can be observed in solution. An increase in fluorescence in solution therefore signals the presence of citrate.

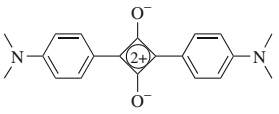
Tests have shown that the sensor is specific for citrate in water in the presence of species that typically interfere in citrate analysis such as phosphates, sugars, carboxylic acids and simple salts. The assay is able to give this level of selectivity because the dye-podand complex has been tuned so that dye displacement is only favourable if citrate competes for the dye's binding site. When other potentially interfering molecules are present the stability of the competitor-podand complex is lower than that of the dye-podand complex so the latter remains intact, until citrate displaces the dye. The modular construction of the podand allows each 'leg' of the tripod to be varied resulting in a vast number of sensor compounds. Using these molecules, either individually or as part of an array, highly specific assays can be devised for a range of compounds. The general principle is illustrated in Fig. 6.7.

### 6.3.1.4 Squaraine Dyes

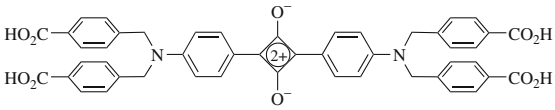
Most biological features are imaged through the direct use of dyes with an affinity for their target sites. Often these dyes are chosen for the different electronic absorption and emission spectra so that tissue, in particular, can by labelled for several different cell types. Techniques can be as sophisticated as the introduction of green fluorescent protein (GFP) by genetic manipulation of the organism or the simple staining by a dye that has empirical specificity for a particular target. As with most molecular imaging approaches, it is desirable to introduce as little of the imaging agent as possible while having it give as strong a response as possible.

Squaraines are derivatives of 3,4-dihydroxycyclobut-3-ene-1,2-dione, or squaric acid, so called because the central structural motif consists of four carbon atoms linked to form a square [8]. Numerous squarylium dyes, such as those in Fig. 6.8,

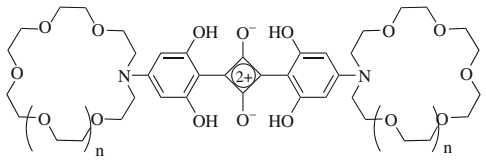
**Fig. 6.8** Squaraine dyes




squaraine dye motif




functionalized squaraine dye




alkali metal responsive squaraine dye

have been prepared in the past four decades and have found uses as optical sensors, modifiers in photocells and, of relevance here, bioimaging.

Hewage and Anslyn have taken their pattern based recognition methods and combined them with the highly responsive squaraine dyes, discussed above, with recognition sites for *d*-block metals [9]. Using a variety of thiols in combination with metals and squarianes it is possible to elicit a number of outcomes which can be subjected to principle component analysis. This results in diagnostic fingerprint responses for specific metal and thiol combinations. One obvious application of this method is the detection of harmful mercury cations but other harmful metals such as palladium and nickel can also be detected by this approach using the same combination of co-analytes.

Simple squaraine dyes incorporate conjugated aromatic groups to modulate the operating range of the fluorescent response or enhance the intensity of that response. The application of supramolecular principles to this field has generated some remarkable compounds with valuable diagnostic properties. Oguz and Akkaya demonstrated that a symmetric squaraine dye containing terminal two azacrown ether groups could be prepared in high yield in a one pot reaction [10]. Although not yet used in sensors, it is clear that changes in the compounds' fluorescent responses would occur when appropriately sized cations were encapsulated by the crown groups.

The Ajayaghosh group has taken the recognition principle one step further [11]. By linking two squaraine dyes together with polyethers of varying lengths they have generated a family of fluorescent foldamer dyes that only function when the

polyether is of the correct size to bind to the metal. For enhanced fluorescence to occur the molecule must induce a folded conformation in which the polyether wraps around the metal to bring the two squaraine dyes in close enough proximity to form an excimer. Cation selectivity was investigated for homologues with different lengths of polyether. In most cases two or more cations were bound with similar binding constants, however, the compound with a pentaethylene glycol spacer only responded to calcium. As well as the fluorescent response, a very pronounced light blue to deep purple optical change was observed.

A far more impressive example where supramolecular principles have been considered in the design of a squaraine-based imaging agent comes from the work of Smith. His group has incorporated squaraines as part of the receptor-spacer-reporter motif to detect of bacteria in vivo [12]. Both Gram-positive and -negative bacteria have anionic regions on their outer surfaces which were targeted by a simple zinc-containing complex. The probe is composed of two dipicolylamine groups bound to zinc ions and linked to a near infrared active squaraine dye. As an exemplification of this potential use for the dyes the Smith group created a squaraine rotaxane that incorporated dipicolylamine groups. These groups act as stoppers on a rotaxane, formed when macrocyclization of two 4-xylylenediamine and two isophthaloyl chloride molecules occurs around the squaraine moiety. The dipicolylamine termini bind zinc very efficiently so that four cationic groups emanate from the rotaxane core. The picolyl-zinc motif bound to *Escherichia coli* bacteria to the extent that their successful binary fission could be followed over thirty minutes by fluorescence microscopy. The same method was used to prepare other squaraine rotaxanes as shown in Fig. 6.9.

The purpose of the rotaxane is to protect the squaraine from nucleophilic attack, likely to occur under physiological conditions, and so enhance the lifetime of the probe in vivo. Indeed, the half-life of the probe is 100 times longer than a non-rotaxane analogue that incorporates a commercial sulfonated carbocyanine dye in place of the squaraine moiety. Furthermore the near infrared spectral nature of the dye means that it gives a strong signal even through centimetres of soft tissue.

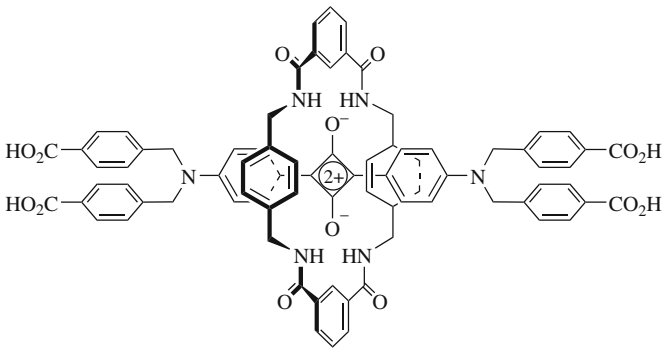



**Fig. 6.9** A rotaxane squaraine dye [13]

To demonstrate the efficacy of the probe, a mouse was infected with both *Escherichia coli* and S*taphylococcus aureus* that had been labelled with the imaging agent. Shortly thereafter the sites of infection and injection were clearly detected by fluorescence microscopy of the anaesthetized animal. The squaraine was safely metabolized and excreted within 18 h allowing for the progress of infection to be followed in the same mouse at later dates without harm to the animal. The ability to follow the bacteria over time in an otherwise healthy animal has important repercussions for the pharmaceutical industry. The method could be used to follow the progress of novel antibiotics and other treatments without having to sacrifice animals at different stages to observe the efficacy of the treatment. It would also avoid any variation in response between individuals. Using a lipophilic analogue of the probe the Smith group was able to show that their general approach can be used to follow cell division in live human breast cancer cells with the probe passed on through several generations of daughter cells over 8 days with no loss of fluorescence intensity [13].

## *6.3.2 Electrochemical Sensors*

An alternative to light-related detection is an electrochemical response. If the sensor and analyte are in solution then cyclic voltammetry can be used to detect changes in redox potential between the free sensor and its complex with the analyte. Supramolecular applications of this approach were pioneered by Beer who linked crown ethers to electrochemically responsive ferrocenium [1] and cobalticinium [14] groups. In the former case a response was detected when cations complementary to the crown ether cavity were added to acetonitrile solutions of the sensors; in the latter, anions were detected by an acyclic receptor.

Alternative strategies include incorporating the sensor molecules as modifiers to solid electrodes. In this case the sensor is often mixed with a carbon paste to form an electrode that responds to the analyte. The electrodes themselves can be miniaturized to be employed invasively in a clinical setting. Such an approach was adopted by O'Hare and colleagues who used a microelectrode incorporating an iron-salen complex to detect nitric oxide as discussed below [15]. Sensors may also be linked to the surface of a metal electrode, often gold, to act as analyte specific signal modifiers. Other techniques where the sensor is linked to a surface include the quartz crystal microbalance where a quartz disc is coated with gold to form electrodes and the gold coated with the sensor. As shown in Fig. 6.10, when the analyte binds to the modified crystal, piezoelectric effects are measured that are proportional to a change in mass. The property is registered continually from the start of the experiment so that concentrations of the target can be determined in real time [16].

### 6.3.2.1 Electrochemical Detection of Nitric Oxide

Until the mid 1980's the simple chemical species nitric oxide, NO, was generally associated with the numerous nitrogen oxides found in photochemical smog and other forms of urban pollution but was suddenly identified as an essential biological

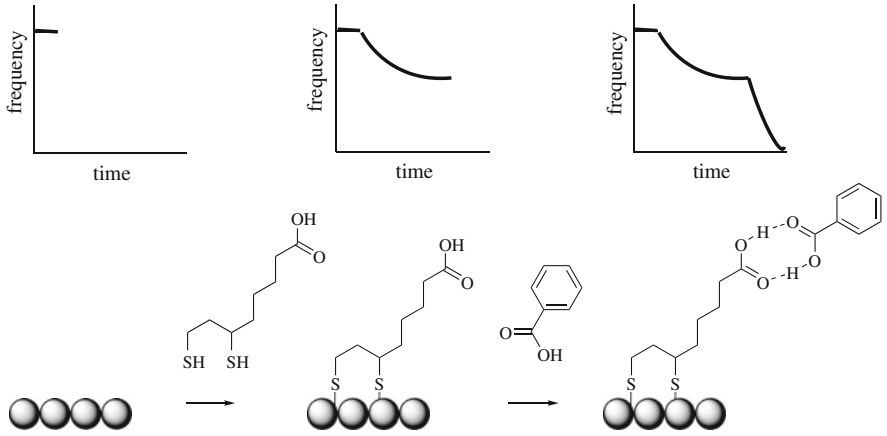



**Fig. 6.10** The principle of the quartz crystal microbalance: gold surface modification (*left*) and analyte detection by frequency modulation (*right*)

messenger. Earlier in that decade it had been postulated by Furchgott that endothelial cells released a compound, of unknown nature, that caused smooth muscle cells in the arteries to relax [17]. The causative agent was named the endothelium-derived relaxing factor, or EDRF, and the search for its true identity started. Initially it was thought that the substance must be a small protein or similar molecule that targeted receptors on the muscle cells but it was shown that both small molecules, in the form of dioxygen, and large iron-porphyrin containing proteins, exemplified by haemoglobin, inhibited the activity of EDRF. How could two such diverse molecules have the same effect on EDRF?

In parallel with this work other groups were investigating the synthesis of a secondary messenger known as cyclic guanosine monophosphate, or cGMP, which was affected by nitric oxide. In 1986 it was proposed that NO was the key, or possibly sole, constituent of EDRF. This made sense as compounds containing oxidized forms of nitrogen such as nitroglycerine and sodium nitroprusside were known to be potent vasodilators. Later work confirmed that NO had the same effects as EDRF and that it was formed naturally in its own biosynthetic pathway [18]. A group of closely related enzymes called nitric oxide synthases transform L-arginine into L-citrulline with NO as the by-product. The rapid progress in the identification of NO led to it being named 'Molecule of the Year' in 1992 and three of those central to the research, Furchgott, Murad and Ignarro, shared the 1998 Nobel Prize in Physiology or Medicine. A fourth contributor to the NO story, Salvador Moncada, was surprisingly overlooked by the Nobel Prize committee despite publishing a seminal paper on the subject in 1987 that has since been cited over 8,000 times in the scientific literature.

Given the importance of NO in muscle function, particularly in the heart, there needs to be a simple method by which it can be measured in vivo. One problem with detection is that NO is a radical species which makes it highly reactive and

short lived. To detect it a sensor needs to operate in the muscle tissue where NO is present and intercept the molecule before it reacts further. Noting that NO forms a stable complex with iron(III) in haem systems and simple coordination compounds, O'Hare and colleagues used a Fe(III)salphen compound to act as an electrochemical sensor embedded in an epoxy-graphite microelectrode [15].

The electrode made use of two supramolecular phenomena: the reversible $Fe^{III}$-NO interaction that changes the iron(II) to iron(III) redox couple, and so can be detected electrochemically, and the π–π interactions between graphite fragments and the aromatic backbone of Fe(III)salphen, to enhance electrical conductivity. In vitro experiments showed that the microelectrodes, which were small enough for tissue and single cell measurements, responded in direct proportion to the concentration of NO up to 200 μM with a detection limit down to 190 nM. The only interfering chemical species that was likely to be present under physiological conditions was ascorbate but it could easily be resolved from the NO related peak. To test the effectiveness of the microelectrode in vivo one was implanted into a rat and was able to detect bursts of NO at physiological levels in response to muscle activity. The NO derived current increased in the presence of L-arginine, a biological precursor for NO, and decreased when a nitric oxide synthase inhibitor was added.

## 6.4 Macrocyclic Complexes for Imaging

Diagnostic sensors based on crown ethers have been designed based on size and symmetry complementarity between host and guest analyte. The focus is therefore generally related to the binding of spherical metal cations or protonated amines, as both the metals and the $RNH_3^+$ group have strong affinities for [18]crown-6. Several examples are shown in Fig. 6.11. These include one from the Gokel group where *N*-(2-methoxyethyl)aza[18]crown-6 demonstrated a response to $Na^+$, but in the non-biological medium of methanol [19]. Gunnlaugsson's group used this as the basis from which to design a $Na^+$ sensor that incorporated a chromophore. The chromophore was based on a 4-nitrophenol azo dye moiety and responded to physiological concentrations of $Na^+$, but no other metals, at the concentrations found in blood [20]. The colorimetric response, which was also pH independent, made the compound an ideal candidate to become the active component in an optode sensor that could be used for real-time testing of $Na^+$ in blood in a clinical environment.

In cases where multiple binding sites are required the reporter group can be inserted between them as part of the spacer. Over many years de Silva's group has used fluorescent anthracene spacers to link crown ethers to a second binding site so that fluorimetric methods may be used to determine guest concentration [21]. This approach can be applied to the simultaneous detection of cations and anions or zwitterionic species such as amino acids. In one example aza[18]crown-6 was linked by an anthracene fluorophore to a guanidinium group. The result was a fluorescent photoinduced electron transfer (PET) sensor for the brain neurotransmitter γ-aminobutyric acid (GABA). As a zwitterion the GABA ammonium group binds

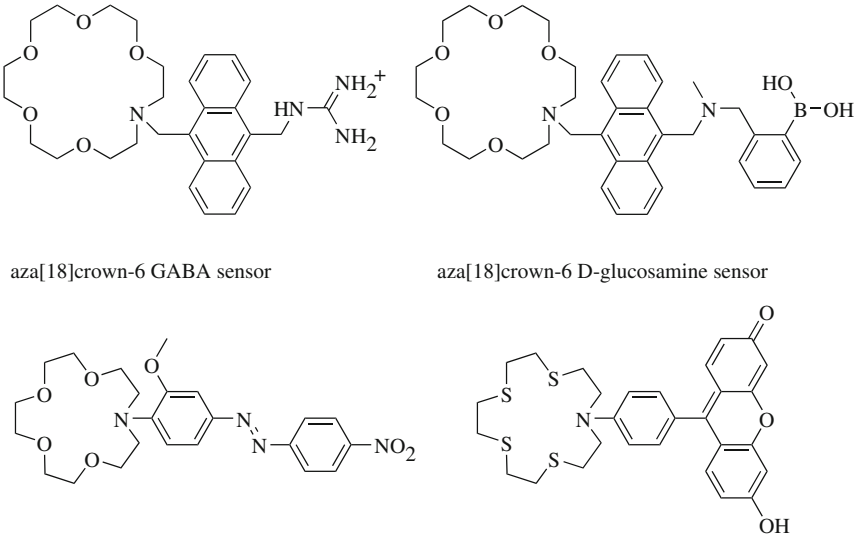



**Fig. 6.11** Examples of sensors based on analyte binding to crown ethers

to the sensor's azacrown ether, capitalizing on the complementarity between the three fold symmetry of [18]crown-6 and GABA's ammonium terminus as well as electrostatic attractions between the ammonium protons and crown ether oxygen atoms. At the same time GABA's carboxylate terminus binds to the guanidinium group. Clever design set the binding sites in the sensor to the correct distance apart to select GABA over other molecules where the acid and amine termini are closer or further apart. James and Cooper later combined an aza[18]crown-6-anthracene fragment with an aryl boronic acid group to detect D-glucosamine hydrochloride [22]. A sulfur analogue of aza[15]crown-5, azatetrathia[15]crown-5, was linked directly to fluorescein by the group of Chang to make a sensor for mercury in edible fish [23].

One constant problem for crown ether based sensors in vivo is the ubiquitous presence of chemical species that will compete with the target analyte for the sensor binding site. For sensors incorporating [18]crown-6 this especially problematic as sodium, potassium, ammonium and hydronium ($H_3O^+$) cations are all attracted to the threefold symmetry of the crown's cavity. Protonated terminal amines, including amino acids and peptides, can also interfere with analyte detection. It is therefore all the more pleasing when a crown ether based sensor is developed that does not bind to biologically common cations. This is the case for a saxitoxin chemosensor reported by Gawley, LeBlanc and co-workers [24].

Saxitoxin is a neurotoxin produced by certain species of dinoflagellates, cyanobacteria and, most famously, many species of both marine and freshwater pufferfish. The toxin is found throughout these fish, from internal organs through

to the skin. The neurotoxin is also found in shellfish contaminated with the bacteria that produce it. The toxin works by rapidly blocking cellular sodium channels which, in turn, leads to respiratory paralysis and, potentially, death from respiratory failure. A sensor for saxitoxin that functions in sea water, with its high concentration of dissolved salts, would therefore be valuable for determining the safety of seas adjacent to algal blooms. Incorporation of this sensor into a simple test for the food industry in counties where pufferfish are often on the menu would be potentially life-saving.

The sensor developed by Gawley and co-workers is based on de Silva's modular approach comprises an azacrown ether linked to a fluorophore by a short link. The molecule combines the aza[18]crown-6 recognition element with a fluorescent coumaryl group attached by a methylene spacer. In tests it was shown to bind to saxitoxin with a binding constant on the order of $10^5$ $M^{-1}$, even in a phosphate buffer at physiological pH and concentrations of sodium and potassium, and could detect levels of the toxin down to $10^{-7}$ M. Subsequent modification of the sensor allowed it to be linked to the surface of a quartz slide to allow the fluorescent response to concentrations of saxitoxin between $10^{-4}$ and $10^{-6}$ M to be detected via a fibre optic system. This level of sensitivity is comparable to the current mouse bioassay that requires the inoculation of a large number of animals to determine the concentration of saxitoxin present in the test sample [25].

## 6.5 In vivo Imaging: Magnetic Resonance Imaging Agents

Practically since their discovery, X-ray imaging methods have been used to detect anomalies in bone and tissue. For many decades this was the only non-invasive method to 'see' inside a patient, or any other organic form for that matter, but since the advent of affordable MRI instruments it has been possible to generate far more data from a single experiment. Magnetic resonance imaging (MRI) has rapidly become a valuable tool in the physician's armoury. The technique allows him or her to look inside the living patient's body but, unlike the older X-ray methods, this generates a 3D image and differentiates between soft structures such as organs, fat and muscle. Furthermore it is possible to design imaging agents that respond to specific tissue types such as cancerous tumours. MRI instruments are expensive to buy and maintain, and so are not as widespread as many other medical devices, yet they can give incredibly sensitive and sophisticated diagnoses. From a research perspective they can be used to monitor a vast number of biological processes in real time. One example where MRI techniques are leading the way to new discoveries is the observation of brain activity in response to particular stimuli which allows the subject's actual thought processes to be followed.

MRI is based on nuclear magnetic resonance (NMR) spectroscopy which is widely used to deduce the structures of organic molecules. Given the public perception of the word 'nuclear' (nuclear bombs, nuclear waste, etc) this has been dropped from the acronym used for the technique. In NMR spectroscopy a sample

of an organic molecule, usually in solution in a deuterated solvent, is subjected to a magnetic field. Hydrogen atoms (but not deuterium atoms) have a specific resonant frequency when placed in a magnetic field which is directly related to their immediate chemical environment. The technique is used to generate a spectrum in which the different environments, and the relative number of hydrogen atoms in each environment, are revealed. Other atomic nuclei, including carbon, also respond to the magnetic field. This allows correlation spectra to be derived where it is possible to determine which hydrogen atoms are formally bound to which carbons and also which hydrogens have weak non-bonded, or 'through space', interactions with other atoms. Analysis of through space interactions can reveal useful information about secondary and tertiary structure, for instance aspects of protein folding or hydrogen bonding. The knowledge that different chemical environments generate different responses for hydrogen atoms can be applied to organisms as they are primarily constructed of hydrogen-rich organic compounds and water.

Although MRI operates on the same principle as NMR spectroscopy, imaging requires an additional sensitizing agent if good resolution is to be attained. As with NMR the target is subjected to a strong magnetic field and pulses of energy at in the same range as radio waves. When the pulse stops, the magnetic field oscillates and decays but at fractionally different rates depending on the local environment. The experiment detects the alignment of protons, particularly in the water molecules that make up on average 80% of body mass, but needs to differentiate between the subtly different environments in which they are found. One way to do this is to incorporate a sensitizing molecule that makes the relative size of the shifts larger. An ideal system is one that incorporates a paramagnetic metal in which each unpaired electron acts to enhance the alignment of protons in its immediate vicinity. The more unpaired electrons there are the greater the degree of sensitization so the lanthanide elements, with up to 14 electrons in their outer shells, are good candidates. Of the lanthanides, gadolinium with seven unpaired *f*-electrons, is the most commonly used metal. Unfortunately the metal alone is highly toxic so it must be delivered as a complex with an organic ligand. Again, there are many choices but most MRI contrast agents use derivatives of the macrocycle DOTA which binds the metal ion strongly whilst being small enough to be eliminated from the body once the experiment has finished.

The macrocycle 1,4,7,10-tetraazacyclododecane-1,4,7,10-tetraacetic acid, abbreviated to DOTA, is a key compound in clinical imaging particularly when coordinated to gadolinium [26]. Using three of its pendent carboxylic acids, the compound binds lanthanide metal cations extremely well: several of these have good fluorescent imaging profiles and one, gadolinium, is an excellent contrast agent for MRI applications. Gadolinium(III) has seven electrons in its *f*-shell so that each orbital is half full. As all the electrons are unpaired they affect the proton relaxation times of all hydrogen containing molecules in their proximity. In practice this means those of local water molecules which vary in concentration with the tissue type. $Gd^{3+}$ has to be bound as a complex that will not decompose under the ranges of pH and concentrations of competitive ions in the body. Fortunately this can be achieved by DOTA and the resulting complex, gadotetric acid, is a highly effective MRI

imaging agent. To make the complex more specific for particular organs and diseased tissue the fourth carboxylate group can be derivatized. It can then be linked to appropriate antibodies or sequences of amino acids known to bind to cell surface receptors that are highly abundant on those cells associated with the targeted condition. As only three acid groups bind the metal these ligands are abbreviated to DO3A derivatives. An example of this is ProHance, where a 2-hydroxypropyl group occupies the fourth pendent site, used to image the central nervous system [27].

