

*Perspectives in
Supramolecular
Chemistry*

FOUNDED BY J.-M. LEHN



**The Lock and Key
Principle**

The State of the Art — 100 Years on

Edited by
Jean-Paul Behr

The Lock-and-Key Principle

Editorial Board

Founding Editor

J.-M. Lehn, Université Louis Pasteur, Institut le Bel, 4 Rue Blaise Pascal,
F-67070 Strasbourg, France

Editors

J.-P. Behr, Faculté de Pharmacie, B.P. 24, F-67401 Illkirch, France

G. R. Desiraju, University of Hyderabad, School of Chemistry, Hyderabad
500134, India

A. D. Hamilton, University of Pittsburgh, Department of Chemistry,
Pittsburgh, PA 15260, USA

T. Kunitake, Kyushu University, Faculty of Engineering, Hakozaki, Fukuoka
812, Japan

D. N. Reinhoudt, University of Twente, Faculty of Chemical Technology, PO
Box 217, NL-7500 AE Enschede, The Netherlands

J.-P. Sauvage, Université Louis Pasteur, Institut le Bel, 4 Rue Blaise Pascal,
F-67070 Strasbourg, France

The Lock-and-Key Principle

*The State
of the Art—
100 Years On*

EDITED BY JEAN-PAUL BEHR

Université Louis Pasteur de Strasbourg, Illkirch, France

Perspectives in Supramolecular Chemistry Volume 1

John Wiley & Sons

Chichester · New York · Brisbane · Toronto · Singapore

Copyright©1994 by John Wiley & Sons Ltd,
Baffins Lane, Chichester,
West Sussex PO19 1UD, England
Telephone: National Chichester (0243) 779777
International +44 243 779777

All rights reserved.

No part of this book may be reproduced by any means,
or transmitted, or translated into a machine language
without the written permission of the publisher.

Other Wiley Editorial Offices

John Wiley & Sons, Inc., 605 Third Avenue,
New York, NY 10158-0012, USA

Jacaranda Wiley Ltd, 33 Park Road, Milton,
Queensland 4064, Australia

John Wiley & Sons (Canada) Ltd, 22 Worcester Road,
Rexdale, Ontario M9W 1L1, Canada

John Wiley & Sons (SEA) Pte Ltd, 37 Jalan Pemimpin #05-04,
Block B, Union Industrial Building, Singapore 2057

British Library Cataloguing in Publication Data

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

ISBN 0 471 93902 1

Typeset in 10/12pt Times by Dobbie Typesetting Limited, Tavistock, Devon
Printed and bound in Great Britain by Biddles Ltd, Guildford, Surrey

Contents

Contributors	vii
Preface	ix
1 Emil Fischer's Lock-and-Key Hypothesis after 100 Years— Towards a Supracellular Chemistry Friedrich Cramer	1
2 Molecular Recognition in Biology: Models for Analysis of Protein–Ligand Interactions Doron Lancet, Amnon Horovitz and Ephraim Katchalski-Katzir	25
3 New Biocatalysts via Chemical Modification Ian M. Bell and Donald Hilvert	73
4 Oligonucleotides: Superspecific Ligands for Targeting Nucleic Acids and Proteins and Development of Molecular Devices V. V. Vlassov	89
5 Macrocycles and Antibodies as Catalysts D. B. Smithrud and S. J. Benkovic	149
6 Lock-and-Key Processes at Crystalline Interfaces: Relevance to the Spontaneous Generation of Chirality Isabelle Weissbuch, Ronit Popovitz-Biro, Leslie Leiserowitz and Meir Lahav	173
7 A Model of the Origin of Life and Perspectives in Supramolecular Engineering H. Kuhn and J. Waser	247
8 Perspectives in Supramolecular Chemistry—From the Lock-and-Key Image to the Information Paradigm Jean-Marie Lehn	307
Index	319

Contributors

Ian M. Bell, Department of Chemistry and Molecular Biology, The Scripps Research Institute, 10666 North Torrey Pines Road, La Jolla, CA 92037, USA

S. J. Benkovic, Chemistry Department, Pennsylvania State University, 152 Davey Laboratory, University Park, PA 16802-6300, USA

Friedrich Cramer, Max-Planck-Institut für Experimentelle Medizin, H. Rein Strasse 3, D-37075 Göttingen, Germany

Donald Hilvert, Department of Chemistry and Molecular Biology, The Scripps Research Institute, 10666 North Torrey Pines Road, La Jolla, CA 92037, USA

Amnon Horovitz, The Weizmann Institute of Science, Rehovot 76100, Israel

Ephraim Katchalski-Katzir, The Weizmann Institute of Science, Rehovot 76100, Israel

H. Kuhn, (formerly Max-Planck-Institut für biophysikalische Chemie, Göttingen) Ringoldwilstrasse 50, 3656 Tschingel ob Gunten, Switzerland

M. Lahav, The Weizmann Institute of Science, Rehovot 76100, Israel

Doron Lancet, The Weizmann Institute of Science, Rehovot 76100, Israel

Jean-Marie Lehn, Université Louis Pasteur, Institut le Bel, 4 Rue Blaise Pascal, F-67070 Strasbourg, France

L. Leiserowitz, The Weizman Institute of Science, Rehovot 76100, Israel

D. B. Smithrud, Chemistry Department, Pennsylvania State University, 152 Davey Laboratory, University Park, PA 16802-6300, USA

V. V. Vlassov, Institute of Bioorganic Chemistry, Lavrentiev Prospekt 8, 630090 Novosibirsk, Russia

J. Waser, (formerly California Institute of Technology) 6120 Terryhill Drive, La Jolla, CA 92037, USA

PREFACE

In 1894, a farsighted (bio)chemist came up with an idea that would provide the key to understanding the phenomena of molecular recognition that underlie life itself. After a century of refinement his concept still remains productive, as is evident from the many scientists who draw on it. This book is their tribute to the man who unlocked our thought.

Chapter 1

Emil Fischer's Lock-and-Key Hypothesis after 100 years—Towards a Supracellular Chemistry

FRIEDRICH CRAMER

Max-Planck-Institut für Experimentelle Medizin, Göttingen, Germany

1. INTRODUCTION

Around 100 years ago Emil Fischer in his famous paper [1] proposed that enzyme and substrate can be compared to lock and key. Since that time this metaphor has been used to describe enzyme action. In this chapter we shall try to explore whether this metaphor still holds. On the one hand, enzymology has made enormous progress in these 100 years, but on the other hand the lock-and-key concept has greatly changed. In Figure 1 three types of 'keys' are shown.

The classical key, which Emil Fischer had in mind, is pushed into the lock and turned clockwise in order to open the lock. Thus, the process is chiral. The key shown in the figure was manufactured in the year 1853, one year after Emil Fischer was born; thus, probably this type of key was imprinted in the young boy's conceptual memory. Today, the keys have become much more refined. Around the turn of the century the safety lock was invented, which does a kind of proof-reading of the key. At present an entirely new system is being installed; namely the magnetic card. This card can unlock money sources and hotel rooms. We shall discuss further these three key concepts and give a few examples of each of them.

2. CLASSICAL LOCKS

One almost classical molecule for lock-and-key studies is cyclodextrin. I entered the field of supramolecular chemistry 45 years ago with this molecule and it still

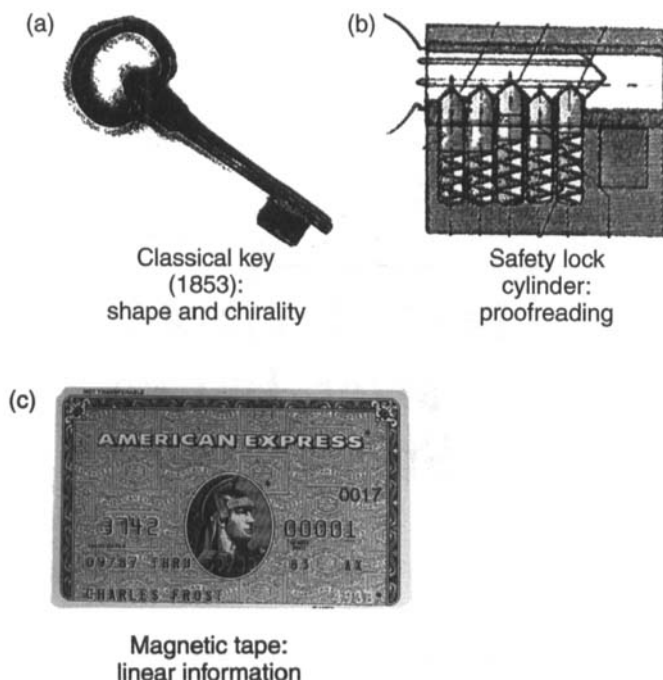


Figure 1 Lock and keys from the last 100 years: (a) key from 1853; (b) safety lock and key; and (c) magnetic tape as a key

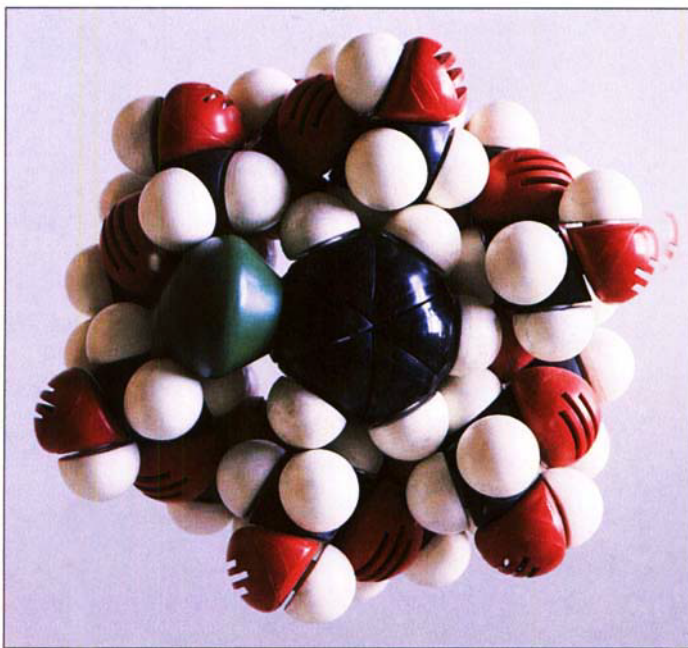
seems to be of interest. It has a well-defined hydrophobic cavity. It readily forms a complex with chlorobenzene but not with bromobenzene (Figure 2). Of special interest with respect to enzyme models are the kinetics of the formation of such inclusion compounds. The kinetics can be measured using azo dyes, which change their spectra on formation of the inclusion compound in solution. In Figure 3, all three molecules independent of the substituent R have the same equilibrium constant. However, the rate constants are different by five orders of magnitude because the stereochemistry of the 'threading-in process' becomes more unfavorable the larger the substituent R.

Numerous enzyme models are based on this principle. For instance, pyrophosphates, if suitably substituted, are hydrolyzed by the cyclodextrin cavity, or rather by the OH group of the glucose in cyclodextrin, 400 times faster than without cyclodextrin (Figure 4).

On a similar basis, we have constructed a chymotrypsin model by furnishing β -cyclodextrin with imidazole groups. Ester hydrolysis is 300 times faster, but does not yet come close to the actual catalytic rate of chymotrypsin (Figure 5).



(a)



(b)

Figure 2 Geometrical fit of chlorobenzene into α -cyclodextrin: (a) empty cavity and (b) cavity filled with chlorobenzene [2]

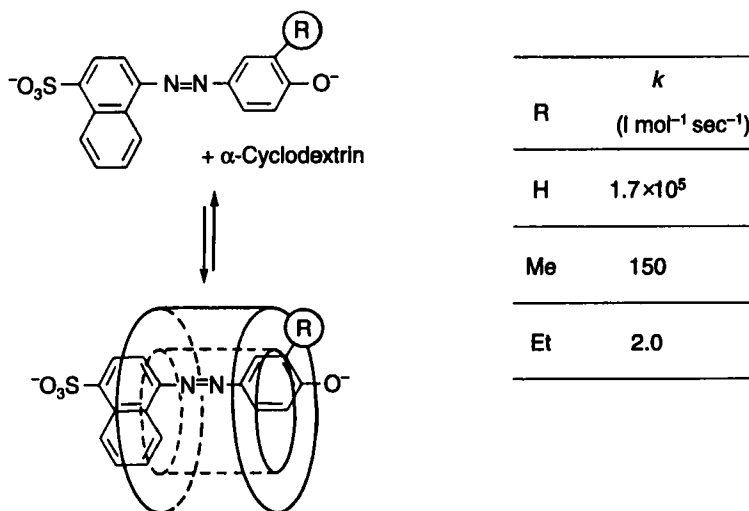


Figure 3 Kinetic recognition of various azo dyes by the cyclodextrin cavity [3]

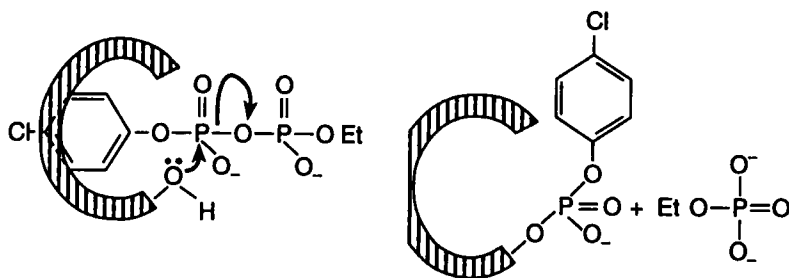


Figure 4 Catalysis of pyrophosphate splitting by cyclodextrin [4]

3. CATENANES

It is tempting to try to synthesize catenanes according to the lock-and-key principle. An attempt made 35 years ago is illustrated in Figure 6. The substituted hydroquinone formed the inclusion complex readily. This was then oxidized in the hope that the ring would close to form a catenane. This, however, was done at a time when NMR was not available, a situation hard to think of nowadays. Therefore we could not finally prove that the catenane had been formed. Around 35 years ago I told as a kind of joke the following story. If one tried to give a rough estimation of the frequency of the molecular vibrations of the two rings against each other, one would arrive at the conclusion that this frequency was in the ultrasonic region between 10 000 and 20 000 Hz, a frequency which the human ear cannot perceive but the ear of a dog can.

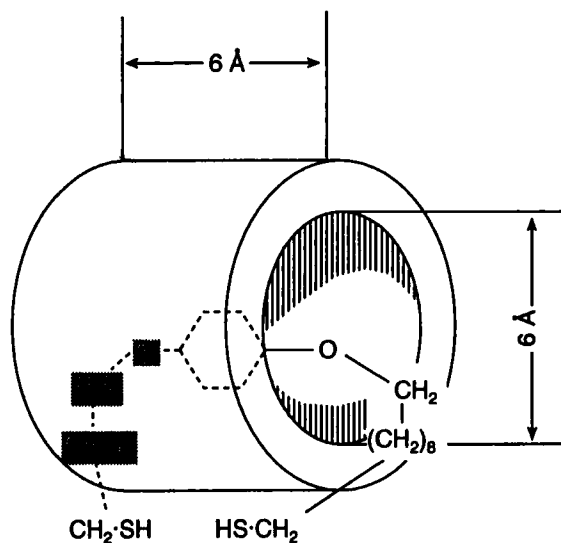


Figure 6 An early attempt to synthesize a catenane by the lock-and-key principle [6]

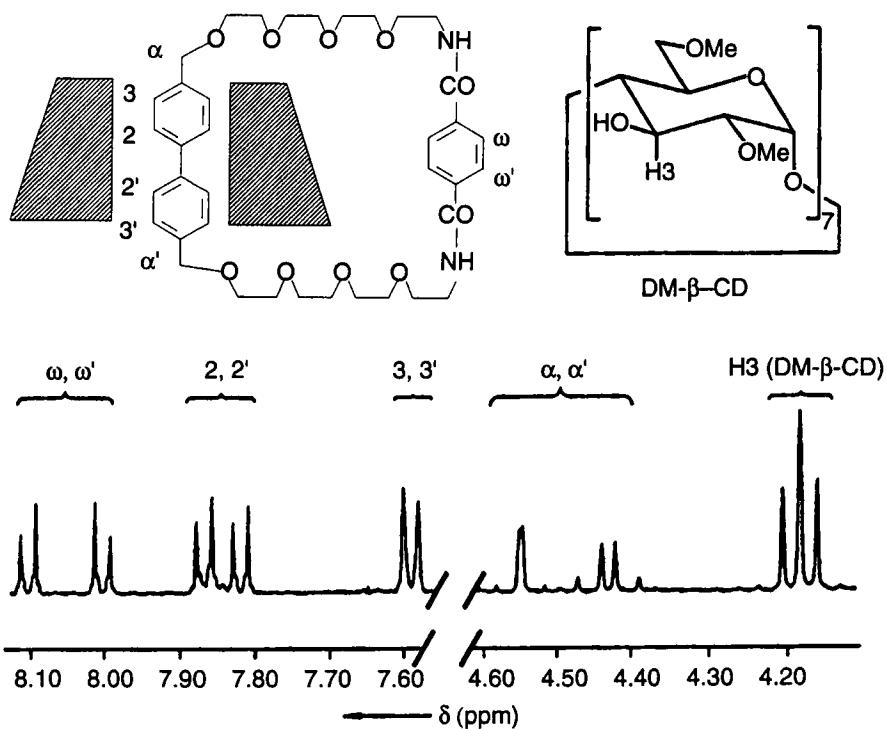


Figure 7 Catenane formed with cyclodextrin according to Harada [7] (DM-β-CD: dimethyl-β-cyclodextrin)

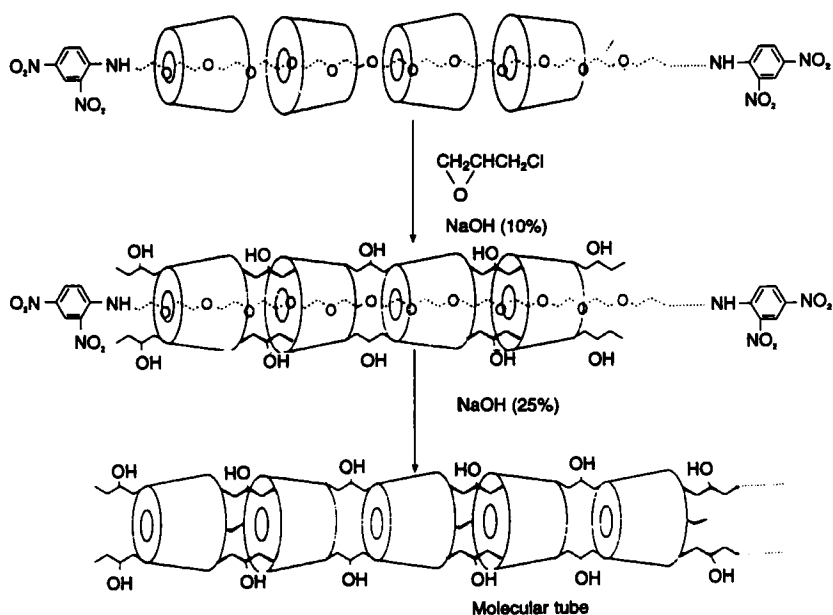


Figure 8 Preparation of a molecular tube according to Harada [8]

of recognition is apparently used when very high precision in biosynthetic processes is required, especially in the synthesis of informational macromolecules. However, we are not going to discuss proofreading in DNA synthesis. We have done some work on the fidelity of protein biosynthesis, in particular on the aminoacylation of tRNA which is the key step in introducing the correct amino acid into its proper place in the protein [9]. The aminoacyl tRNA synthetases have a twofold task in recognizing their substrates. At first they must recognize and activate the amino acid to form the adenylated amino acid, then in a second step this activated amino acid must be transferred to its proper tRNA (Figure 9). The three-dimensional structure of yeast phenylalanine tRNA is shown in Figure 10.

Amino acids are, in principle, very similar, and the differentiation between them, e.g. isoleucine and valine, seems to be hopeless using simple lock-and-key principles alone (Figure 11). One can estimate that the process should be erroneous by 20%; in other words, every fifth amino acid should be wrongly placed. In contrast, we can show that the isoleucyl synthetase selects its proper substrate against the similar valine with a fidelity of one error in 40 000. How is this possible? The recognition process goes through a selection cascade, as shown in Figure 12.

The primary selection, as measured by direct binding studies, is indeed very poor. There is only a threefold preference for isoleucine over valine. In a second selection step, this fidelity is improved by one error in 240. These two steps