In many cases it is sufficient to deliver a contrast agent tagged to a site-specific probe in order to image targets: the buildup of the agent in the body over time is easy to follow. However, for some uses an alternative method is required. Imaging aspects of the vasculature through MRI require that the probe be relatively large otherwise it would leak out of the system. Imaging of vasculature is important in determining any blockages or any areas of increased vascularization associated with solid tumours. Increasing the size of the complex can be achieved by increasing the mass of the probe or linker between the probe and contrast agent. The main problem with this is in finding compounds with safe metabolic pathways by which the complex can be excreted.

A second problem lies in the attachment of a single contrast agent to a relatively massive molecule: the signal may be buried in the bulk of the probe. Finally there is the question of solubility throughout the transport process from delivery to active site. One alternative is to have a contrast agent that binds to a biomolecule that is ubiquitous but concentrated at the site of interest. Human serum albumin is a common target for this approach but there is no guarantee that multiple probes will bind successfully. A completely different approach has been taken by the Bayer company in its development of blood pool agents. A lanthanide-DOTA complex was attached via a linker, chosen following a complex optimization process, to a perfluorinated alkyl chain. The compound was small enough to be easily administered but once inside the bloodstream it formed micelles through the interaction of the fluorocarbon chains. The resulting supramolecule was large enough to be stable in the vasculature and had a high density of contrast agents on its exterior, ideal for imaging. Furthermore, once the imaging was complete the micelles dissociate and are removed by normal kidney function [28].

Another method to increase the number of metals associated with each macromolecule is to attach them, in a chelated form, to a carrier molecule with many sites of attachment. An example of this can be found in the work of Meade [29]. His group has synthesized a β-cyclodextrin from which seven Gd-DOTA complexes were attached using highly efficient ‘click’ chemistry. The contrast agent was observed to accumulate in primary cancers but then localize in secondary metastases over the subsequent 48 hours. This is a particularly valuable property as one of the main problems with cancer treatment is the inability to detect these small secondary tumours in their early stages.

The range of imaging experiments that can be carried out using MRI is vast, from real time observation of brain function in response to physical or mental stimuli to the more mundane, though no less important, detection of diseased tissue in advance

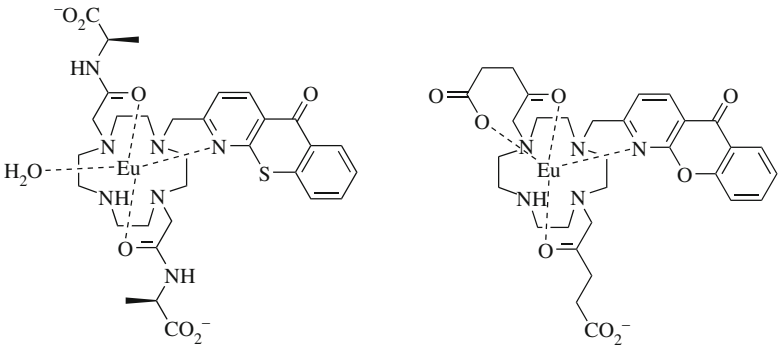


**Fig. 6.12** DOTA derivatives to detect citrate and lactate [30]

of surgery. Regardless of the final objective one factor remains essential: linking a target specific probe to an MRI contrast agent.

One excellent application comes from the Parker group which has pioneered the use of DOTA derivatives as metal binding agents for MRI applications. As a development of this work azaxanthone or azathiaxanthone groups, shown in Fig. 6.12, have been appended to the macrocycles together with two other acid or amide containing groups [30]. The central cavity holds a single europium cation that binds both to the nitrogen atoms in the main cavity and to the sidearms. In aqueous solution there is competition to bind to the europium cation between the anions present and the macrocyclic side arms. If the anions are able to displace the side arms then a change in the fluorescence is observed. Three anions have been tested: bicarbonate, citrate and lactate. Bicarbonate is commonly encountered as a naturally occurring buffering agent produced by the enzyme carbonic anhydrase. Lactate is a key indicator of liver function and tumour status. The level of citrate, central to the Krebs (citric acid) cycle, is an indicator of kidney function and, of greater diagnostic use, the progression of prostate cancer. Existing methods that test for citrate concentration in prostate or other bodily fluids rely on enzymatic bioassays such as those incorporating citrate lyase. As with many bioassays extensive sample pre-treatment is essential to remove species that may interfere with the assay and high molecular weight compounds, such as proteins, that may foul the instrumentation used. The time taken for the assay and its sensitivity were also crucial factors. Consequently the development of a new assay based on the highly selective and sensitive nature of lanthanide fluorescence was a major advance. In test solutions that simulated the balance of ions present in prostate fluids the best europium complexes were able to discriminate in favour of citrate over lactate by factors of 30:1 or 40:1. These complexes were then used to check the citrate levels of volunteers with a dual determination using a standard citrate lyase kit. Results showed a remarkable correlation over almost two orders of magnitude. Significantly the fluorescent method was far more rapid, with fewer pre-treatment steps, than the existing bioassay. More importantly the sample volume required was 25 times less.

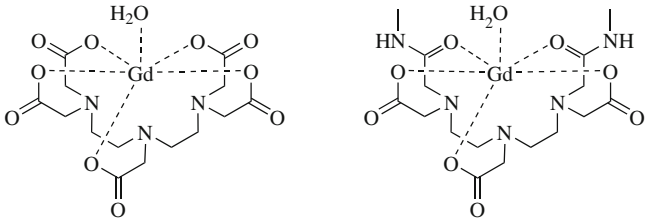


**Fig. 6.13** Omniscan® (*left*) and Magnavist® (*right*)

Other non-cyclic ligands can be used to bind $Gd^{3+}$ with most based on the diethylenetriaminepentaacetic acid (DTPA) skeleton. These include GdDTPA, or Magnevist® [31], and a di(methyl)amide derivative GdDTPA-BMA, or Omniscan® [32], illustrated in Fig. 6.13.

These non-specific contrast agents become distributed throughout the body as they accumulate in plasma and all extracellular fluids. More organ specific imaging is possible with related benzyl containing GdBOPTA complex that accumulates in hepatocytes [33]. As these are the predominant cells in the liver the agent can be used to focus on that organ. Other agents target the vasculature and have great potential in angiography performed using MRI rather than the more conventional X-ray contrast methods. This promise needs to be coupled to the recent observations that a small number of patients receiving gadolinium MRI contrast agents have suffered from acute renal toxicity having taken the agents. The toxicity appears to be related to the presence of free gadolinium. It is therefore vitally important that all gadolinium contrast agents are checked to determine their binding constants in vivo to reassure patients that the metal is always bound in the complex.

## 6.6 Other Supramolecular Sensors

The use of supramolecular diagnostics is not limited to medicine. Some of the most mundane applications can have far reaching effects. Analytical targets include explosives, bioterrorism agents such as anthrax, illegal drugs and other controlled substances. Sessler and co-workers have reported on the binding ability of calixpyrroles towards conventional explosives, including trinitrotoluene, where the binding event is signalled by a distinctive colour change [34].

One simple question that can be answered by supramolecular chemistry relates to consumer safety: is this food safe to eat? Many methods are used to ensure that produce is fresh when sold. Freezing or vacuum packing preserve food from deterioration and ‘sell by’ dates let both buyer and seller know when the food should be consumed but none of these precautions actually checks the food directly. A group led by Lavigne has tackled this problem directly by firstly determining which compounds are most closely associated with food spoilage and then by designing a colorimetric test that responds to elevated levels of these compounds [35].

In most meats an obvious indicator that the food is no longer safe to eat is the smell of putrefaction. This results from bacterial toxin build-up and the breakdown of amino acids which leads in turn to an increase in production of volatile biogenic amines such as histamine, spermine, putrescene and cadaverene. Amines, especially polyamines, have a good affinity for carboxylic acids so the Lavigne group developed a thiophene-based polymer with carboxylic acid substituents that bind to amines and signals the binding event with a colour change.

In a fortunate coincidence the polymer can be attached to cheap plastic dipsticks and used directly by placing against the food. The volatile amines rise up the sticks, in a similar manner to thin layer chromatographic plates, changing colour as they go. The extent of the colour change can then be used directly as a measure of the target analyte. Variation in the polymer generates subtly different colours and allows pattern recognition methods to elicit a finely tuned array of responses that monitor several different decomposition products. Experimental detection of these compounds for different meats over time, by a technique such as gas chromatography, yields a series of compound ratios that should be broadly the same regardless of the meat's origin. Matching the sensor output to an experimental trace determines if the levels of biogenic amines are safe or if the meat poses a health risk.

## 6.7 Summary

Many biomedical diagnoses are based on assays in which dyes bind to features specific to particular cell types. Supramolecular diagnostic methods differ by applying the receptor-spacer-reporter design paradigm to target analytes with high specificity. To do so the key step is to identify, or synthesize, the receptor so that it can bind through a combination of weak interactions associated with supramolecular systems. The binding event leads to signal transduction and a measurable output. Depending on the transduction mechanism the response may be optical, fluorescent or electrochemical. The range of targets is limited only by the recognition element and extends to optical detection of critical analytes in blood, toxins, bacterial infections, small intracellular signalling molecules, and even fish past their 'sell by' date.

Supramolecular chemistry is now at the forefront of clinical practice. By linking lanthanide metal-macrocyclic complexes to tissue selective receptors the use of magnetic resonance imaging has become far more useful and widespread. This technique, which relies on the receptor-spacer-reporter motif, has become one of the most powerful medical diagnostics tools available in hospitals worldwide.

## References

1. Beer PD et al (1990) Redox-responsive crown ethers containing a conjugated link between the ferrocene moiety and a benzo crown ether. Inorg Chem 29:378–381
2. Leevy WM et al (2008) Quantun dot probes for bacteria distinguish *Escherichia coli* mutants and permit *in vivo* imaging. Chem Commun 2331–2333

3. Magri DC et al (2006) Communicating chemical congregation: A molecular AND logic gate with three chemical inputs as a “lab-on-a-molecule” prototype. J Am Chem Soc 128: 4950–4951
4. Tusa JK, He H (2005) Critical care analyzer with fluorescent optical chemosensors for blood analytes. J Mat Chem 15:2640–2647
5. He H et al (2003) A fluorescent chemosensor for sodium based on photoinduced electron transfer. Anal Chem 75:549–555
6. Kumar A et al (1988) Chromogenic ionophore-based methods for spectrophotometric assay of sodium and potassium in serum and plasma. Clin Chem 34:1709–1712
7. Piatek AM et al (2004) Threshold detection using indicator-displacement assays: An application in the analysis of malate in Pinot Noir grapes. J Am Chem Soc 126:6072–6077
8. Park JD, Cohen S, Lacher JR (1962) Hydrolysis reactions of halogenated cyclobutene ethers: synthesis of diketocyclobutenediol. J Am Chem Soc 84:2919–2922
9. Hewage HS, Anslyn EV (2009) Pattern-based recognition of thiols and metals using a single squaraine indicator. J Am Chem Soc 131:13099–13106
10. Oguz U, Akkaya EU (1998) One-pot synthesis of squaraine fluoroionophores. J Org Chem 63:6059–6060
11. Ajayaghosh et al (2008) Controlled self-assembly of squaraines to 1D supramolecular architectures with high molar absorptivity. Chem Commun 969–971
12. Johnson JR et al (2007) Squaraine rotaxanes: superior substitutes for Cy-5 in molecular probes for near infra-red fluorescence cell imaging. Angew Chem Int Ed Engl 46: 5528–5531
13. Gassensmith JJ et al (2007) Self-assembly of fluorescent inclusion complexes in competitive media including the interior of living cells. J Am Chem Soc 129: 15054–15059
14. Beer PD et al (1992) Acyclic redox responsive anion receptors containing amide linked cobalticinium moieties. J Chem Soc Chem Commun 270–272
15. Leung E et al (1996) A novel in vivo nitric oxide sensor. Chem Commun 23–24
16. Fucassi F et al (2001) Characterisation of small molecule binding to DNA using a quartz crystal resonant sensor. Chem Commun 841–842
17. Furchgott RF, Zawadzki JV (1980). The obligatory role of the endothelium in the relaxation of arterial smooth muscle by acetylcholine. Nature 288:373–376
18. Palmer RMJ, Ferrige AG, Moncada, S (1987) Nitric-oxide release accounts for the biological-activity of endothelium-derived relaxing factor. Nature 327:524–526
19. Schultz RA et al (1985) 12-, 15-, and 18-Membered-ring nitrogen-pivot lariat ethers: synthesis, properties, and sodium and ammonium cation binding properties. J Am Chem Soc 107:6659–6668
20. Gunnlaugsson T et al (2001) A novel optically based chemosensor for the detection of blood $Na^+$. Tetrahedron Lett 42:4725–4728
21. de Silva AP et al (1996) Fluorescent signalling of the brain neurotransmitter γ-aminobutyric acid and related amino acid zwitterions. Chem Commun 2191–2192
22. Cooper CR, James TD (1997) Selective D-glucosamine hydrochloride fluorescence signalling based on ammonium cation and diol recognition. Chem Commun 1419–1420
23. Kim SH et al (2006) $Hg^{2+}$-selective fluoroionophoric behavior of pyrene appended diazatetrathia-crown ether. Tetrahedron Lett 47:497–500
24. Kele P et al (2002) Coumaryl crown ether based chemosensors: selective detection of saxitoxin in the presence of sodium and potassium ions. Tetrahedron Lett 43:4413–4426
25. Harada K, Kondo F, Lawton L (1999) Laboratory analysis of cyanotoxins. In: Chorus I, Barthram J (eds) Toxic cyanobacteria in water: a guide to their public health consequences, monitoring and management. World Health Organization: New York, NY, pp 368–405
26. Hermann P et al (2008) Gadolinium(III) complexes as MRI contrast agents: ligand design and properties of the complexes. Dalton Trans 3027–3047
27. Tweedle MF (1997) The ProHance story: the making of a novel MRI contrast agent. Eur Radiol 7:S225–S230

28. Platzek et al (2005) Use of perfluoroalkyl-containing metal complexes as contrast media in MR-imaging for visualization of plaque, tumors and necroses US patent 2005/0074409 A1
29. Song Y, Kohlmeir EK, Meade TJ (2008) Synthesis of multimeric MR contrast agents for cellular imaging. J Am Chem Soc 130:6662–6663
30. Pal R, Parker D, Costello LC (2009) A europium luminescence assay of lactate and citrate in biological fluids. Org Biomol Chem 7:1525–1528
31. Relix R, Felix R, Heshiki A (1997) Magnevist. Blackwell Scientific: London
32. Sze G et al (1993) Use of the magnetic-resonance contrast agent gadodiamide in the central-nervous-system – results of a multicenter trial. Invest Radiol 28:S49–S55
33. Kirchin MA, Pirovano GP, Spinazzi A (1998) Gadobenate dimeglumine (Gd-BOPTA): an overview. Invest Radiol 33:798–809
34. Park JS et al (2010) Positive homotropic allosteric receptors for neutral guests: annulated tetrathiafulvalene-calix[4]pyrroles as colorimetric chemosensors for nitroaromatic explosives. Chemistry 16:848–854
35. Maynor MS et al (2007) A food freshness sensor using the multistate response from analyte-induced aggregation of a cross-reactive poly(thiophene). Org Lett 17:3217–3220

# Chapter 7
# Supramolecular Therapeutics

## 7.1 Therapeutic Applications of Supramolecular Chemistry

In keeping with the basic tenet of supramolecular chemistry, that it is 'chemistry beyond the molecule', any therapy that claims to work through a supramolecular mechanism must involve multiple target specific interactions. Furthermore the nature of the interactions should be non-covalent.

As with diagnostic supramolecules, those designed for therapy are often multifunctional. Just as imaging agents have often been based on the 'receptor-spacer-reporter' paradigm so therapeutic agents often link a specific probe to a highly potent chemical agent. Medicinal chemists have been using this concept for many decades through the 'prodrug' and 'magic bullet' approaches.

Prodrugs are molecules that have been formulated so that one region is therapeutic and another has been designed to transport the drug to its site of action. In this form the drug characteristics are often masked to give the prodrug low general toxicity. Once the target is reached a particular quality of the site, usually an intrinsic metabolic process, activates the drug by cleaving it from the remainder of the molecule. This allows a prodrug to be inactive when ingested but to activate when cleaved by particular digestive enzymes. Examples of this approach include the analgesic aluminium bis(acetylsalicylate), illustrated in Fig. 7.1, which hydrolyses in the stomach to release two aspirin molecules and the antacid aluminium hydroxide

The idea of a medicinal magic bullet was first described by Ehrlich in the 19th century [1]. He theorized that if a drug could be made to be highly toxic against a particular bacterium, in his case it was an arsenic compound with specificity for *Treponema pallidum*, then it would represent a therapy that was both highly potent and highly selective. The main benefit of this is that much smaller doses of drugs would need to be administered because they would not be distributed throughout the body. As a consequence the toxicity of the drug could be extremely high as it would not lead to systemic poisoning. The main drawback with the magic bullet route is that no potential candidate has shown perfect specificity. Despite this, the search for magic bullets continues with many examples utilizing interactions that can easily be identified as supramolecular.

P.J. Cragg, *Supramolecular Chemistry*, DOI 10.1007/978-90-481-2582-1_7,

© Springer Science+Business Media B.V. 2010

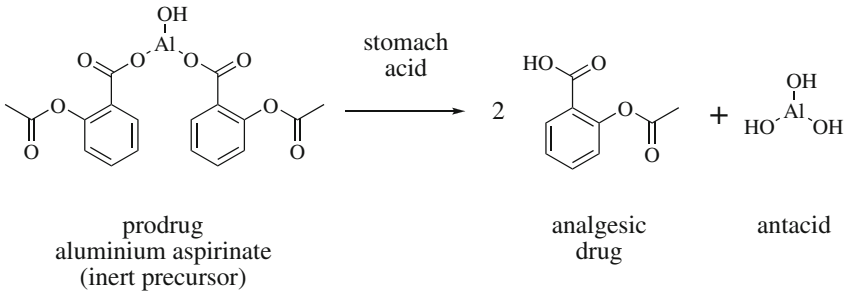



**Fig. 7.1** The prodrug principle

Of all the examples known to date only one can be considered to have approached success in terms that Ehrlich would understand: monoclonal antibody therapy. Monoclonal antibodies are raised against known antigens and could therefore, in a future involving personalized medicine, be tailored to individuals and their therapeutic needs. Ideally both antigens and antibodies would come from the patient but in practice they are harvested from mouse, rabbit, sheep or chimeric mouse-human hybrid cells. Thus the specificity to the patient is never ideal although this approach has been used to generate clinically useful compounds. The antibodies interact with their target antigens through the numerous weak forces associated with protein-protein interactions and can therefore act directly to block particular receptors or otherwise interfere with biochemical pathways. A more sophisticated use for the antibodies is to link them to a drug molecule and deliver that drug to its desired target with high specificity thus fulfilling the magic bullet promise.

## 7.2 Chelation Therapy

Designing receptor molecules to target particular chemical species is at the core of supramolecular chemistry. Often the target is a small molecule, however, the ability to bind transition metal ions is also important both in the industrial extraction and isolation of commercially valuable metals and, of relevance here, in chelation therapy. This branch of medicine exists to correct imbalances in homeostatic levels of essential trace metals. The body ingests varying amounts of transition metals and many of these are essential components of enzymes and other biomolecules. While surplus metals are usually excreted, excessive amounts of certain metals can build up to toxic levels. The reasons for this may be through acute ingestion of a large amount of the metal or through conditions that affect the body's ability to transport particular elements. The reverse also occurs: the body may have difficulty absorbing essential metals and so have reduced levels of certain enzymes. This in turn affects biochemical pathways that rely on those enzymes. To treat both cases a molecule is required that can bind the target metal and transport it effectively. Examples of chelators are shown in Fig. 7.2. Where an excess of the metal exists

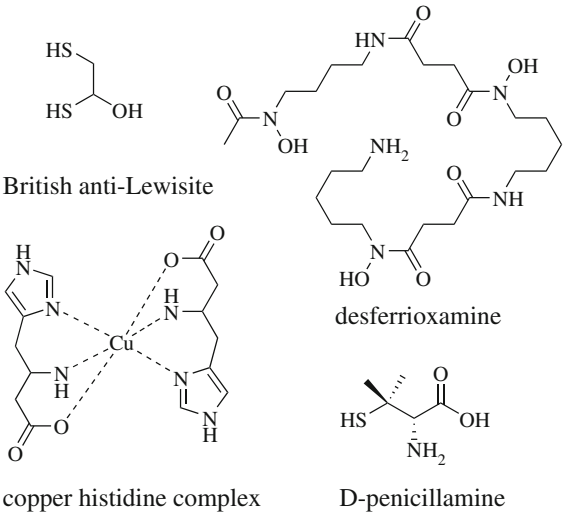



**Fig. 7.2** Therapeutic chelating agents

the chelating molecule needs to seek it out in vivo and form a stable complex that can be excreted. Where the metal needs to be smuggled into the particular cells the complex has to be stable enough to deliver the metal but to allow decomplexation at the appropriate time and leave a non-toxic ligand that can be safely metabolized. Preparation of ligands for these purposes builds on the wealth of knowledge that has arisen from decades of research on transition metal complexation. The size selectivity of transition metal ions, related to charge density, together with geometric preferences and donor group affinities can all be used to design highly specific complexing agents. Once synthesized the stability of the complex can be determined under varying conditions of pH, temperature and interference from competitor ions.

The origins of chelation therapy can be traced back to the treatment of First World War soldiers who had suffered from gas attacks that used the arsenic-based toxin, Lewisite. A dithiol, British anti-Lewisite (BAL), was developed to remove the toxic metal.

During the Second World War there was a need to treat workers who had been exposed to lead in the paint, particularly 'white lead' or lead(II) carbonate, used on military vehicles and ships. This was achieved with the well known chelating agent ethylenediamine tetraacetic acid (EDTA). Since then other chelating agents have been identified or synthesized for the purpose of binding specific metals.