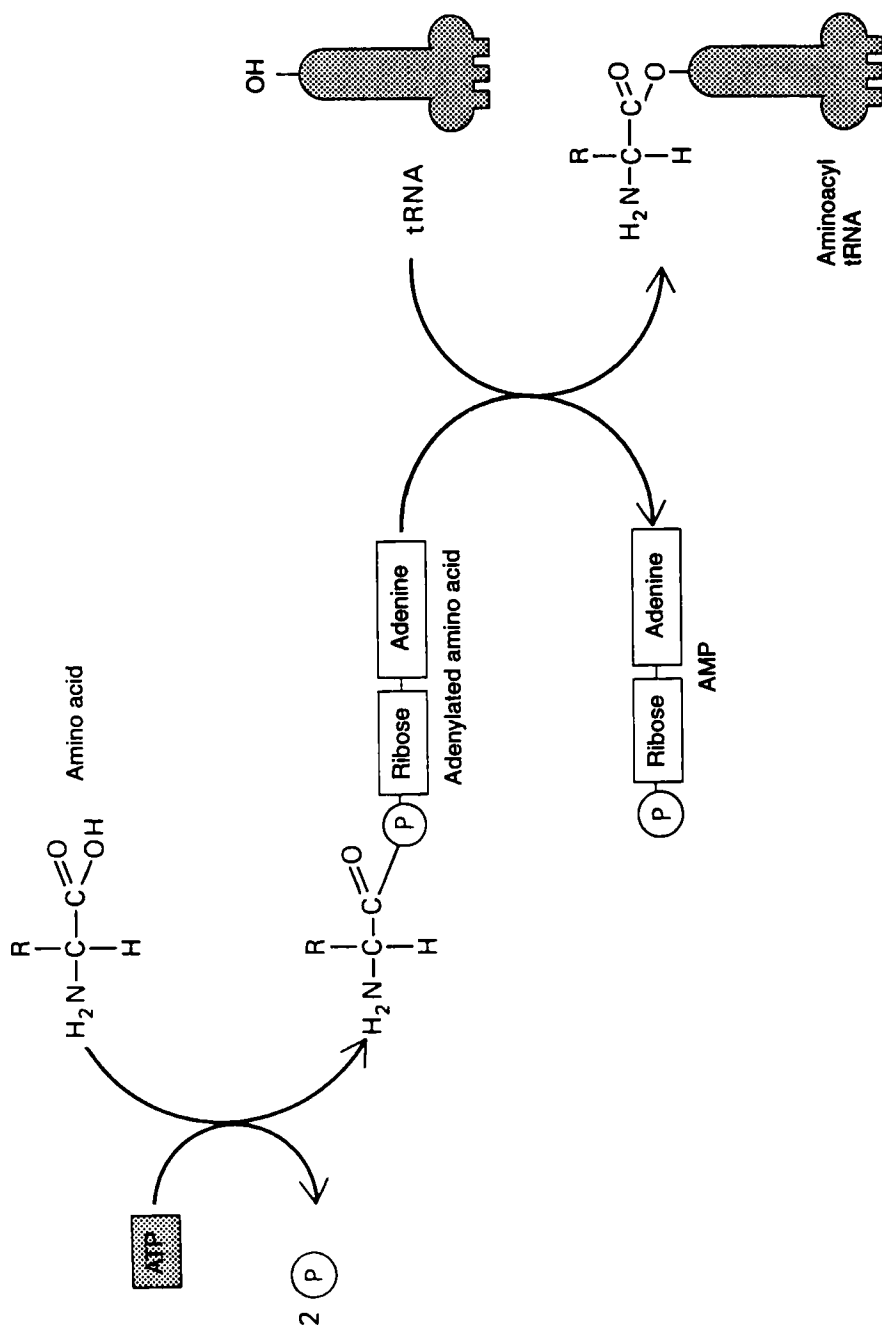


Figure 9 Course of the aminoacylation of tRNA

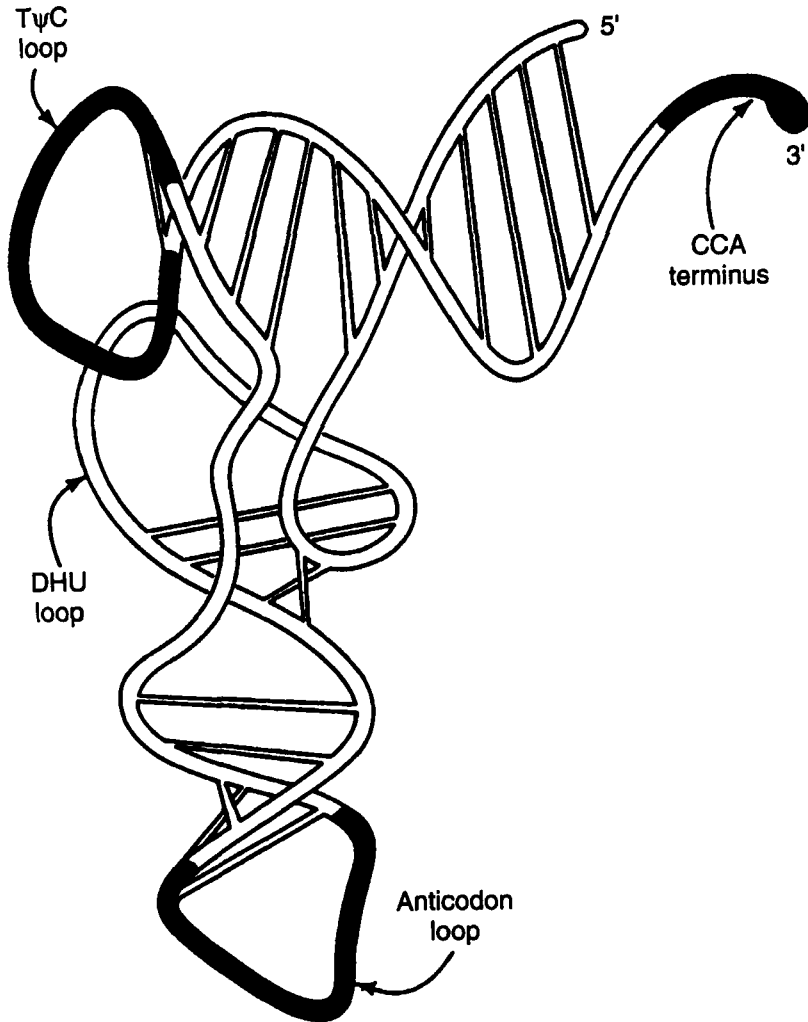


Figure 10 Three-dimensional structure of yeast phenylalanine tRNA (after S.-H. Kim)

are of the classical Michaelis–Menten type of recognition; that is, the recognition occurs in a reversible manner. One error in 240 is by far not good enough, because it would mean that each protein has more than one error. Therefore two proofreading steps follow in which the incorrect product is hydrolyzed preferentially. However, the correct product also has to be partially sacrificed. This occurs twice and finally one arrives at a fidelity of one error in 38 000. In this proofreading about 80% of the originally activated, correct isoleucyl AMP is sacrificed. Therefore this is an energy-consuming recognition. Five molecules of ATP are used for recognition and one molecule of ATP for

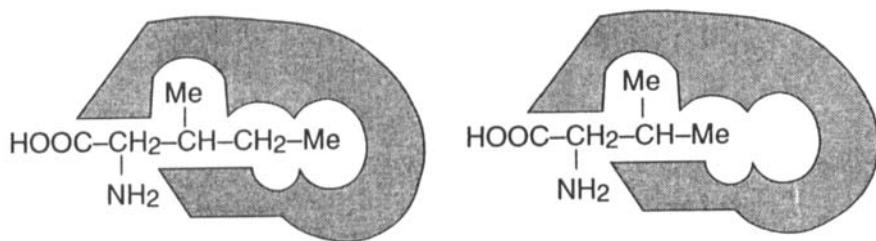


Figure 11 Discrimination between isoleucine and valine [9]

activation. This reaction is therefore an irreversible dissipative process in which ATP energy is pumping up the specificity (Figure 13).

With other synthetases there are similar mechanisms [10]. In general, the fidelity of amino acid recognition is of the order of one error in 10 000, thereby consuming energy. The aminoacyl synthetases are not classical enzymes with Michaelis–Menten microreversibility, rather they work in a unidirectional way.

What do we know about the structural details of the recognition process? Some synthetases have been crystallized and their structures been analyzed, e.g. seryl synthetase (Figure 14) [11].

Aspartyl and glutamyl tRNA synthetases have also been co-crystallized with their tRNAs and analyzed, thus giving a picture of how these molecules are attached to each other. But this does not give any mechanistic details about the dynamics of the recognition process. Such information comes from kinetic studies. We have measured the relative $\Delta\Delta G$ values for the activation of various amino acids, as compared with the substrate isoleucine, and plotted these values (Figures 15 and 16) against the available surface areas of the various amino acids, which one can take from the literature. The data very nicely lie on one line, at least those for amino acids which are shorter than isoleucine (glycine, leucine, alanine, valine). As the amino acids become longer a break of this line occurs and leucine joins the longer ones on a different line. This we interpret as a conformational change in the enzyme, like a plug controlling the substrate that must be removed. After that, the longer amino acids can also be accommodated, albeit in a less favorable situation. This applies for the first binding step (Figure 15). If one does the same for the second binding step (Figure 16) the story becomes more complicated, but still the various amino acids lie on defined lines. The very short ones, glycine and alanine, fit into the entrance part of the cavity, while the longer ones can only be accommodated if again two plugs are pushed aside. We arrive at such conclusions from our kinetic data. Only upon preparing this chapter was it realized that this is exactly the principle according to which the safety lock works, as shown in Figure 1.

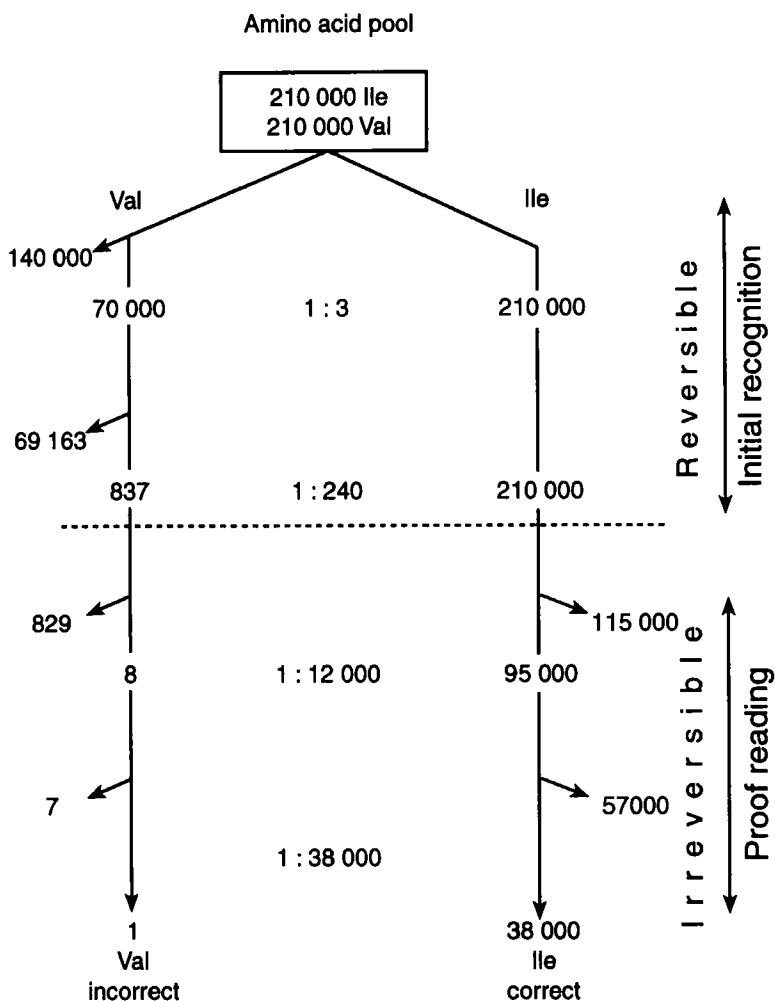


Figure 12 Proofreading cascade in the selection of isoleucine by isoleucyl synthetase [9]

5. MAGNETIC LOCKS

One can foresee that in a few years the mechanical locks will belong to the past. The magnetic card is coming. On this, the information is inscribed on magnetic tape in a linear fashion and read off in sequence. This is the principle of information laid down in DNA and RNA. This information can be transcribed in various ways such as through reverse transcriptase and DNA polymerase and the polymerase chain reaction (PCR). The PCR has become very important and was the subject of the 1994 Nobel Prize in Chemistry (Figure 17).

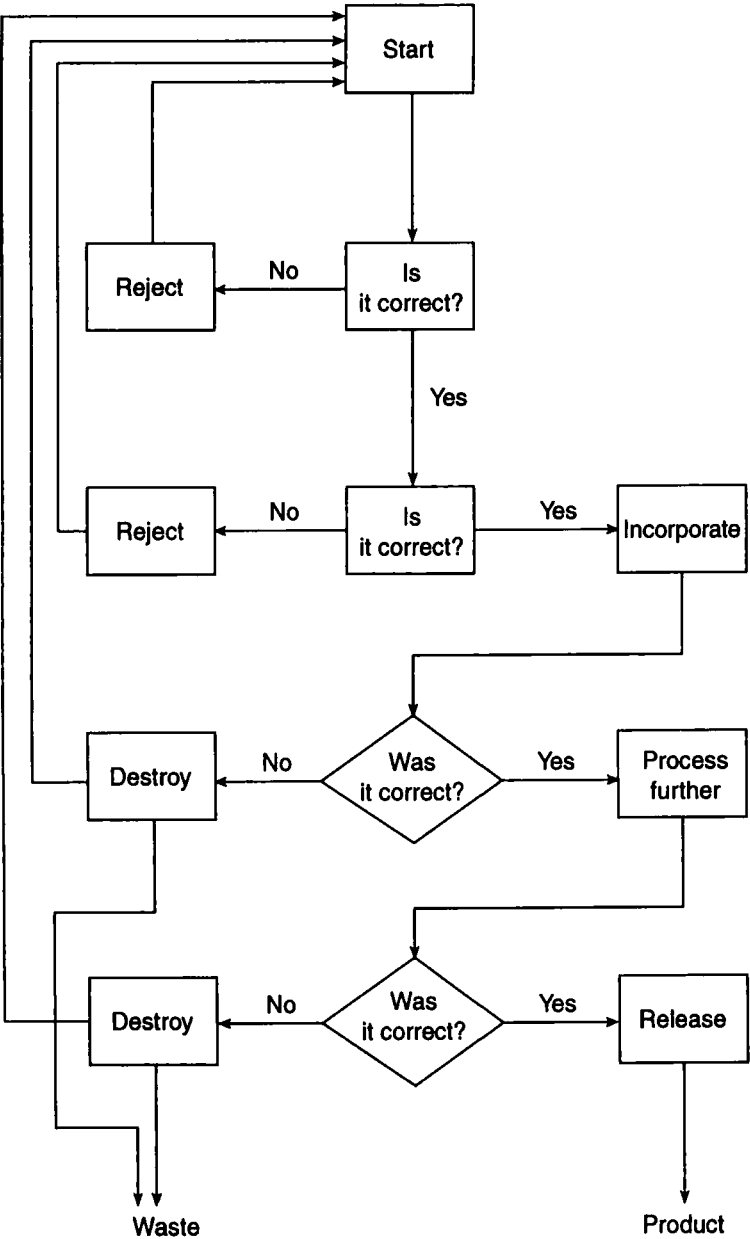
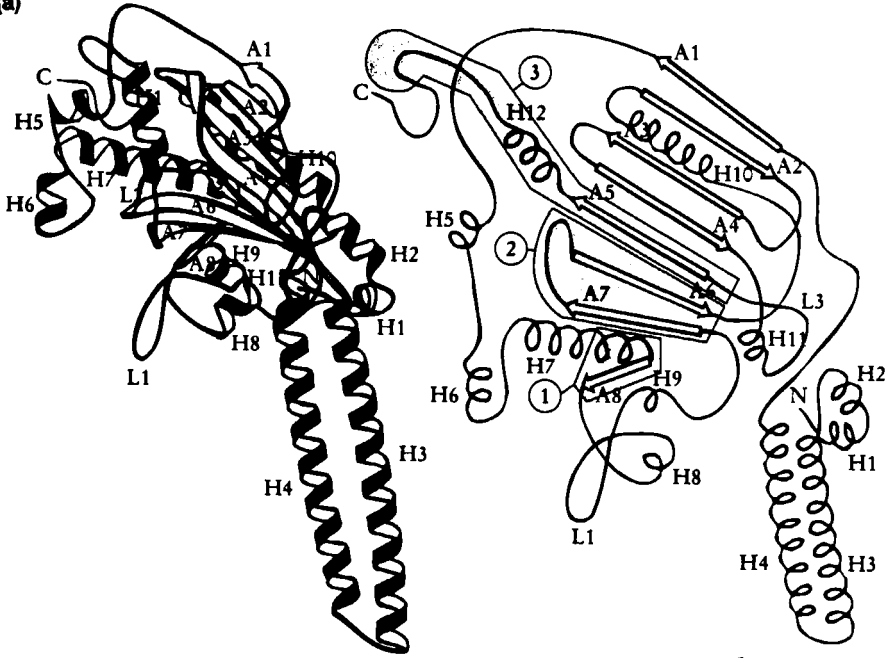


Figure 13 Dissipative recognition process of isoleucine by isoleucyl synthetase [9]

(a)



(b)

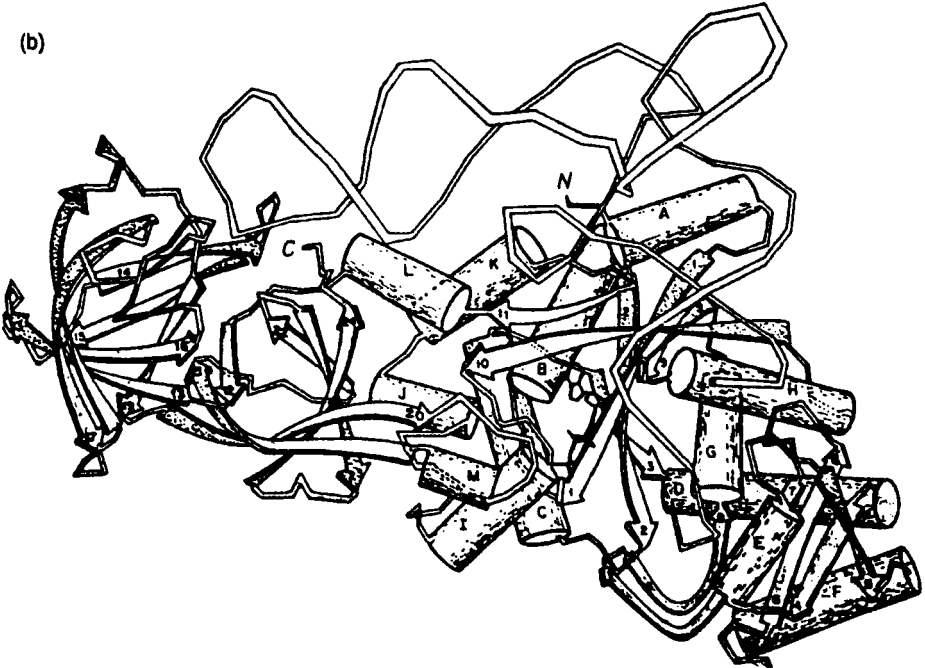


Figure 14 Three-dimensional structures of (a) seryl synthetase in two different representations and (b) glutamyl synthetase as co-crystals with their substrate tRNAs [11]

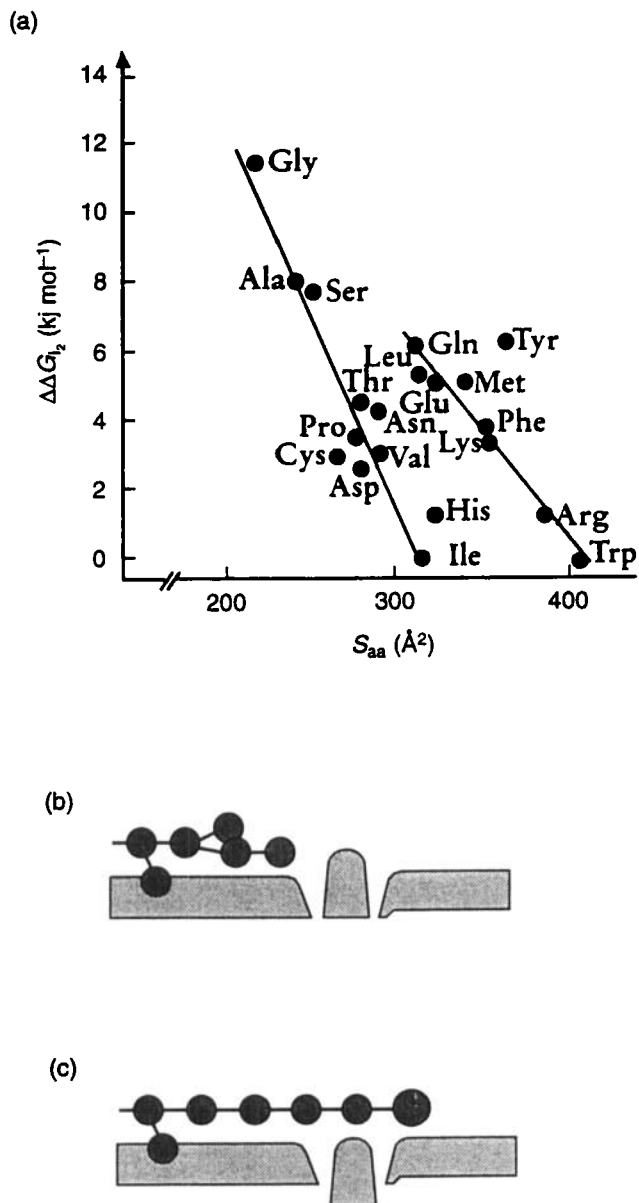


Figure 15 (a) Plot of $\Delta\Delta G_{12}$ versus available surface area for the first activation step of several amino acids. Recognition mechanism for smaller (b) and larger amino acids (c) [11, 12]

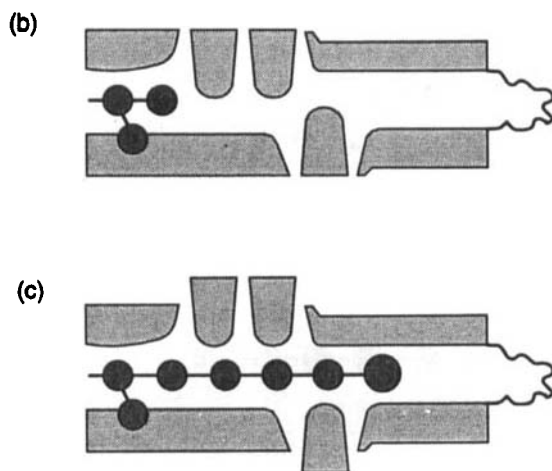
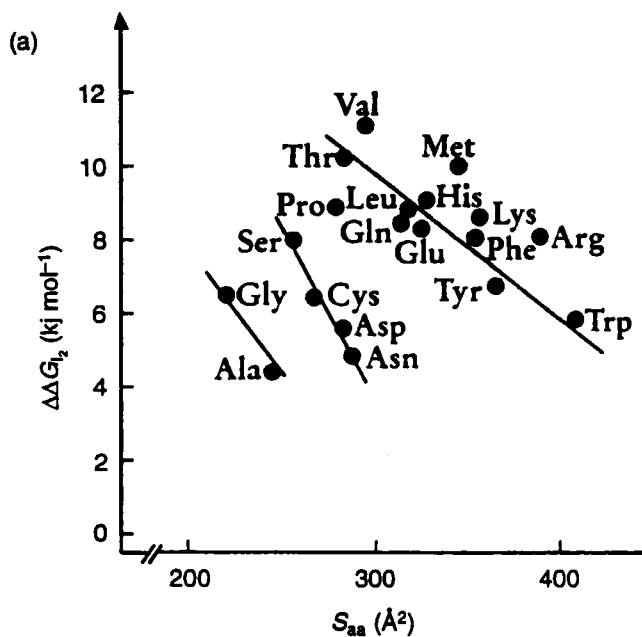


Figure 16 (a) Plot of $\Delta\Delta G_{II}$ versus available surface area for the second activation step of several amino acids. Recognition mechanism for smaller (b) and larger (c) amino acids [11, 12]

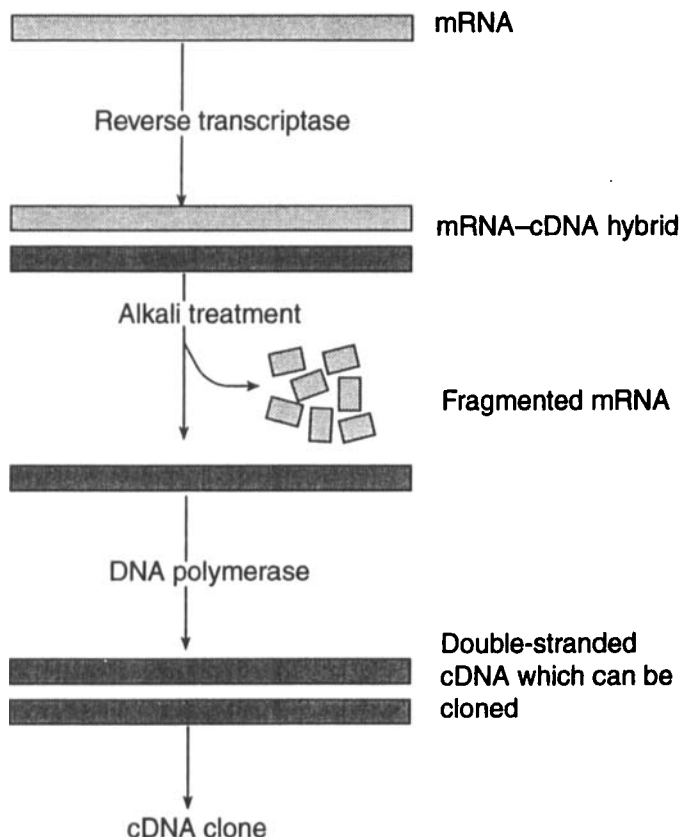


Figure 17 Mechanism of polymerases in cloning procedures

A very important area of research involves the synthesis and application of antisense RNA and DNA. The idea is to synthesize pieces of nucleic acids which are complementary to certain genes or parts of them. By hybridization the piece of nucleic acid is attached to the gene and rigidly fixed, thereby blocking the reading of this gene. Blocking the reading of information is the key to this technique. We have done some experiments in this direction. An oligonucleotide was furnished with psoralene and then hybridized with another oligonucleotide (Figure 18). Psoralene has the capacity to intercalate the nucleic acid and on irradiation it reacts with the nuclear base (Figure 19). A reaction was carried out with an antisense oligonucleotide directed against the α -chain of hemoglobin, and indeed the *in vitro* biosynthesis of the α -chain in the rabbit reticulocyte system was inhibited with this antisense oligonucleotide (Figure 20).