## *7.2.1 Desferrioxamine*

Desferrioxamine B is a siderophore from the *Streptomyces pilosus* bacterium which binds iron(III) [2]. As discussed in Chapter 5 the evolution of siderophores, literally 'iron-carrying' molecules, is as a consequence of organisms' reliance on iron in

many enzymes and metal containing proteins. In an oxidizing environment, such as air or oxygenated water, iron is readily oxidized from the water soluble iron(II) to the sparingly soluble iron(III) species. Bacteria, fungi and some grasses accumulate iron by releasing small iron(III)-specific ligands that seek out iron(III) in solution and bind it with high selectivity. The siderophores bind high spin $Fe^{3+}$ in an octahedral environment and use N- or O-incorporating ligands as these have higher affinities for $Fe^{3+}$ than $Fe^{2+}$. Binding motifs may be cyclic with converging metal binding sites as seen in enterobactin or may be flexible linear molecules that incorporate metal binding sites, found in desferrioxamine B.

Two binding sites are commonly found: catecholate, as in enterobactin, and hydroxamate, the motif in desferrioxamine B. The resulting complex is targeted by a membrane-bound receptor and captured by the organism. The complex is transported across the cell membrane where the iron is reduced to iron(II), which has a lower affinity for the siderophore, and subsequently decomplexed.

The high affinity for oxidized iron makes the siderophores ideal candidates for chelation therapy where the body is becoming overwhelmed by iron(III) either through acute poisoning or conditions like haemochromatosis that can occur when patients receive frequent blood transfusions. While enterobactin would seem to be the primary choice it has two major drawbacks: its synthesis is complicated and, although both isomers bind iron(III) to the same extent, only the L-isomer has activity in vivo. Consequently desferrioxamine B is the agent of choice.

### *7.2.2 Copper Imbalance: Wilson's Disease and Menke's Syndrome*

In vivo tolerance to copper is quite high, however, deficiency and excess are serious problems. Infants are particularly vulnerable as they take time to assimilate the correct levels and it is known that trace copper from cooking utensils or water pipes can cause childhood cirrhosis. Copper deficiency leads to arterial weakness and heart enlargement. This is probably caused by a decrease in catecholamine neurotransmitters derived from the biosynthesis of adrenaline which requires the copper-containing enzymes phenylalanine hydroxylase, dopamine β-monooxygenase and tyrosinase.

Wilson's disease is a pathological accumulation of copper in tissue which is later released into the bloodstream, leading to anaemia, and final accumulation of copper in liver and brain. It is the result of a mutation in the Wilson's disease gene in chromosome 13 which ordinarily codes for a cation transporting ATPase so that copper can be incorporated into ceruloplasmin prior to excretion. Also known as ferroxidase, in acknowledgement of its primary function as an oxidoreductase responsible for electron transfer, this enzyme contains iron and, more importantly, six copper atoms. It accounts for the transport of 90% of copper in the plasma so any impairment in its production or efficacy has a major impact on copper homeostasis. The greatly reduced concentration of ceruloplasmin in the blood of Wilson's disease sufferers correlates with their inability to metabolize copper effectively. It leads to chronic liver disease, for which the only real cure is a liver transplant,

but the condition can be treated with copper specific chelating agents such as D-penicillamine or triethylenetetramine which, when administered as its salt, is known as trientine hydrochloride [3].

These agents bind the copper so that it is solublized and excreted in the urine. They are effective because of their high affinity for copper(II) over other metals in the body. Although both isomers of penicillamine bind copper equally well it has been found that the L-form is toxic. Interestingly D-penicillamine is also used to treat rheumatoid arthritis where it acts to reduce collagen crosslinking: the enzymes responsible for this process are largely copper centred.

Menke's syndrome is a hereditary dysfunction, through a defect localized on the x-chromosome, of intracellular copper transport leading to insufficient copper storage which leads to lethal degradation of the central nervous system. The opposite effects to those of Wilson's disease are observed: low concentrations of copper in brain but high concentrations in the kidneys and intestines indicating poor uptake and high levels of excretion. The low levels of copper affect numerous copper containing enzymes notably those involved in neurotransmitter production such as dopamine β-monooxygenase It can be treated with copper acetate injections and administration of copper histidine complexes.

## 7.3 Macrocyclic Complexes for Radiotherapy

Ligands designed to bind metals as imaging agents can also be used to deliver metals capable of cytotoxic β-emission as the preferred coordination environments of both groups of metals are very similar. The yttrium isotope $^{90}$Y has a half life of 64 h and decays through β-emission. It can be bound by macrocyclic chelating agents such as the DOTA derivative DOTA-tyr3-octreotide which in turn binds to murine monoclonal antibodies raised against cell surface receptors on specific malignant cancer cells [4]. Once the metal-antibody complex has accumulated in the tumour the cells are destroyed by β-emission before the body's natural immune response to the foreign murine antibodies speeds their removal. As illustrated in Fig. 7.3 and discussed below, other complexes with multiple functionality can also include a radiotherapeutic aspect.

## 7.4 Photodynamic Therapy

Porphyria, the disease from which Britain's King George III is believed to have suffered, arises through the accumulation of porphyrin decomposition products in the skin due to impaired enzyme function in the haem biosynthetic pathway [5]. In addition to many other unpleasant side effects, porphyria renders the individual highly sensitive to light. The effects of porphyria, if they could be controlled and directed towards particular diseased tissue, would have the potential as a powerful therapeutic method. As the mechanism involves local, light-initiated generation of

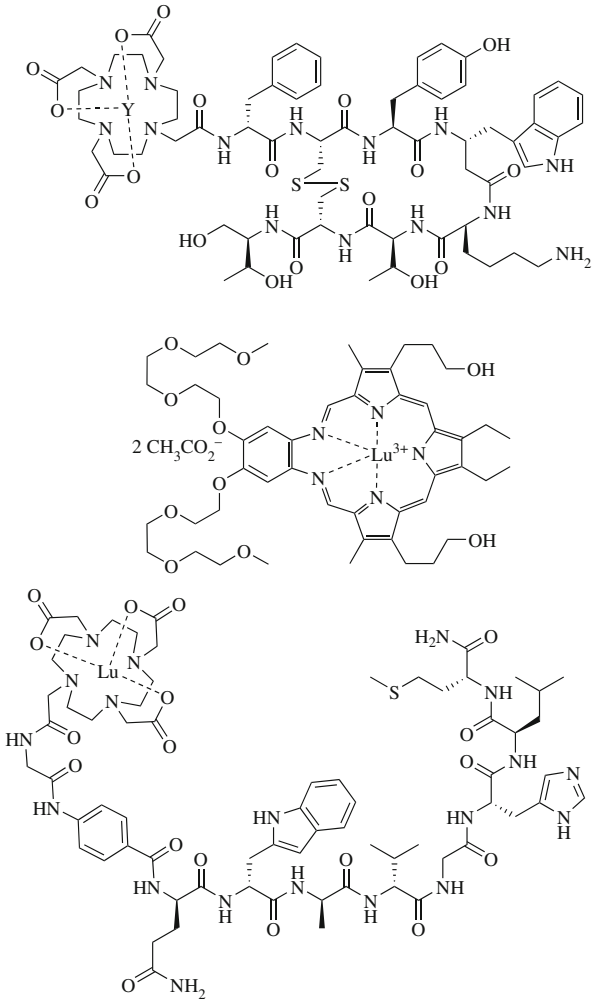



**Fig. 7.3** Radiotherapeutics: (from *top*) DOTA—tyr3-octreotide, LUTRIN® and Lu-177-AMDA

toxic radicals the effects of porphyria can be replicated by simple model compounds with similar chemical structures to natural porphyrins. The success of this research is evident in the clinical application of photodynamic therapy (PDT).

PDT originates in the work of Finsen at the end of the 19th century [6] but was only successfully demonstrated in 1961 using haematoporphyrin [7]. Other porphyrin derivatives, and compounds which share a similar structural motif, have since been shown to act as photosensitizing agents [8]. In PDT metal-free macrocycles are used to target tumour cells, for which they are quite selective, then, with the compounds in situ, the affected area, often the skin, is irradiated with red light. As shown in Fig. 7.4 the molecules are energized from the ground state to an excited

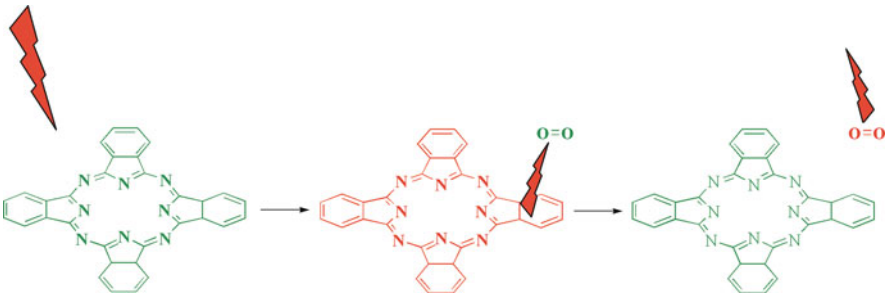



**Fig. 7.4** Generation of therapeutic $^1O_2$ for photodynamic therapy

state which then undergoes intersystem crossing to and eventual energy release. The sensitizer returns to the ground state by reacting with dioxygen. In doing so the highly reactive singlet oxygen species is formed which returns to the triplet ground state through reaction with nearby molecules. As a result a cascade of cytotoxic free radicals is formed that damage local biomolecules, including DNA, leading to destruction of the targeted cell. It is possible to use this therapy in a highly accurate manner through the use of a laser light source. Unfortunately the macrocycles are not magic bullets for cancer cells and circulate throughout the body until destroyed by natural processes and excreted. The patient therefore remains highly sensitive to light, including sunlight, for some time after therapy. In the search for alternatives to porphyrin-based PDT one group of macrocycles has shown a great deal of promise: the phthalocyanines.

The early history of phthalocyanines is a quite bizarre and tragic. They are easily synthesized by a transition metal templated cyclocondensation of phthalonitrile, however, this was not how they were discovered. It is probable that a metal-free phthalocyanine was successfully prepared by Braun and Tcherniac in the early 1900s though it was not characterized [9]. Given their synthetic methods it is also probable that de Diesbach and von der Weid were the first to prepare copper phthalocyanine but, again, their 1927 paper gives no characterization details [10].

In the 1920s the Scottish Dyes Ltd Grangemouth works was producing phthalimide from phthalic anhydride and ammonia in iron vats. Dandridge, an analytical chemist, noticed that a blue deposit had formed around the vats and blamed it on contamination from a neighbouring reactor. Grieg from the analytical department showed this not to be so and, in further analysis on metal complexes prepared by Drescher, determined the empirical formula. During these tests he noticed that the compounds contained iron that, significantly, could not be removed by hydrochloric acid extraction. The iron had presumably come from through cracks in the reaction vessel's enamel coating. Dandridge, Drescher, Thomas and Scottish Dyes Ltd filed a patent in 1928 detailing the synthesis of iron phthalocyanine and covering the general process [11]. The compound became a valuable addition to the range of synthetic dyestuffs but it was soon discovered that the copper complex made much more stable pigment and it has since become the basis of many colorants. Drescher

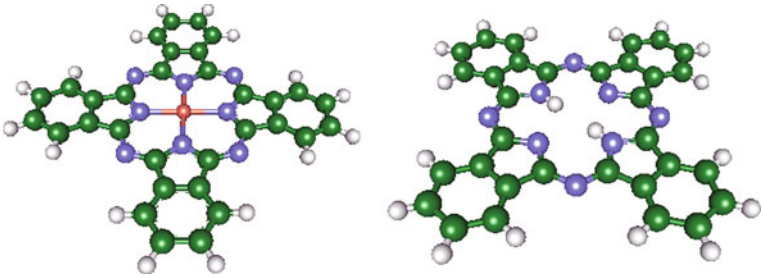


**Fig. 7.5** Phthalocyanine crystal structures: a copper complex [12] (*left*) and the metal free macrocycle [13] (*right*)

unfortunately did not enjoy his success for much longer: he died when his Sunbeam motorcycle collided with a steamroller on his way to work.

The insolubilities of phthalocyanines made their analysis difficult and it took some time before a satisfactory structure was elucidated. Initial work was undertaken by the Linstead group at Imperial College in the 1930s that culminated in a series of six back to back papers published in 1934 [14]. It was also Linstead who named the compounds in recognition of their synthesis from phthalic anhydride and similarity to the blue cyanine dyes. Definitive characterization of the nickel, platinum and copper phthalocyanine complexes, together with the metal-free compound, was revealed in 1935 following the publication of their X-ray structures by Robertson [15]; the copper and metal-free compounds are illustrated in Fig. 7.5.

## 7.5 Texaphyrins

The search for innovative pharmaceuticals can come about for a variety of reasons; in response to global crises such as the treatment of HIV/AIDS, in anticipation of catastrophic pandemics such as that posed by new strains of influenza, or from companies wishing to find ways around patented processes that generate best selling medicines. Sometimes the research is driven by groups and individuals with a keen personal interest in finding a cure for a particular disease as exemplified by the targeted research funding by charities. Rarely is a single individual responsible for the research and development of a novel treatment yet this is exactly how texaphyrins originated. It is perhaps the most inspiring story there is to tell in contemporary chemistry.

While completing his undergraduate degree at the University of California, Berkeley, a chemistry major was diagnosed with Hodgkin's lymphoma. After radiotherapy he went into remission, completed his degree and moved on to undertake postgraduate research. The lymphoma reappeared but was successfully treated, this time by chemotherapy. The researcher, Jonathan Sessler, wondered if there was a better drug to treat the lymphoma to which his oncologist responded that, as a

chemist, he should make one. To anyone else this may seem like a joke in very poor taste but to this particular researcher it became a challenge.

When Sessler joined the faculty at the University of Texas at Austin he initially followed up his doctoral work on the chemistry of porphyrins but was soon struck by how much bigger everything seems to be in Texas. This led to the groundbreaking insight that if the porphyrin structure could be enlarged the 'expanded porphyrins' would be able to bind metals larger than iron. Theoretically the cavity would be a good match for lanthanide metals including gadolinium, widely used in MRI contrast agents. In due course the Sessler group synthesized a porphyrin derivative in which one pyrrole group had been replaced by 1,2-diaminobenzene [16]. Incorporation of this group resulted in the formation of a diimine to give a molecule with five donor atoms and an expanded metal-binding cavity. After their success it was only fitting that the group named the compound. The only appropriate choice was 'Texaphyrin' because, of course, everything is bigger in Texas, even the molecules. What was more, the five nitrogen atoms that define the molecule's central cavity describe the apices of a five-pointed star which coincidently is the emblem of the Lone Star state. Indeed, the Texas sized, Lone Star shaped molecule was proposed as the 'State Molecule of Texas', only to be beaten by $C_{60}$ in recognition that two of its co-discoverers worked at the state's Rice University.

As shown in Fig. 7.6, texaphyrins have a larger cavity than porphyrins so they can form complexes with lanthanide metals such as gadolinium (XCYTRIN®), that enhances the efficacy of treatment for certain brain tumours, and lutetium (LUTRIN®), used as a sensitizer for photodynamic therapy of recurrent breast cancer [17]. Crucial to their success is the increased number of donor atoms available as the more lanthanide binding sites that a ligand can satisfy, the more stable the complex.

There are two main modes of action depending on whether imaging or cell destruction is the target. For imaging purposes the potential for high coordination complexes is central. The metal-texaphyrin complex is completed by a ligating water molecule which improves MRI relaxivity and improves image resolution. The therapeutic mechanism is typical of porphyrin-type systems. Thiol containing proteins such as glutathione are oxidized by the metal-texaphyrin species which in turn is reduced to the radical. The texaphyrin radical reacts with molecular oxygen to generate DNA-damaging superoxide which directly affects the tumour. As oxygen is found throughout the body this would appear to be an indiscriminate method,

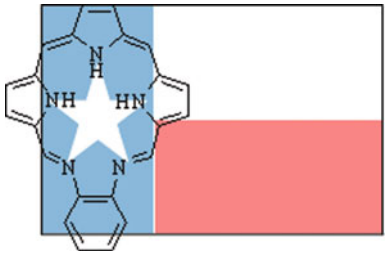



**Fig. 7.6** The Texas-sized texaphyrin

however, a secondary effect of the metal-texaphyrin complex appears to involve upregulation of hypoxia (increased levels of oxygenation) in tumours. The molecule therefore has two complementary effects that sensitize and target tumours.

The jury is still out on the clinical efficacy of texaphyrin-based therapeutics: the commercialization of novel pharmaceutics requires careful negotiation through the minefield of clinical trials and the subsequent analysis of the data. Nevertheless this example should serve as an inspiration to chemists everywhere.

## 7.6 Targeting Cancer with Peptides

In the treatment of cancer the magic bullet approach may have finally come of age. Cancer cells often have highly specific protein receptor sites on their surfaces that can be targeted by prodrugs composed of a protein and therapeutic component. One such approach takes the protein bombesin, originally isolated from frog skins, and couples it to the familiar DOTA ligand. Bombesin has been chosen because the receptor for its tetradecapeptide sequence, pyr-GRLGNGWAVGHLM-$NH_2$, has been found to be overexpressed in many common cancer cells such as those associated with lung, prostate and breast cancer [18]. The therapeutic agent Lu-177-AMDA adds the site specificity of bombesin to the destructive power of 177m-lutecium DOTA. The resulting complex can be used in radiotherapy through the decay of the metal by β-emission to 177-hafnium. The success of this approach is related not only to the magic bullet concept but also because the complex is broken down in vivo to small molecules that are successfully excreted by the body [19]. Other versions of this system incorporate 68-gallium for PET imaging. Two issues must be considered when designing this, or any, new imaging agent. Firstly, the technology to image the metal must be in place and, secondly, the drug conjugate must be in a suitable dosage form. In this example the radiotracer emission is in the same range as existing detectors for 99m-technetium decay so the same detectors can be used. Here, though, a stable formulation is still under investigation because the radiochemical decay affects the peptide stability.

## 7.7 Drug Delivery and Controlled Release

Dendimers, from the Greek ‘δενδρον’ for tree, are polymers that branch from a central point. Early synthesis, pioneered by the Vögtle group [20], was done in a stepwise manner. Subsequent work by the groups of Tomalia [21], on ‘starburst’ dedrimers, and Fréchet [22], on convergent growth methods, led to the current cascade method of synthesis. As shown in Fig. 7.7 the dendrimer may start from a simple organic compound with divergent reactive sites or a macrocycle such as a porphyrin. This core molecule is known as ‘generation 0’ of the dendrimer which can be treated with an excess of a bifunctional reagent that, in turn, grows another layer of reactive branches from each point of attachment. Once all the sites of the generation 0 compound have reacted remaining reagents are washed off to leave a

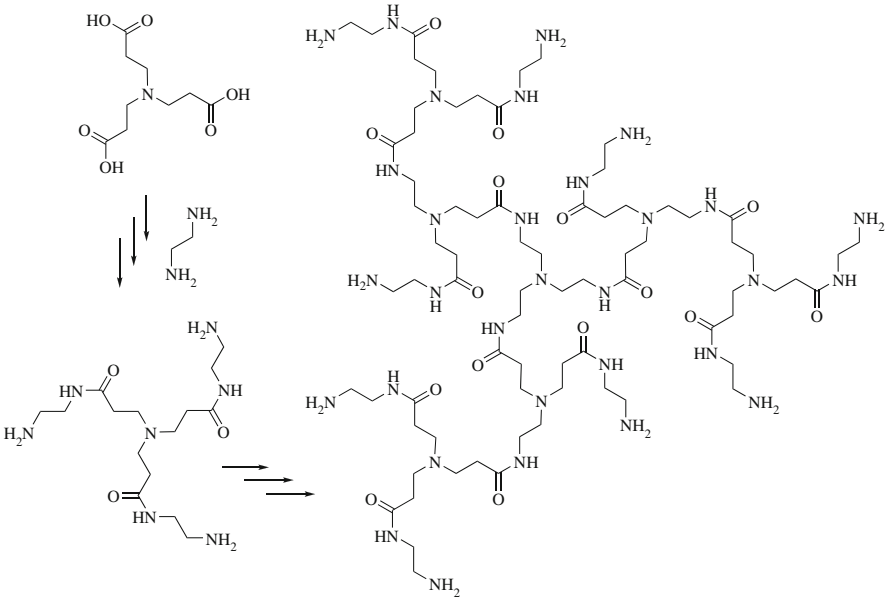


**Fig. 7.7** Synthesis of a dendrimer

'first generation' dendrimer. If the groups added to form this first generation dendrimer have multiple functionalized termini then the next generation will have even more branches. Later generations result from repeating this process leading to a rapid increase in the dendrimer size. Each generation may use the same branch or, by analogy to co-block polymer synthesis, different compounds may be used to vary the dendrimer properties at each stage.

Tuning the properties of each generation to complement those of particular guest molecules makes it possible to bind those guests and then let them dissociate at a known rate. This 'controlled release' property makes dendrimers potential candidates as drug delivery agents as illustrated in Fig. 7.8 [23].

Perhaps of greater interest is the incorporation of therapeutic compounds, such as the anticancer drugs cisplatin [24] and doxorubicin [25], into the growing dendrimer which undergo triggered release in response to their environment. In essence these are dendrimeric prodrugs capable of accurately delivering multiple doses. The only concern is in the biological fate of the remaining dendrimer which may not be excreted through the usual pathways due to its size and physical properties [26].

## 7.8 Cyclams as Anti-HIV Agents

Cyclam, 1,4,8,11-tetraazacyclotetradecane, is a cyclic amine readily synthesized from a linear tetraamine precursor, 1,5,8,12-tetraazaduodecane, which is preorganized by a metal template that promotes square planar complex formation, such as nickel, followed by the addition of ethanedione. Reduction of the Schiff base by a

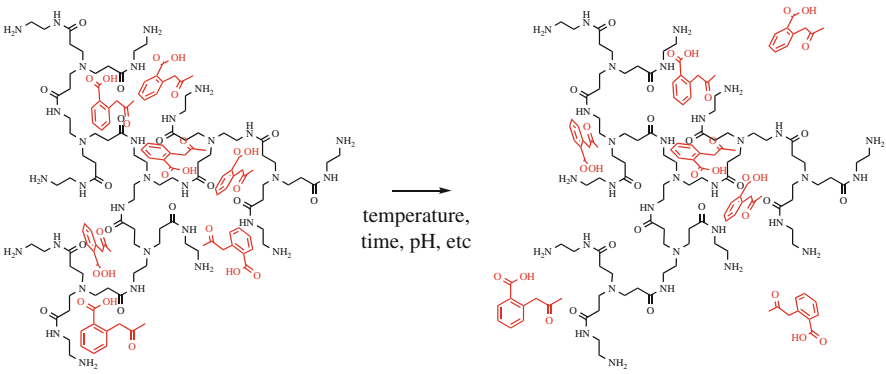



**Fig. 7.8** Controlled release of therapeutic drugs by a dendrimer

metal borohydride and removal of the metal by precipitation with a more strongly binding ligand, such as cyanide, results in the isolation of the metal-free cyclam. The macrocycle can be used to bind other metals in its amine form or further derivatized through the nitrogens. The derivatives about which there has been considerable interest are bridged bis(cyclam)s, illustrated in Fig. 7.9.