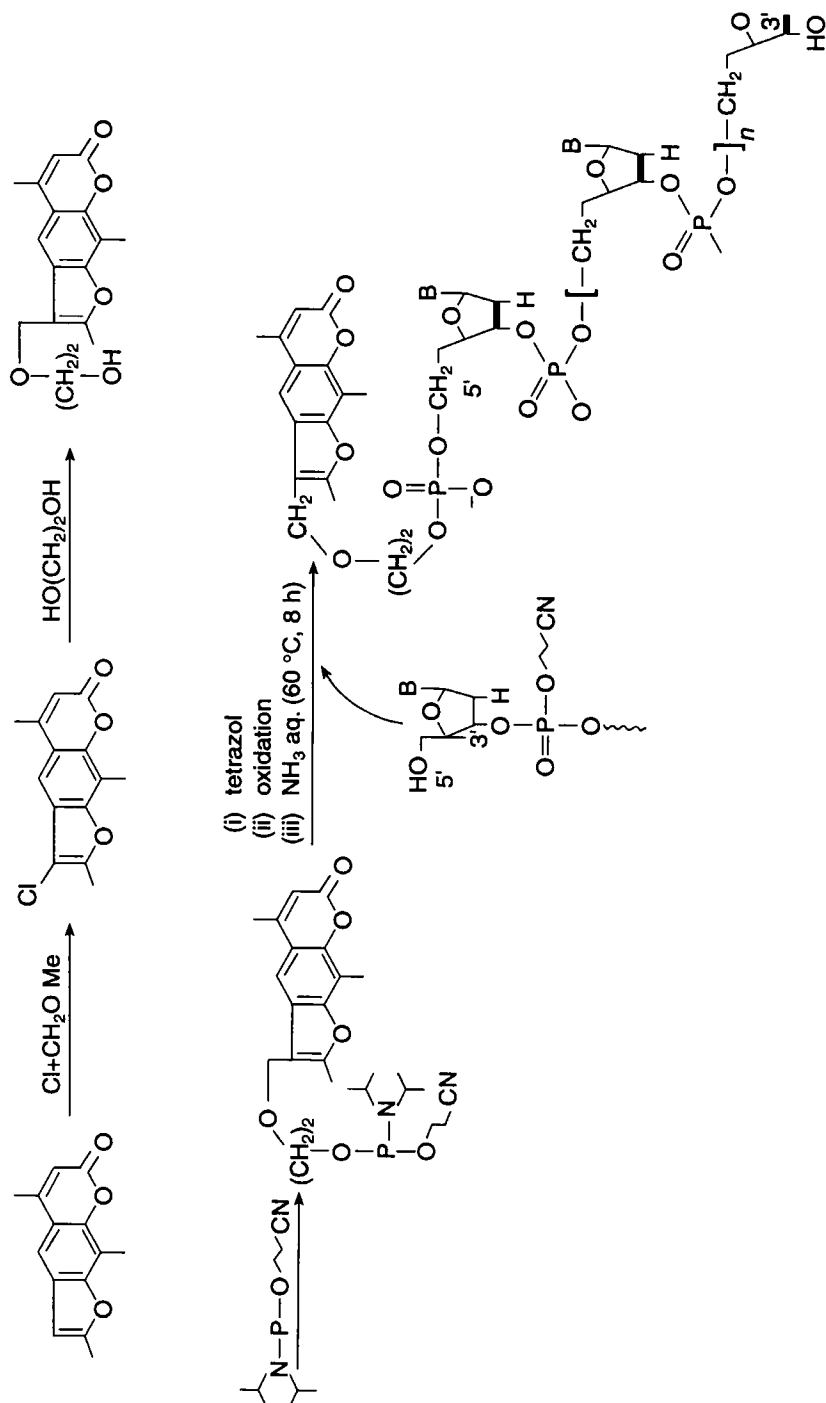


Figure 18 Synthesis of an antisense oligonucleotide furnished with psoralene [13]

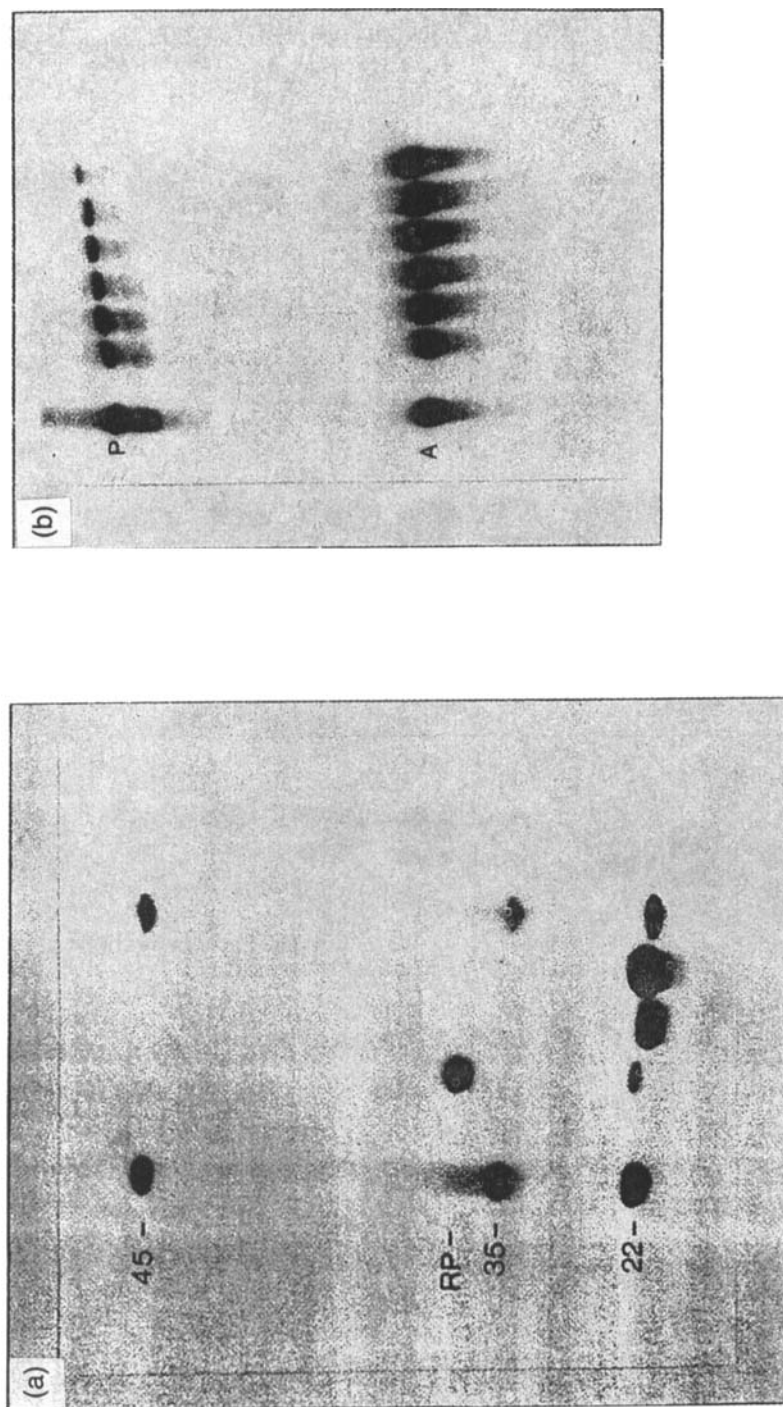


Figure 19 Reaction of psoralene oligonucleotides *in vitro*: (a) linking of the psoralene oligonucleotide with a complementary oligonucleotide by irradiation at 350 nm; RP: reaction product, the numbers are molecular weights; and (b) destroying the covalent bond by irradiation at 254 nm. A: starting material of the original reaction; P: Product which is degraded

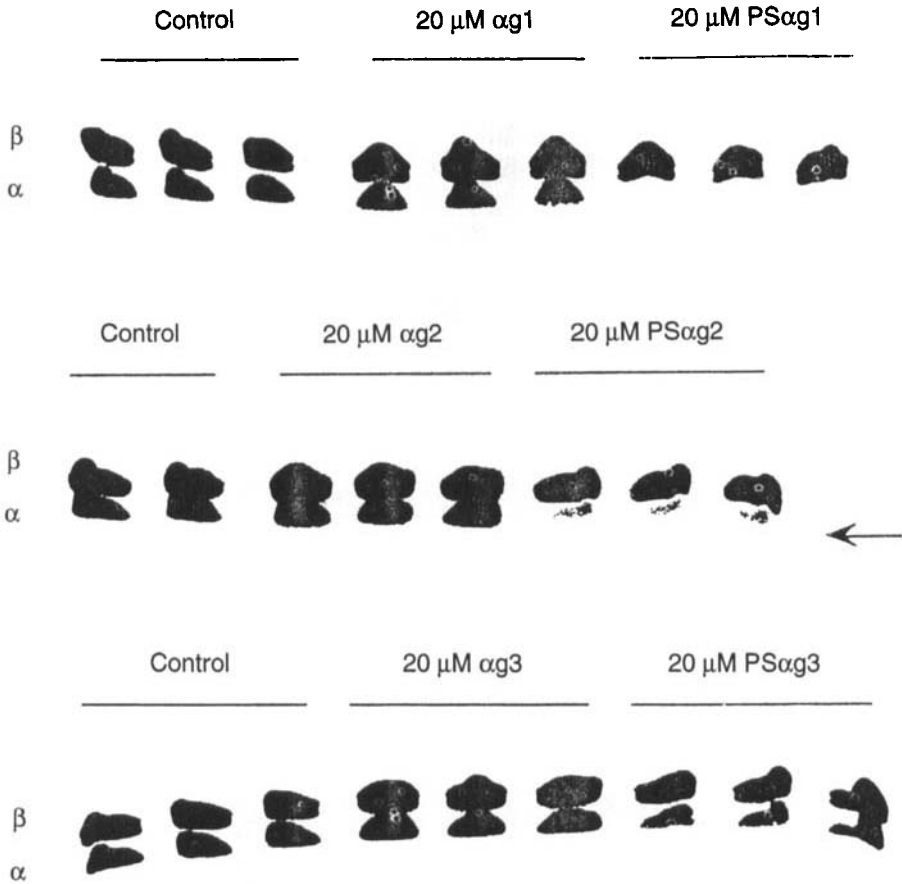


Figure 20 Inhibition of the α -chain of hemoglobin in the rabbit reticulocyte system [13]

6. CARBOHYDRATE-LECTIN INTERACTIONS: SUPRACELLULAR CHEMISTRY

Finally, let us look at a new field which is just beginning to be exploited. As one knows, cell membranes are composed of various components, and among these the glycosyl residues of the glycoproteins are very important. These have been somewhat neglected so far. In a recent publication [14] after an excellent description of these glycosyl residues it is stated that 'the functions of the sugar residues in glycoproteins as yet are unknown' (Figure 21).

The oligosaccharides of the glycopeptides are very complicated and are synthesized in a genetically determined way, whereby each step in the synthesis is carried out by a separate enzyme. It is difficult to believe that the effort made

by nature is for something which is not very meaningful. If the sugars can serve as kinds of keys, then there should be locks for these keys. Accordingly, we have searched in various cellular materials of higher animals and in human tissue for sugar-binding proteins which can serve as locks for these 'sugar keys', and

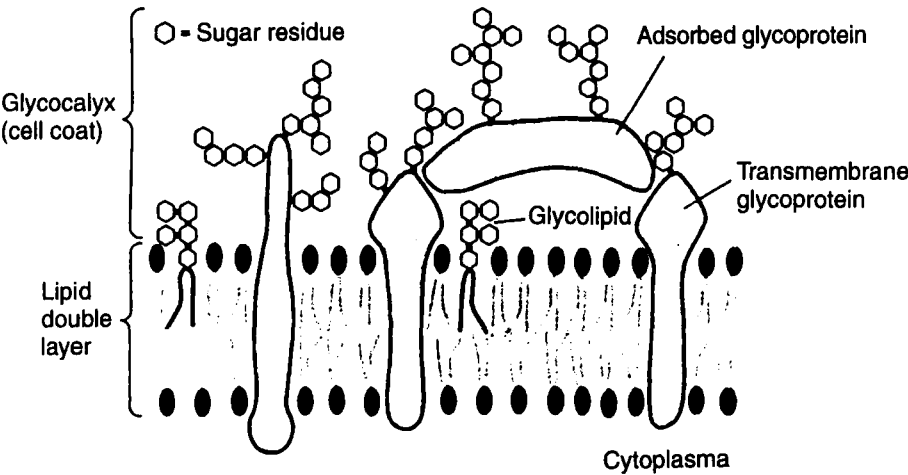


Figure 21 Schematic representation of the glycoproteins in the cellular membrane [14]

Table 1 Carbohydrate-binding proteins of human liver and primary liver tumors^a

	Lac		Asi	Mel		Man	Fuc	
Liver	56	14	27	54	54	28	29	54
	58	54	27.5	56			80	
			28				180	
	44	14	28	27		28	180	
	46			28				
	56			54				
Hepatoblastoma				58				
				62				
	66	15	28			28	66	
Hepatoma	80							
	45	15						
	21	14	28	54	28	54	28	54
	22	26		54				
	28	54						
	43	64						
	58	100						
	66							
	28	14	28			28		

^aSource: H.-J. Gabius, R. Engelhardt and F. Cramer, 1986.

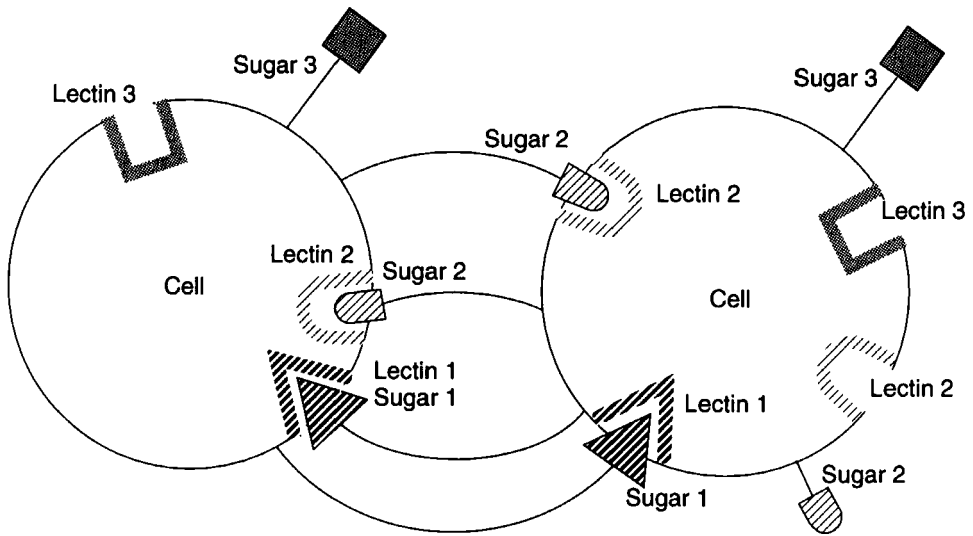


Figure 22 Supracellular chemistry cell-cell interaction through endogenous lectins (H.-J. Gabius and F. Cramer, 1988)

indeed there are such proteins. Using affinity chromatography we indeed found such sugar-binding proteins [15], namely lectins, which up to now were only known from plant material [16]. The lectin pattern is highly specific for every cell type—that for liver tissue is given in Table 1 [17]. The lectin-carbohydrate pattern plays a decisive role in cell adhesion and cell aggregation. We have named these endogeneous lectins Cell Adhesion Lectins (CALs) [15]. With this we enter the new field of *supracellular chemistry*. This is depicted in Figure 22.

Now knowing that cellular membranes not only carry carbohydrates but also carbohydrate receptors, we were able to use these recognition elements in order to characterize particular cells or cell lines. By constructing neoglycoproteins we could bind these in a specific manner to particular cells and thereby characterize these cells in various ways (Figure 23) [18].

Consider an example of drug targeting. If 5-bromouridine is attached to a neoglycoprotein, it acts on transformed rat fibroblast in a specific manner depending on the sugar residue. Thereby, the action of the drug becomes at least 10 times more specific with respect to this particular cell line (Table 2).

The histochemistry according to this principle has been further developed in the Pathology Department of Göttingen University and is almost a routine method for the differentiation of various tissues. For example, using a fluorescent neoglycoprotein one can even differentiate in a sample of bull testis between the various parts of the spermatozoa (Figure 24). This is a new and

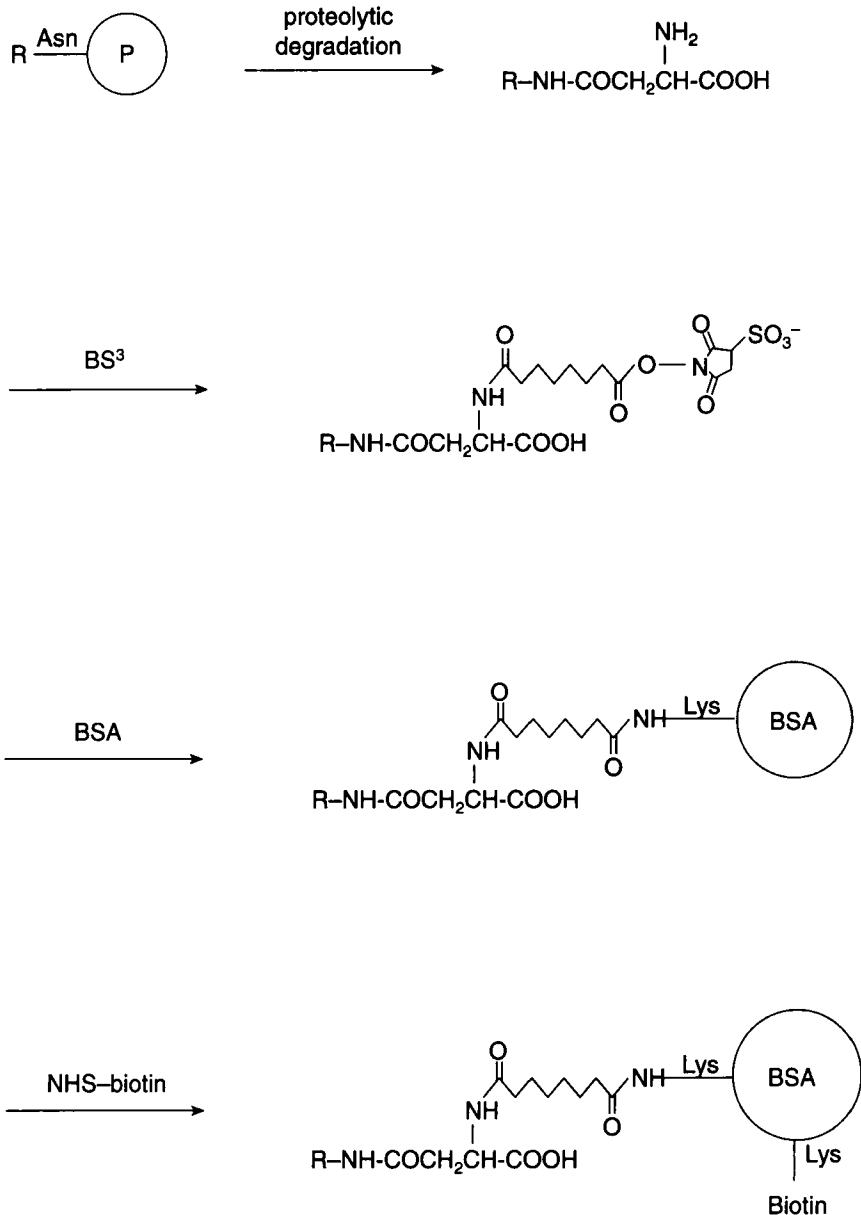


Figure 23 Synthesis of neoglycoprotein as carrier for histochemistry and drug targeting [18]. The sugar residue R is isolated from the glycoprotein P. R is then attached to bovine serum albumin (BSA) via the crosslinking reagent BS3. BSA is further derivatized with biotin

Table 2 Growth inhibition of SV-40-transformed rat fibroblasts by 10^{-6} M 5-bromouridine (5BrU)^a fixed on bovine serum albumin (BSA) together with different sugar residues

	H ³ -Thymidine uptake (%)
5BrU	60
5BrU-BSA	5
5BrU-BSA-cell	5
5BrU-BSA-Mal	8
5BrU-BSA-Lac	12
5BrU-BSA-Fuc	22
5BrU-BSA-Man	60

^aSource: H.-J. Gabius *et al.* [19].

rapidly developing field, and in fact sugars are the ideal partners for a lock-and-key recognition mechanism because they possess a rather rigid structure and many possibilities for hydrogen bonding.

7. CONCLUSION

The metaphor of lock and key still holds after 100 years and is as vivid as at Emil Fischer's time. With the development of more-refined types of keys this metaphor will become even more refined and coevolve with modern technology.

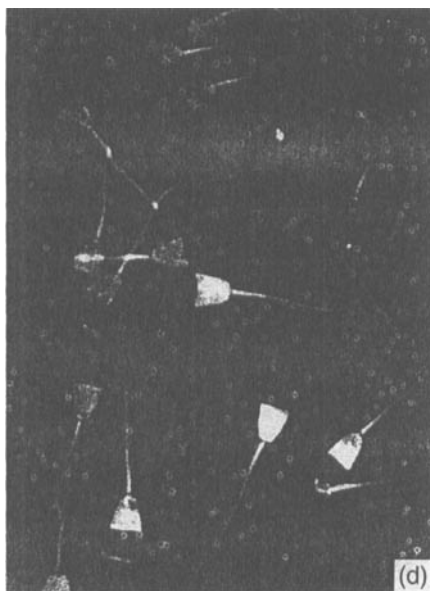


Figure 24 Histochemical staining of bull testis with a fluorescent neoglycoprotein [18]. Only the part of the spermatozoa with the suitable lectin is stained

8. REFERENCES

1. E. Fischer, *Chem. Ber.*, **27**, 2985 (1894).
2. F. Cramer, Inclusion compounds (Einschlussverbindungen, Heidelberg, 1954; F. Cramer, *Chaos and Order—The Complex Structure of Living Systems*, VCH, Weinheim, 1993).
3. F. Cramer, W. Saenger and Ch. Spatz, *J. Am. Chem. Soc.*, **89**, 14 (1967).
4. N. Hennrich and F. Cramer, *J. Am. Chem. Soc.*, **87**, 1121 (1965).
5. F. Cramer and G. Mackensen, *Chem. Ber.*, **103**, 2138 (1970).
6. A. Lüttringhaus, F. Cramer, H. Prinzbach and F. M. Henglein, *Liebigs Ann. Chem.*, **613**, 185 (1958).
7. A. Harada, J. Li and M. Kamachi, *Nature*, **356**, 325 (1992).
8. A. Harada, J. Li and M. Kamachi, *Nature*, **364**, 516 (1993).
9. F. Cramer and W. Freist, *Acc. Chem. Res.*, **20**, 79 (1987).
10. F. Cramer and W. Freist, *Angew. Chem., Int. Ed. Engl.*, **32**, 190 (1993).
11. W. Freist, *Chem. in unserer Zeit*, **27**, 25 (1993).
12. W. Freist, H. Sternbach and F. Cramer, *Eur. J. Biochem.*, **204**, 1015 (1992).
13. A. A. Wolf, Dissertation, University of Göttingen, 1990; F. Cramer (ed.), *Erkennen als geistiger und molekularer Prozess*, VCH, Weinheim, 1991.
14. B. Alberts, D. Bray, J. Lewis, M. Raff, K. Roberts and J. D. Watson, *Molecular Biology of the Cell*, Garland, New York, 1983, p. 285.
15. F. Cramer and H.-J. Gabius, in *Interrelation among Aging, Cancer and Differentiation* (eds B. Pullmann, P. O. P. T'so and E. L. Schneider), Reidel, Dordrecht, 1985, p. 187.
16. H.-J. Gabius, *Angew. Chem., Int. Ed. Engl.*, **27**, 1267 (1988).
17. H.-J. Gabius and S. Gabius, *Lectins and Cancer*, Springer, New York, 1991.
18. H.-J. Gabius, S. Gabius, T. V. Zemlyanukhina, N. V. Bovin, U. Brinck, A. Dunguy, S. S. Joshi, K. Kayser, J. Schottelisu, F. Sinowatz, L. F. Tietze, F. Vidal-Vanaclocha and J.-P. Zanetta, *Histol. Histopath.*, **8**, 369 (1993).
19. H.-J. Gabius, R. Engelhardt, J. Kasper, D. Reile, S. Schumacher, H. J. Schmoll, G. Graupner and F. Cramer, *Tumour Biol.*, **6**, 145 (1985).