These compounds have been demonstrated to be potent anti-HIV-1 and HIV-2 agents and it is presumed that they target the uncoating process that occurs early in the virus' replication cycle [27]. The $IC_{50}$ values (the concentrations which inhibit 50% of cytopathicity of the virus) for derivatives linked through two of the carbon atoms, or a propyl bridge between two nitrogens, were determined to be less than one micromolar. Not only were the compounds potent, they were also highly selective for HIV-1. Several phases are involved in viral infection by HIV-1 and it transpired that these compounds target the initial fusion event between virus and cell surface. Later work identified that when two cyclams were bridged by a 1,4-xylyl spacer between two nitrogen atoms the potency increased further and was improved when administered as the zinc complex [28]. One zinc containing cyclam, in which the ligand adopts a *cis* configuration, simultaneously binds aspartate and glutamate carboxylate groups in the CXCR-4 co-receptor, while a second cyclam adopts a *trans* conformation which can only bind a single anion, in this case an aspartate group in another region of the protein. In the HIV-1 infection cycle a viral glycoprotein recognizes the receptor proteins CD4 and CCR5 or CXCR-4, a

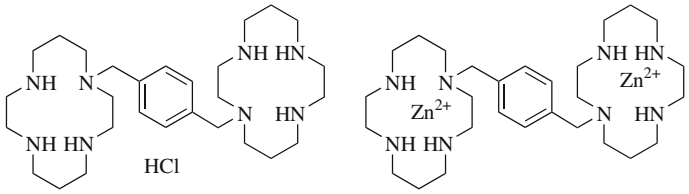



**Fig. 7.9** Bis(cyclam)s with anti-HIV activity

G-coupled protein, on the surface of white blood cells. After this interaction has commenced it sets off a conformational change that reveals another binding site on the virus to strengthen the virus-cell affinity. Once the virus-membrane interaction is cemented in this way it paves the way for the virus to transfer its genetic material into the host cell.

Although xylyl linked bis(cyclam)s would appear to have a future as anti-HIV agents the history of the lead candidate tells a different story. It was developed as the octahydrochloride hydrate to mobilize stem cells, known as plerixafor or Mobozil$^{TM}$, but became associated with cardiac disturbances in early trials [29]. The compound is also a highly selective receptor antagonist for CXCR-4 and acts as a non-peptide cytokine. Cytokines are small proteins secreted by cells that initiate cell movement in response to chemical signals, what is termed chemotaxis, and may be released in response to infection or to maintain the normal growth of the organism. The use of small synthetic molecules in this regard is quite novel and plerixafor may have a clinical role to play in the regulation of stem cell migration. Research has shown that it is able to mobilize both haematopoietic stem cells, responsible for blood cell formation, and progenitor cells. It has been used therapeutically alongside stem cells that have been implanted following surgery where large amounts of tissue have been removed. One key finding has been that CXCR-4 regulates both primary breast cancer and associated metastatic tumours which suggests that bis(cyclam)s could be a powerful weapon against breast cancer acting both to slow the rate of growth and repair damaged tissue. The compound also seems to slow down the spread of cancerous ovarian cells.

## 7.9 A Supramolecular Solution to Alzheimer's Disease?

Alzheimer's disease, named after the psychiatrist who first described the condition in the early 20th century, is a progressive neurodegenerative form of dementia. It is most prevalent in the aging population and is associated with loss of memory and cognitive functions. The causes of Alzheimer's disease are not well understood and may be a combination of environmental and genetic risk factors. A genetic link to chromosome 21 is supported by data on individuals who have an extra copy of the chromosome which indicates a greatly increased likelihood of developing the condition in late adulthood. Other theories point to a reduction in the production of acetylcholine, an essential neurotransmitter. This is observed in other degenerative conditions, notably the motor neurone disease amyotrophic lateral sclerosis, where it has been linked to impairment of superoxide dismutase (SOD) activity. The form of SOD affected contains zinc and copper and loses function without copper. Copper deficiency can lead to the reduction of neurotransmitters through other routes as it is central to the biosynthesis of adrenaline which requires the copper-containing enzymes phenylalanine hydroxylase, dopamine β-monooxygenase and tyrosinase.

The consensus opinion is that, regardless of the causes of Alzheimer's disease, amyloid-β proteins are deposited in the brain and, together with tau proteins, form

entangled structures called β-amyloid plaques. Plaque formation is initiated by small protein fragments in the region of 40 amino acids long and requires metal ions to induce the folds that form the plaque's tertiary structure. A major focus has been the aggregation of the amyloid-β proteins. These occur in different isoforms and it seems that while the $A\beta_{40}$ form is relatively benign, the slightly longer $A\beta_{42}$ form is responsible for initiating plaque formation. The different lengths of the proteins reflect the points at which the enzymes β- and γ-secretase cleave APP, the amyloid precursor protein. Mass spectrometric studies indicate that $A\beta_{42}$ forms a linear tetramer then a hexamer before doubling up as a dodecamer. One particular amino acid sequence, NNQQNY, appears to be responsible for complementary β-sheet formation as shown in Fig. 7.10 [30].

The dodecamers rearrange slowly to give protofibrils that aggregate to give the dangerous β-fibrils that ultimately form the β-amyloid plaques. The $A\beta_{40}$ form of the protein forms tetramers that very slowly transform into β-fibrils directly and does not form the higher order species that lead to protofibril formation. Not only is the slower rate of fibril formation by the $A\beta_{40}$ proteins important but also the way in which those fibrils form. It may be that intervention at the protofibril stage is key to the aggregation of β-fibrils into plaques. Other health related issues are associated with plaque formation as the metals present are also able to cause oxidative stress through the generation of reactive oxygen species that attack nearby biomolecules including nucleic acids. Other biochemical pathways are disrupted, including calcium uptake and transport, and inflammatory responses are common.

Pharmaceuticals currently in use to combat Alzheimer's disease are mainly acetylcholine esterase (ACE) inhibitors designed to slow down the removal of acetylcholine so that neurons keep active, however, they do not reverse plaque formation. The only other approach is to inhibit the activities of the secretases, in particular, to favour the formation of $A\beta_{40}$ over $A\beta_{42}$.

Considering Alzheimer's disease from a supramolecular perspective there are several observations to be made. Firstly there is protein entanglement. Short

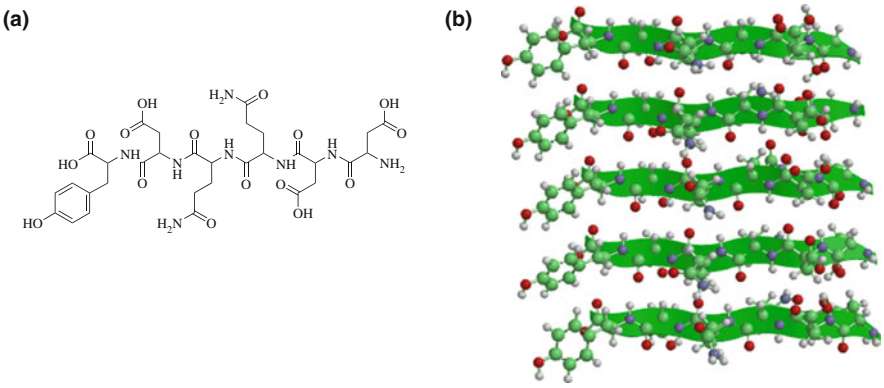



**Fig. 7.10** The β-sheet structure formed by the NNQQNY sequence (**a**) in amyloid-β proteins that leads to plaque formation [30]

sequences of amino acids on their own are unlikely to cause many problems so the process of entanglement needs investigation. As discussed in a previous chapter, protein-protein interactions are predominantly through hydrogen bonding between complementary functional groups, such as carbonyl oxygen atoms and amide protons in α-helices and β-sheets, or through hydrophobic effects. Some π–π interactions may be important if the sequences include tyrosine or phenylalanine. If sulfur-containing amino acids are present, and given the presence of transition metals this is probable, then disulfide bridges are also likely to form between proteins. Transition metal coordination is also a causative factor as it helps to bind proteins together and form the tertiary structure. To reduce protein aggregation these interactions must be inhibited. Secondly there is the issue of free radical generation. This can be addressed by the removal of transition metals bound up within the plaques. Although there may be some covalent bond formation it would seem that all the main binding forces involved in plaque formation and generation of toxic radicals are reversible supramolecular phenomena.

Orvig's group has come up with a potential therapeutic solution that targets several of the problems outlined above [31]. As with many compounds designed from a supramolecular point of view, those prepared by the Orvig group combine several functions. The first issue to be addressed is the removal of transition metals. This is accomplished by a tetrahydrosalen, a derivative of the salens described in an earlier chapter as nitric oxide sensors. Salens are excellent ligands for transition metals and their reduced forms make them more flexible, to maximize the range of metals that they can sequester, and imparts additional functionality through the nitrogen bridgeheads.

Aqueous solutions containing tetrahydrosalen ligands and amyloid proteins were exposed to zinc or copper and, based on turbidity assays, showed reduced protein aggregation. It seems reasonable to assume that the flexible ligands alone are unable to disrupt inter-protein binding forces but that the addition of metals induces a conformational change in the ligands. If this is analogous to the behaviour of salen complexes then the resultant structures will have both aromatic rings in the same plane. The planar aromatic structural motif allows salen-metal complexes to intercalate between planar nucleosides. Having metal cations at their core they are also attracted to the phosphate backbone of nucleic acids but cannot get close enough to form covalent bonds. Tetrahydrosalen metal complexes presumably have similar effects on proteins: they can intercalate between aromatic groups to disrupt π–π interactions and interact electrostatically with anionic side chains to break up regions of hydrogen bonding. As the ligands are administered in the metal-free form they will bind any metals available in the metal-rich plaque environments. In doing so, not only are inter-protein disrupting complexes formed, but the numbers of protein-folding and radical generating metal ions are reduced.

There are two final problems: the ligand cannot traverse the blood-brain barrier and will bind any metal it encounters before it approaches the target region. These have been addressed by glycosylating the ligands through the phenolic oxygens to make a prodrug, shown in Fig. 7.11, that can be smuggled into the brain. In doing so the phenolic oxygens are blocked from binding metals until the substituents are

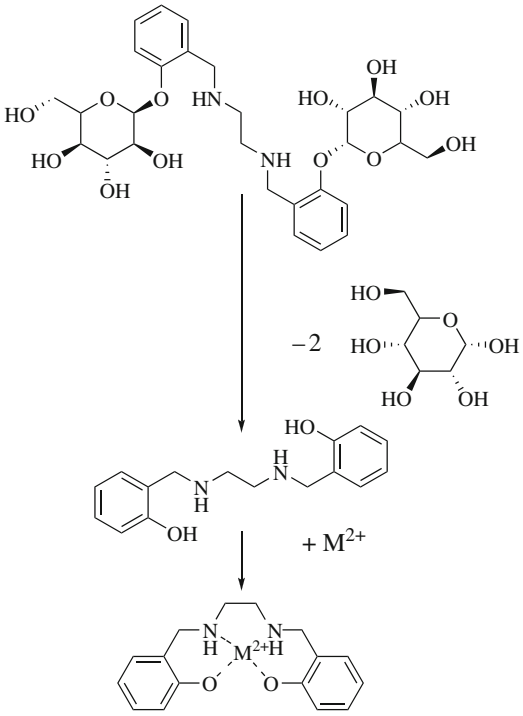



**Fig. 7.11** A tetrahydrosalen-based therapy to target Alzheimer's disease [31]

removed. Once inside the brain the sugar substituents are expected to be removed by β-glucosidase enzymes to reveal the metal sequestering ligand in an active form. At the time of writing this compound has not yet been tested in vivo, nevertheless, the combination of several functions in one molecule together with an elegant drug delivery strategy makes this a truly supramolecular approach to treating the causes, rather than the manifestations of Alzheimer's disease.

## 7.10 Calixarenes as Therapeutic Agents

There is some evidence that particular calixarene derivatives have direct therapeutic properties. The antitubercular action of 4-octylcalix[8]arenes with pendent polyether chains was determined by Cornforth's group in the early 1950s where the compound, then of unknown composition and called *Macrocyclon*, was shown to be generally non-toxic to mammals and yet more potent than streptomycin in the treatment of tubercular bacilli [32].

The compound's activity was later reinvestigated by Tascon and colleagues who confirmed the original findings [33]. Both macrophages and live mice were infected with *Mycobacterium tuberculosis* to determine the effects at the level of the cell and organism levels, respectively. *Macrocyclon*, illustrated in Fig. 7.12, was found

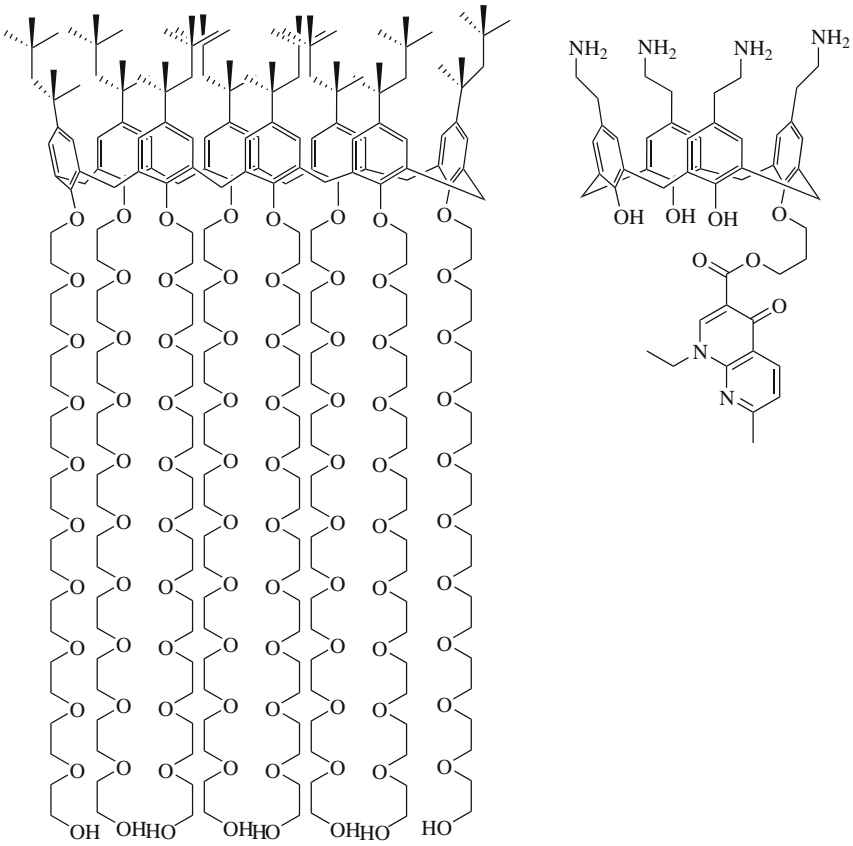


**Fig. 7.12** Calixarene-based therapeutics: anti-tubercular *Macrocyclon* [32] (*left*) and a nalidixic acid delivering prodrug [34] (*right*)

to work differently to other antitubercular drugs by enhancing the innate defence mechanisms in murine macrophages. This is important because strains of *M. tuberculosis* are starting to show signs of the multidrug resistance which has become such a problem with other infective species such as MRSA.

How do calixarenes enter cells? To answer this question Matthews, Mueller and co-workers prepared a soluble calix[4]arene with quaternary ammonium groups on the upper rim and a fluorescent label on the lower rim [35]. Following the progress of the compound by fluorescence confocal microscopy it was apparent that molecules were not transported across the membrane by normal endocytosis mechanisms, nor did they penetrate the cell nucleus but remained in the cytoplasm. This would seem to point to a membrane bound channel as the point of entry and paves the way for an array of therapeutic interventions.

One potential application is to incorporate a calixarene into a prodrug designed to deliver a pharmaceutical to a destination within a bacterium. To make a prodrug the

pharmaceutically active molecule is linked to a carrier molecule and the conjugate travels to the target area. Once in the target area the local conditions initiate bond cleavage to release the active drug whilst the carrier is either chemically degraded or excreted.

Often it is only through conjugation that the drug can reach its target either because of solubility problems, that are alleviated through physical properties of the carrier, or through the carrier's affinity for a particular feature of the target. For example, the carrier portion of the prodrug could be a small peptide that binds specifically to a particular site of a protein where the drug is released. Without the molecular recognition element provided by the protein the drug may be entirely non-specific whereas with the peptide it is released exactly where it is needed. This type of magic bullet approach is extremely important when highly toxic therapeutic agents are concerned.

The group of Regnouf-de-Vains has achieved this goal, using calix[4]arenes as a platform for drug delivery by making the calixarene part of a nalidixic acid delivering prodrug [34]. Nalidixic acid is an antibacterial agent that works not as a biocide, by killing the host, but as a bacteriostatic agent that blocks reproduction by interfering in the DNA replication mechanism. The group coupled nalidixic acid to an aminocalix[4]arene through an ester linkage. In biological media the prodrug decomposes, releasing about 30% of the bound drug within the first 24 h, and has an inhibitory effect on the growth of Gram negative *Escherichia coli* and Gram positive *Staphylococcus aureus* although Gram negative *Pseudomonas aeruginosa* and Gram positive *Enterococcus faecalis* were relatively unaffected.

## 7.11 Supramolecular Antibiotics

Many naturally occurring antibiotics act through mechanisms that involve penetration of cell membranes. Once this has been achieved the balance of ions and water between the intra- and extracellular media can be compromised and lead to the destruction of the cell. Other antibiotics may work as simple phase transfer agents, moving small destructive chemical species across the membrane, as in the case of valinomycin that transports potassium. Alternatively they may aggregate to form larger channels through which the cellular structure is compromised. This barrel stave aggregation is commonly adopted by natural macrocycles such as alamethicin, monensin or amphotericin B. The formation a supramolecule within a membrane cannot be random: the exterior of the structure must match the hydrophobicity of the central region of the phospholipid bilayer as well as the polar external surfaces.

Furthermore the pore must be lined with atoms that impart qualities which are attractive to ions and small molecules with the potential to pass through. The types of channels formed by natural antibiotics are often permeable to many species that diffuse through the central pore, attracted to the polarized hydrophilic regions within. Examples of these compounds are shown in Fig. 7.13.

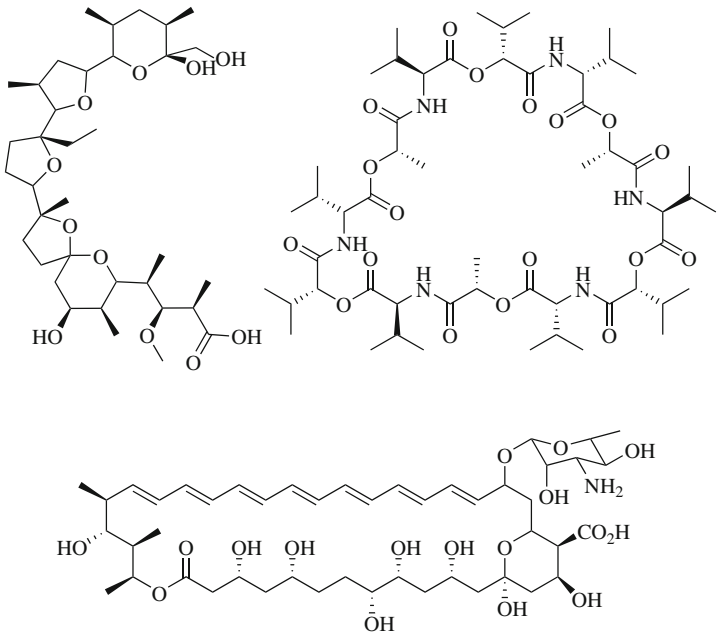


**Fig. 7.13** Some naturally occurring antibiotics and related compounds: monensin (*top left*), valinomycin (*top right*) and amphotericin (*bottom*)

Simple proteins can also act as potent agents for cell destruction. The main active component of bee sting venom is a protein composed of 26 amino acids called mellitin which adopts a helical, and potentially membrane puncturing, conformation. The protein is also an inhibitor of bacteria that cause Lyme disease and suppresses infections caused by *Candida albicans*, *Mycoplasma hominis* and *Chlamydia tracomatis*.

In a completely novel approach the research group of Ghaderi prepared cyclic D,L-α-peptides (Fig. 7.14) that self-assembled, through a combination of complementary hydrogen bonding and electrostatic interactions, into nanotubes that pierced cell membranes [36]. Alternating sequences of D- and L-amino acids give rise to compounds such as cyclo[(L-arg-D-leu)$_4$–] and cyclo[(L-glu-D-leu)$_4$–] which have central pore sizes of about 0.75 nm. Variation of these sequences elicited different responses with some having have antibacterial activity against methicillin-resistant *Staphylococcus aureus* (MRSA) and *Escherichia coli*.

As shown in Fig. 7.15, it has been proposed that the cyclopeptides stack and then insert in the bacterial cell membrane whereupon they form multitube aggregates, through favourable side chain interactions, and tilt to rupture the lipid bilayer resulting in species-dependent cell destruction.

Several years after publishing their results the Ghaderi group returned to these cyclopeptides, prompted by the discovery that antibacterial compounds isolated from *Streptomyces hydroscopicus*, known as mannopeptimycins, were also based on

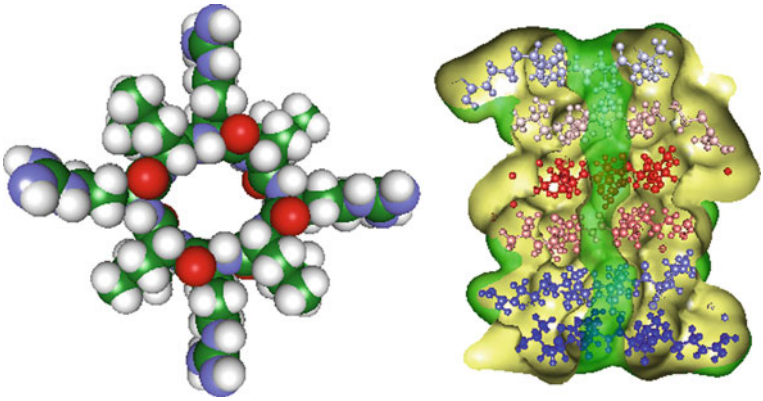


**Fig. 7.14** Self-stacking cyclopeptides [36]

a cyclic peptide architecture [37]. The compounds were first discovered at Lederle Labs in the 1950s and, although they showed activity against Gram-positive bacteria, were thought too complex at the time to analyse. Subsequently researchers at Wyeth Research have determined their structures [38].