Chapter 2

Molecular Recognition in Biology: Models for Analysis of Protein–Ligand Interactions

*DORON LANCET, AMNON HOROVITZ
AND EPHRAIM KATCHALSKI-KATZIR*

The Weizmann Institute of Science, Rehovot, Israel

1. INTRODUCTION

Biological reactions, such as enzyme catalysis, the immune response and signal transduction, which involve specific recognition by proteins of their cognate ligands are among the most characteristic processes of living organisms. Some proteins function as building blocks for complex cellular macromolecular structures, while others are intricate macromolecular machines capable of carrying out highly sophisticated biochemical reactions. All of these activities involve specific interactions on a molecular level.

Proteins associate with each other in different ways. Some associations between proteins are relatively long lived, as, for example, in multimeric proteins, multienzyme complexes, ribosomes and virus coat proteins. Other associations are more transient and dynamic, such as the interaction between polypeptide hormones and their receptors, the association and dissociation of actin and tubulin polymers, and the interaction of regulatory proteins such as calmodulin and protein kinases with their respective substrates. The interaction between proteins and their corresponding ligands is often highly specific, as in many antibody–antigen interactions or in the various interactions of protein hormones with their receptors. Protein–protein interactions can, however, show broad specificity, as in the case of the interaction of molecular chaperones with a great variety of protein substrates.

Proteins may also exhibit narrow or broad specificity towards low-molecular-weight ligands. Some enzymes interact specifically with their corresponding low-molecular-weight substrates and inhibitors, whereas others can interact with a rather broad ensemble of reactants. Of particular interest is the finding that class I and II molecules of the major histocompatibility complexes (MHCs) can bind a large number of peptides that are eight to nine and 13–17 amino acids long, respectively [1].

It is the task of the experimentalist and the theoretician to elucidate the structures of ligand–protein complexes, their energy of binding, the conformational alterations involved and the kinetics of complex formation.

2. BIOLOGICAL SPECIFICITY AS REVEALED BY A SINGLE RECEPTOR

Biochemists usually quantify the strength of binding between a receptor (R) or protein and its corresponding ligand (L) by means of a binding constant K describing the equilibrium attained in the system



where RL stands for the receptor–ligand complex. The free energy of the interaction ΔG can be readily calculated from the expression

$$\Delta G = -R_g T \ln K$$

where R_g is the gas constant and T is the temperature in kelvin. A high binding constant is often associated with a highly specific interaction, whereas a low binding constant usually suggests a lack of specificity.

The specificity concept as it relates to protein–ligand binding seems, however, to require a somewhat more elaborate consideration. The system to be considered usually contains a well-defined protein or receptor and an ensemble of potential ligands. When the binding of each of these ligands to the receptor protein is examined and if an *a priori* binding constant is chosen, it is usually found that some of the ligands will bind to the receptor with a higher binding constant and others with a lower binding constant than the chosen one. If the number of ligands with binding constants above the chosen one is relatively small, this suggests high specificity for the receptor under consideration. It seems, therefore, that the concept of specific binding, when applied to proteins and receptors, requires a description of the protein or receptor, a description of the ensemble of ligands being tested and a preselected binding constant determining the free energy of binding below which ligand binding is ignored.

Let us assume, for example, that two receptors, A and B, are exposed to 10^7 hexapeptides synthesized at random, and that the first (A) binds 10 peptides with binding constants above 10^8 M^{-1} , whereas the second (B) binds only one peptide with $K > 10^8 \text{ M}^{-1}$. It is obvious, then, that receptor B shows a higher binding specificity than receptor A.

In seeking a quantitative definition of specificity (see also Section 3.5) in a system of the type discussed above, consisting of a well-characterized receptor at a concentration $[R]$ and a limited number n of different ligands (i) each at a concentration $[L_i]$, one can proceed as follows. Let us first choose ligand m , which binds to the receptor with the highest binding constant K_m . When the concentration of this free ligand $[L_m]$ equals $1/K_m$, half of the receptors in the system will be saturated. However, each of the other ligands i , characterized by a binding constant K_i , when exposed to a concentration of free ligand of $1/K_m$ will saturate only a fraction of the receptor, to be given by the term $K_i/(K_i + K_m)$. When $K_i \ll K_m$, this term approaches the value K_i/K_m . By repeating the experiment for many different ligands, each at a free concentration of $1/K_m$, one can calculate the average fraction of receptors occupied by a ligand (excluding the 'best' ligand with the highest binding constant)

$$\frac{1}{n-1} \sum_{i=1, i \neq m}^n K_i/(K_i + K_m) \quad (2)$$

The specificity of the receptor S_R in the given ligand system can be defined quantitatively by

$$S_R = 0.5 \left[\frac{1}{n-1} \sum_{i=1}^n K_i/(K_i + K_m) \right]^{-1} \quad (3)$$

representing the ratio between the fraction of receptor bound to the 'best' ligand, the free concentration of which equals $1/K_m$, and the average fraction of receptor bound to a 'poor' ligand applied at the same free concentration of $1/K_m$.

In those cases in which all the K_i are much smaller than K_m , equation (3) is transformed into

$$S_R = 0.5 K_m \left(\frac{1}{n-1} \sum_{i=1, i \neq m}^n K_i \right)^{-1} \quad (4)$$

This formula is analogous to that developed for one ligand with many receptors (see equation 15).

In the simple model adopted, specificity is therefore defined by the ratio between K_m , the binding constant of the receptor with the 'best' ligand, and the average binding constant of the receptor with the rest of the ligands present in the system.

Finally, it is worth reemphasizing that among the various receptors analyzed so far, one can distinguish between those with narrow specificity and those with broad specificity. Returning to our original system consisting of one well-characterized receptor and an ensemble of potential ligands, one might attribute the properties of narrow specificity to a receptor that binds one or a very small number of ligands with binding constants above a suitably chosen high value, as opposed to a receptor that has broad specificity and is capable of binding a relatively large number of such ligands with the same or higher K values. In specific cases, such as in the analysis of the binding characteristics of antibodies, the use of peptide libraries [2] might be particularly helpful in distinguishing between two such classes of antibodies, i.e. those with narrow specificity and those with broad specificity.

2.1 The Kinetics of Ligand Binding

In the simplest cases, the ligand binding constant is uniquely determined by the ratio between the rate constants of the association and dissociation reactions, i.e. k_+ and k_- , respectively. In many cases, such as those presented in Table 1 [3], the value of k_+ differs by only one to three orders of magnitude from that of a diffusion-controlled bimolecular process. The thermodynamic meaning of this is that there is no large free energy of activation linked to the association of the ligand with the receptor. Relatively large values of k_+ thus seem to indicate that even in the absence of the ligand the receptor protein molecules possess a conformation appropriate for its binding.

The changes by up to 12 orders of magnitude in the ligand binding constants displayed in Table 1 thus primarily reflect the different rates of dissociation of the ligand from the receptor. Taking $10^7 \text{ mol}^{-1} \text{ s}^{-1}$ as typical for k_+ in the examples given in Table 1, k_- would vary from 10^{-5} s^{-1} , i.e. one dissociation per month, to 10^6 s^{-1} , or one dissociation per microsecond.

In considering the very large differences in the rates of dissociation of low- and high-molecular-weight ligands from their corresponding receptors, one

Table 1 Association free energies of some representative complexes of proteins and ligands^a

Protein	Ligand	DG ^b
Avidin	Biotin	-20
Lac repressor	DNA	-18
Antibody	Fluorescein	-7, -14
Horse liver ADH	NADH	-9
Beef heart LDH	NADH	-8.8
Yeast ADH	Ethanol	-2.5
	Acetaldehyde	-4.7
Fumarase	l-Malate	-2.5
	Fumarate	-3

^aFrom Weber [3].

^bBinding energies in kcal mol^{-1} .

reaches the inescapable conclusion that these rates are determined by functional requirements. For example, the formation of stable protein structures, such as fibrous actin, ribosomes or viral protein coats, from their corresponding protein subunits is characterized by relatively low values of k_- and high free energies of binding. However, in labile protein–ligand complexes, for example in some of the enzyme–substrate complexes, relatively high dissociation constants and low free energies of binding have been observed, most probably in order to facilitate efficient catalytic activity.

2.2 Some Thermodynamic Considerations

The free energy of interaction ΔG between a given receptor and its corresponding ligand characterized by the binding constant K is given, as indicated above, by the expression $\Delta G = R_g T \ln K$. However, since ΔG at any given temperature T is the result of the corresponding changes in enthalpy ΔH and entropy ΔS ($\Delta G = \Delta H - T\Delta S$), quantitative estimation of these two parameters gives valuable information on the thermodynamics of the reaction taking place between ligand and receptor. The alterations in entropy and enthalpy accompanying the receptor–ligand binding can be readily derived from the variation of K with temperature at a constant pressure since

$$\Delta S = - \left[\frac{d(\Delta G)}{dT} \right]_p \quad (5)$$

and

$$\Delta H = \left[\frac{d(\Delta G/T)}{d(1/T)} \right]_p \quad (6)$$

The literature contains a considerable amount of data on the changes in entropy and enthalpy characterizing the binding of different ligands to different receptors [4].

A conceptual model of protein association was proposed [5] according to which an interaction is thought to comprise two steps: a mutual penetration of hydration layers causing disordering of the solvent, and the subsequent domination of short-range forces such as van der Waals interactions and hydrogen bonds. The net ΔG for the complete association process is primarily determined by the positive entropy change accompanying the first step and the negative enthalpy change of the second step. That this is not always true is inducible by the recent findings of Poljak on the structure and thermodynamics of an antigen–antibody reaction [6]. After careful X-ray analysis, Poljak concluded that water molecules represent an important part of the complex between mouse monoclonal antibody D1.3 and hen egg-white lysozyme. Water molecules bound to the free combining site are conserved and additional ones

are bound to the complex, providing a further bridge between antigen and antibody.

2.3 Force Field Determining the Energy of Binding Between Receptor and Ligand

When binding of a ligand to a protein receptor does not lead to significant conformational changes either in the receptor or in the ligand, the approximate interaction energy between receptor and ligand (force field, FF) can be given by the following expression representing the sum of the van der Waals (VDW) and electrostatic energy components [7]

$$FF_{\text{score}} = \sum_{i=1}^{\text{lig}} \left(\sqrt{A_{ii}} \sum_{j=1}^{\text{rec}} \frac{\sqrt{A_{jj}}}{r_{ij}^{12}} - \sqrt{B_{ii}} \sum_{j=1}^{\text{rec}} \frac{\sqrt{B_{jj}}}{r_{ij}^6} + 332.0 \sum_{j=1}^{\text{rec}} \frac{q_i q_j}{Dr_{ij}} \right) \quad (7)$$

In this expression each term is a double sum over ligand atoms i and receptor atoms j , A and B are VDW repulsion and attraction parameters, respectively, r_{ij} is the distance in Å ($1 \text{ Å} = 0.1 \text{ nm}$) between atoms i and j , q_i and q_j are the point charges on atoms i and j , D is the dielectric constant and 332.0 is a factor that converts the electrostatic energy into kcal mol^{-1} ($1 \text{ cal} = 4.2 \text{ J}$).

The two VDW terms, corresponding to atom repulsion and atom attraction, indirectly represent the closest practical contact between atoms i and j . Assuming, therefore, that atoms i and j are represented by their corresponding van der Waals radii, the VDW term of equation (7) indirectly represents the 'geometric' fit between the atoms of the receptor and the ligand. Furthermore, as the energy of binding between receptor and ligand is determined by a sum of two independent terms, VDW and electrostatic, it is permissible to consider first the geometric fit between the atoms of the receptor and the ligand, and thereafter to screen, or correct, the data obtained by introduction of the appropriate electrostatic terms.

To facilitate theoretical considerations concerning receptor–ligand recognition on a molecular level, as well as molecular docking of ligands, rigid conformations of both receptor and ligand were assumed in many of the theoretical papers published. We have made similar assumptions in the section of our article dealing with the geometric fit between protein molecules and their ligands (Section 4). It should be noted that even when conformational alterations occur in the protein receptor during ligand binding ('induced fit' [8, 9]), they lead to the formation of a complex in which the geometric fit between the reacting molecules is closer than that expected between the individual intact molecules. It thus seems that induced fit occurs in order to form a complex structure with a clearly discernible, final 'lock-and-key' structure.

In the following, we first extend our formal analysis of biological specificity on the molecular level to systems consisting of receptor repertoires, and present a probabilistic model of molecular recognition in such systems (Section 3).

We then concentrate on some molecular aspects of protein–ligand recognition, indicate the importance of surface recognition and describe in some detail the possible determination of the geometric fit between proteins and their ligands by correlation techniques (Section 4). Finally, we will discuss the need to take account of energy considerations in any of the recognition models developed; in particular, we will survey the double-mutant cycle approach which enables the measurement of receptor–ligand amino acid residue interactions by site-directed mutagenesis carried out in one or both interacting proteins (Section 5).

3. ANALYSIS OF SYSTEMS CONSISTING OF A REPERTOIRE OF RECEPTORS

3.1 The Concept of Probabilistic Interactions

Molecular recognition is the end product of a long evolutionary process in which the pairwise interaction between receptor and ligand is gradually honed. However, just as the perfect fit of an organism to an ecological niche can best be understood as stemming from the accumulation of slight selective advantages arising by chance, so a flawless ligand–receptor fit may be assumed to be rooted in far from perfect, weak interactions. In some systems epitomized by immune recognition, and including also olfactory receptors and mixed function monooxygenases (cytochrome P-450s), probabilistic recognition is a present-day rule rather than a relic of past evolution. Thus, an organism continuously encounters numerous novel ligands for which no unique and specific receptors have evolved. These include antigens, odorants and toxic compounds biotransformed by monooxygenases. Recognition of such diverse ligands is made possible by the prior evolution of large and varied repertoires of receptors. Every novel ligand ‘selects’ the best recognition counterpart, which by definition becomes its ‘specific’ antibody, receptor or enzyme.

The founder of the selective theory of antibody action, Sir Macfarlane Burnet, described the relationship between an arbitrary antigen and the entire immune receptor repertoire (see Figure 1) as follows [10]:

Most of the molecules will show a minimal absorptive affinity . . . while only an occasional binding site pattern will show a high affinity. The rarity of the high affinity immunoglobulins is for precisely similar reasons that four aces are rarely dealt in a poker hand.

This view, if generalized, implies that given a collection (or repertoire) of potential receptors (e.g. all the globular proteins in the cytoplasm of a cell), a randomly selected ligand will have a defined affinity value towards any member of the repertoire. Those repertoire members that happen to have a high affinity are those whose binding of the ligand might carry functional implications, while the rest will constitute ‘background’ or ‘nonspecific’ binding.

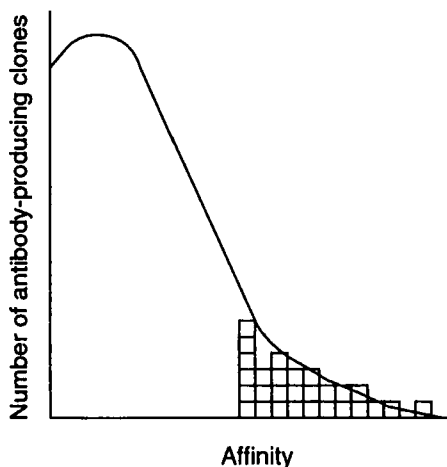


Figure 1 'Most of the molecules will show a minimal absorptive affinity . . . while only an occasional binding site pattern will show a high affinity. The rarity of the high affinity immunoglobulins is for precisely similar reasons that four aces are rarely dealt in a poker hand' [10]

An important consequence of such a scenario is that recognition is often imperfect, as it arises from chance complementarity. Furthermore, under such conditions one would not expect that each ligand would have one defined receptor in a repertoire, and vice versa. Rather, a given ligand may have several functional receptors within a repertoire, and each receptor might bind functionally more than one ligand. This concept is at the root of the notion of 'antibody multispecificity' (Figure 2) [11, 12].

3.2 Affinity Distributions

A natural outcome of the notion of probabilistic interactions is the concept of receptor affinity distributions [13]. If any member of a protein repertoire binds an arbitrary ligand with a certain affinity, then one may ask what is the probability that any affinity value will be represented within the repertoire. A relevant *gedanken* experiment would be to measure the binding constant of a given antigen (or hapten) towards all the different immunoglobulins present in the serum of an unimmunized animal. By 'binning' similar values, one obtains a discrete distribution of affinities. Each bar in the histogram of $P(K)$ represents the probability that the affinity will be within a defined interval around the value K . As argued by Burnet [10], it makes sense that low affinities will usually be more prevalent than high ones, as also seen in the intuitive distribution drawn by Burnet.

Formally [13], we define a repertoire of receptors (R) with N receptor types ($R_1, R_2, \dots, R_N$). Such a repertoire may be the family of different

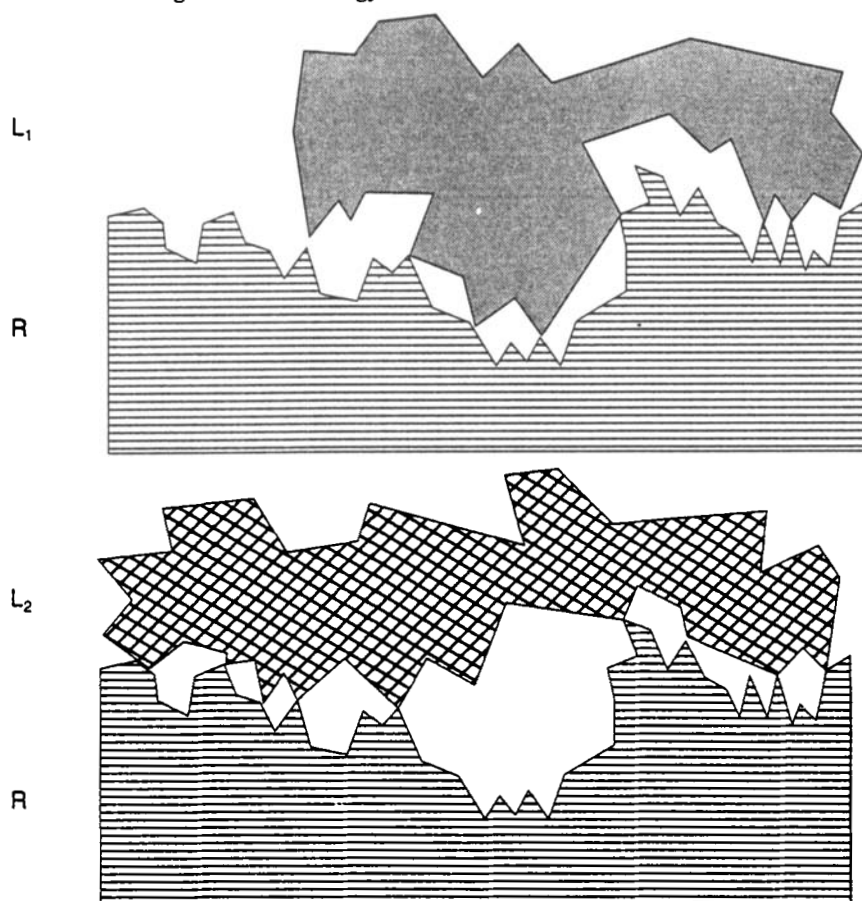


Figure 2 A diagrammatic representation of two ligands (L_1 and L_2), unrelated in structure, being bound alternatively in the same receptor binding region. This view is based on Inman's delineation of the hypothesis of antibody general multispecificity [12]

immunoglobulins produced by many B lymphocytes, or different odorant receptors in the olfactory neuroepithelium. An arbitrary ligand (L) binds to the receptors noncovalently, a reaction described by



where RL is the ligand–receptor complex. We define K_i to be the equilibrium association constant (affinity) of the i th receptor in the repertoire towards the ligand, and $\Psi(K)$ as the distribution of affinities for a ligand L towards all the different receptor types in the repertoire. $\Psi(K)$ is a probability function denoting the fraction, out of the total number of receptor types, with an affinity towards L between K_i and $K_i + dK_i$ (Figure 3). It is also possible to consider a discrete approximation for $\Psi(K)$.

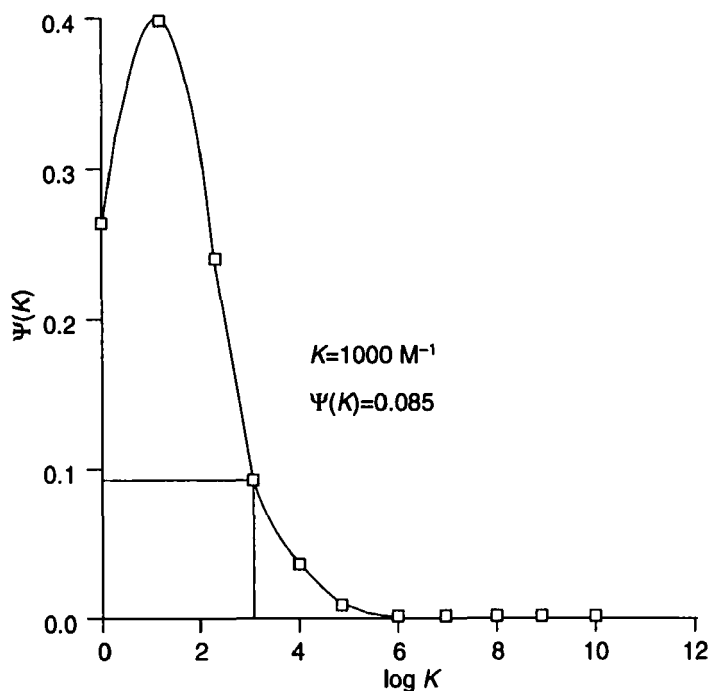


Figure 3 An example of a probability function $\Psi(K)$ is depicted on a linear vertical scale, with a marked example of one specific value: the probability for $K=10^3 \text{ M}^{-1}$ is 0.085, i.e. 8.5% of all receptors have an affinity value in this neighborhood

(a) Properties of the affinity distribution $\Psi(K)$

A few points need to be clarified with respect to $\Psi(K)$.

Generality Any biological macromolecule may assume the role of a receptor (including proteins, nucleic acids or polysaccharides). Actually, many of the arguments presented here may also apply to inorganic surfaces and small molecules.

Symmetry The definitions of receptor and ligand are symmetric, and their roles may be reversed. Thus, all the arguments about a collection of receptors and one ligand apply equally to a collection of ligands with one receptor.

Size constraints A receptor need not be a macromolecule and a ligand need not be a small molecule—any size pairing should conform to the formalisms developed here.

Multiple binding sites Strictly speaking, one receptor macromolecule may bind a ligand at many different sites (overlapping or nonoverlapping) with different

affinities. One can either consider the highest-affinity binding site as representing that receptor, or treat the collection of binding possibilities on the surface of one receptor as part of the affinity distribution.

(b) Open questions about the affinity distribution $\Psi(K)$

Given that for any collection of receptors and a ligand there exists an affinity distribution $\Psi(K)$, we may ask whether it is possible to make any predictions about the mathematical form of such a distribution. In addition, the following pertinent questions may be asked.

- (1) Can $\Psi(K)$ be experimentally determined?
- (2) For a given ligand, are the parameters in $\Psi(K)$ identical or similar for different receptor repertoires?
- (3) For a given receptor repertoire, are the parameters in $\Psi(K)$ identical or similar for different ligands?
- (4) Is the functional form of $\Psi(K)$ universal?
- (5) What are the biologically relevant inferences resulting from knowledge of $\Psi(K)$?