Mannopeptimycin α is a cyclic hexapeptide with alternating D- and L-amino acids that have mono- and disaccharide side chains and two highly unusual β-hydroxyenduracididine amino acids, one of which acts as a link between the

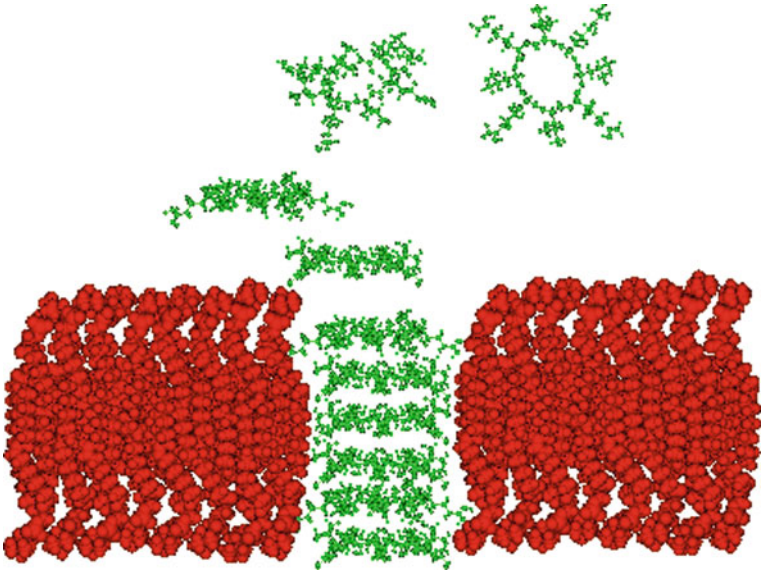


**Fig. 7.15** Proposed mode of antibiotic action by cyclopeptides: tubes (*green*) form and aggregate to disrupt bacterial membranes (*red*)

cyclic structure and an appended mannose group. Even more unusually both isomers of this amino acid are present. As model compounds for the mannopeptimycans the Ghadiri group prepared cyclic octapeptides with D-galactose, D-mannose and D-glucosamine substituents based on their highly active cyclo[L-trp-D-leu-L-trp-D-lys-L-ser-D-lys-L-ser-D-lys-] where the one of the last five amino acids has been replaced by a serine derivative linked to a monosaccharide [39]. The researchers were able to show that all variants had significant activity against MRSA, *Bacillus cereus* and vancomycin-resistant *Enterococcus faecalis* often within 2 hours. The position of the sugar had a significant impact on the compounds' haemolytic effects against both human and mouse red blood cells though in the best cases the values were much higher than the required therapeutic concentration. More importantly the $LD_{50}$ values for these compounds, as assessed against mouse fibroblasts, were up to 10 times higher than the therapeutic dose required to kill the bacteria. This finding means that the cyclopeptides are safe to use at the therapeutic level without affecting either blood cell integrity or overall toxicity. They are also much easier to synthesize and modify than the mannopeptimycans.

## 7.12 Summary

The scope of supramolecular therapeutic compounds, or those with the potential to be used in a clinical setting in the near future, is exceedingly diverse. Most of the compounds described in this chapter either bind guest metal ions or use weak forces to interact with features of their biological targets. In chelation therapy ligands are used to remove or deliver metals for which they have high affinities. These ligands are often macrocyclic in nature with donor atoms strategically placed to provide optimal metal binding. Macrocycles also form the core of complexes of radiopharmaceuticals such as DOTA-tyr3-octreotide, Lu-177-AMDA and texaphyrin derivatives. Other macrocycles, the phthalocyanines, are able to use energy from a light source to excite oxygen species in proximity to tumours and so destroy them leading to the widely used anticancer treatment known as photodynamic therapy. Zinc complexes of other small nitrogen-containing macrocycles, the cyclams, are able to disrupt the initial fusion step between HIV and a cell surface, thereby halting its transmission into the cell and subsequent replication.

Multiple weak interactions allow small pharmaceutically active molecules to be trapped in polymeric dendrimers. Over time the molecules slowly break free and are released in a controlled manner to the sites where they are required. The formation of dendrimer prodrugs allows for covalent attachment of many copies of a drug to a single dendrimer. Under predetermined conditions, such as pH or oxygen content, the bonds break and the drug is released in a concentrated burst. Weak interactions can also be harmful. The entanglement of small amyloid proteins appears to be responsible for the formation of plaques associated with Alzheimer's disease. The presence of transition metals poses an added complication as they have been implicated in protein crosslinking. A multifunctional molecule that can both disrupt

protein-protein interactions whilst simultaneously binding transition metal cations and hence remove them may be the ideal therapy for the condition.

Macrocycles with an affinity for cell membranes and the ability to pack, or stack, together have been shown to mimic the behaviour of natural antibiotics by piercing the cell walls of both Gram-positive and -negative bacteria with different degrees of specificity. Most of the examples discussed are either in use currently or undergoing some form of trials. Of the remainder, most show promise as future therapeutics.

## References


1. Winau F, Westphal O, Winau R (2004) Paul Ehrlich – in search of the magic bullet. Microb Infect 6:786–789
2. Kerberle H (1965) The biochemistry of desferrioxamine and its relation to iron metabolism. An New York Acad Sci 119:758–768
3. Walshe JM (1982) Treatment of Wilson's disease with trientine (triethylene tetramine) dihydrochloride. Lancet, 319:643–647
4. De Jong M et al (1998) Internalization of radiolabelled [DTPA(0)]octreotide and [DOTA(0), Tyr(3)]octreotide: Peptides for somatostatin receptor-targeted scintigraphy and radionuclide therapy. Nucl Med Com 19:283–288
5. Cox T et al (2005) King George III and porphyria: an elemental hypothesis and investigation. Lancet 366:332–335
6. Finsen NR (1901) Phototherapy. London: Edward Arnold
7. Lipson RL, Baldes EJ, Olsen AM (1961) The use of a derivative of hematoporphyrin in tumor detection. J Natl Cancer Inst 26:1–11
8. Bonnett R (1995) Photosensitizers of the porphyrin and phthalocyanine series for photodynamic therapy. Chem Soc Rev 24:19–33
9. Braun A, Tcherniac J (1907) Über die Produkte der Einwirkung von Acetanhydrid auf Phthalamid. Chem Ber 40:2709–2714
10. de Diesbach H, von der Weid E (1927) Quelques sels complexes des o-dinitriles avec le cuivre et la pyridine. Helv Chim Acta 10:886–888
11. Dandridge AG, Drescher HAE, Scottish Dyes Ltd (1928) Improvements in and relating to the manufacture and use of colouring materials. GB patent, 322169
12. Brown CJ (1968) Crystal structure of β-copper phthalocyanine. J Chem Soc A 2488–2493
13. Zugenmaier P et al (1997) Diffraction studies on metal free phthalocyanines (beta-$H_2$Pc and X-$H_2$Pc). J Mat Sci 32:5561–5568
14. Linstead RP (1935) Phthalocyanines. Part I. A new type of synthetic colouring matters. J Chem Soc 1016–1017 *et seq*
15. Robertson JM (1935) An *X*-ray study of the structure of the phthalocyanines. Part I. The metal-free, nickel, copper, and platinum compounds. J Chem Soc 615–621
16. Sessler JL et al (1993) Synthesis and structural characterization of lanthanide(III) texaphyrins. Inorg Chem 32:3175–3187
17. Sessler JL, Miller RA (2000) Texaphyrins – New drugs with diverse clinical applications in radiation and photodynamic therapy. Biochem Pharmacol 59:733–739
18. Anastasi A, Erspamer V, Bucci M (1971) Isolation and structure of bombesin and alytesin, 2 analogous active peptides from skin of European amphibians *Bombina* and *Alytes*. Experientia 27:166–167
19. Maddalena ME et al (2009) Lu-177-AMBA biodistribution, radiotherapeutic efficacy, imaging, and autoradiography in prostate cancer models with low GRP-R expression. J Nucl Med 50:2017–2024
20. Buhleier E, Wehner W, Vögtle F (1978) "Cascade"- and "nonskid-chain-like" syntheses of molecular cavity topologies. Synthesis 155–158

21. Tomalia DA et al (1985) A new class of polymers: starburst dendritic macromolecules. Polymer J 17:117–132
22. Hawker CJ, Fréchet JMJ (1990) Preparation of polymers with controlled molecular architecture. A new convergent approach to dendritic macromolecules. J Am Chem Soc 112: 7638–7647
23. Aulenta F, Hayes W, Rannard S (2003) Dendrimers: a new class of nanoscopic containers and delivery devices. Eur Polymer J 39:1741–1771
24. Malik N, Evagorou EG, Duncan R (1999) Dendrimer-platinate: a novel approach to cancer chemotherapy. Anti-Cancer Drugs 10: 767–776
25. Padilla De Jesus OL et al (2002) Polyester dendritic systems for drug delivery applications: *in vitro* and *in vivo* evaluation. Bioconjug Chem 13:453–461
26. D'Emanuele A, Attwood D (2005) Dendrimer-drug interactions. Adv Drug Deliv Rev 57:2147–2162
27. Inouye Y et al (1994) Inhibition of human-immunodeficiency-virus proliferation by macrocyclic polyamines and their metal-complexes. Biol Pharm Bull 17:243–250
28. Liang XY et al (2002) Structure and dynamics of metallomacrocycles: Recognition of zinc xylyl-bicyclam by an HIV coreceptor. J Am Chem Soc 124:9105–9112
29. De Clercq E (2009) The AMD3100 story: the path to the discovery of a stem cell mobilizer (Mozobil). Biochem Pharmacol 77:1655–1664
30. Goldschmidt L et al (2010) Identifying the amylome, proteins capable of forming amyloid-like fibrils. Proc Natl Acad Sci USA 107:3487–3492
31. Stor T et al (2009) Glycosylated tetrahydrosalens as multifunctional molecules for Alzheimer's therapy. Dalton Trans 3034–3043
32. Cornforth JW et al (1955) Antituberculous effects of certain surface-active polyoxyethylene ethers Brit J Pharmacol 10:73–86
33. Colston MJ et al (2004) Antimycobacterial calixarenes enhance innate defense mechanisms in murine macrophages and induce control of Mycobacterium tuberculosis infection in mice. Infect Immun 72:6318–6323
34. Diamba HM et al (2009) Towards calixarene-based prodrugs: drug release and antibacterial behaviour of a water-soluble nalidixic acid/calix[4]arene ester adduct. Bioorg Med Chem Lett 19:2679–2682
35. Lalor R et al (2008) Cellular uptake of a fluorescent calix[4]arene derivative. J Am Chem Soc 130:2892–2893
36. Fernandez-Lopez S et al (2001) Antibacterial agents based on the cyclic D,L-α-peptide architecture. Nature 412:452–455
37. Singh MP et al (2003) Mannopeptimycins, new cyclic glycopeptide antibiotics produced by *Streptomyces hygroscopicus* LL-AC98: antibacterial and mechanistic activities. Antimicrob Agents Ch 47:62–69
38. Ruzin A et al (2004) Mechanism of action of the mannopeptimycins, a novel class of glycopeptide antibiotics active against vancomycin-resistant gram-positive bacteria. Antimicrob Agents Ch 48:728–738
39. Motiei L et al (2009) Antibacterial cyclic D,L-alpha-glycopeptides. Chem Commun 3693–3695

# Chapter 8
# Bionanotechnology, Nanomedicine and the Future

## 8.1 Bionanotechnology

Nanotechnology is a broad description that encompasses the use of objects on the nanoscale, that is between 0.1 and 100 nm, and methods for their chemical synthesis or manufacture. In Nature it is the scale of protein synthesis, DNA replication, enzyme catalysis and the chemistry involved in the immune response, to mention just a few processes. In the synthetic world it is the scale of the molecule and of supramolecular assembly. Bionanotechnology is where nanotechnology and biology interact. This may be in the analysis of a bacterial flagellum in terms of conventional rotary motors or in the use of biological material as nanoscale components in unnatural constructions. The interplay between the two domains is essential because existing synthetic methods are unable to function reproducibly and accurately at the nanoscale whereas biological systems routinely do so rapidly and with very few errors.

Currently two approaches are used to construct manmade objects on this scale, as was illustrated in an earlier chapter. The bottom up approach starts with atoms and small molecules (0.1–5 nm) that are extended using conventional chemical synthesis. This is time consuming and frequently results in errors. Until an error checking mechanism is invented there will always be defects in the final object. Many techniques involve laying down sequential courses of molecules that self-assemble in a specific orientation but the process relies on absolute purity of each chemical layer, absolute precision in applying each layer, and absolute fidelity in every chemical reaction. Not only is this impossible to achieve but it is also very inefficient. The alternative is to use a top down lithographic method where bulk material is etched by a finely focused laser, or similar method, enabling lower layers to be revealed and functionalized. The problems here are twofold. Firstly, the method is only really appropriate for flat surfaces: complex three dimensional structures would be hard to manufacture on this scale. Secondly it is impossible to focus down to achieve perfect atomic level resolution leading to inaccurately rendered, blurred edges of any feature prepared using this method. Biological processes function far better and can produce an amazing complexity of nanoscale objects that, in turn, can aggregate to form macroscopic structures. Unfortunately these processes have only recently

P.J. Cragg, *Supramolecular Chemistry*, DOI 10.1007/978-90-481-2582-1_8,

© Springer Science+Business Media B.V. 2010

been opened up to the types of unnatural modification that allow us to reprogram their natural functions.

## 8.2 The Unnatural Chemistry of DNA

One of the problems with DNA is that it is essentially a linear code that stores information. The functional groups (nucleotides) only interact with each other and, while this can lead to the elegant double helix, it limits the ability of DNA to form different secondary and extensive tertiary structures. RNA is rather more amenable to forming other structural motifs, hence the RNA World theory of molecular evolution, but it appears that only proteins with their varied side chains are able to adopt truly complex structures.

However, DNA has one important advantage over proteins: nucleotide interactions are very specific and the hydrogen bonded interactions, particularly for G-C, are very strong. This has led to some extremely interesting non-biological uses for DNA as a structural motif including Seeman's demonstration that DNA can be used to construct nanoscale objects [1]. He showed that sequences could be synthesized, using standard laboratory techniques, so that they are complementary for most of their lengths before terminating in several nucleotides with no complementary partner. Duplexes constructed from complementary strands made in this way therefore had 'sticky ends' that could be spliced to complementary ends on other DNA molecules. Because branching occurs when the two DNA strands unravel in preparation for replication or transcription these sticky ends can be used to initiate the formation of junctions. Permanent branches form when the enzyme ligase is added to join the free sticky ends together.

Careful design was used by Seeman to prepare complementary DNA strands that self assemble into a cube with 20 nucleotide pairs along each edge. Other derivatives include 'nylon DNA', through the synthesis of peptide nucleic acids (PNA), and polyethylene glycol derivatized DNA. Figure 8.1 illustrates the DNA knots that can be prepared using these techniques.

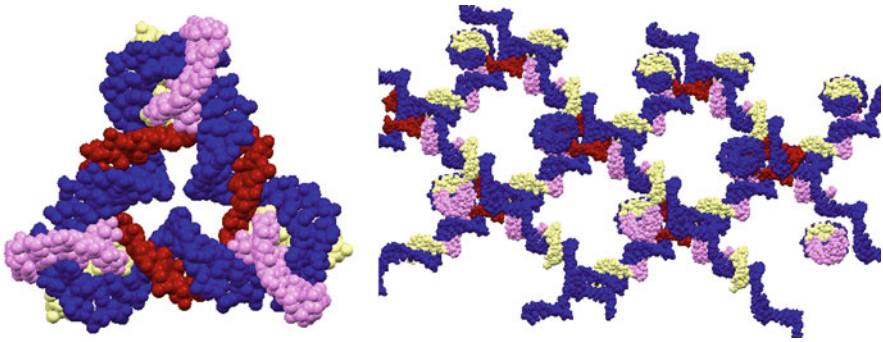


**Fig. 8.1** Knots from DNA [2]

Since Seeman's box and similar nanoscale DNA structures were prepared in the 1990's the manipulation of DNA has taken many different directions including applications in computing, molecule level bar coding, and nanomachines. As with many areas of supramolecular chemistry, the inspiration for this work remains firmly rooted in the biological sciences. Modification of DNA for non-biological purposes requires technology borrowed from biochemistry, notably PCR, and natural enzymes. Seeman uses the term biokleptic, literally stealing from biology, when referring to this approach to supramolecular construction. Returning to the idea of nanostructures, the Shih group demonstrated that a 1,669 base long single strand DNA sequence, when heat denatured in the presence of five 40 base strands, had a planar connectivity that was programmed to fold in only one way to generate an octahedron [3].

The Turberfield group developed a method whereby four complementary DNA strands rapidly self-assemble into tetrahedra [4]. Given the specific sequences involved only one product was possible. The Mao group has used strands of different lengths to construct three pointed star shaped molecules that can join at their apexes thereby forming a number of different three dimensional structures [5]. A long central strand contains three loops of varying lengths to impart three fold symmetry. The lengths of the loops control the degree of flexibility that the final construct can possess. Simple stars are flat and form sheets of hexagons with astonishing regularity over several hundred nanometres. When the loop extends to five bases the highly flexible structure that results can form a corner of a cube, tetrahedron, octahedron or any number of prisms. Cooling mixtures of one centrally looped, three medium, and three short DNA strands from 95ºC to room temperature resulted in the formation of three dimensional tetrahedral structures. These were revealed by atomic force microscopy and cryogenic transmission electron microscopy to be 10 nm on edge. Why tetrahedra rather than cubes? Presumably the three pointed star is too flexible to form a stable cube. A cube would also require eight corners to self assemble rather than the four necessary to make a tetrahedron. Although the strain on the star's central branch would be greater when triangular (60º) rather than square-based (90º) corners are formed, with fewer stars needed to make a stable polyhedron, it is likely that the tetrahedron is favoured by thermodynamics over other species in a dynamic combinatorial library of kinetically possible structures.

When the central threefold motif was rigidified, with only three bases in the flexible loops, 25 nm dodecahedra were formed. Dodecahedra are composed of five membered rings sharing their edges. Given that the arms of the DNA stars are at 120º to each other it would seem that hexagonal lattices ought to form but these would result in a planar sheet structure akin to graphite. In forming five membered rings the remaining arm of each component star projected at an angle to the pentagon's plane imposing a slight curvature on the growing structure. When the concentration of stars was increased tenfold the predominant structure was that of a 42 nm soccer ball: a truncated icosahedron comprised of five- and six-membered rings.

What can be done with these nanostructures? Seeman's group has already prepared more complex and rigid DNA arrays using methods which impart stiffness to the lattices that form. The same methods have been used to introduce a junction at

90º to the plane of the DNA lattice which allows more robust molecular building blocks to be prepared. By splicing different forms of DNA it is possible to assemble nanotriggers and tweezers that can be used to activate RNA [6]. More importantly, triggers exist in on- and off-states which suggests that arrays of DNA could form the basis of molecular computers. Of course, if these developments are to be truly useful they must be easily reproducible on a large scale. For this they need to be amenable to methods of mass production; in the case of DNA this is undertaken by the polymerase chain reaction (PCR), a technique that 'amplifies' a chosen base sequence by artificial means.

PCR was originally developed by Mullis while working for the Cetus Corporation, for which he received the 1993 Nobel Prize in Chemistry, and is now a routine laboratory method [7]. The principle uses a single strand DNA template, a pool of nucleoside triphosphates, the enzyme DNA polymerase and DNA primers. The latter are the oligonucleotides required to initiate DNA replication. Once a strand has been copied, heat is used to separate the resultant double strand DNA so that both can replicate. Normally the enzymes involved would not survive the experimental conditions but polymerase enzymes from particular bacteria, such as *Thermus aquaticus*, are heat stable and can undergo cycles of heating and cooling. Once the first duplex has been reproduced each strand can go on to replicate exponentially. PCR therefore allows a specific DNA sequence to replicate under unnatural conditions until the pool of nucleotides has been exhausted. The process is designed to work for linear sequences of DNA as it uses the natural polymerase process of replication. At present, synthetic DNA containing junctions cannot be replicated which limits the application of this technique as a route to unnatural DNA structures. Where a three dimensional structure forms as a result of linear DNA folding in a programmed manner, as occurs in Shih's octahedron and to some extent with the components of Mao's stars, multiple copies of the nanopolygons can be made directly by PCR. Once the sequence has been replicated it will fold into the desired shape. For this to be successful the folding should not occur as soon as the sequence has been synthesized, or further replication cannot occur, but should be dependent on a specific stimulus. The ultimate goal would be for a bacterium or plant to be modified so that it synthesizes the polygonal sequence as part of its normal biochemistry.

Manipulation of DNA in such a precise manner demonstrates what can be achieved when the nature of nanoscale interactions is well understood and can be controlled. As more and more becomes known about similar molecular recognition motifs in the natural world the potential to use other 'biomaterials' as structural components will increase. Other structural biomolecules that could be targets for unnatural manipulation include collagen [8] and clathrin proteins [9], shown in Fig. 8.2. The former is a polypeptide that forms a triple helical coil in which the individual strands are left handed but the resulting supercoil is right handed. It provides structure to bone and cartilage and has many other rigidifying functions.

The latter is another complex protein that instead of adopting rigid helices forms a three legged, or triskelion, structure. In this it is not unlike Mao's three pointed DNA stars, however, each triskelion arm is bent allowing molecules to self assemble. The result is a lattice that can fold in on itself to form a truncated icosahedron:

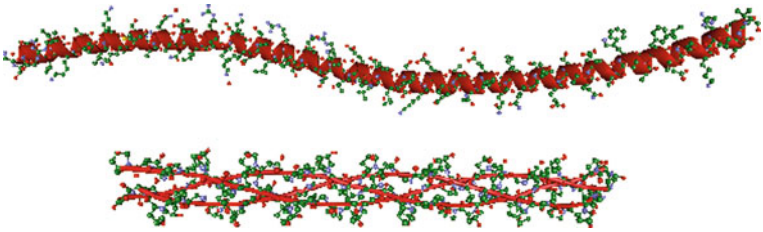


**Fig. 8.2** The molecular fibres of clathrin [10] (*top*) and collagen [11] (*bottom*)

the same stable arrangement of five- and six-membered rings seen in buckminsterfullerene. As discussed in Chapter 4, these cages can encapsulate proteins and transport them across cells. Not only would manipulation of these systems generate valuable nanostructural materials but a greater understanding of the natural systems would offer novel methods of protein-assisted drug delivery.

## 8.3 Molecular Muscles

Muscles use energy to contract and expand in response to stimuli. They may act as an unconscious 'knee jerk' reaction, as a consequence of neural activity arising from touching a hot surface, or in any number of conscious movements. Regardless of the reason for muscle activity the fundamental mechanism is a cycle between contraction and extension.

On the human scale it is possible to copy this behaviour with pistons and hydraulic systems but can this be mimicked on the nanoscale? If it can, then therein lies the potential to manufacture molecular scale pumps that could be incorporated into nanoscale devices. Furthermore if the contraction and expansion cycle could be linked to an external signal, such as the change in chemical environment associated with diseased tissue or tumour, the pump could form the basis of an in vivo drug delivery system that identified and treated diseases before they became manifest to a medical practitioner.