(c) Potential experimental measurement of $\Psi(K)$

Measuring the affinity of a ligand towards all members of a receptor repertoire is not feasible, as at least a few hundred separate binding curves need to be determined. There is, however, a simple alternative: since antibody preparations from an unimmunized animal contain a very large number of different immunoglobulins, measuring the saturation curve $\theta(L)$ for an arbitrary ligand L in such a preparation may yield information on the value of $\Psi(K)$ in a single titration. Evidently, given $\Psi(K)$, $\theta(L)$ may be calculated: for any value of L , the law of mass action allows a simple calculation of the saturation for any subpopulation of the receptor repertoire, and the overall saturation value $\theta(L)$ is obtainable by summation or integration. It is less obvious that $\Psi(K)$ may be calculated uniquely from $\theta(L)$, but this may be possible [13].

A binding curve for nonimmune immunoglobulins may well be characterized by a rather low affinity, in the realm of what is normally regarded as nonspecific binding. Nevertheless, we propose that it may contain important information. The binding curve for a ligand with normal serum immunoglobulins is often measured as nonspecific binding, to be subtracted from a specific binding curve. The present reasoning is that what is normally regarded as 'noise' contains useful information.

The analysis of actual binding results is best done in the context of a mathematical model for $\Psi(K)$, and is therefore deferred to a later section of this chapter.

(d) Quantitative approaches to stereospecificity: shape space

Although stereospecific recognition is a fundamental phenomenon governing all life processes, surprisingly little has been done to formulate general quantitative models describing it. Much effort has been directed to the question of predicting ligand–receptor docking through formulating appropriate energy surfaces and energy minimization procedures. These usually relate to particular recognition counterparts with defined shape and functional groups (see Section 4). The more general approach, involving an attempt to analyze the general behavior of noncovalent stereospecific binding, has received much less attention.

One important set of formalisms developed are the shape space models [14–18]. In these models, each receptor binding site or ligand epitope is represented by a point in a B -dimensional shape space. Each shape is represented by B coordinates characterizing the size and location of bumps and grooves, as well as charges etc. Overall complementarity is defined in terms of a distance in shape space: if the points representing the receptor and the ligand are closer than a minimal distance ϵ , then binding occurs with an affinity K greater than a minimal constant $K_{\min}$. Thus, the basic shape space model defines a continuous interaction metric, but has a discontinuity in the resultant affinity distribution at the shape space distance value of ϵ . B , the shape space dimensionality, is predefined, and values of $B = 1, 2, 3$ and 4 have been discussed [19].

(e) Models for the affinity distribution $\Psi(K)$

The physicochemical forces governing stereospecific recognition are universal, as indicated elsewhere in this review. It is thus not unreasonable to seek a more general model for $\Psi(K)$, with the aim of obtaining a continuous affinity distribution function that may faithfully reflect physical reality.

In order to obtain a quantitative description of $\Psi(K)$, it is necessary to relate complementarity to the free energy of binding [13]. This may be facilitated by a concretization of the shape space model, which regards each coordinate as a *subsite* within the receptor binding site [13, 20] capable of forming an elementary interaction with its counterpart on the ligand. The strength of the interaction is then calculable from the free energy contributions from the different subsites.

3.3 The Receptor Affinity Distribution (RAD) Model [13]

In order to develop a coherent and complete model for affinity distributions in receptor repertoires, the following physical parameters need to be determined:

- (1) the size of the interaction surface;
- (2) the variety of elementary interaction types;

- (3) the free energy magnitude and sign of elementary interactions; and
- (4) the additivity of the free energy contributions from elementary interactions (see Section 5). These points are dealt with below.

(a) Vector representation of the interaction surfaces

A vector representation of the subsite approach is shown in Figure 4. The two interacting species are represented by the vectors $\mathbf{r}$, which represents the receptor, and $\mathbf{l}$, which represents the ligand (hapten). The vectors $\mathbf{r}$ and $\mathbf{l}$ are termed the interaction vectors. The dimensionalities of $\mathbf{r}$ and $\mathbf{l}$ are respectively B_r and B_l . The elements of each vector may be regarded as representing subsite configurations. Such configurations may include the types of potential elementary interaction on one of the interacting surfaces (e.g. hydrogen bond donor), as well as geometric/steric attributes (see Section 4).

In a more general interpretation, similar to that prevailing in the shape space model, the vectors may be looked upon as an abstract mathematical portrayal of the complexity of the interaction surfaces. The dimensionality of the vector bears a direct analogy to the number of dimensions of the shape space. It is assumed that complementarity is somehow related to the 'fit' of the respective elements in the vectors.

(b) Assigning values to the components of the interaction vectors

A crucial step towards developing the RAD model is the decision about which values are allowed for the components of the interaction vectors. The complexity of the two interacting surfaces (represented by the vectors), which may be reflected in the likelihood that a strong interaction will arise, obviously depends on how many different values are allowed for each of the vector components. S , the subsite variability, has been defined as the number of different values allowed for any of the components of the vector [13]. A vector of dimensionality B therefore has S^B different values.

A special case of the present model is the 'bit-string' model [14] where $S = 2$. Here the components of the interaction vectors may assume only two different values, 0 and 1.

(c) Computing the 'goodness of fit'

It is now necessary to turn to the question of how to compute an affinity for a given pair of interaction vectors. The basic premise is that a pairwise interaction occurs between respective components of the $\mathbf{r}$ and $\mathbf{l}$ vectors. In our model, each of the subsites assumes a value symbolizing one possible subsite configuration. For example, it is possible to construct a lookup table that determines what values of the components r_i and l_i lead to a 'fruitful' interaction. A simple scenario is that an interaction occurs if and only if $r_i = l_i$.

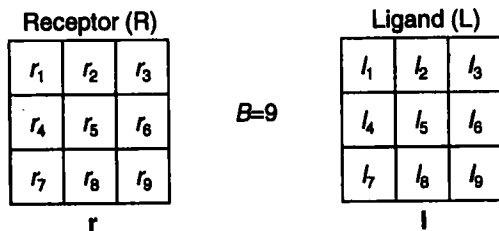


Figure 4 The vector representation of receptor and ligand which underlies the receptor affinity distribution (RAD) model. A vector element represents the potential elementary interaction on the receptor (R) or ligand (L). In this example, the dimensionality of each vector is $B=9$

In order to gain some intuitive view of the foregoing argument, consider the case of $B_r = B_l = B$ and $S_r = S_l = S$, i.e. the receptor and ligand have the same interaction vector dimensionality and the same subsite complexity. Another point that needs to be stressed is that all the subsites are assumed to be independent and their disposition completely random, i.e. the receptor repertoire represented by the collection of all possible r vectors is not 'biased' towards particular ligands or subsite configurations. Examples of such deviations in the real world include the elimination of self-reacting immune cells [19] or a possible bias of the olfactory repertoire towards environmentally abundant odorants.

The probability $P(M)$ that the receptor and ligand will form M successful interactions ($M = 1, \dots, B$) is given by a binomial distribution. Since there are S different possible values for any of the vector components, the distribution is governed by a probability of $1/S$. The value of $P(M)$ is therefore

$$P(M) = \frac{B!}{M!(B-M)!} \left(\frac{1}{S}\right)^M \left(1 - \frac{1}{S}\right)^{B-M} \equiv \text{Bin}(1/S, B, M) \quad M=0, \dots, B \quad (9)$$

(d) Modeling the free energy of interaction

The final step in deriving a probability function for the affinity requires the introduction of free energy terms. A straightforward way to achieve this is to assign an elementary free energy change $\Delta g = -\alpha \text{ kcal mol}^{-1}$ ($\alpha > 0$) to each successful elementary interaction. In a more general case, each elementary interaction may have a different Δg value associated with it. It is then also possible to consider $\Delta g > 0$ (an unfavorable free energy change) for some of the interactions.

(e) The additivity of elementary free energies

A basic conjecture that may be addressed theoretically and experimentally is the additivity of elementary free energies (see Section 5). Assuming that additivity

holds, the total free energy change for the interaction between receptor and ligand is given by

$$\Delta G = \sum g_i \quad (10)$$

In a case where M successful interactions are formed and if all elementary interactions contribute an equal free energy change $g_i = -\alpha$

$$\Delta G = -\alpha M \quad (11)$$

The association constant K is then given by

$$K = \exp(\alpha M / R_g T) \quad (12a)$$

or

$$M = \frac{2.3 R_g T}{\alpha} \log K \quad (12b)$$

Then $\Psi(K)$, the probability distribution (or histogram) of the association constants K , is

$$\Psi(K) = \text{Bin} [1/S, B, (2.3 R_g T / \alpha) \log K] \quad (13)$$

This formula embodies a receptor affinity distribution (RAD) model, in which the functional form is that of a binomial distribution. A typical form of $\Psi(K)$ is shown in Figure 5.

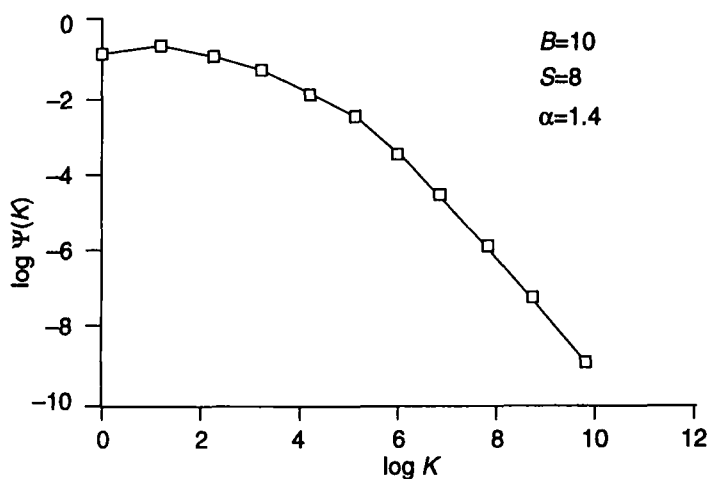


Figure 5 The functional form of $\Psi(K)$ as predicted by the RAD model. This is based on a binomial distribution (equation 9) with the parameter values $B=10$, $S=8$ and $\alpha=1.4 \text{ kcal mol}^{-1}$

A generalization of the RAD model with a lookup matrix that overcomes the artificial restriction $r_i = l_i$ has been proposed [13]. This is intended to represent the reality that any given chemical group may take part in more than one type of elementary interaction.

3.4 Experimental Fitting and Physiological Significance of the RAD model

(a) Measuring $\Psi(K)$ experimentally: parameter fit

As mentioned above, it is possible to determine the affinity distribution $\Psi(K)$ by binding measurements. We have actually done this, utilizing equilibrium dialysis suitable for measurement of the relatively weak binding expected between a ligand and a repertoire of 'nonspecific' receptors [13]. An experiment with iodovanillin, an odorant and a structural homologue of immunological haptens, gave a binding curve with half-maximal saturation at 0.1–0.3 mM ligand (Figure 6). Using the RAD model, a parameter fit for $\Psi(K)$ was carried out. The rounded parameter values obtained were $B = 10$, $S = 8$ and $\alpha = 1.4 \text{ kcal mol}^{-1}$ (Figure 5). These were used to derive the affinity distribution curves of Figure 5. These parameters may be thought of as describing the real world through the phenomenological RAD model. Interestingly, low values of S , approaching the bit-string limit ($S = 2$) and rendering the binomial distribution symmetrical (approximately normal), led to an appreciable deterioration in the goodness of fit.

(b) The universality of $\Psi(K)$

Among the open questions in Section 3.2(b) above, we asked whether for a given receptor $\Psi(K)$ is similar for different ligands. We tested two other ligands with different chemical structures [Lancet and Great, unpublished results] (Figure 6). One, iodotyrosine, shows a saturation curve that practically overlaps the one found for iodovanillin. By inference, the parameters for $\Psi(K)$ are very similar for both ligands. The third ligand tested, iodoanthranilic acid, shows somewhat lower binding. A very tentative conclusion is that it is possible for a structurally disparate ligand to have the same $\Psi(K)$ parameters, suggesting a possible universality of the affinity distribution function for different ligands in a broadly defined class.

3.5 Inferences from the RAD Model: The Meaning of Biological Specificity

(a) Repertoire size and maximal receptor affinity

Knowledge of the affinity distribution $\Psi(K)$ may help answer some basic questions related to receptor repertoires. For instance, how large should a

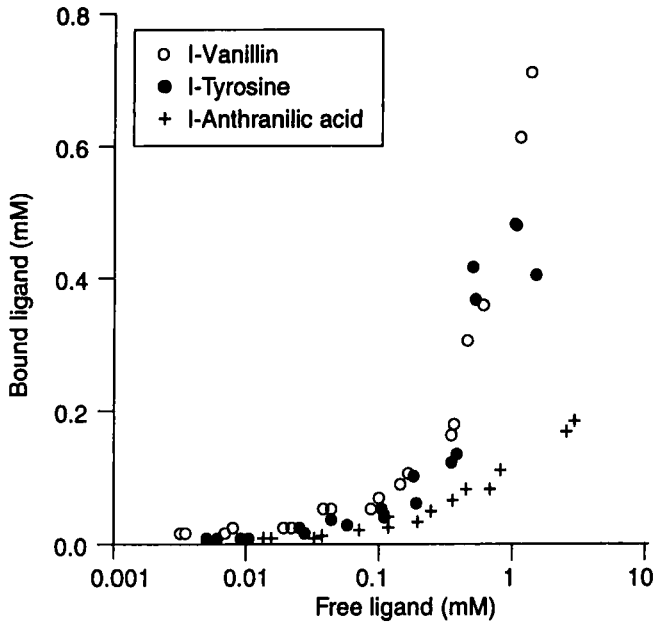


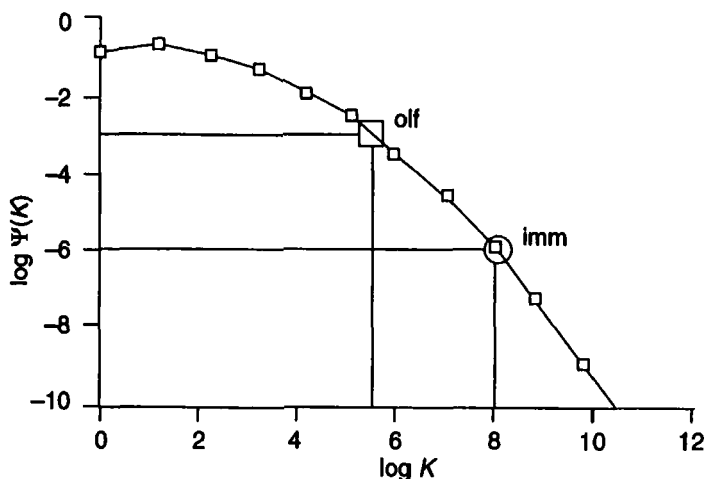
Figure 6 Binding curves for three radioiodinated ligands to normal bovine immunoglobulins, measured by equilibrium dialysis

repertoire be to assure proper representation of receptors with a minimal affinity K towards any ligand? The present model considers repertoires with a maximal size of S^B (8^{10} , or about 10^9 receptors for the fitted values of B and S). However, in reality, biological repertoires may be considerably smaller. Their size, as related to the required minimal affinity, may be computed as follows. The discrete probability function that among N receptors there will be n_K receptors with an affinity K towards an arbitrary ligand is itself a binomial distribution whose mean is equal to $\Psi(K)N$. If we require that on the average one receptor will have an affinity of K , then

$$N \approx 1/\Psi(K) \quad (14)$$

Thus, on the average, K will be the highest affinity represented in a repertoire of size N .

We may then calculate, for example, that for the distribution function $\Psi(K)$ in Figure 7 the probability of an affinity of $10^{7 \pm 0.5} \text{ M}^{-1}$ is about 3×10^{-5} . In other words, in a random repertoire of about 3×10^4 receptors there will be, on the average, one receptor with an affinity of 10^7 M^{-1} towards an arbitrary ligand. This inference may be true for a wide variety of ligands, provided that they conform with certain constraints regarding molecular size and chemical nature. Using the derived form of $\Psi(K)$, an actual value of maximal K may



$$N \approx 1/\Psi(K_{\max}) \quad K_{\max}^{\text{olf}} \approx 10^5 \text{ M}^{-1} \longleftrightarrow N_{\text{olf}} \approx 1000$$

$$K_{\max}^{\text{imm}} \approx 10^8 \text{ M}^{-1} \longleftrightarrow N_{\text{imm}} \approx 10^6 \text{ M}^{-1}$$

Figure 7 One of the consequences of the RAD model is that a relationship exists between the size of a repertoire (N) and the value of the average maximal affinity ($K_{\max}$) that it displays towards ligands. This relationship is given in equation (14). For a repertoire such as that of olfactory receptors with perhaps 1000 receptors, it is predicted, based on the function $\Psi(K)$ with parameter values as in Figure 5, that the average maximal affinity displayed towards an arbitrary ligand will be $K_{\max}^{\text{olf}} = 10^5 \text{ M}^{-1}$. In comparison, for an immune-like repertoire with 10^6 receptors the corresponding predicted value is $K_{\max}^{\text{imm}} = 10^8 \text{ M}^{-1}$

be computed for the immune system (Figure 7). In an immunoglobulin repertoire of $N=10^7$ [21], the maximal value is around $K=10^{8.5} \text{ M}^{-1}$, not unlike the known high-range affinities of polyclonal antibodies. A similar analysis may apply to the T-cell receptor repertoire.

(b) Specificity versus nonspecificity in biological recognition

The affinity distribution $\Psi(K)$ probably describes the background of nonspecific binding that exists whenever a specific receptor is studied in the presence of a large ensemble of unrelated proteins. The RAD model allows one to develop a quantitative relationship between specific and nonspecific binding for a ligand-receptor pair with an affinity K' [13]. An aim may be set to have specific binding f times larger than nonspecific binding at a free ligand concentration

$L' = 1/K'$ that half saturates the specific receptor. This will require that the relative abundance of the receptor be W times larger than $\Psi(K)$. For a given f , it can be shown that the value of W is

$$W = \frac{2f}{\Psi(K')} \sum_{K < K'} \Psi(K) \frac{KL'}{1 + KL'} \quad (15)$$

(see Figure 8). Because of symmetry, this formalism is similar to that for one receptor with many ligands (Section 2). As an example, it may be calculated (using the above parameter values) that an antibody with $K' = 10^7 \text{ M}^{-1}$ should constitute 0.01% of the serum protein in order to show a signal-to-noise ratio

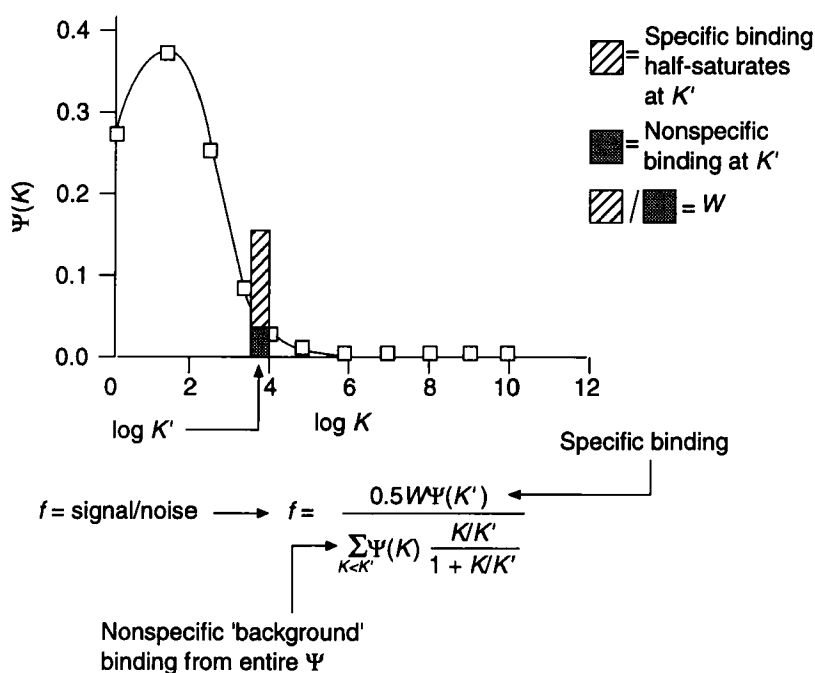


Figure 8 The RAD model allows one to define the signal-to-noise ratio between specific and nonspecific binding. It is assumed that the collection of nonspecific binders obeys the distribution $\Psi(K)$. Using a rearranged form of equation (15) as shown at the bottom of the figure, the signal-to-noise ratio f is given as a ratio of the numerator to the denominator. The numerator is the concentration of ligand L bound to the specific receptor (which has an affinity K') at a particular free ligand concentration L' that half saturates that receptor (i.e. $L' = 1/K'$). The expression $W\Psi(K')$ is the concentration of the specific receptor, since we assume that it is present at a concentration W times higher than that of proteins with similar affinity contained within the nonspecific distribution $\Psi(K)$ (see text). The denominator is the sum of bound ligand concentrations for all the proteins in the nonspecific distribution at the ligand concentration $L' = 1/K'$

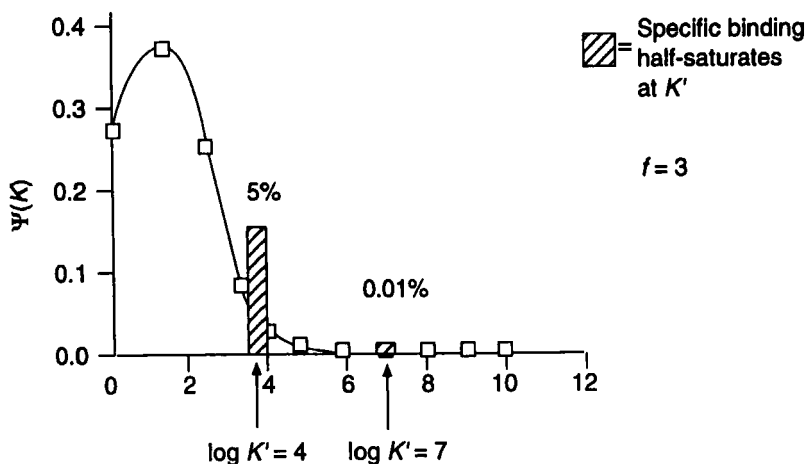


Figure 9 Examples of ‘emergence’ from nonspecific binding in the immune system, based on the RAD model with the parameter values of Figure 5. At a high value of $K' = 10^7 \text{ M}^{-1}$ (typical for immune sera to haptens), a specific antibody should constitute only a small fraction (0.01%) of the total serum Ig to be discernible from a nonspecific background with a signal-to-noise ratio $f = 3$. A ‘weak’ individual immunoglobulin species with $K' = 10^4 \text{ M}^{-1}$ towards the same hapten should be present at a very high concentration (5%) to be similarly detectable. Analogous affinity/quantity considerations apply to the ability of one receptor to recognize a particular ligand among many ‘noisy’ ligands

$f = 3$, while if $K' = 10^4 \text{ M}^{-1}$ the same ratio between specific and nonspecific binding would be achieved only at an exceedingly high abundance of 5% (Figure 9). Because of symmetry, similar arguments may also describe how multiple unrelated ligands interfere with the binding of a specific ligand to its receptor [12]. In general, the present formalism may be viewed as a quantitative tool for describing nonspecific binding, and may be used in the future to describe evolutionary scenarios leading to the emergence of specific receptors out of the background of nonspecific binding.

3.6 Insight Gained from the RAD Model into a Particular System

(a) Predicting the size of the olfactory receptor repertoire

Assuming universality, the same $\Psi(K)$ function allows one to estimate the size of the olfactory receptor repertoire. The maximal affinity of such receptors for ‘general’ odorants has been estimated to be in the region of 10^5 M^{-1} because of kinetic constraints [22]. This value is lower than might be expected on the basis of olfactory thresholds, typically $10^{-9 \pm 1} \text{ M}$ in man [23]. This apparent discrepancy may be explained if one takes into account the variety of amplification mechanisms operating in the olfactory pathway [22]. Using the above estimate of olfactory receptor affinity and equation (14), an olfactory

receptor repertoire $N_{\text{olf}} = 300\text{--}1000$ is obtained. A very similar number was recently determined experimentally [24].