It appears that rotaxane supramolecular complexes, discussed in Chapter 1, are able to perform this task. Examples described earlier were of a simple type where a cyclic molecule was threaded onto a linear molecule. This shuttle was permanently locked in place around the thread by attachment of bulky stoppers at the ends of the latter. The shuttle is free to slide along the thread and rotate but, unless it is able to transmit energy or information depending on its position, remains a chemical curiosity. However, if a shuttle is attached to a stoppered thread two rotaxanes can be interlinked such that extension and contraction along the axis of the thread can occur. An example of an interlocked rotaxane dimer has been reported by the group of Sauvage [12]. Each strand of the dimer consisted of three binding sites and a stopper. A macrocycle incorporated a metal binding 1,10-phenanthroline group and a polyether region forms the shuttle. This was linked to a thread, comprising another 1,10-phenanthroline and a terpyridine group, which in turn was

linked to a bulky 4-tris(t-butylbenzene)methylphenol stopper. The interlinked system was formed through symmetric copper binding between a macrocycle in one ligand and a 1,10-phenanthroline group in the thread of a second. The complex was built up as first the terpyridine groups were added and then the stoppers. In the dicopper complex the macrocycles bound to neighbouring 1,10-phenanthrolines to represent the 'extended' state of muscles. When the copper(I) cations were removed and replaced with zinc(II) the thread slipped, as shown in Fig. 8.3 so that the zinc cations bound between the macrocycle and more distant terpyridine groups to give an approximation of the 'contracted' state. Ideally the switch between states would

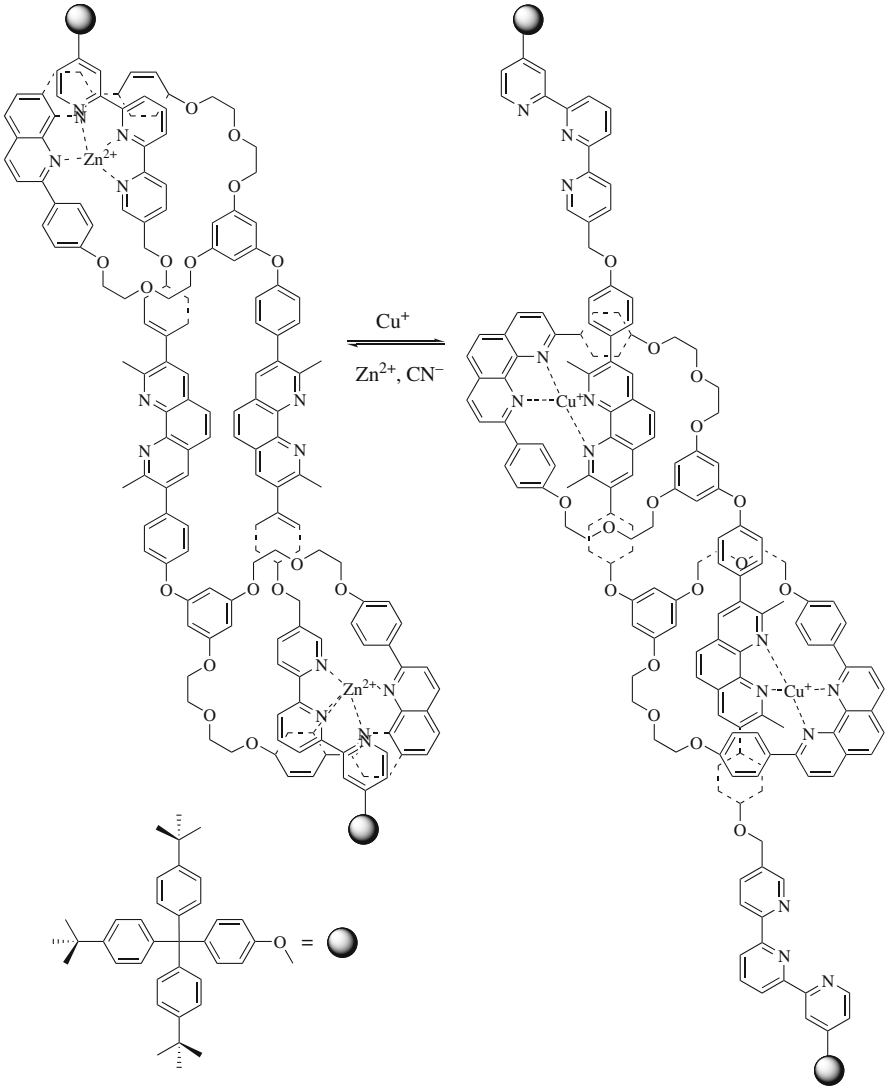



**Fig. 8.3** A molecular muscle in contracted (*left*) and extended (*right*) forms [12]

be triggered by something more subtle than the removal of one ion and replacement with another but the molecular muscle demonstrates the principle of controlled contraction and extraction admirably. In addition the entire system is very flexible: the degree of extension and contraction would be far more useful if all the components were rigid. Nevertheless, this complex example demonstrates the distinct possibility of synthesizing nanoscale muscles and pumps.

On a simpler scale the Rybak-Akimova group found that when 4′-(aminomethylene)benzo[18]crown-6 was treated with an acid salt, mass spectrometric evidence indicated that dimers were the most stable species [13]. This was backed up by an X-ray crystal structure of the compound crystallized from methanol which revealed pairs of mutually interlocked crowns. The amine termini had become protonated and the resultant ammonium group, illustrated in Fig. 8.4, formed a complex with a second crown. Although yet to be attempted, it would be intriguing to react the complex with a linear component containing a second amine and a bulky stopper group to generate a metal-free, interlocked pair of rotaxanes with two amine 'stations'. The contraction and extension of the complex could then be controlled as a function of pH.

The foregoing examples are highly artificial analogues of muscles operating on the molecular level. What if a non-biological artificial muscle was required to operate on the nanoscale? Guan and co-workers have reported a biomimetic modular polymer which does just this [14]. The inspiration behind the research was titin, a multidomain skeletal muscle protein also known as connectin. Titin is the largest known protein with a mass of 3 MDa or greater and is composed of an A-band and an elastic I-band. Each titin molecule spans about half of the sarcomere, the fundamental striated muscle building block that links thick myosin filaments with Z-discs which define the boundaries of the muscle. The protein's N-terminus attaches to the Z-disc and its C-terminus interacts with the thick filaments. It acts like a spring, initially in a shortened form where the protein is able to fold into β-sheets to give a more globular, β-sandwich structure. When slack the titin section of the sarcomere is about 50 nm in cardiac muscle rising to 200 nm in soleus muscle in the lower leg. Initial tension rapidly 'unzips' the β-sandwich section of the I-band which is followed by the unfolding of the so called PEVK region, named because of the regular occurrence of the proline-glutamic acid-valine-lysine sequence. The PEVK region appears to have a 28-residue long proline rich repeat structure that gives rise to a helical structure. Once extended this presumably springs back rapidly to contract the muscle microfibre unlike the β-sandwich section which reforms more slowly.

Guan modelled the multidomain of titin by synthesizing a polymer composed of large macrocycles linked by alkene spacers, as seen in Fig. 8.5, that could

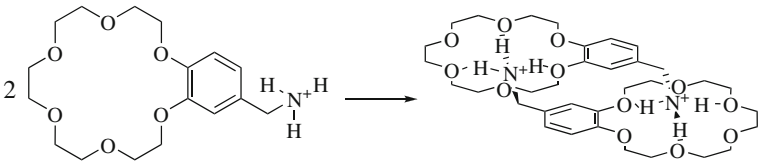


**Fig. 8.4** Self-interlocking crown ethers [13]

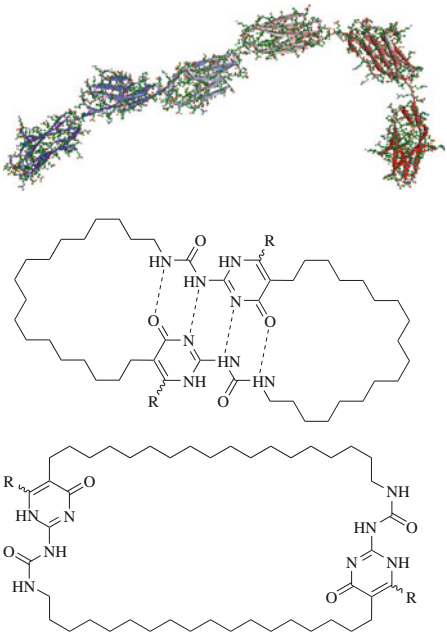


**Fig. 8.5** Part of the the muscle-forming protein, titin [15] (*top*), and a polymeric model in contracted (*centre*) and extended (*bottom*) forms [14]

form internal hydrogen bonds. Starting from a donor-acceptor-donor-acceptor motif present in 2-ureido-4[1*H*]-pyrimadone (UPy) the researchers constructed a macrocycle with two UPy groups separated by 19 -$CH_2$- repeat units. The UPy groups were set symmetrically within the macrocycle and were further functionalized to include divergent esters terminating in alkene groups. The alkene groups underwent acyclic diene metathesis in the presence of a second generation Grubbs' catalyst to link the macrocycles together like a string of beads. The polymer that forms has an average length of 7–8 nm and responds to stress and temperature changes. Under stress the polymer deforms over a series of cycles but reverts almost to its original length overnight; analysis revealed the material to have good shape memory.

The key to the material's behaviour, and its similarity to titin, lies in the duality of its construction. The alkyl bridges in the macrocycle are large and highly flexible, allowing the polymer a great deal of contraction and extension, but these are controlled by the two complementary UPy groups in each macrocycle. When these are held together through hydrogen bonding the polymer is in a contracted state but if the bonding is disrupted, either through stress or increased temperature, it can extend. The process, especially under temperature control, is largely reversible mimicking the extended and contracted forms of titin.

## 8.4 Nanomedicine

In 2001 Alivisatos, writing in *Scientific American*, identified several clinical targets for bionanotechnology [16]. As with conventional medicine there were two main threads: diagnosis and therapy. Many existing and proposed applications of

supramolecular chemistry in medicine address the former problem, that of detecting and signalling the presence of a particular disease state. The challenge here is to develop a molecule which has high specificity for the target, and to reduce the incidence of false positive results, without compromising sensitivity. Furthermore the reporting mechanism for the assay must be amenable to standard detection methods. Far more challenging is the development of a novel therapeutic regime. This is hard enough in the context of conventional medicine, and exercises the minds of many of the brightest scientists in both the pharmaceutical industry and academia, but to invoke a supramolecular connotation is particularly hard. Unlike the diagnostic approach there is no requirement to report the presence of the target, it merely needs to be destroyed or seriously compromised. In theory it would be ideal for a therapeutic molecule to also signal its presence to ensure that it has reached the correct target but this is rarely possible in practice. Research in conventional drug therapy that uses small molecules usually starts with the development of a molecule known to affect a key biochemical pathway. The progression of the molecule and determination of its effects can be easily followed in vitro by tried and tested techniques, whether biological assays or chemical analyses. Once the principal of action has been established, trials at a number of levels follow to ensure that the drug does what it is designed to do without harmful side effects. Ultimately these result in large scale human trials where not only can quantitative data be gathered but also qualitative information based on the subjects' reported experiences. If these drugs were able to signal their presence and effects on their targets, the efficacy of the treatment would be known with much greater certainty at a much earlier stage. As will be shown, the application of supramolecular principles can achieve just such an outcome.

### *8.4.1 Labelling with Nanoparticles*

Attachment of magnetic nanoparticles to antibodies that label specific proteins, as shown in Fig. 8.6, is now a possibility [17]. Once the particles bind to proteins through highly specific antibody protein complementarity the entire nanoassembly becomes magnetically aligned in the presence of an external magnetic field:

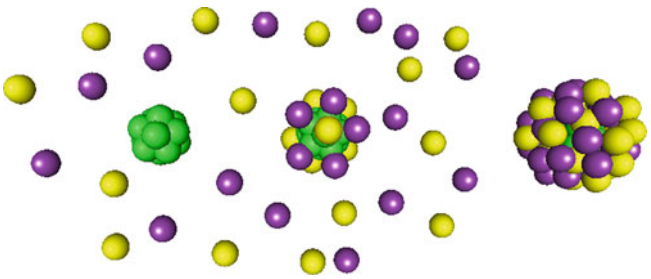


**Fig. 8.6** Magnetic nanoparticle synthesis: (*left* to *right*) a nucleating core is surrounded by the first shell of components followed by a second shell of antibodies

unattached particles will have random orientations and give much weaker responses. This allows proteins that are bound to the particles to be removed with great accuracy. There are clear advantages of this approach over current assays which should eventually make them viable in vivo.

Similarly, synthetic sequences of DNA, complementary to target sequences diagnostic of particular diseases, can be attached to gold nanoparticles [18]. When the complementary strands attach the nanoparticles become linked by a web of DNA and form supramolecular aggregates which scatter light in a different way to the original particles leading to a distinct, and diagnostic, colour change.

### *8.4.2 DNA Fingerprinting*

Recognition of complementary sequences of DNA as a mechanism for detecting specific genes associated with a number of clinical conditions has seen great interest in recent years [19]. The power of the method is that it uses well known techniques based upon biological recognition. As has been discussed earlier, advances following Mullis' development of PCR allow small samples of DNA to be amplified to the point of reliable detection. This has led to the widespread use of DNA fingerprinting, a technique whereby individuals' DNA is treated with restriction enzymes that reduce the entire sequence to much smaller oligomers. These can then be analysed based on fragment length by Southern blotting, using conventional gel electrophoresis, and stained to ease visualization. The result of the experiment is that sequences can be compared between individuals based on how far they travel along the electrophoretic gel. Although this method is widely used in biomedicine and forensics it cannot distinguish individual bases and cannot therefore be used in genome sequencing.

### *8.4.3 Full Genome Sequencing*

To determine the sequence of bases in a single strand of DNA it is necessary to detect each one of the four nucleotides and register the order in which they appear. At present the method of dideoxy chain termination based on Sanger's original protocol [20], shown in Fig. 8.7, is used. The DNA fragment to be tested is denatured into single strands and a primer added that starts complementary strand synthesis at the 3′ end. The reaction medium includes all four of the deoxoribonucleotides necessary for replication together with their dideoxy analogues, each of which has been labelled with a different fluorescent marker. The polymerase process builds up complementary DNA sequences but when a dideoxyribonucleotide adds it terminates the sequence as it lacks the 3′ hydroxyl group necessary for further reaction. Although incorporation of a dideoxyribonucleotide occurs at random, eventually every possible labelled sequence will be formed. The products are passed through a capillary tube filled with acrylamide gel so that they elute in order of mass with

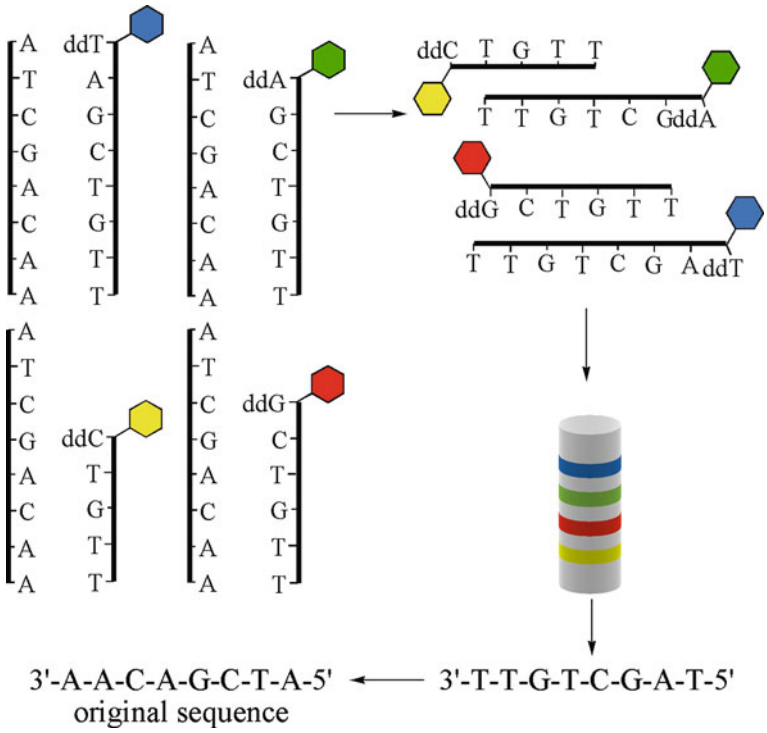



**Fig. 8.7** The principle of full genome sequencing using labelled dideoxy (dd) nucleotides: labelled fragments are eluted through a gel and detected in order of length to determine the complementary DNA sequence from which the original can be deduced

the shortest strand first. They pass though a fluorescent detector which registers the wavelength of each fragment and, as this is specific to one of the four tagged dideoxyribonucleotides, the order in which the bases are eluted can be determined. This corresponds directly to the complementary sequence to the DNA template fragment.

Since this technique was pioneered the speed of automated DNA sequence determination has increased dramatically to the point that the same sequence is routinely determined many times so that a consensus sequence results with as few errors as possible.

The key to making such approaches useful is to link the enhanced concentration of a specific sequence with an output that can be measured. More importantly the technology used to detect and report the output signal must be relatively cheap and easy to use if it is to be widely adopted. The cost of a personal genome determination is now relatively low, to the extent individuals' underlying medical conditions have been diagnosed based on their symptoms and rapid sequencing.

Researchers are actively pursuing a method of reading the sequence of bases on DNA in a rapid manner. Some advances have already been made in this area due to

its obvious importance in future healthcare: full genome sequencing could become a routine test for many newborn children in the developed world in the next decade. With a greater understanding of genetic markers of disease it will become desirable to screen for DNA sequences associated with a predisposition to these diseases so that the individuals can make informed choices about their lifestyles.

### *8.4.4 DNA Sequencing in Real Time*

To sequence DNA the order and identity of each base needs to be determined precisely. The methods discussed above involve the synthesis of fragmentary labelled sequences followed by chromatographic separation and detection. How much easier, and faster, it would be if each base could be detected in situ without having to be labelled? Such a process would remove any errors that occurred during the polymerase reaction ensuring complete fidelity with the sample being tested. This concept assumes that properties specific to each base can be measured and recorded in the order they are detected. To achieve this, the DNA strand must pass through a detector with either the 3′ or 5′ end entering first. One way to do this is to pass it through a nanoscale slit, in a known orientation, which can measure some property of each base. The strand can be induced to enter the slit using techniques well known in electrophoresis and, as it passes through the slit, the specific force that each base exerts can be related to the original sequence. Advances in nanotechnology may now make this a reality. At present two approaches are being investigated in parallel.

#### 8.4.4.1 DNA Sequencing Through Natural Nanopores

Howorka and co-workers have pioneered the use of a natural nanopore, α-haemolysin, placed in a lipid bilayer [21]. The benefit of this method is that it uses a natural pore-forming protein which has been characterized crystallographically at high resolution and about which much is known, including the pore size and its normal conductance. When a known oligonucleotide sequence is attached to the rim of the α-haemolysin pore it recognizes and binds to a complementary sequence and, in the process, blocks the current of ions passing through the pore. Using standard planar bilayer methods it is simple to detect these binding events because the usual conductance steps, which result from a single channel opening, no longer appear. If an oligonucleotide with a mismatched sequence is present the steps are replaced by spikes of brief duration indicating that the pore is partially blocked. This method has been used to detect specific sequences of DNA from a drug resistance conferring mutation in part of the HIV reverse transcriptase gene and in single nucleotide polymorphisms in the HIV protease gene responsible also for drug resistance.

#### 8.4.4.2 DNA Sequencing Through Manufactured Nanopores

While Howorka's approach is relatively simple and reproducible at the atomic scale, it cannot yet record a sequence of bases passing through the pore. To do this it

is necessary to design a nanopore that reacts differently to each base that passes through and can relay that information somehow. The problems here are twofold.

Firstly the sizes and non-covalent interactions that could be used to discriminate between the bases are very similar, indeed, there are almost no detectable differences between the dimensions or hydrogen bonding requirements of the purines, adenine and guanine, or between the pyrimidines, thymine, cytosine and uracil. The second problem is one of manufacture: how can nanopores with reproducible dimensions be formed and linked to a reporting mechanism? These problems have been answered to some degree by Bashir and co-workers, as seen in Fig. 8.8 [22].

Rather than using a biological pore they prepared thin silicon wafers with microfabricated square holes 100 nm on each edge. Silicon dioxide was grown on the surface in two stages to shrink the pores down to 50 nm and then tapering to 5 nm across. The wafers were approximately 250 nm thick with the 5 $nm^2$ region extending for about 60 nm. Both ends of the pores opened into funnels allowing DNA to be channelled into the pores and to egress easily. Electrophysiological methods were again used to monitor the passage of DNA oligomers though in this case, rather than a lipid membrane, it was the silicon/silicon dioxide wafer that formed the barrier through which the DNA had to pass. Movement of the oligomers, 200 base pair dimer fragments of the human CRISP-3 gene from a prostate cancer source, was

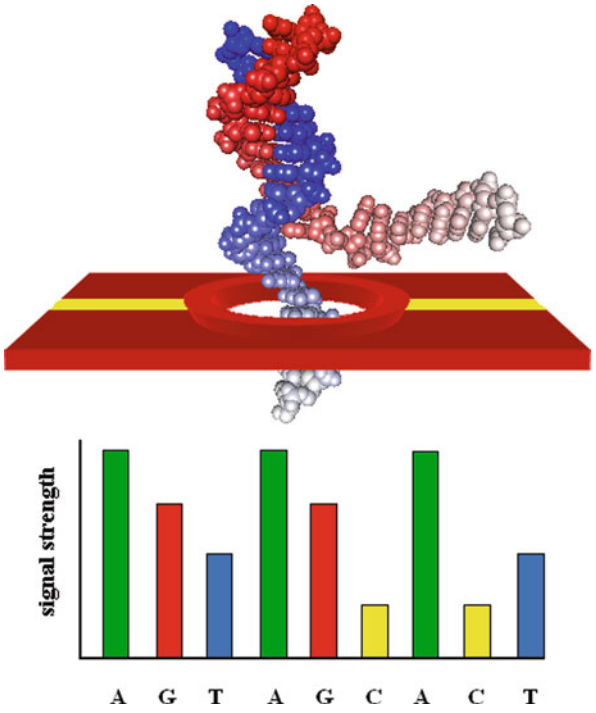



**Fig. 8.8** The future of real-time DNA sequencing: a single DNA strand passes through a nanopore and base-specific properties are measured [22]

indicated by a current pulse. The profile of pulse widths against frequency was used to determine the charge and length of the oligonucleotide passing through. So far the technique is somewhat limited, however, it may be possible to use the silicon wafer itself as the electrode that reports the passage of each individual nucleotide. The benefit of such a technique is that it would be a direct measure of the sequence and could be extremely rapid. An advance in this direction would make personal genomics cheap and readily available to all.