(b) Implications for olfactory coding—why olfactory receptors are ‘broadly specific’

The olfactory bulb acts upon incoming sensory information in the form of a vector of neuronal activation values [25, 26]. Our model allows us to make a general statement on the functional form of such an activity vector. This prediction is based on the notion that in a repertoire of size N , the highest-affinity receptor will be accompanied by $N\Psi(K')$ receptors with a lower affinity K' , in agreement with the known activation of multiple receptor types by a single odorant. It may be calculated (using the above value of Ψ and $N = 300$) that when the highest-affinity receptor for a given odorant is 90% saturated, 12 more receptors (one in 25) are at least 10% saturated; and when the highest-affinity receptor is 99% saturated, 120 additional receptors (more than a third) are above the 10% threshold. In contrast, for an immune system repertoire of 10^7 receptors, the cross-reactivities at 90% and 99% saturation of the highest-affinity receptor are merely one in 2.5×10^4 and one in 1500, respectively (Figure 10). A Hill plot [27] based on the affinity distribution Ψ has a slope $n_H \approx 0.5$, in line with the relatively low value of the slope (0.3–0.5) for the log–log dose–response curves seen in olfactory electrophysiological recordings [28] and psychophysical functions [29]. The activation of multiple receptors by one odorant is thought to be important for olfactory coding, as it provides the central nervous system with patterns of activity with high information content [22, 30].

(c) Human olfactory threshold polymorphisms

It has been proposed that an olfactory threshold, the minimal concentration at which a given odorant is perceived, is determined by the highest-affinity receptor for this odorant [31]. The human population displays a genetically correlated distribution of olfactory thresholds [32, 33], including cases of specific anosmia (inability to perceive particular odorants) [34, 35]. Such genetic polymorphism of thresholds may be explained if each individual expresses a subset of n receptors drawn out of the larger number of alleles in the population. For any odorant, the affinity distribution among all the receptors in the population is likely to be described by Ψ . The distribution of individual thresholds will then be related by a maximal value distribution for the discrete affinity distribution $\Psi(K)$, as previously described [13]. By exploring the relation between the standard deviation of a Gaussian approximation for such a distribution and the values of n , an independent estimate for the size of the human olfactory repertoire $n = N_{\text{olf}} = 500$ was obtained, in very good agreement with the above prediction (Figure 11). The prevalence of specific anosmias may also be analyzed by this formalism [Seidemann *et al.*, in preparation].

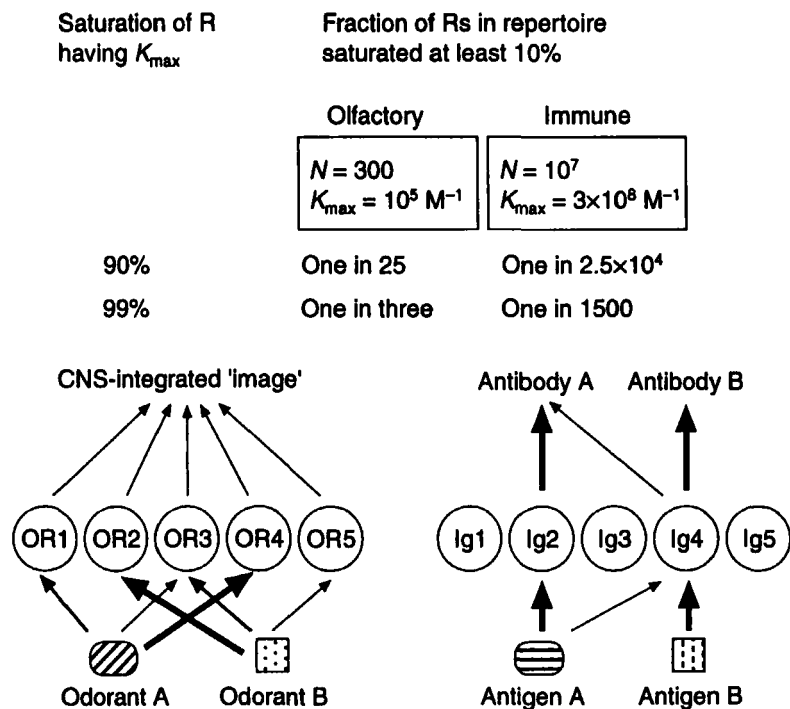
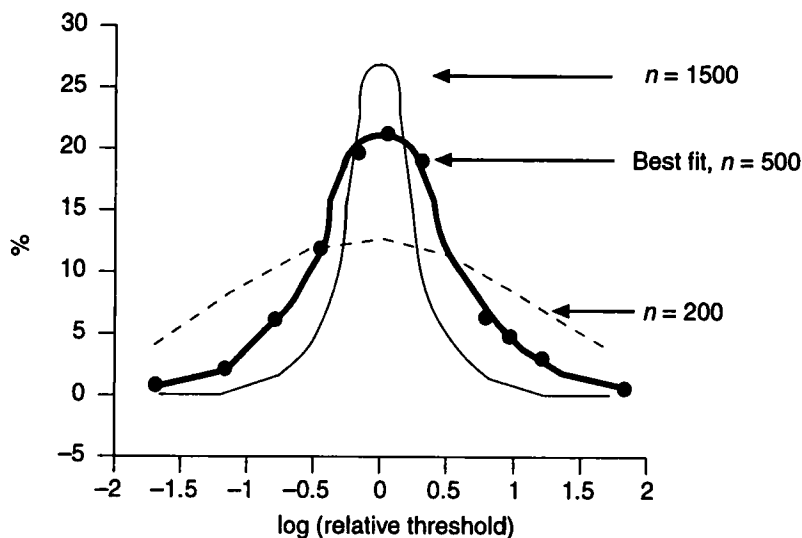


Figure 10 According to the RAD model, the size of a receptor repertoire influences its 'band width' or cross-reactivity value. Given that the highest-affinity receptor is saturated with ligand to a certain value (e.g. 90% of all its molecules are occupied with ligand), how many other receptors in the distribution are appreciably saturated (e.g. at $>10\%$)? As an example, it may be calculated that if the receptor repertoire is small ($N=300$, similar to the estimate for the olfactory receptor repertoire), one in 25 of all the other 'nonspecific' receptors will be saturated. In contrast, for a much larger repertoire, e.g. $N=10^7$, typical of the immune system, only one in 2.5×10^4 of all other receptors will be saturated under similar conditions. Respective numbers for 99% specific receptor saturation are also given

4. THE GEOMETRIC FIT BETWEEN PROTEINS AND THEIR LIGANDS

4.1 The Importance of Molecular Surface Recognition in Specific Protein-Ligand Interactions

The association of proteins with their ligands involves intricate inter- and intramolecular interactions, solvation effects and conformational changes. Because of this complexity, we do not yet have a comprehensive and efficient approach for predicting the formation of protein-ligand complexes from the structures of their free components. However, if certain assumptions are made, such predictions become feasible, and attempts at such prediction, based on



$$P(K_{\max} = K) = \left[\sum_{K \leq K_{\max}} \Psi(K) \right]^n - \left[\sum_{K \leq qK_{\max}} \Psi(K) \right]^n$$

Extreme value distribution of $\Psi(K)$

Figure 11 One more corollary of the RAD model is the extreme value distribution of $\Psi(K)$. This addresses the distribution of values of $K_{\max}$ in many small repertoires of identical size, taking into account statistical fluctuations arising in small samples. Such an extreme value distribution is given by the equation in the figure (equation 10 in [13]), where n is the size of the small repertoire. The extreme value distribution may be approximated by a Gaussian function whose width increases as n decreases. The experimental points in the figure represent a real distribution of olfactory thresholds for a group of human individuals, averaged for nine odorants. The thick curve is fitted with $n = 500$, while the thin and broken curves show what the distribution would look like for other n values

energy minimization, have been partially successful [36–41]. Another simplifying approach, which could alleviate some of these difficulties, is based on geometric considerations.

The three-dimensional structures of most protein complexes reveal a close geometric match between those parts of the surfaces of the protein and the ligand that are in contact. Indeed, the shape and other physical characteristics of the surfaces largely determine the nature of the specific molecular interaction in the complex. Furthermore, in many cases the three-dimensional structures of the components in the complex closely resemble those of the molecules in their free, native states. Geometric matching thus seems to play an important role in determining the structure of a complex.

Crystallographic data showing a close fit between the contiguous molecular surfaces of proteins and their corresponding ligands have been recorded for protein-protein and protein-ligand complexes that could be crystallized. Molecular surface complementarity was observed, for example, in complexes consisting of trypsin and its corresponding pancreatic trypsin inhibitor [42, 43], for the two subunits of citrate synthetase [44], for the α -subunit and β -subunit of human hemoglobin [45], for lysozyme and its corresponding F_{ab} fragments [46], for aspartic proteinase and its corresponding peptide inhibitor [47, 48] and for tyrosyl tRNA synthetase and tyrosine adenylate [49].

Furthermore, it is worth noting that in those cases in which conformational alterations occur in the separate components of the complex before final complexation (by an induced fit or any other molecular mechanism), they take place in order to produce complex components possessing geometrically complementary contiguous surfaces.

Several investigators have exploited the advantages of a geometric approach to study shape complementarity between a given protein and its ligand [50–62]. They considered the geometric match between molecular surfaces as a fundamental condition for the formation of a specific complex [56]. In this approach, which treats proteins as rigid bodies, the complementarity between surfaces is estimated. The geometric analysis can serve as the basis for a more complete approach that includes energy considerations.

4.2 Geometric Algorithm for Identification of Molecular Surface Complementarity

A geometric recognition algorithm to identify molecular surface complementarity was developed recently [63]. It is based on a purely geometric approach and employs techniques applied in the field of pattern recognition. The algorithm involves an automated procedure, which includes (1) a digital representation of the molecules (derived from atomic coordinates) by three-dimensional discrete functions which distinguish between the surface and the interior; (2) the calculation, using Fourier transformation, of a correlation function which assesses the degree of molecular surface overlap and penetration upon relative shifts of the molecules in three dimensions; and (3) a scan of the relative orientations of the molecules in three dimensions. The algorithm provides a list of correlation values indicating the extent of geometric match between the surfaces of the molecules; each of these values is associated with six numbers describing the relative positions (translation and rotation) of the molecules. The procedure is thus equivalent to a six-dimensional search, but is much faster by design, and the computation time is only moderately dependent on the molecular size.

In the development of our algorithm we began with a geometric description of the protein and ligand molecules derived from their known atomic coordinates. The two molecules, denoted by **a** and **b**, are projected onto a

three-dimensional grid of $N \times N \times N$ points, where they are represented by the discrete functions

$$a_{l,m,n} = \begin{cases} 1 & \text{inside the molecule} \\ 0 & \text{outside the molecule} \end{cases}$$

and

$$b_{l,m,n} = \begin{cases} 1 & \text{inside the molecule} \\ 0 & \text{outside the molecule} \end{cases}$$

where l , m and n are the indices of the three-dimensional grid ($l, m, n = 1, \dots, N$). Any grid point is considered to be inside the molecule if there is at least one atomic nucleus within a distance r from it, where r is of the order of the van der Waals atomic radius. Examples of two-dimensional cross sections of these functions are presented in Figure 12(a) and 12(b).

Next, to distinguish between the surface and the interior of each molecule, we retain the value of 1 for the grid points along a thin surface layer only, and assign other values to the internal grid points. The resulting functions thus become

$$\bar{a}_{l,m,n} = \begin{cases} 1 & \text{on the surface of the molecule} \\ \rho & \text{inside the molecule} \\ 0 & \text{outside the molecule} \end{cases}$$

and

$$\bar{b}_{l,m,n} = \begin{cases} 1 & \text{on the surface of the molecule} \\ \delta & \text{inside the molecule} \\ 0 & \text{outside the molecule} \end{cases}$$

where the surface is defined here as a boundary layer of finite width between the inside and the outside of the molecule. The parameters ρ and δ describe the values of the points inside the molecules, and all points outside are set to 0. Two-dimensional cross sections of these functions are shown in Figure 12(c) and 12(d).

In our method, matching of surfaces is accomplished by calculation of correlation functions. The correlation between the discrete functions $\bar{a}$ and $\bar{b}$ is defined as

$$\bar{c}_{\alpha,\beta,\gamma} = \sum_{l=1}^N \sum_{m=1}^N \sum_{n=1}^N \bar{a}_{l,m,n} \bar{b}_{l+\alpha, m+\beta, n+\gamma} \quad (16)$$

where α , β and γ are the numbers of grid steps by which molecule **b** is shifted with respect to molecule **a** in each dimension. If the shift vector $\{\alpha, \beta, \gamma\}$ is such

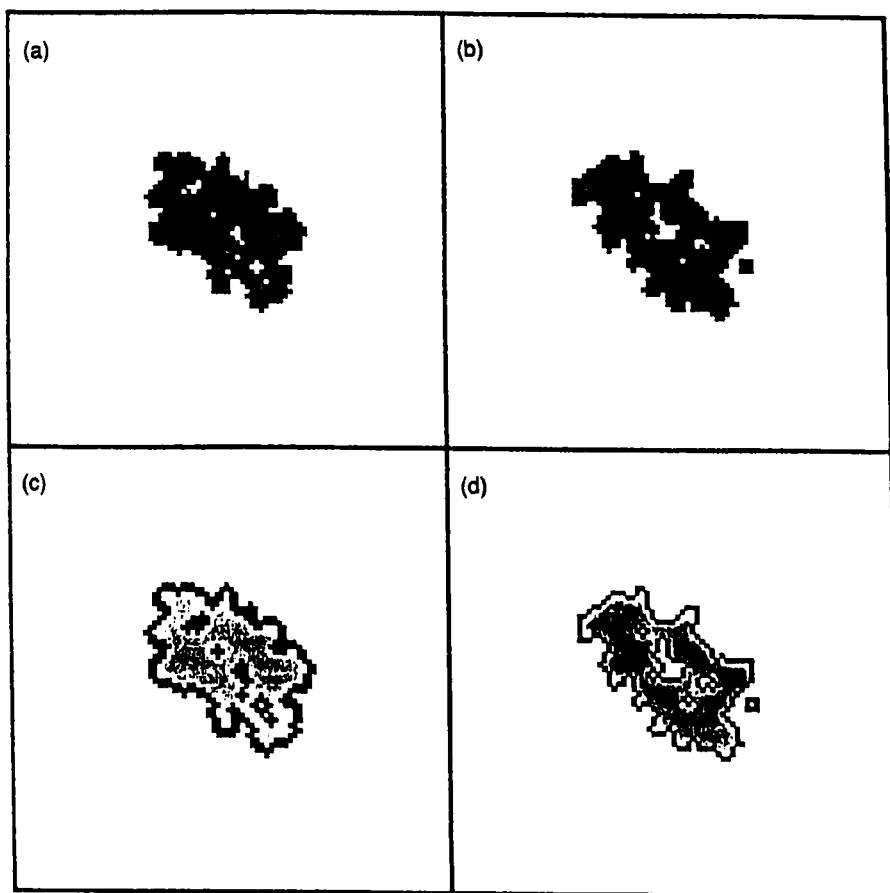


Figure 12 Typical cross sections through the three-dimensional grid representations of the molecules. (a) Cross section (at $l=46$) through the function $a_{l,m,n}$ derived by projection of the α -subunit of hemoglobin (from 2HHB) onto a three-dimensional grid ($N=90$). The values 0 and 1 are represented in white and black, respectively. (b) The cross section $b_{46,m,n}$ was similarly derived for the β -subunit (from 2HHB). Other details are as in (a). (c) The cross section (at $l=46$) through the function $\bar{a}_{l,m,n}$ obtained by distinction of the surface layer from the interior of the molecule in the function $a_{l,m,n}$. The large negative value for r is represented in gray. (d) Cross section $\bar{b}_{46,m,n}$ similarly derived from $b_{l,m,n}$. The small positive value for d is represented in a different shade of gray. The values for r and h were 1.8 Å and 1.2 Å, respectively

that there is no contact between the two molecules (see Figure 13a), the correlation value is 0. If there is contact between the surfaces (Figure 13b), the contribution to the correlation value is positive. Nonzero correlation values can also be obtained when one molecule penetrates into the other (Figure 13c). Since such penetration is physically forbidden, a distinction between surface contact and penetration must be clearly formulated. Accordingly, we assign a large

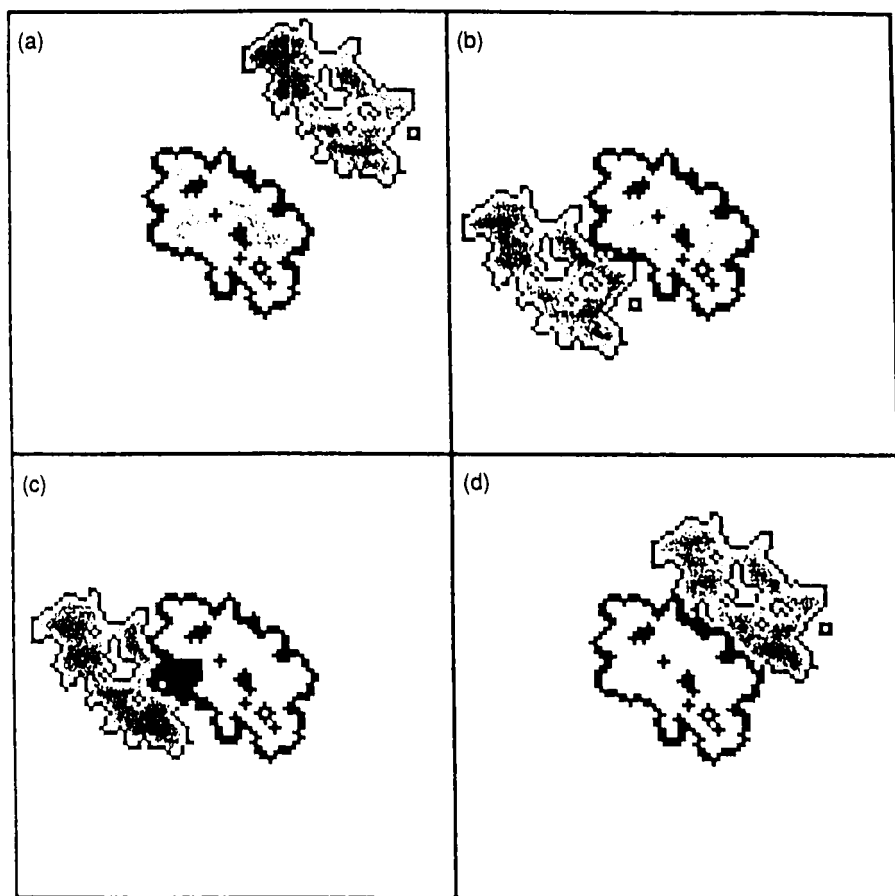


Figure 13 Different relative positions of molecules **a** and **b** illustrated by the cross sections $\bar{a}_{46,m,n}$ and $\bar{b}_{46,m,n}$ from Figure 12. The relative orientations of the molecules are as in the known α - β dimer: (a) no contact; (b) limited contact; (c) penetration (the penetrated area is represented in black); and (d) good geometric match, as indicated by the extensive overlap of complementary surface layers

negative value to ρ in $\bar{a}$ and a small positive value to δ in $\bar{b}$. Thus, when the shift vector $\{\alpha, \beta, \gamma\}$ is such that molecule **b** penetrates molecule **a**, the multiplication of the negative numbers (ρ) in $\bar{a}$ by the positive numbers (1 or δ) in $\bar{b}$ results in a negative contribution to the overall correlation value. Consequently, the correlation value for each displacement is simply the score for overlapping surfaces corrected by the penalty for penetration.

Positive correlation values are obtained when the contribution from surface contact outweighs that from penetration. Thus, a good geometric match (such as in Figure 13d) is represented by a high positive peak, while low values reflect a poor match between the molecules. A cross section of a typical correlation

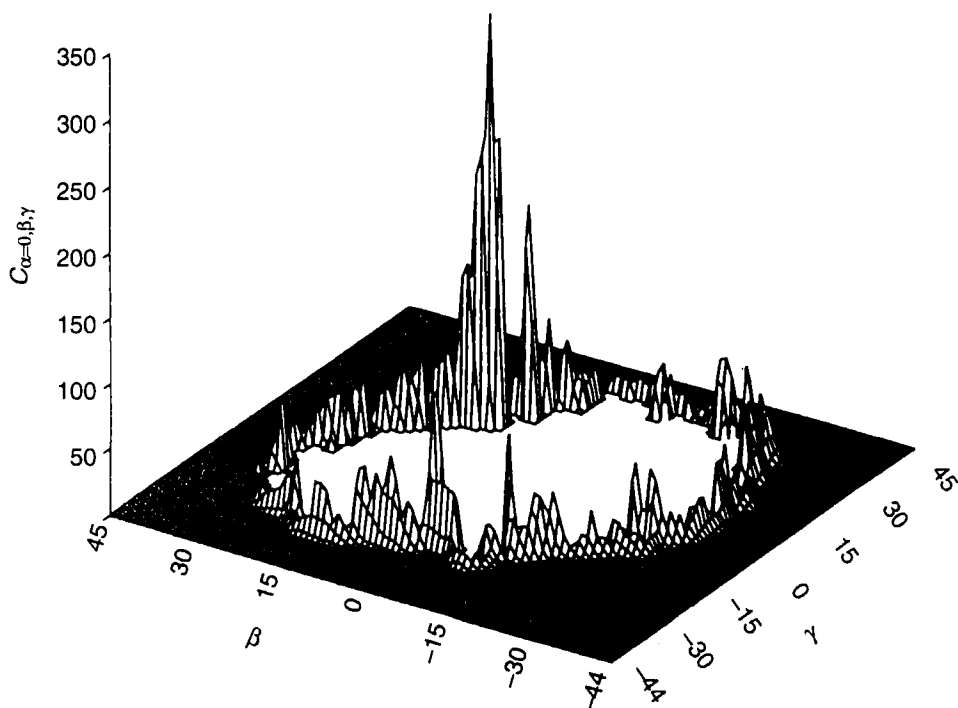


Figure 14 Cross section (at $\alpha=0$) through a three-dimensional correlation function $\bar{c}_{\alpha, \beta, \gamma}$. The correlation function shown was calculated for the α -subunit and β -subunit of hemoglobin oriented as in the dimer (from 2HHB, see Figures 13c and 13d). The correlation value at each shift vector $\{0, \beta, \gamma\}$ is represented by the height of the graph. The prominent peak at $\{\alpha=0, \beta=14, \gamma=17\}$ corresponds to the correct match between the molecules (see Figure 13d). Other intermolecular surface contacts (such as in Figure 13b) give rise to the low positive correlation values around the center of the graph. The negative correlation values caused by penetration (see Figure 13c) are omitted, leaving the empty area at the center

function for a good match is presented in Figure 14. The coordinates of the prominent peak denote the relative shift of molecule **b** yielding a good match with molecule **a**. The locations of the recognition sites on the surface of each molecule can readily be determined from these coordinates. In addition, the width of the peak provides a measure of the relative displacement allowed before matching is lost.

Direct calculation of the correlation between the two functions (see equation 16) is rather a lengthy process, since it involves N^3 multiplications and additions for each of the N^3 possible relative shifts $\{\alpha, \beta, \gamma\}$, resulting in N^6 computing steps. We therefore chose to take advantage of Fourier transformations, which allowed us to calculate the correlation function much more rapidly.