### *8.4.5 Therapeutic Multimodal Nanoparticles*

One method to deliver drugs accurately to a target, as has been illustrated by many examples, is to prepare a multifunctional prodrug that has an affinity for the target and decomposes to release the active agent. Most examples given so far involve a single drug molecule linked to the targeting group with the assumption that many molecules will attack the target but another way to accomplish this is to cover a nanoscale object with the therapeutic agent and find a way to direct the modified nanoparticle to its target. Nanoscale composites of metal atoms and organic binders used to seed quantum dots are small enough to be used as tags for any number of targets on the cellular scale and below. As their optical and magnetic properties can be tuned many such tags could be used to follow a number of biological processes simultaneously. Thus a single blood, urine or biopsy sample could be subjected to a whole battery of tests at once.

One advantage of using nanoparticles is that they can be injected close to the target site, such as a tumour, so that less of the active agent is lost on its way to the target. Cheon and colleagues used this approach to design a multimodal nanoparticle that targets cancer cells [23]. They used magnetic manganese-doped iron oxide nanoparticles as the core because their magnetic properties enhance the contrast of MRI images. The 15 nm magnetic core was covered with bovine serum albumin (BSA) which was subsequently modified to leave approximately 700 pyridyldisulfide groups on the surface. Tumour endothelial and metastastic tumour cells are known to overexpress $\alpha_v\beta_3$ integrins so a complementary tripeptide, the cyclic aginine-glycine-aspartate (cRGD) sequence, was linked through a biocompatible poly(ethyleneglycol) spacer to a thiol terminus. This reacted with surface pyridyldisulfide groups to introduce a guidance system for the nanoparticle. Through random attachment all over the nanoparticle surface there is a very high chance that at least one cRGD group will encounter an integrin binding site and attach to a cancerous cell. Once the particle has encountered a cancerous cell it sticks to the cell surface through multiple cRGD-$\alpha_v\beta_3$ integrin interactions. These interactions are the familiar weak, reversible supramolecular interactions common to proteins: hydrogen bonding, electrostatics, $\pi-\pi$ interactions and, probably most importantly, hydrophobic interactions. The final component of this multifunctional system is the drug itself. Cheon chose a small interfering RNA (siRNA) fragment designed to disrupt mRNA and attached it to the nanoparticle. Again a thiol group was used as the linker because the resultant disulfide bond would be cleaved by

glutathione once inside the cell. The siRNA chosen was siGFP, an inhibitor of green fluorescent protein (GFP), and was terminated with a contrasting red cyanine dye, Cy5. About 40 siRNA strands were found to bind to each nanoparticle. Obviously normal tumours do not express GFP but the use of this particular gene-silencing RNA allowed the researchers to assess its effectiveness on GFP-producing cell lines and follow the inhibition by confocal fluorescence microscopy rather than measure the size of tumour or count of viable cells.

The functionalized nanoparticles successfully bound to breast cancer cells and were shown to enter the cell through endocytosis by the Cy5 fluorescent response and MRI signal enhancement. In a separate experiment cells with surface $\alpha_v\beta_3$ integrins that also expressed GFP were treated with the nanoparticles. Again the nanoparticles entered the cells whereupon the disulfide bonds holding the siGFP strands were cleaved to release the gene-silencing RNA. A reduction in GFP-related fluorescence, coinciding with co-located fluorescence from Cy5, indicated that the siGFP-Cy5 fragments had been released and targeted the mRNA that led to GFP production.

The multimodal approach to targeting, detection and siRNA release used by Cheon and co-workers, illustrated in Fig. 8.9, paves the way for other similar therapeutics. The general location of the nanoparticles can be determined by MRI of a patient. They can also be detected with greater accuracy by fluorescence confocal microscopy, for example in a biopsy sample to check the extent of the cancer, through the terminal fluorescent tags on the siRNAs. So far the nanoparticles have only been tested on cells but they hold much promise as future stealth, or preemptive, therapeutic agents for numerous diseases. Other targeting groups could be devised, such as antibodies, and, assuming that they only stick to cells when specific features are at levels associated with diseased cells, therapeutic nanoparticles could

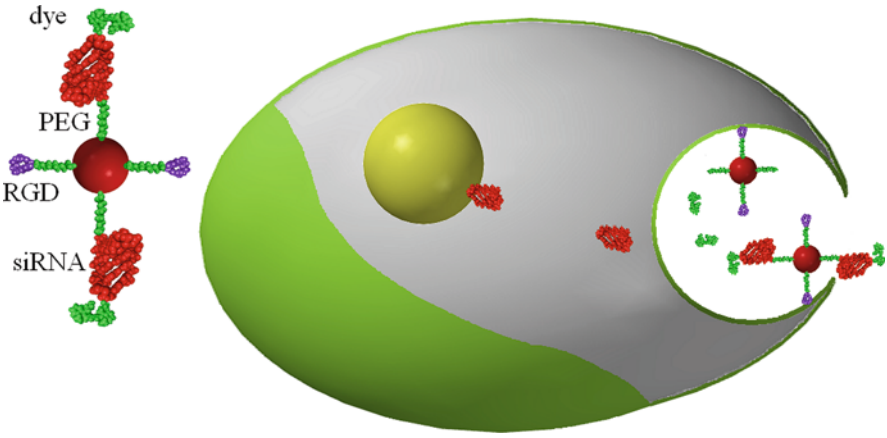



**Fig. 8.9** A multifunctional nanoparticle: the cRGD sequence binds to a cell surface, the nanoparticle is absorbed by endocytosis and can be tracked by a fluorescent dye before the particle disintegrates to deliver short interfering RNA (siRNA) sequences [23]

be injected into a healthy subject to cycle around in the bloodstream without having any effect unless diseased cells were encountered.

## 8.5 Cell Mimics as Drug Delivery Vehicles

Viruses are able to infect hosts by hiding within protein capsules, fusing with cells and delivering foreign genetic material without initial detection. It is no surprise that recent approaches to drug delivery have employed variations on this natural drug delivery vehicle to introduce their payload to an unwitting host organism. One basic method is to prepare a nanoparticle containing a drug and introduce it to the target cells. This can be achieved using vesicles or micelles, prepared in the presence of the drug, which then decompose under specific conditions. Alternatively, polymeric nanoparticles can be created that contain a drug which can be released in response to a particular stimulus such as a change in concentration of a key chemical, as illustrated in Fig. 8.10. Cancer cells are often targeted, not just because of the importance of tackling the disease but because cancer cells are often well defined as solid tumours and the progress of the drug can be monitored by numerous imaging techniques. Cancer cells also have other properties which can be used against them, notably the necessity for a complex blood system, or vasculature, and the rapid replication of generic material. Interfering with either of these can lead to a therapeutic regime.

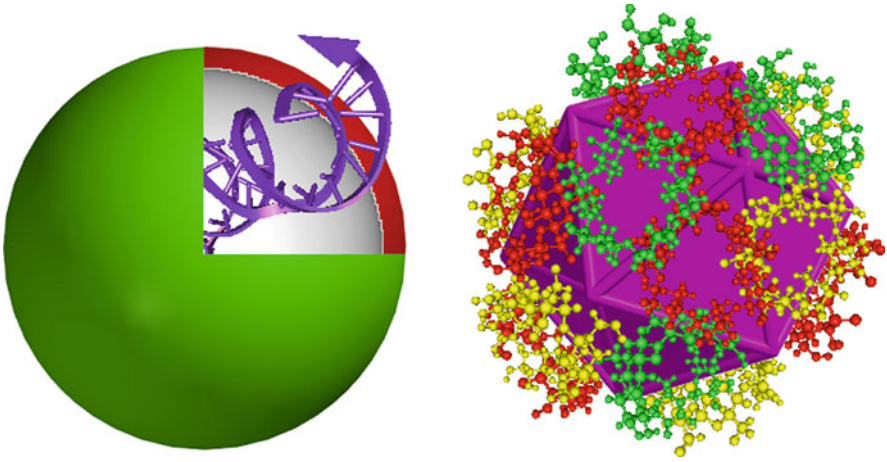


**Fig. 8.10** Stealth nanoparticles: RNA delivery by a liposome (*left*) and drug delivery by an icosahedral cyclodextrin aggregate

### *8.5.1 Polymer Encapsulated siRNA Delivery*

The first example of a bionanopharmaceutical that exemplifies the range of supramolecular chemical motifs that can be harnessed to deliver drugs in a

biological setting has recently been developed. The general principle combines siRNA delivery by controlled release from a surface functionalized capsule which mimics a capsid virus or cell.

The complex multicomponent system, called CALAA-01, has been developed by Davis for Calando Pharmaceuticals and is the first of many possible RONDEL$^{TM}$ therapeutics based on a biomimetic approach to drug delivery [24]. The approach combines a linear cationic polymer that incorporates cyclodextrins, a therapeutic payload (siRNA strands that target a specific process), and adamantane molecules modified with biocompatible polyethylene glycol chains (PEGs) or complementary proteins that bind to the target cell types.

Supramolecular interactions are key to the success of the system. The positively charged polymer wraps around the negatively charged siRNA strands through mutual electrostatic attraction to produce spherical particles about 100 nm in diameter. Adamantane is hydrophobic and so is held within a protective cyclodextrin cavity that shields it from the surrounding aqueous medium. The protein or PEG substituents extend from the polymer core and interact with surface features on cells. This elegant system has been used to target cancer cells and destroy them by delivering siRNAs that target a subunit of the ribonucleotide reductase. To achieve this some of the adamantane groups have transferrin proteins attached. The nanoparticles are injected into the blood stream where they circulate harmlessly, without eliciting an immune system response due to the PEGs which confer biocompatibility, and eventually enter tumours. This is possible for two reasons: firstly, the PEG groups prevent particle aggregation so that they can enter cells through very small openings and, secondly, the vasculature of cancer cells is particularly leaky, far more so than that of healthy tissue. Once through the tumour vasculature the transferrin proteins bind to receptors on cell surfaces and the particle is captured by endocytosis. The low pH of the tumour (another specific peculiarity of cancer cells) triggers the unravelling of the particle's polymer shell and releases the siRNA molecules. These in turn block the tumour's protein production abilities and kill the cell.

The importance of this approach to drug delivery is that it is modular so that variations can be made to each component. Thus the adamantane substituents can be varied to elicit different biological responses or bind to other protein receptors. The polymer can be varied so that it binds to different guest molecules, incorporates different macrocycles or unravels under different conditions. Finally, and most importantly, the guest can be varied from the large, charged siRNAs down to small drug molecules.

### *8.5.2 Drug Delivery by Particle Disintegration*

A similar process for preparing functional cyclodextrin nanospheres has been reported by the Harada group [25]. β-Cyclodextrins were exhaustively thiolated at the lower rim and assembled around colloidal gold nanoparticles. Following assembly around the gold template the cyclodextrins were linked through the formation

of disulfide bridges. The composite was treated with iodine to remove the gold and the remaining spherical structure purified by dialysis. Analysis showed that the gold was completely removed by iodine and that the remaining nanoparticle had a mass of 12 kDa, corresponding to 20 linked cyclodextrin units. This fits with the icosahedral geometry found in many pseudospherical structures such as viral capsids. The hydrodynamic radius of the gold complex was 2 nm; once the gold had been removed this shrank slightly to 1.8 nm. δ-Valerolactone was polymerized by the nanoparticles and it was found that several of the oligovalerolactones could bind to secondary hydroxyl sites on the cyclodextrin framework. In a further step it was possible to thread more cyclodextrins on these oligovalerolactone 'spokes' to form pseudorotaxanes. While the therapeutic potential of these compounds was not considered, the similarity with spherical viruses was noted. The nanoparticles have the potential to release entrapped materials from their cores, analogous to viral RNA release, and also to have a coating of macrocycles, such as cyclodextrin-based prodrugs, that can be released from the pseudorotaxane threads in response to external stimuli.

### *8.5.3 Minicells as Drug Delivery Systems*

While large synthetic capsules make good delivery systems they do have problems, notably they have to avoid triggering a response from the patient. Liposomes can be filled with a therapeutic payload in a Trojan horse approach to drug delivery but have the complication that the external membrane must be broken to allow the contents to leave. One way around this is to use cells as the delivery vehicle. Brahmbhatt and co-workers have used 'minicells' and filled them with chemotherapeutics, siRNAs or a combination of the two [26]. The minicells are buds, around 400 nm in diameter, from genetically engineered *Salmonella typhimurum*. The buds have a complete bacterial cell wall and are more robust than liposomes and similar carriers. As with other nanoparticles these minicells can be modified but through alterations made to the lipopolysaccharides on the surface of the outer membrane rather than covalent links to metal or oxygen atoms. Attachment of antibodies to particular tumour types gives the minicells target specificity. The minicells can be filled with therapeutic agents which, given their size, could include extensive nucleic acid sequences to target particular genes.

Minicells have been used to address one of the most troubling issues in cancer treatment, the increase in drug resistant tumours. By first targeting the gene responsible for drug resistance with disabling siRNA then, a few days later, a second treatment with chemotherapeutics to which the tumours were previously resistant, the Brahmbhatt group has successfully treated resistant tumours in mice, dogs and primates. Two aspects of the therapy are crucial to its success. Treatment with siRNA should be possible in theory but siRNAs are hard to administer and tend not to survive long enough to reach their targets inside cells. The minicells are able to carry the siRNAs directly to the tumour where their outer surfaces are bound

through antibody-antigen interactions. At this close proximity the siRNA fragments are able to leave the minicell through endogenous transmembrane protein channels and affect the target. Normal liposomes are unable to recognise cancer cells nor can they release their contents in a controlled manner. The second key feature of the minicells is their size. They are too large to pass through the walls of healthy blood vessels but can pass through the tumour vasculature. This, in concert with surface antibodies, allows the minicells to target only tumour cells. As a final point the minicells, despite being derived from *S. typhimurum* do not seem to trigger an immune response so that patients are unlikely to become resistant to therapy.

## 8.6 Supramolecular Protein Engineering

Supramolecular protein engineering, a term coined by Katsura in 1987, refers to the synthesis of proteins from DNA with artificial sequence deletions or duplications [27]. He was able to show that by deleting or duplicating parts of gene *H* the protein 'tail' of bacteriophage Λ could be sequentially shortened. The concept has more recently been applied to the artificial synthesis of peptides with useful properties. For example, short peptides with complimentary β-strand sequences can self-assemble into structures dictated by the inter-peptide interactions. By placing such a sequence at each terminus of a short flexible linker a bend is induced leading to the formation of a β-hairpin structure. These structures can aggregate to form fibres reminiscent of the amyloid protein entanglements that characterize plaques associated with prion diseases, Alzheimer's disease and similar conditions. The motif is also found in some spider silks. Here it is possible to see how applying supramolecular principles to the design of simple peptides can lead to structures that resemble those that form in diseases for reasons that are not well understood. Analysis of the factors that lead to the formation of artificial peptide fibres may well lead to clinical interventions to inhibit the formation of unwanted plaques in vivo.

## 8.7 Antimicrobial Limpet Mines

A central aspect of supramolecular chemistry is the ability to assemble groups of molecules each able to do a different tasks such that the functionality of the whole is greater than the sum of its parts. The same principle can be applied to nanoscale objects designed with therapeutic functions in mind. While we are a long way from the day when intelligent nanorobots travel through the bloodstream to deliver nanogenes or magic bullets to treat disease or injury, the potential exists to use multifunctional micron sized objects in medicine and allied fields. The use of surface modified nanoparticles for recognition or imaging is relatively straightforward. To use such vectors as targeted drug delivery systems is also well explored: porous drug saturated polymer dendrimers with peptide recognition sequences at their tips can accomplish this. Advances in materials are leading to more complex systems that

can perform multiple linked tasks. One such example is the development of zeolite L as a bateriocide by De Cola and colleagues [28], illustrated in Fig. 8.11.

Zeolites are porous aluminosilicate minerals, formed naturally or in the laboratory, that contain channels of different dimensions. They can be used to trap small molecules, including water, and have found uses as 'molecular sieves'. The channel framework, in which guest molecules can become entrapped, usually runs through all three dimensions but the recently synthesized zeolite L has channels that run in parallel down one dimension that are large enough for aromatic dye molecules. Crystals of zeolite L about 50 nm along each edge were filled with a green fluorescent dye so that the otherwise colourless particles could be imaged. Subsequently they were treated with silicon phthalocyanine derivatives that adhered to the particles' external surfaces. As discussed earlier these molecules are used in photodynamic therapy where upon irradiation they generate cytotoxic $^1O_2$ from atmospheric $^3O_2$. Finally the remaining surface of the nanocrystal was coated with a short chain amine. Untreated crystals have a negative surface and would be repelled from bacteria but the addition of amine groups gave them an affinity for the external bacterial membrane, allowing them to stick like a navel limpet mine. The multifunctional nanocrystals were observed, by normal and fluorescence microscopy, to adhere to antibiotic resistant *Escherichia coli* and *Neisseria gonorrhoeae* where they photoactivated oxygen that killed all the bacteria over a 2 h period of irradiation at wavelengths between 570 and 900 nm.

What has been developed is a therapeutic system that targets a specific biological feature, bacterial membranes, and can be imaged while it treats the source of the illness, by killing the bacteria, through photochemistry. It is not hard to think of many other applications for this system. Replacing the surface amines with peptide sequences associated with any number of biological targets and flooding the zeolite channels with drugs to be released over time would result in a highly specific nanosyringe which could deliver very potent drugs to the point where they were needed with submicron accuracy.

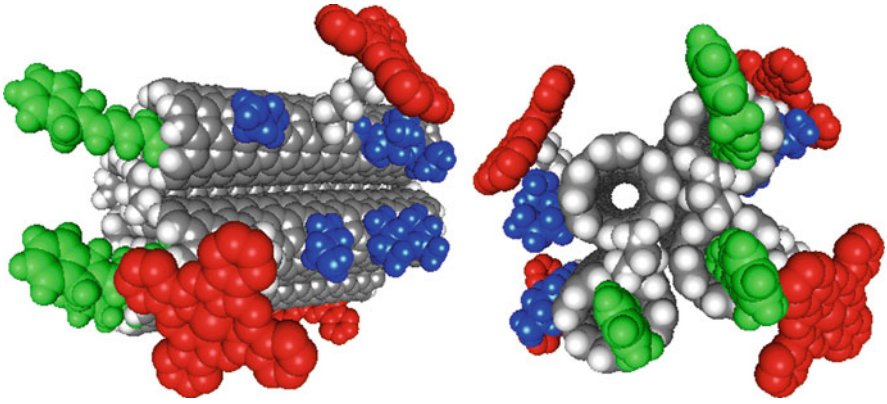


**Fig. 8.11** A nanoscale limpet mine decorated with bacteria recognizing amine groups (*blue*), tracking dyes (*green*) and therapeutic superoxide-generating phthalocyanine groups (*red*) [28]

Nanocrystals or supramolecular capsules containing potent drugs could have their surfaces modified to incorporate specific biomolecules that react to chemical signals within the body. Chemotaxis, physical movement of cell or bacterium in response to an increasing concentration of chemicals, could be mimicked by such a nanoscale drug delivery system through its surface interactions with peptides and other species associated with a particular condition. The surface interactions would determine the trajectory towards the target and, once there, the drug payload would be delivered exactly where required. It would not even be necessary to have had the condition diagnosed as these nanocontainers could be administered as a precaution and activate only when responding to the specific chemical that is released as the condition progresses from benign to harmful.

## 8.8 Future Directions

Science is often inspired by the creative ideas of science fiction writers and nanomedicine is no exception. Functional medical nanodevices have entered into popular culture through the fantasy worlds of *Doctor Who*, as nanogenes, *Star Trek*, as nanites and nanoprobes, and numerous other sources, as nanorobots and nanobots. Similar themes exist in films such as *Fantastic Voyage* and *Inner Space*. In all these examples writers envisaged a future in which healthy bodies are maintained through the intervention of microscopic machines which are able to enter and circulate through the body fixing whatever problem is encountered. How realistic are these scenarios? To answer the question several issues need to be considered. How large, or small, does the device need to be if it is to be functional? How is it to be powered? How will it work and what will it do? How do we know that it works? How can it be controlled?

### *8.8.1 Medicinal Nanodevices*

Pharmaceutical molecules such as aspirin are at the bottom end of the nanoscale and have one function, to interact with the appropriate biological receptors when they are encountered, but they lack the multifunctional design element implicit in the application of nanotechnology to medicine. A medical nanodevice must be expected to have multiple functions assuming that its purpose is to seek out and treat an abnormal condition. The device must comprise a recognition element, a payload, a method of delivery, and possibly a method of propulsion. Recognition may require the device to be coated with a peptide or molecule complementary to the target surface. The payload must be encapsulated within the device, just as in viral capsids, so the size of the device depends upon the size and number of molecules it holds. Delivery may be as simple as device disintegration or may require some mechanochemical response. Propulsion could come from a mechanical source, by analogy to bacterial flagellae, or through some form of chemical reaction that creates

a jet of ions or molecules to move the device in the desired direction. Some of the possibilities are discussed below. Given all these requirements it is likely that a medical nanodevice would have to be in the region of 50–500 nm, the size of a small virus. The problem is that structures on this scale feel far greater effects from water resistance and changing chemical composition of solutions than much larger entities. Consequently there may be little point in trying to control their motion and leave them to the mercy of natural currents. If larger nanodevices were used then they would encounter problems when trying to pass through small vessels like the vasculature of newly formed tumours. If the devices were any smaller they risk being attacked by immunoresponsive cells and removed by endocytosis and excretion.

### *8.8.2 Powering Nanodevices*

Although some progress is being made in nanoscale propulsion methods it may be easier to use an alternative approach, the most obvious being to simply use the circulation of blood to move the device. If this were coupled to chemotaxis, so that the device was attracted to an increasing concentration of chemicals associated with the target site, it could circulate freely until it picked up the target's chemical signature.

Assuming that the devices were required to move then Nature has shown the way with bacterial flagellae. To reproduce this effect the Dreyfus group showed that a red blood cell with an appended chain of colloidal magnetic particles can be propelled by the same mechanism [29]. Golestanian and co-workers have considered the possibility of asymmetric chemical release from a nanodevice [30], essentially a variation on the jet propulsion methods used by squids, which has led to other chemically induced propulsion mechanisms [31]. The simplest method of propulsion is by direct chemical reaction on the surface of the device. Currently this means that the surface must catalyse a reaction that generates molecules which form the jet. To date the best examples involve the breakdown of hydrogen peroxide, which is not commonly found in vivo, so an alternative reaction needs to be found. The main biological power source is the decomposition of ATP but this requires numerous enzymes and complex protein interactions to transform the reaction itself into motion. Alternatively, nanodevices could be manipulated externally if they respond to magnetic or electrical signals but to be effective the operator must know in advance where they need to be sent. This may be appropriate for some applications but not if the devices were to be used to attack targets in unknown locations.