Finally, in order to complete a general search for the match between the surfaces of molecules **a** and **b**, the correlation function $\bar{c}$ has to be calculated

for all relative orientations of the molecules. In practice, molecule **a** is fixed, whereas the three Euler angles defining the orientation of molecule **b** (*xyz* convention in [64]) are varied at fixed intervals of Δ° . This results in a complete scan of $360 \times 360 \times 180/\Delta^3$ orientations for which the correlation function $\bar{c}$ must be calculated.

Each high and sharp peak found by this procedure indicates a geometric match, and thus represents a potential complex. The relative positions and orientations of the molecules within each potential complex can readily be derived from the coordinates of the correlation peak and from the three Euler angles at which the peak was found.

4.3 Implementation of the Algorithm and its Application to Several Known Complexes

In order to implement our algorithm, it is necessary to assign specific values to the various parameters involved, i.e. the surface layer thickness r , Δ , ρ , δ , N and the grid step size denoted by η . The choice of these values is based on a number of considerations, as outlined in our original article [63].

The values assigned to these parameters in our computer program were $r = 1.8 \text{ \AA}$, $\Delta = 20^\circ$, $\rho = -15$, $\delta = 1$, $N = 90$ ($\eta \approx 1.0\text{--}1.2 \text{ \AA}$) for the scan stage and $N = 128$ ($\eta \approx 0.7\text{--}0.8 \text{ \AA}$) for the discrimination stage. The program was run on a Convex C-220 computer using the VecLib Fast Fourier Transform subroutine. The computation time for each iteration (steps 3–8 in the summarized algorithm) in the scan stage was 9 s. The total computation time for matching two molecules in the region of 1110 atoms each, including both the initial scan and the discrimination stage, was typically 7.5 h.

Our algorithm was applied to several known complexes whose coordinates are given in the Brookhaven Protein Data Bank (Brookhaven National Laboratory, Upton, NJ) in order to test its ability to predict the correct structures of protein complexes. We chose complexes that represent a wide variety of relative sizes for molecules **a** and **b**. These were two hemoglobin variants: human deoxyhemoglobin [45] (2HHB) and horse methemoglobin [65] (2MHB), representing naturally occurring heterodimers; and three complexes: tyrosyl tRNA synthetase–tyrosine adenylate [49] (3TS1), aspartic proteinase–peptide inhibitor [47] (3APR) and trypsin–trypsin inhibitor [43] (2PTC). In these tests, we treated the component molecules as separate entities by using their respective atomic coordinates within the complex. Additional tests were performed with native aspartic proteinase [66] (2APR) and its peptide inhibitor, as well as with trypsin and native trypsin inhibitor [67] (4PTI). The relative positions of the molecules yielding the best geometric fit in a complex, as determined by the algorithm, were finally compared with the corresponding known complex.

The results are summarized in Figure 15, which shows histograms of the 10 correlation peaks for each pair of molecules. The left side of each panel presents the 10 highest peaks obtained at the scan stage, whereas the right side shows

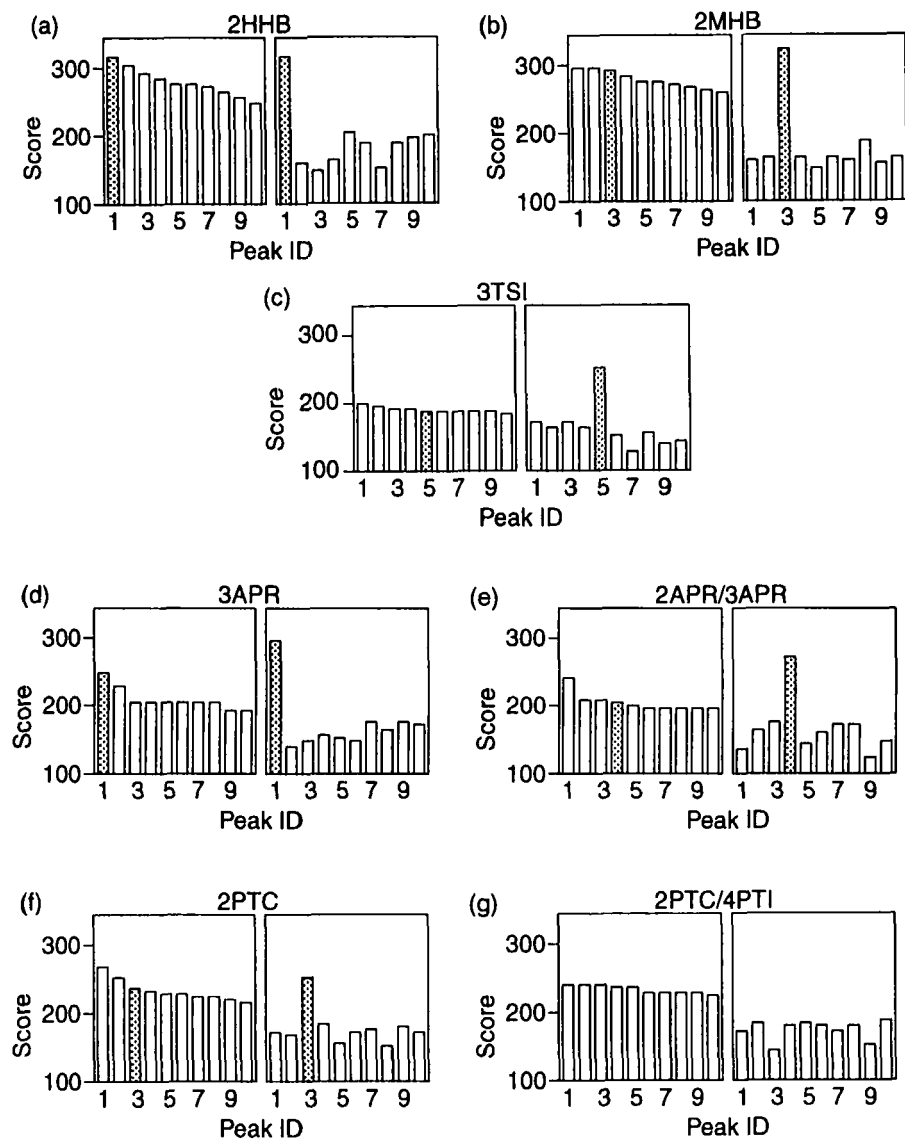


Figure 15 Correlation results for different pairs of molecules. The pairs are identified by their respective codes (see text). In each panel, the histogram on the left shows the 10 highest correlation peaks obtained at the scan stage ($\eta = 1.0-1.2 \text{ \AA}$), sorted by their score. Each of these peaks was obtained at different relative orientations of the molecules and corresponds to a potential geometric match. The shaded peak in each histogram corresponds to the known complex between the molecules. The histogram on the right side of each panel shows the scores obtained at the discrimination stage ($\eta = 0.7-0.8 \text{ \AA}$) for the 10 orientations singled out at the scan stage

the peaks reevaluated for the same 10 orientations in the discrimination stage. It can be seen that the correlation peak for the known complex (shaded) is not necessarily the highest at the scan stage. However, the highest peak that was obtained after discrimination represents the correct orientation and position of molecule **b** with respect to **a**, and is significantly higher than the other peaks.

Application of the algorithm to the α -subunit and β -subunit of human hemoglobin (2HHB, Figure 15a) revealed that the highest peak at the scan stage (score 312) corresponds to the well-known α - β dimer. In the horse methemoglobin variant (2MHB, Figure 15b), however, the correct position for the dimer is represented by the third peak (score 290) in the sorted histogram for the scan stage. Nevertheless, both of these peaks became predominant at the discrimination stage (scores 302 and 347 for 2HHB and 2MHB, respectively).

We next applied the algorithm to the tyrosyl tRNA synthetase-tyrosine adenylate pair (3TS1, Figure 15c), which serves as an example of a complex between a high-molecular-weight protein and a small ligand. In this case, the correlation peak corresponding to the correct position of the ligand in the complex was not the highest one at the scan stage. Discrimination, however, yielded the expected result, i.e. the correct orientation was associated with a peak distinctly higher than the other peaks.

Further assessment of the procedure was carried out by analysis of the complex between aspartic proteinase and its peptide inhibitor (3APR). This system illustrates a case in which the structure of the protein in the complex closely resembles the structure of the native protein [47, 48]. It is thus possible to look for the best match between the structure of the complexed peptide and the protein in either its complexed (3APR) or its native (2APR) structure. With the complexed protein the correct relative position of the ligand yielded the highest peak already at the scan stage (Figure 15d), whereas with the native protein the peak describing the correct position was only the fourth in the sorted list (Figure 15e).

Analysis of the complex trypsin-trypsin inhibitor (2PTC) was chosen because the native structure of one of the components, namely the inhibitor, differs from its structure in the complex. Specifically, conformational changes involving the side chains of three amino acids, located in the binding site of the inhibitor, occur upon complex formation [43, 67]. When the structure of the inhibitor in the complex was used (Figure 15f), the highest peak after discrimination corresponded to the correct position of the inhibitor in the complex. However, when the native structure of the inhibitor (4PTI) was used (Figure 15g), the algorithm did not yield a distinct correlation peak either at the scan stage or at the discrimination stage. This result indicates that the extent of the conformational change occurring at the surface of the inhibitor upon binding to trypsin exceeds that tolerated by the algorithm.

4.4 Future Considerations

Further theoretical considerations indicate that it might be possible to include hydrophobic and electrostatic interactions in our algorithm in addition to the

van der Waals interactions determining the molecular geometric fit discussed above. This would increase the reliability of prediction for the molecular structure of complexes formed between proteins and their ligands whose three-dimensional structures are known and which do not undergo a marked change in their conformations on complexation. One might therefore expect that in the future it will be possible to predict the structure of complexes formed between proteins with a known structure. It is also reasonable to assume that extension of the algorithm will facilitate the prediction and understanding of the formation of multiprotein structures composed of a considerable number of subunits whose three-dimensional structures are known.

When complex formation is accompanied by conformational alterations in the participating components, energy considerations have to be taken into account. One should bear in mind, however, that in most of the cases investigated so far there is a good geometric fit between the contiguous molecular surfaces in the final complexes.

Van der Waals and electrostatic interactions and hydrogen-bonding and hydrophobic interactions within the region of contact between a protein and its corresponding ligand determine the energy of complex formation. In a system consisting of a given protein or receptor and an ensemble of ligands, as discussed above, it is the energy of binding that determines the specificity of the protein or receptor under the conditions specified. As van der Waals forces play a major role in many of the complexes investigated, it is plausible that molecular surface recognition, i.e. the geometric fit between proteins and their ligands, plays an important role in determining biological specificity.

5. ADDITIVITY AND NONADDITIVITY IN PROTEIN-LIGAND INTERACTIONS

5.1 Analysis of Additivity in Protein-Engineering Experiments

Complex systems are difficult to study, largely owing to the mutual interdependence and multiplicity of interactions that give rise to them. A reductionist approach to studying such systems is severely limited by the fact that their properties are a function of the entire system and not of the sum of its parts. Experimental intervention usually aims to remove or perturb only one component of a system in order to determine the effect this has on the system as a whole. In the case of complex systems, however, such intervention often results in other changes not directly related to the component under study, thus preventing clear-cut analysis. For example, in the social sciences one sometimes wishes to study interactions between specific individuals within the context of a larger group. The obvious experimental approach is to study such a group both in the presence and in the absence of the individual(s) of interest. One problem in such experiments, however, is that interactions between other

members of the group may change in the absence of the individual of interest. For the biochemist, proteins are an example of such complex systems. The structural and functional properties of proteins are determined by the many interdependent interactions between their constituent amino acids. In what follows we describe a powerful experimental and theoretical approach to the study of proteins and protein–ligand complexes which circumvents some of the problems that arise owing to their complexity. Although the subject matter of this chapter is protein–ligand interactions, part of the discussion and some of the examples in this section will concern intramolecular interactions. This is justified by the physical and formal analogies between the two types of interactions (see below) and the lack of accurate experimental data on the contributions of defined intermolecular interactions to binding energies (see Section 5.4).

An additivity-based reductionist approach to the study of proteins has been previously suggested [68, 69]. Since the ‘folding problem’ of predicting the structure of a protein from its sequence has yet to be solved, it was proposed that direct sequence-to-reactivity algorithms would be of considerable value [68, 69]. Such algorithms would bypass the protein folding problem by predicting the activity of a protein directly from its sequence. The underlying assumption in such algorithms is that the contributions of amino acids to the activity of a protein are additive. Additivity exists when a certain functional or structural property of a protein is found to be a sum of the independent contributions of a number of different residues. For example, when only two positions in a protein sequence are considered there are 20×20 possible amino acid combinations. In cases where the additivity assumption holds it should be sufficient to determine experimentally the activity of only 39 specific combinations ($20 + 20 - 1$, i.e. one row and one column in the corresponding matrix), allowing the activity (e.g. the contribution to the binding energy of a ligand) of the other 361 combinations to be predicted. In other words, the effect on the activity of each double mutation may, in such cases, be predicted from the effects of the two corresponding single mutations.

Consider, for example, two residues, X and Y, that contribute to a functional or structural property of a protein. Additivity (or nonadditivity) in the contributions of these residues may be detected by construction of a double-mutant cycle [70] comprising wild-type protein, the two single mutants in which either X or Y is replaced and the double mutant where both residues are replaced (Figure 16). The change in free energy upon mutation of X, associated with a functional or structural property of the protein (such as ligand binding or unfolding), may be expressed relative to wild-type protein as $\Delta G_{P-X \rightarrow P-Y}$. Similarly, the change in free energy of this property upon mutation of X, when Y has already been mutated, is $\Delta G_{P-X \rightarrow P}$. The free energy changes $\Delta G_{P-XY \rightarrow P-X}$ and $\Delta G_{P-Y \rightarrow P}$ are similarly defined. If the effects of the mutations are not independent of each other, then $\Delta G_{P-XY \rightarrow P-Y} \neq \Delta G_{P-X \rightarrow P}$ and $\Delta G_{P-XY \rightarrow P-X} \neq \Delta G_{P-Y \rightarrow P}$. The free energy of coupling $\Delta \Delta G_{\text{int}}$ between residues X and Y is given by

$$\Delta \Delta G_{\text{int}} = \Delta G_{P-XY \rightarrow P-Y} - \Delta G_{P-X \rightarrow P} = \Delta G_{P-XY \rightarrow P-X} - \Delta G_{P-Y \rightarrow P} \quad (17)$$

In the case of additivity $\Delta\Delta G_{\text{int}} = \Delta G_{\text{P-XY} \rightarrow \text{P-Y}} - \Delta G_{\text{P-X} \rightarrow \text{P}} = \Delta G_{\text{P-XY} \rightarrow \text{P-X}} - \Delta G_{\text{P-Y} \rightarrow \text{P}} = 0$. The free energies corresponding to the transitions in the cycle (Figure 16) cannot be determined directly from experiment. They are calculated by assuming the following relations

$$\Delta G_{\text{P-XY} \rightarrow \text{P-Y}} = \Delta G_{\text{P-XY}} - \Delta G_{\text{P-Y}} \quad (18)$$

$$\Delta G_{\text{P-X} \rightarrow \text{P}} = \Delta G_{\text{P-X}} - \Delta G_{\text{P}} \quad (19)$$

where $\Delta G_{\text{P-XY}}$, $\Delta G_{\text{P-Y}}$, $\Delta G_{\text{P-X}}$ and ΔG_{P} are the free energies of some process (e.g. ligand binding or unfolding) that the wild-type protein, the two single mutants and the double mutant, respectively, are undergoing. By combining equations (17)–(19) one obtains

$$\Delta\Delta G_{\text{int}} = \Delta G_{\text{P-XY}} - \Delta G_{\text{P-Y}} - \Delta G_{\text{P-X}} + \Delta G_{\text{P}} \quad (20)$$

The free energy of coupling $\Delta\Delta G_{\text{int}}$ is a measure of the extent of nonadditivity and is determined from equation (20). It follows from equation (20) that in the case of additivity when $\Delta\Delta G_{\text{int}} = 0$, the determination of the free energies of some process for any three of the four proteins in the cycle enables one to predict the free energy of that process for the fourth protein.

It is important to note that equations (18) and (19) do not hold unless certain assumptions are made (see [71] and [72] for a discussion of these assumptions).

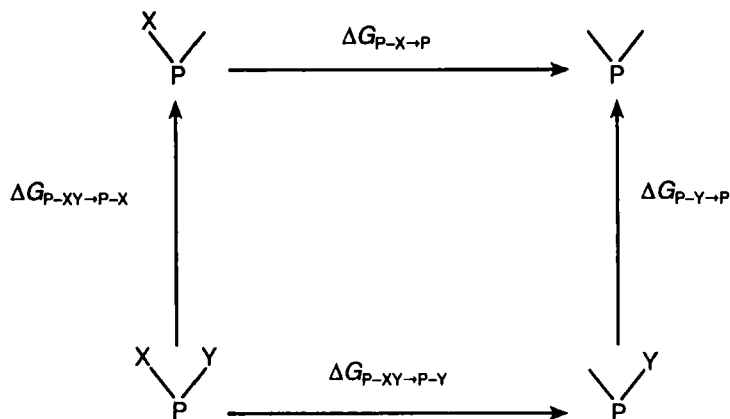


Figure 16 Scheme for a double-mutant cycle. P stands for the protein, X and Y for the two residues being mutated and $\Delta G_{\text{P-XY} \rightarrow \text{P-X}}$, $\Delta G_{\text{P-Y} \rightarrow \text{P}}$, $\Delta G_{\text{P-XY} \rightarrow \text{P-Y}}$ and $\Delta G_{\text{P-X} \rightarrow \text{P}}$ for the appropriate transitions in the cycle. Each of the free energies corresponding to a transition in the cycle is determined by measurement of the difference between the free energies associated with the functional property of the proteins in the initial and final states in that transition. Thus, for example, when the functional property is unfolding, $\Delta G_{\text{P-XY} \rightarrow \text{P-X}}$ is taken to be equal to the difference between the unfolding free energies of P-XY and P-X

It may be shown that, in general, the free energy of coupling determined from double-mutant cycles when the functional property is ligand binding is a sum of intramolecular and intermolecular contributions [71]. The intramolecular contribution exists in the absence of ligand and can be determined from protein-unfolding experiments (see below), whereas the intermolecular contribution depends on having the ligand present. By measuring both contributions one can show the existence of 'induced fit' in limiting cases when the intramolecular contribution is small relative to the intermolecular contribution or vice versa.

5.2 Additivity in Protein-Protein Interactions

Examples of additivity in protein-protein interactions have been reviewed and analyzed previously [73, 74]. Additivity in peptide-protein interactions has also been reported. One example of additivity in peptide-protein interactions, reported by Berger and Schechter [75, 76] as early as 1967, is the interaction of papain with various synthetic peptides. A striking case of additivity in protein-protein interactions is the interaction of ovomucoid third-domain inhibitors with elastase, chymotrypsin and subtilisin [69, 77, 78]. As mentioned above, Laskowski and coworkers selected to study the latter interaction in an attempt to develop an additivity-based algorithm that bypasses the intermediate protein-folding step. This protein proteinase inhibitor family was thought to be very well suited for the development of such an algorithm for the following reasons: (1) ovomucoid third-domain inhibitors represent a highly homologous protein family with much structural similarity [79, 80] but with differences in specificity for different enzymes owing to the variability of their amino acid sequences; (2) owing to the rigid structure of ovomucoid third-domain inhibitors, in particular in the area of their reactive sites, there is little structural change upon their binding to proteinases; and (3) the inhibition constant K_i of an enzyme-inhibitor interaction is a true equilibrium constant from which the free energy of binding may be readily obtained. Equilibrium constants for the interaction of a large number of ovomucoid third-domain inhibitors with elastase, chymotrypsin and subtilisin have been accurately determined [69, 78].

Additivity in the interaction of ovomucoid third-domain inhibitors has been demonstrated by use of double-mutant cycles [77]. Here, we describe in more detail a somewhat different approach, based on the mathematical method of Free and Wilson [81] that was originally developed for the design of drugs. An advantage of this method is that it is applicable not only to data corresponding to double-mutant cycles but also to those of more complex mutational flow diagrams (Figure 17). Following the method of Free and Wilson [81], one may write an equation for the free energy ΔG_i that is associated with a functional or structural property of a protein as follows

$$\Delta G_i = \frac{1}{N} \sum_{i=1}^N \Delta G_i + \sum_n \Delta \Delta G_{i,n} \quad (21)$$

where i stands for the i th protein in a family comprising N homologous proteins and n stands for the n th amino acid residue in any of the N proteins. The first term in equation (21) is the overall average of the experimentally determined free energies associated with a functional or structural property of the N proteins. $\Delta\Delta G_{i,n}$ is the deviation from the average contribution to ΔG_i of the n th amino acid. Thus, for any position in the N aligned sequences, the sum of these deviations for all N proteins is 0 ($\sum_i \Delta\Delta G_{i,n} = 0$). Each of these deviations, whether positive or negative, is defined as the 'contribution' to ΔG_i of the n th amino acid residue in the i th protein. Equation (21) assumes additivity since the 'contributions' depend on the position and type of the amino acid but not on other neighboring residues, i.e. interactions are neglected. This method was applied [82] to the data of Empie and Laskowski [69] for the proteins in Figure 17. The figure shows only the residues at the seven positions at which there is variability in the amino acid sequence. The 'contributions' of the different amino acid residues to ΔG_i , the free energy of association of the i th inhibitor to the respective proteinase, were calculated from a set of 16 simultaneous linear equations with 16 unknowns. Of these, nine equations (of the form of equation 21) are for different inhibitor sequences and seven equations are the constraints, $\sum_i \Delta\Delta G_{i,n} = 0$, for each of the seven positions. From these contributions we were able to predict [82] the free energies of association for inhibitors that were excluded from the data set. Excellent agreement was found between the predicted and the experimentally determined values. In addition, it was possible to suggest a sequence for a new, more powerful and selective ovomucoid third-domain inhibitor of subtilisin. To the best of our knowledge, this suggestion still remains to be tested experimentally.

5.3 Nonadditivity in Protein–Ligand Interactions and the Hammett Plot

Structure–reactivity relations of physical organic chemistry [83] may be employed to analyze mutational effects in proteins. Fersht *et al.* [84] showed that the Brønsted linear free energy relation [85] can be applied to data for engineered proteins, thereby providing information on their reaction mechanisms. In this section, we describe a previous application [86] of another structure–reactivity relation of physical organic chemistry, the Hammett equation [87], to ligand binding data for engineered proteins.