### *8.8.3 Functional Nanodevices*

As noted above, it is most probable that future therapeutic nanodevices will operate in a prophylactic manner. They will circulate within the blood stream until attracted to a target through a chemical signal. After following the gradient of the signal they

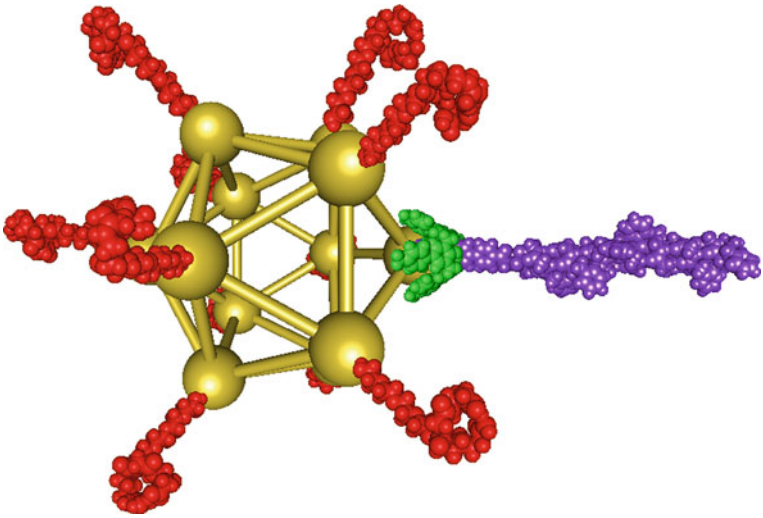


**Fig. 8.12** A functional nanoscale medical device incorporating targeting protein sequences (*red*), a flagella-type propulsion mechanism (*green* and *purple*) and drug delivery vehicle (*gold*)

will attach to the target where they will deposit their payload before being excreted. Ideally the devices will circulate safely until they have been activated which makes activation coupled to decomposition into molecules that can be metabolized the preferred strategy.

The targets for therapeutic nanodevices could be a recently formed malignant cell or amyloid plaque, to be destroyed by a localized high dose of a highly specific drug, or a compromised biological function, such as impairment of transmembrane ion transport as occurs in cystic fibrosis, cellular uptake of copper as in Wilson's disease, or problems with an impaired immune system. Larger scale examples may be the regrowth of bone following a fracture and repair of tissue following spinal cord injury. All of these could be addressed through functional devices homing in on chemical targets as shown in Fig. 8.12.

Clearly many aspects of medicinal chemistry, nanoengineering and supramolecular chemistry will have to be brought together if functional nanodevices are to be manufactured and used in vivo. The main challenge will be to verify that the devices work as anticipated and, most importantly, cause no harm.

### *8.8.4 Verification of Treatment*

If, hypothetically, a patient was given a dose of therapeutic nanodevices for a specific condition how could success be judged? Obviously if the symptoms and underlying condition were treated successfully it would count as a success but it may take a considerable time to find this out. If the devices were prophylactic in nature, perhaps to seek out and destroy a particular form of cancer to which an

individual had a genetic susceptibility, then it may never be obvious that they were effective. The only way to assess success would be to couple a secondary function to the device, preferably one that allowed imaging. Radiotracers would be unsuitable so magnetic and near infrared responses would be required. The former would allow identification of the devices positions by MRI, assuming that they accumulated in reasonable numbers anywhere, the latter would allow imaging through the skin, but only if they accumulated subcutaneously. Neither solution is ideal which leaves direct sampling of blood or tissue. Until this issue is resolved it will be almost impossible to determine if therapeutic nanodevices have any acute effects.

### *8.8.5 Nanodevice Control*

One of the biggest problems facing the nanotech industry is the safety of nanoscale materials. The effects of simple nanoparticles on humans and the environment are relatively unknown so the issues surrounding any proposed multifunctional nanoscale therapeutic device on patients will undoubtedly have to be given great scrutiny. These issues do not relate to the naive idea of self-replicating nanomachines reducing the world to 'grey goo' that has arisen in some circles, as the nanodevices will have no such potential, but to the adverse effects they may have on the public. Control of non-replicating nanodevices relates to the ability to track their progress and biological fate.

## 8.9 Supramolecular Chemistry and Nanomedicine

Throughout this book it has been the intention to show how the field of supramolecular chemistry has drawn inspiration from biology. The weak, reversible interactions central to the former are also essential features of the latter whether they are associated with molecular recognition, supramolecular (or protein) aggregation, natural or artificial photosystems, or drug receptor interactions. Where Nature has pioneered systems, supramolecular chemists have sought to design analogues for non-biological purposes. The same principle that governs the allosteric signalling of a binding event can be used to design a molecule that will detect disease by MRI. It is therefore entirely appropriate that supramolecular chemistry is at the heart of nanomedicine whether it is in the design of receptors that respond to the specific chemistry of a tumour or a multifunctional nanocapsule that releases its payload in response to external stimuli.

## References

1. Seeman N (2004) Nanotechnology and the double helix. Sci Am 290(6):34–43
2. Zheng J et al (2009) From molecular to macroscopic via the rational design of a self-assembled 3D DNA crystal. Nature 461:74–77 PDBID:3GBI

3. Shih WM, Quispe JD, Joyce GF (2004) A 1.7-kilobase single-stranded DNA that folds into a nanoscale octahedron. Nature 427:618–621
4. Goodman RP et al (2005) Rapid chiral assembly of rigid DNA building blocks for molecular nanofabrication. Science 310:1661–1665
5. He Y et al (2008) Hierarchical self-assembly of DNA into symmetric supramolecular polyhedra. Nature 452:198–201
6. Seeman NC (2005) From genes to machines: DNA nanomechanical devices. Trends Biochem Sci 30:119–125
7. Mullis K (1999) Dancing naked in the mind field. Bloomsbury, London, pp 3–14
8. Ramachandran GN, Kartha G (1955) Structure of collagen. Nature 176:593–595
9. Kwahara K et al (2005) Effect of hydration on the stability of the collagen-like triple-helical structure of [4(R)-hydroxyprolyl-4(R)-hydroxyprolylglycine.$_{10}$. Biochem 44:15812–15822 PDBID:1WZB
10. Ybe JA et al (2007) Crystal structure at 2.8 Å of the DLLRKN-containing coiled-coil domain of huntingtin-interacting protein 1 (HIP1) reveals a surface suitable for clathrin light chain binding. J Mol Biol 367:8–15 PDBID:2NO2
11. Schmid SL (1997) Clathrin-coated vesicle formation and protein sorting: an integrated process. An Rev Biochem 66:511–548
12. Jiménez MC, Dietrich-Buchecker C, Sauvage JP (2000) Towards synthetic molecular muscles: contraction and stretching of a linear rotaxane dimer. Angew Chem Int Ed 39: 3284–3287
13. Kryatova OP et al (2002) Stable supramolecular dimer of self-complementary benzo-18-crown-6 with a pendent protonated amine arm. Chem Commun 3014–3015
14. Kushner AM et al (2009) A biomimetic modular polymer with tough and adaptive properties. J Am Chem Soc 131:8766–8768
15. von Castelmur E et al (2008) A regular pattern of Ig super-motifs defines segmental flexibility as the elastic mechanism of the titin chain. Proc Natl Acad Sci USA 105:1186–1191 PDBID:3B43
16. Alivisatos PA (2001) Less is more in medicine. Sci Am 285(3):58–65
17. Ito A et al (2005) Medical application of functionalized magnetic nanoparticles. J Biosci Bioeng 100:1–11
18. Storhoff et al (1998) One-pot colorimetric differentiation of polynucleotides with single base imperfections using gold nanoparticle probes. J Am Chem Soc 120:1959–1964
19. Jeffreys AJ, Wilson V, Thein SL (1985) Individual-specific 'fingerprints' of human DNA Nature 1985 316:76–9
20. Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74:5463–5467
21. Borsenberger V, Mitchell N, Howorka S (2009) Chemically labeled nucleotides and oligonucleotides encode DNA for sensing with nanopores. J An Chem Soc 131:7530–7531
22. Chang H et al (2004) DNA-mediated fluctuations in ionic current through silicon oxide nanopore channels. Nano Lett 4:1551–1556
23. Lee JH et al (2009) All-in-one target-cell-specific magnetic nanoparticles for simultaneous molecular imaging and siRNA delivery. Angew Chem Int Ed 48:4174–4179
24. Schluep T et al (2009) Pharmacokinetics and tumor dynamics of the nanoparticle IT-101 from PET imaging and tumor histological measurements. Proc Natl Acad Sci USA 106: 11394–11399
25. Osaki M et al (2009) Nanospheres with polymerization ability coated by polyrotaxane. J Org Chem 74:1858–1863
26. MacDiarmid JA et al (2009) Sequential treatment of drug-resistant tumors with targeted minicells containing siRNA or a cytotoxic drug. Nature Biotech 27:643–651
27. Katsura I (1987) Determination of bacteriophage-lambda tail length by a protein ruler. Nature 327:73–75

28. Strassert CA et al (2009) Photoactive hybrid nanomaterial for targeting, labeling, and killing antibiotic-resistant bacteria. Angew Chem Int Ed 48:7928–7931
29. Dreyfus R et al (2005) Microscopic artificial swimmers. Nature 437:862–865
30. Golestanian R, Liverpool TB, Ajdari A (2005) Propulsion of a molecular machine by asymmetric distribution of reaction products. Phys Rev Lett 94:Article 220801
31. Paxton WF et al (2006) Chemical locomotion. Angew Chem Int Ed 45:5420–5429

# Index


**A**
Alamethicin, 162, 179, 224
Aldolase, 21–22
Alkaline phosphatase model, 135
Alzheimer's disease, 219–222, 227, 449
Amino acid, 2, 9, 12, 14, 36, 50–55, 58, 62, 64, 69, 71–72, 76, 78, 83, 87, 104, 115, 128, 132–134, 138, 142–143, 147, 158, 163, 174, 197–198, 201, 204, 220–221, 225–227
Amphotericin, 157, 173, 179, 224–225
Amyloid plaque, 220, 253
Amyloid protein, 221, 227, 249
Antibiotic, 155, 157–158, 195, 224–227, 250
Aquaporin, 162
Artificial ion channel, 173, 175, 178
ATPase model, 144, 167, 210
Aza[18]crown-6, 26–28, 176, 178, 197–199

**B**
Baas-Becking, LGM, 3
Bilayer–12, 36, 66, 81, 86, 91–92, 105, 108–109, 122, 137, 154–155, 157–158, 160–162, 168–174, 176–180, 224–225, 242
  electrochemical techniques, 42–43
Biomineralization, 102–103
Bionanotechnology, 231–254
Biosynthesis, 21–22, 53–54, 103, 108, 210, 219
  amino acid, 53–54
Bonds
  coordinate, 9–10
  covalent, 3, 5, 8–12, 67, 78–79, 96, 99, 221
  disulfide, 10, 80, 244–245
  hydrogen, 2–5, 9, 12–13, 16, 22, 30, 40–41, 44, 54–56, 60–61, 65, 67–80, 84, 93, 96, 104, 106, 135–136, 140, 146, 158, 168, 200, 221, 225, 232, 238, 243–244
  imine, 9–10, 16, 69, 76, 117
  ionic, 11, 172
  reversible, 7
Borromean rings, 35–36

**C**
$C_{60}$, 24, 92, 215
Calix[4]arene, 20, 22, 73, 135, 177–179, 223
Calix[6]arene, 129
Calix[8]arene, 22, 222
Calixpyrrole, 23, 203
Calixtube, 96, 178
Cancer, 2, 25, 65, 127, 195, 199, 201–202, 211, 213, 215–217, 219, 243–249, 253
Carbonic anhydrase, 117, 132–135, 146, 202
Carbonic anhydrase model, 117, 132–135, 146, 202
Carboxypeptidase, 132–134, 143
Catalase, 124–125, 127
Catenane, 30–37
Cavitand, 23, 74, 96
Cell
  aggregation, 109–110
  artificial, 91–110
Chelation therapy, 208–211
Cisplatin, 217
Clathrin, 92–94, 234–235
Coat protein, 94
Coccolith, 91, 102–103
Codon, 36, 55, 62, 64, 72
Collagen, 211, 234–235
Combinatorial
  chemistry, 78, 144
  library, 9, 145, 233
Controlled release, 104, 216–218, 247
Cowpea virus, 94, 107
Cram, D. J., 4, 23, 25, 74, 96, 100
Crick, F. C. H, 60–61
Critical analyte, 191


P.J. Cragg, *Supramolecular Chemistry*, DOI 10.1007/978-90-481-2582-1,
© Springer Science+Business Media B.V. 2010

[18]Crown-6, 26, 41, 171–172, 175–176, 197–198
[21]Crown-7, 174–175
Crown ether, 18, 22, 25–27, 30, 32, 34, 42, 74, 119, 143–144, 171–172, 174–178, 187, 191, 193, 195, 197–199, 237
Cryptand, 27–29, 95, 171, 191
Cyanuric acid, 75
Cyclam, 17, 134, 217–219
Cyclodextrin, 22, 28–29, 34, 122, 142–143, 176–177, 246–248
  esterase model, 122, 143
Cyclophane, 24–25
Cyclotriveratrylene, 23–24
Cytochrome, 25, 121–123, 128–131, 138
  model, 123
Cytochrome c oxidase, 128, 130–131
  model, 131

**D**
Dendrimer, 145, 216–218, 227, 249
Desferrioxamine, 209–210
Diaza[18]crown-6, 27, 176, 178
Dibenzo[18]crown-6, 26
Displacement assay, 191–192
DNA
  A, 64–65
  B, 64–65
  fingerprinting, 240
  polymerization, 11, 87
  replication, 7, 38, 62, 77–78, 84, 104, 224, 231, 234
  sequencing, 37, 242–244
  transcription, 12, 62, 64, 69, 107–108, 133, 232
  Z, 64–65
DOTA, 200–202, 211–212, 216
Doxorubicin, 217

**E**
Electrochemical techniques, 42
Enzyme, 4, 7, 9, 12, 21, 28, 37, 53, 55, 62–64, 66, 78, 81, 83–84, 104, 108, 113–148, 153, 173, 196, 202, 207–208, 210–211, 219–220, 222, 231–234, 240, 252
  mimics, 21, 113–148
Evolution
  enzyme, 146
  molecular, 38, 79–80, 232
  replicator, 68–69

**F**
Fieser, L.E., 4
Fischer, E., 4, 55–56, 81, 113–115

**G**
Galliher, E. W., 3
Genome, 37, 108–109, 240–242
Glycoprotein, 58–59, 94, 218
G-quadruplex, 65, 171
Gramicidin, 157–159, 162, 178
Grid, 14, 37–38

**H**
Haemoglobin, 10–11, 25, 44, 55, 119–121, 146, 196
  model, 119–121
α-Helix, 56, 78, 134, 162, 166, 173–174
α-Hemolysin, 36, 109
HIV, 64, 214, 217–219, 242
Holliday junction, 65–66
HSAB (hard and soft acid and base), 15
Hydraphiles, 176
Hydropathy, 52–53, 168, 180
Hydrophobic effect, 4–5, 13–14, 55, 59, 68, 74, 80, 221

**I**
Induced fit, 56, 115
Interaction
  π-π, 2, 13, 32, 146, 197, 221, 244
  anion-π, 13
  cation-π, 13
  dipole-dipole, 12–13
  host-guest, 41
  ion-dipole, 11
  ion-ion, 11
  non-covalent, 2, 5, 53, 243
  protein-protein, 55, 221, 228
  van der Waals, 2, 13
Ionophore, 30, 155–156, 171–172, 174, 177, 180, 191
Iron-sulfur world, 83

**J**
Jack bean urease, 114

**K**
Katapinand, 27

**L**
Lariat ether, 26–28
Lehn, J-M, 3–5, 9, 27, 31, 37, 96, 143, 145, 175
Life, 49–87, 103, 105, 153, 194, 199, 211
Ligand, 4–5, 10, 14, 16, 18, 30, 37, 40, 98–99, 117–118, 122, 135, 141, 146, 156, 158, 200, 209, 215–216, 218, 221–222, 236

Lipid
bilayer, 12, 36, 92, 105, 108, 122, 137, 154, 157, 162, 169, 171–174, 176, 178–179, 225, 242
capsule, 105–106
world, 80–83, 103, 105
Liposome, 81, 109, 170–171, 246, 248–249
Lock and key, 4, 55, 113, 115
Logic gate, 190
Luria, S. E., 4

**M**
Macrocycle, 17, 19–31, 33–34, 42, 96–97, 120, 129, 134, 136, 142–143, 146, 173, 175, 177, 185, 200, 202, 212–214, 216, 218, 224, 235–238, 247–248
*Macrocyclon*, 222–223
Magic bullet, 207–208, 213, 216, 224, 249
Magnetic resonance imaging (MRI), 207–208, 213, 216, 224, 249
Mannopeptimycin, 225–226
Mass spectrometry, 39–40
Maxwell's demon, 34–35
Melamine, 75
Mellitin, 162, 225
Menke's syndrome, 210–211
Micelle, 12, 66, 80, 82, 105, 201, 246
Molecular
imprinted polymer, 144
mechanics, 43, 168
modelling, 43–44
muscles, 235–238
recognition, 4, 57–58, 91, 153, 185, 191, 193, 224, 234, 254
Monensin, 155, 179, 224–225
MRSA, 223, 225, 227
Myoglobin, 10, 25, 119–121

**N**
Nalidixic acid, 223–224
Nanodevice, 6, 251–254
Nanofabrication, 8
Nanogene, 249, 251
Nanomachine, 6, 233, 254
Nanomaterial, 6, 66
Nanomedicine, 231–254
Nanoparticle, 239–240, 244–249, 254
Nanopore, 242–244
Nanorobot, 249, 251
Nanoscale, 1, 6, 32, 36, 91, 231–235, 237, 242, 244, 249–254
Nanotechnology, 5–8, 231, 242, 251
Negative entropy, 49
Nitric oxide (NO), 123, 195–197, 221
detection, 195–197

**O**
Osmometry, 41–42

**P**
Palade, G. E., 4, 36
Pedersen, C. J., 4, 22, 26–27
Penicillamine, 209, 211
Photodynamic therapy, 211–214, 250
Photosynthesis, 57, 83, 136–137, 139
Photosystem mimic, 16, 139
Phthalocyanine, 25, 213–214, 250
Podand, 18–19, 171, 191–192
Polyamine, 17, 30, 204
Polyether, 17–19, 26–28, 30–31, 34, 176, 178–180, 193–194, 222, 235
Polymerase chain reaction (PCR), 87, 107, 118, 233–234, 240
Porphyrin, 25, 34, 116–117, 119–122, 125, 140, 146, 196, 211–213, 215–216
deactivation, 120
Prion, 49, 249
Prodigiosin, 156–157, 172
Prodrug, 207–208, 216–217, 221, 223–224, 227, 244, 248
Protein, 1–4, 7, 9–14, 25, 28, 30, 36–39, 42–43, 49–52, 54–62, 64, 69, 78–81, 83–84, 86–87, 91–95, 103–104, 107–109, 113–118, 120, 122, 124, 126–129, 132–134, 138, 141, 146–147, 153–155, 157–159, 161–170, 174, 185–186, 192, 196, 200, 208, 210, 215–216, 218–221, 224–225, 227–228, 232, 234–235, 237–240, 242, 244–247, 249, 252–254
synthesis, 12, 38, 61, 69, 108, 153, 231
Purine, 12, 54, 60–63, 72, 243
Pynchon, T., 3
Pyramidine, 12, 60–61, 72, 85–86, 243
synthesis, 85
Pyrogallolarene, 13, 17, 19, 24, 26, 31, 99–100, 105, 135

**Q**
Quartz crystal microbalance, 195–196

**R**
Radiotherapeutics, 212
Replication
artificial, 57, 102, 108, 234
biological, 7, 69, 153, 224
Replicator, 66–72, 79, 153
Resorcinarene, 20, 23

Ribosome, 36, 38, 54, 61–62, 64, 69, 72, 84, 87, 108–109
Ribozyme, 84
RNA
  messenger (mRNA), 1, 36, 54, 62, 64, 107–108, 244–245
  model, 129
  ribonuclease (RNAse), 129, 131, 136, 142
  ribosomal, 64, 108
  short interfering (siRNA), 244–249
  transfer (tRNA), 36, 54, 64, 107
  vault (vRNA), 95
Rotaxane, 29–31, 33–37, 194, 235, 237, 248
  natural, 36–37

**S**
Salen, 16, 117, 126–127, 221
Salphen, 197
Sarcophagene, 27–28, 95
Schiff base, 16–17, 19, 98, 117, 126, 217–218
Self-assembly, 5, 14, 44, 73–74, 76, 78, 107
Self-replication, 57, 66–79, 83–86, 145
Sensor
  citrate, 191–192
  electrochemical, 195–197
  fluorescent, 187–195
  GABA–212, 198
  Mercury, 198
  optical, 187, 191, 193
  Saxitoxin, 198
  Sodium, 198
Sepulchrate, 27–28, 95
β-Sheet, 12, 55–56, 158, 162, 220–221, 237
Siderophore, 156, 162, 171–172, 180, 209–210
Spectroscopy
  infrared (IR), 40, 194
  nuclear magnetic resonance (NMR), 40–41, 115, 169–171, 199
  ultraviolet/visible (UV/vis), 40, 188
Squaraine, 192–195
π-π Stacking, 5, 13, 84, 127, 143
Sugars
  disaccharide, 57–58
  lipopolysaccharide, 57, 248
  monosaccharide, 57
  polysaccharide, 57, 108, 248
Supermolecule, 3
Superoxide dismutase
  catalytic cycle, 125–126
  model, 126
Supramolecular
  biology, 50, 86–87
  capsule, 40–41, 74, 91, 93, 251
  complex, 2, 7, 14–19, 38, 40, 235
  protein engineering, 249
Supramolecule, 5, 7–9, 12, 14, 36, 39, 44, 73, 96, 145, 201, 207, 224

**T**
Technetium, 216
Texaphyrin, 214–216
Titin, 237–238
Transcription
  artificial, 108
  biological, 108
Translation
  artificial, 108–109
  biological, 108
Transmembrane
  anion transport, 156, 172
  cation transport, 178, 210
Transmembrane channel
  acid sensitive, 164–165
  AQP, 162
  ChR1, 161
  ChR2, 161
  ClC, 163
  gating, 157–161
  KcsA, 158, 164, 166–168
  $M_2$, 159, 161
  MscL, 160, 162
  RyR1, 169
  selectivity, 157–161
Trefoil knot, 31–32

**U**
*Übermolecül*, 3–4, 40

**V**
Valinomycin, 155, 167, 224–225
Vault protein, 28, 94–95, 110
Velcraplex, 74
Vesicle, 66, 81–82, 87, 92–93, 104–106, 109, 122, 159, 246
Viral capsid, 93–94, 248, 251
Viral capsid mimic, 251
Virus, 5, 49, 64, 93–94, 104, 107–108, 110, 114, 159, 161, 218–219, 246–248, 252
Vitamin $B_{12}$, 25, 49, 116

**W**
Watson, JD, 60–61
Wilson's disease, 210–211, 253
Wolf, KL, 3, 40

**X**
X-ray crystallography, 41, 130

**Z**
Zeolite, 250
Zinc finger, 133