The Hammett equation of physical organic chemistry is in fact analogous to double-mutant cycles in protein engineering. It can be readily shown from equation (20) that in the case of additivity ($\Delta\Delta G_{\text{int}} = 0$) one has

$$\frac{K_{\text{P-XY}}}{K_{\text{P-Y}}} = \frac{K_{\text{P-X}}}{K_{\text{P}}} \quad (22)$$

where $K_{\text{P-XY}}$, $K_{\text{P-X}}$, $K_{\text{P-Y}}$ and K_{P} are the respective equilibrium ligand binding constants of the four proteins comprising the double-mutant cycle. Equation (22)

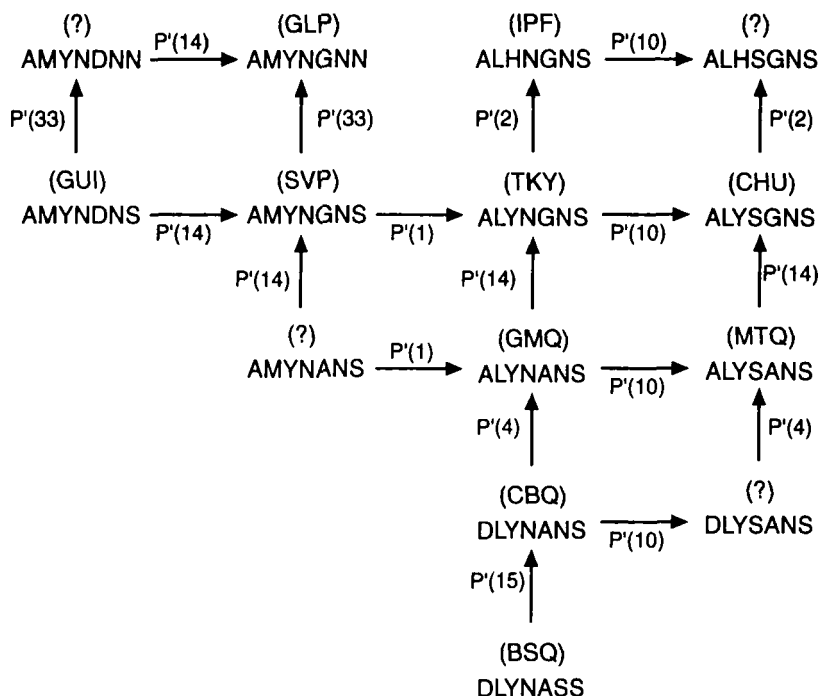


Figure 17 Mutational flow diagram for ovomucoid third-domain inhibitors. The positions in the sequence (P_n , P'_n) are according to Schechter and Berger [75]. The names of the inhibitors (in parentheses) are according to their source, as follows: TKY, turkey; SVP, silver pheasant; GMQ, Gambel's quail; IPF, Indian peafowl; GUI, Guinea fowl; MTQ, mountain quail; GLP, golden pheasant; CHU, chukar; BSQ, blue-scaled quail; CBQ, chestnut-bellied scaled quail. Question marks indicate that the corresponding sequences were not known at the time the predictions were made [82] (after Empie and Laskowski [69])

is a particular case (where $\beta = 1$) of the following more general equation

$$\frac{K_{P-XY}}{K_{P-Y}} = \left(\frac{K_{P-X}}{K_P} \right)^\beta \quad (23)$$

Equation (23) is of the form of the Hammett equation [87], which was originally formulated to describe the ionization of ring-substituted aromatic acids. The 'substitution site' and 'reaction site' in the aromatic ring are analogous to the two mutated positions in the protein sequence. The original formulation of the Hammett equation expresses the logarithm of the dissociation constant of the substituted aromatic acid as a product of the contributions from the 'reaction' and 'substitution' sites

$$\log K_{XY} = \log K_{0Y} + \rho_Y \sigma_X \quad (24)$$

In this equation, X and Y stand for the substituent and acidic groups at the substitution and reaction sites, respectively. K_{XY} and K_{0Y} are the respective dissociation constants of the substituted and unsubstituted molecules both with the same acidic group Y, ρ_Y is the reaction constant and σ_X is the substitution constant. By taking the logarithm of both sides of equation (23) one obtains equation (24), where ρ_Y and σ_X correspond to β and $\log(K_{P-X}/K_P)$, respectively.

It may be seen that the value of β in equation (23) (and of ρ in equation 24) depends on the extent of nonadditivity and is related to $\Delta\Delta G_{\text{int}}$ by the following relation

$$\Delta\Delta G_{\text{int}} = (K_P/K_{P-X})^{(1-\beta)} = (K_{P-Y}/K_{P-XY})^{(1-\beta)/\beta}$$

Additivity, as discussed in previous sections of this chapter, is thus a particular case of the more general Hammett equation. It is important to note that equation (23) enables one to predict the equilibrium binding constants of some proteins from the data of other related proteins not only in cases of additivity (Section 5.2) but also in cases of nonadditivity when $\beta \neq 1$. Since β is an additional unknown, whose value must be determined from the experimental data, predictions based on equation (23) or (24) must consider simultaneously more than one double-mutant cycle.

Application of the Hammett equation requires the choice of appropriate 'reference' states. In physical organic chemistry the hydrogen atom is usually the reference state for the 'substitution site' and the identification of a 'reference' acidic group does not usually pose a problem. In the case of proteins, one possibility for a reference state is the amino acid glycine, which lacks a side chain. Owing to the special conformational properties of glycine, the amino acid alanine would probably be a better candidate for a reference state. The problem of choosing reference states may be circumvented, however, by deriving equation (24) through a different route using factor analysis [88]. The use of factor analysis also makes it possible to determine the value of ρ (β in equation 23) by simultaneous analysis of an array of double-mutant cycles. In order to do this, a matrix **d** is formed where each element d_{XY} ($d_{XY} = -\log K_{XY}$) stands for the activity of a protein with amino acid X at one position of its sequence and amino acid Y at another position. The entries in each column of the matrix are for proteins with a particular substitution at one of the positions. Each row corresponds to proteins with a particular substitution at the second position. Thus, any 2×2 submatrix of this data matrix corresponds to one double-mutant cycle. Using matrix algebra one can show that each element of the data matrix **d** of dimensions $m \times n$ for $m \leq n$ can be written as

$$d_{XY} = \sum_{\alpha=1}^m \rho_Y^{(\alpha)} \sigma_X^{(\alpha)} \quad (25)$$

Each element of the matrix is therefore expressed as a sum of m weighted factors. The maximal number of factors needed to reproduce the matrix depends on its size (dimensions) and not on the entries (data). It may happen, however, that fewer factors are required to reproduce the data within experimental error. This indicates some form of interdependence between the elements of the matrix.

Wells *et al.* [89] constructed a series of mutations at two positions in subtilisin and measured for each mutant its K_M for different short-peptide substrates. It was possible to organize the experimental data reported in their paper [89] in matrix form ($m \times n = 12$) as described above. Factor analysis showed [86] that the data matrix could be reproduced using equation (25) with only two terms (i.e. two factors). The Hammett equation (equation 24) has a two-factor form (i.e. equation 25 with only two terms), with one of the factors, $\log K_{OY}$, being a constant for each column. The two-factor result for the data of Wells *et al.* [89] was therefore rewritten in the form of equation (24) by imposing the constraint $\rho_X^{(\alpha=1)} = 1$. It was found that the data could also be reproduced within experimental error using an equation of the form of equation (24). Predictions were made by deleting one element at a time from the $m \times n$ data matrix and then reproducing the matrix from equation (24) using the remaining $m \times n - 1$ elements plus a trial value for the missing element. The predicted value for the missing element was the trial value for which the best fit was obtained between the $m \times n - 1$ elements plus the trial value and the reproduced matrix.

5.4 Intra- and Intermolecular Interactions: Detection and Measurement by Invoking Double-Mutant Cycles

We now turn to applications of the double-mutant cycle method in cases of nonadditivity. Double-mutant cycles were first constructed, using protein engineering methods, by Carter *et al.* [70] in order to detect possible interactions between active site residues of the enzyme tyrosyl tRNA synthetase from *Bacillus stearothermophilus*. The crystal structures of the enzyme-tyrosine and enzyme-tyrosyl adenylate complexes of the mutant Cys35→Gly subsequently confirmed the prediction made from the cycle involving the mutations Cys35→Gly and Thr51→Pro [90]. The side chain of Thr51 was found to move so as to occupy the cavity formed by the mutation Cys35→Gly. These results demonstrated the ability of the double-mutant cycle method to detect interactions. It is important to stress that, in principle, the double-mutant cycle method may also allow detection of interactions that are not observed using classical structural methods such as X-ray crystallography and NMR.

Double-mutant cycles allow one to determine not only whether an intramolecular interaction between two residues exists but also, in certain favorable cases, to measure accurately the energy of their interaction. Owing to proteins being highly cooperative structures, mutation of one residue will often lead to disruption of several interactions. Single-mutation experiments are therefore

difficult to interpret, since not only is the interaction of interest disrupted but also other interactions. This problem can be partly overcome by judicious choice of the mutation [91]. A much better approach is to invoke double-mutant cycles. The free energy of interaction between the two residues is determined by measurement of the unfolding free energies of each of the proteins comprising the cycle. If the change in the free energy of unfolding upon mutation of one of the residues depends on whether the other residue has been mutated, i.e. the changes are not additive, then those two residues interact. An important advantage of the double-mutant cycle method is that in ideal circumstances, when the mutations cause no structural rearrangements, subtraction of the free energy change for the mutation $P-X \rightarrow P$ ($\Delta G_{P-X \rightarrow P}$) from the free energy change for $P-XY \rightarrow P-Y$ ($\Delta G_{P-XY \rightarrow P-Y}$) will cause the interaction energies of X and Y with the rest of the protein to cancel out. In such cases the measured interaction energy $\Delta\Delta G_{\text{int}}$ is given by

$$\Delta\Delta G_{\text{int}} = G_{X \dots Y} - \Delta G_{X \dots w} - \Delta G_{Y \dots w} \quad (26)$$

where $G_{X \dots Y}$ is the interaction energy between X and Y, $\Delta G_{X \dots w}$ is the increase in the solvation energy of X on removal of Y and $\Delta G_{Y \dots w}$ is the increase in the solvation energy of Y on removal of X [92]. A further advantage of the double-mutant cycle method is that even if there are structural changes upon mutation of, for example, X in P-XY, similar changes might occur upon mutation of X in P-X. Artifacts from structural rearrangements thus tend to cancel out so that their importance is reduced to second-order effects. It should be noted that $\Delta\Delta G_{\text{int}}$ is the interaction energy between X and Y in the folded state of the protein relative to any residual interaction that might exist between X and Y in the unfolded state.

Commonly occurring stabilizing interactions in proteins have been investigated in barnase, a ribonuclease from *Bacillus amyloliquefaciens*, by use of the double-mutant cycle method. Measurements of changes in the free energies of unfolding upon mutation were accompanied by structural experiments to show that no significant reorganization resulting in new interactions had occurred. An (*i*, *i* + 4) surface electrostatic interaction in an α -helix was engineered between residues Asp12 and Arg16 in barnase by making the mutation Thr16 $\rightarrow$ Arg [92]. The strength of this interaction was quantitatively estimated by invoking a double-mutant cycle involving the wild-type enzyme (Asp12, Thr16), the single mutants Asp12 $\rightarrow$ Ala and Thr16 $\rightarrow$ Arg and the double mutant Asp12 $\rightarrow$ Ala, Thr16 $\rightarrow$ Arg. At low ionic strength ($I = 0.01$ M), the changes in the unfolding free energies were not additive with an interaction energy of -0.33 kcal mol $^{-1}$. At high salt concentrations, however, the changes were additive, indicating that the interaction between these residues is purely electrostatic. The single mutant Thr16 $\rightarrow$ Arg, containing the ion pair Asp12-Arg16, was found to be more stable than the wild-type enzyme by 0.5 kcal mol $^{-1}$. The discrepancy between the interaction energy obtained from the cycle and the increased stability of the

single mutant Thr16→Arg shows that only a part of the increased stability of this mutant is attributable to the electrostatic interaction, and that conclusions based on single-mutation experiments can be misleading. The electrostatic interaction energies of other ion pairs at positions ($i, i + 4$) in the α -helices of barnase have been determined by use of double-mutant cycles with similar results [93, 94]. These experiments showed that electrostatic interactions between exposed side chains contribute little to protein stability, probably because of their location on the surface rather than in the interior of the protein.

The contributions to protein stability of other types of interactions have also been determined by double-mutant cycle analysis. An aromatic–aromatic interaction between the edge of Tyr17 and the face of Tyr13 was found to contribute $1.3 \text{ kcal mol}^{-1}$ to protein stability [95]. A charge–aromatic interaction between His18 and Trp94 was found to contribute about 1 kcal mol^{-1} to protein stability [96]. The contributions to stability of two exposed salt bridges between the C-terminal residue Arg110 and two aspartic residues at positions 8 and 12 in the first α -helix of barnase were found to be about 1 kcal mol^{-1} and $1.25 \text{ kcal mol}^{-1}$, respectively [94]. The proximity of the two salt bridges to each other suggested that possible coupling effects between them should be measured, as described in the next section.

The double-mutant cycle method can also be used to detect and measure the strength of *intermolecular* interactions. The analysis of intermolecular interactions, for example, between enzyme and substrate by means of mutant enzymes and modified substrates is analogous to the analysis of intramolecular interactions using the double-mutant cycle method. Recently [97], an intermolecular interaction between the enzyme RNase T1 and its dinucleoside phosphate substrate (GpN, N = A, C or U) was analyzed by this approach. By truncation of a side chain in the protein and elimination of a functional group in the substrate it was possible, using a cycle involving the four possible complexes of wild-type and mutant enzymes with the two different substrates, to determine the contribution to the binding energy of that particular enzyme–substrate interaction. It is also possible, using the double-mutant cycle approach, to analyze intermolecular interactions between different protein molecules (Figure 18). In such a case, one mutation would be at site X in the first protein molecule and the second mutation at site Y in the other protein molecule. The double-mutant cycle to be considered consists of the complex of the two different wild-type proteins, the complex of each mutant protein with the other wild-type protein and the complex of the two mutant proteins. By measuring the equilibrium constant for the formation of each complex in the cycle (Figure 18), one may estimate the contribution of a defined intermolecular interaction to the total free energy of the protein–protein interaction.

5.5 Multiple-Mutant Cycle Analysis

In many cases of intermolecular and intramolecular interactions in proteins, the interactions between several amino acids are coupled to one another and

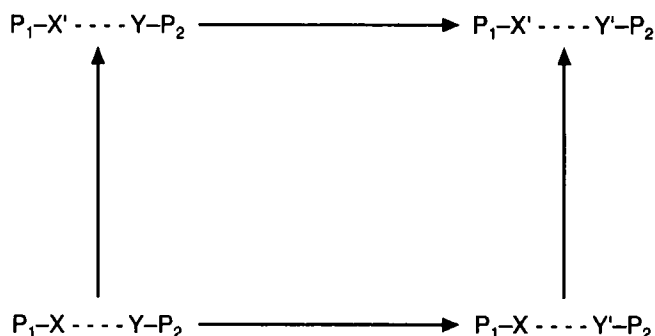


Figure 18 Double-mutant cycle for a protein–protein interaction. In this scheme, P_1 and P_2 stand for two proteins that form a complex, X and Y for the mutated residues in P_1 and P_2 , respectively, and prime stands for the mutation

may not be reduced to a sum of pairwise interactions. It is not possible to determine from one double-mutant cycle whether the observed pairwise interaction is dependent on other neighboring residues. A general method for analysis of the interactions between m residues ($m \geq 2$) has been described [72, 98]. When two residues, X and Y , are targeted and mutated, the construct required to detect an interaction between them is the two-dimensional double-mutant cycle. When three residues are replaced, the construct required to detect the effect of one residue on the interaction between the other two, i.e. a triple-mutant box, is three-dimensional. The interactions between the salt-linked triad Asp8, Arg110 and Asp12 in barnase were therefore analyzed by construction of a triple-mutant box comprising all the single, double and triple mutants at these positions [94]. The mutations were in all cases to alanine. Opposite faces of the box correspond to the same double-mutant cycle with respect to two positions but with different residues at the third position. By comparison of the interaction energies measured by cycles corresponding to opposite faces of the box, it is possible to determine the effect of a third position. If these interaction energies are the same, there is no effect from the third position. It was found that the strengths of the two salt bridges are coupled. The energy of each salt bridge is reduced by $0.77 \text{ kcal mol}^{-1}$ when the other is absent.

The multidimensional cycle approach should be useful in analysis of the energetics of different structural motifs, such as γ -turns and β -turns, that consist of a number of residues. By applying this approach one should be able to obtain experimental evidence for the effects of different neighboring residues in protein–ligand interactions.

Recently, the COSMIC (combination of sequential mutant interaction cycles) technique [99] was introduced, in which a series of double-mutant cycles is constructed. Such a series was formed to leapfrog along the first α -helix of barnase. The combined kinetic and equilibrium measurements for the proteins in that series of cycles showed that this helix is intact from its C-terminus to

residue 12, but begins to unwind towards the N-terminus both in the transition state for unfolding and in a folding intermediate [99]. The COSMIC approach is similar in nature to two-dimensional nuclear Overhauser enhancement spectroscopy (NOESY) experiments in which two residues may be shown to be within an interaction distance. In principle, it should therefore be possible to solve unknown structures with the COSMIC technique by generating a large number of mutants and using distance geometry constraints as in NMR spectroscopy. The COSMIC approach can be applied more readily to a segment of secondary structure to see if the interactions between side chains correspond to those expected in a turn, α -helix or β -sheet. The structures of states which are difficult to study by NMR or X-ray crystallography, such as transition states and folding intermediates, may also be investigated using the COSMIC strategy. In addition, in cases where the structure of a protein–ligand complex is unknown, application of the COSMIC approach may facilitate ligand docking.

6. CONCLUDING REMARKS

The meaning of biological specificity on a molecular level has been discussed here in some detail. We indicated that similar considerations can be applied to the analysis of biological specificity as revealed by a single receptor with multiple ligands and as represented by systems consisting of a repertoire of receptors with one ligand. In both cases one can derive quantitative expressions for specificity, in the first case when the binding constants of a single receptor with every one of a given set of ligands are known, and in the second case when the binding constant of a given ligand with each of the receptors is known. Probabilistic considerations, embodied in a receptor affinity distribution (RAD) model, enabled the derivation of expressions that define quantitatively the specificity of ligand binding in both cases.

Kinetic and thermodynamic descriptions of ligand–receptor binding can readily be obtained by classical theoretical and experimental procedures. These are based on statistical mechanical considerations, since one is dealing with a large number of ligand and receptor molecules without making use of specific information about their molecular structures. Because of advances in the elucidation of the three-dimensional structures of proteins by X-ray and NMR techniques, and the theoretical elucidation of the intra- and interatomic forces associated with small and large molecules, molecular biologists have turned their attention to the three-dimensional structures of receptors, ligands and their complexes. Their work has highlighted the importance of molecular surface recognition in specific protein–ligand interactions, and has demonstrated molecular surface complementarity in many of the complexes studied. The original lock-and-key concept has thus been transformed into a more modern version involving molecular surface complementarity.

We have shown in the above that when the formation of a complex between receptor and ligand is not accompanied by a significant change in their conformations, and their separate three-dimensional structures are known, the molecular structure of the expected complex can be predicted. However, when complex formation is accompanied by conformational alterations, energy considerations must be taken into account in any attempt to predict the molecular structure of the complex. A method for the detection and measurement of the energetics of intra- and intermolecular interactions in proteins using double-mutant cycles has been described in some detail.

Finally, it should be noted that ligand recognition and binding by a single receptor or a repertoire of receptors often leads to a cascade of reactions of paramount biological importance. Specific binding of an enzyme to a substrate initiates a characteristic set of catalytic reactions leading to product formation; binding of ligands to cell receptors leads to signal transduction and the initiation of a whole set of reactions in the cell, such as cell division and differentiation. In each case, binding of the ligand to its receptor leads to a corresponding chemical or conformational alteration that induces a cascade of interactions between the modified receptor and various high- or low-molecular-weight components of the cell. We are beginning to understand the nature of the interaction between the 'lock' and the 'key', although little information is yet available on the detailed molecular interactions that result from opening of the lock. Future studies will need not only to identify the complex reactions occurring as a result of receptor-ligand interactions, but also to unravel, on a molecular level, the conformational and chemical modifications occurring at each step.

7. REFERENCES

1. D. R. Madden and D. C. Wiley, *Curr. Opin. Struct. Biol.*, **2**, 300 (1992).
2. J. K. Scott and G. P. Smith, *Science*, **249**, 386 (1990).
3. G. Weber, *Protein Interactions*, Chapman and Hall, New York, 1992, p. 272.
4. H.-J. Hinz, *Annu. Rev. Biophys. Bioeng.*, **12**, 285 (1983).
5. P. D. Ross and S. Subramanian, *Biochemistry*, **20**, 3096 (1981).
6. R. Poljak, in *10th Int. Symp. Biorecognition and Affinity Technology*, Gwatt/Thun, Switzerland, 1993.
7. E. C. Meng, D. A. Gschwend, J. M. Blaney and I. D. Kuntz, *Proteins: Struct. Function Genet.*, **17**, 266 (1993).
8. D. E. Koshland Jr, *Proc. Natl. Acad. Sci. USA*, **44**, 98 (1960).
9. W. L. Jorgensen, *Science*, **254**, 954 (1991).
10. F. M. Burnet, *Cold Spring Harbor Symp. Quant. Biol.*, **32**, 1 (1963).
11. R. W. R. Rosenstein, R. A. Musson, M. Y. K. Armstrong, W. H. Konigsberg and F. F. Richards, *Proc. Natl. Acad. Sci. USA*, **69**, 877 (1977).
12. J. K. Inman, in *Theoretical Immunology* (eds G. I. Bell, A. S. Perelson and J. H. J. Pimbley) Dekker, New York, 1978, p. 243.
13. D. Lancet, E. Sadovsky and E. Seidemann, *Proc. Natl. Acad. Sci. USA*, **90**, 3715 (1993).

14. J. D. Farmer, N. H. Packard and A. S. Perelson, *Physica*, **22D**, 187 (1986).
15. A. S. Perelson and G. F. Oster, *J. Theor. Biol.*, **81**, 645 (1979).
16. L. A. Segel and A. S. Perelson, in *Theoretical Immunology*, Part 2 (ed. A. S. Perelson), Addison-Wesley, Redwood City, CA, 1988, p. 321.
17. R. J. De Boer and A. S. Perelson, *J. Theor. Biol.*, **149**, 381 (1991).
18. A. S. Perelson, in *Lectures in Complex Systems* (ed. E. Jen), Addison-Wesley, Redwood City, CA, 1990, p. 465.
19. J. K. Percus, O. E. Percus and A. S. Perelson, *Proc. Natl. Acad. Sci. USA*, **90**, 1691 (1993).
20. G. Weisbuch, *J. Theor. Biol.*, **143**, 507 (1990).
21. C. Berek and C. Milstein, *Immunol. Rev.*, **105**, 5 (1988).
22. D. Lancet, *Annu. Rev. Neurosci.*, **9**, 329 (1986).
23. M. Devos, F. Patte, J. Rouault and P. Laffort, *Standardized Human Olfactory Thresholds*, IRL Press, Oxford, 1990.
24. L. Buck and R. Axel, *Cell*, **65**, 175 (1991).
25. D. Lancet, D. Ben Simon and Z. Chen, *Chem. Senses*, **8**, 255 (1984).
26. D. Schild, *Biophys. J.*, **54**, 1001 (1988).
27. A. Levitzki, *Receptors: A Quantitative Approach*, Benjamin/Cummings, Menlo Park, CA, 1984.
28. W. Senf, B. P. Menco, P. H. Punter and P. Duyvesteyn, *Experientia*, **36**, 213 (1980).
29. R. Gross-Isseroff and D. Lancet, *Chem. Senses*, **13**, 191 (1988).
30. G. M. Shepherd and S. Firestein, *Curr. Biol.*, **1**, 204 (1991).
31. D. Lancet, in *Sensory Transduction* (eds D. P. Corey and S. D. Roper), Rockefeller University, New York, 1992, p. 73.
32. C. J. Wysocki and G. K. Beauchamp, *Proc. Natl. Acad. Sci. USA*, **81**, 4899 (1984).
33. R. Gross-Isseroff, D. Ophir, A. Bartana, H. Voet and D. Lancet, *Neurosci. Lett.*, **141**, 115 (1992).
34. J. E. Amoore, *Ann. N.Y. Acad. Sci.*, **237**, 137 (1974).
35. J. E. Amoore and S. Steinle, in *Chemical Senses* (eds C. J. Wysocki and M. R. Kare), Dekker, New York, 1991, p. 331.
36. S. J. Wodak and J. Janin, *J. Mol. Biol.*, **124**, 323 (1978).
37. P. J. Goodford, *J. Med. Chem.*, **28**, 849 (1985).
38. M. Billeter, T. F. Havel and I. D. Kuntz, *Biopolymers*, **26**, 777 (1987).
39. J. Warwicker, *J. Mol. Biol.*, **206**, 381 (1989).
40. D. S. Goodsell and A. J. Olsen, *Proteins*, **8**, 195 (1990).
41. S.-Y. Yue, *Protein Eng.*, **4**, 177 (1990).
42. A. Wlodawer, J. Deisenhofer and R. Huber, *J. Mol. Biol.*, **193**, 145 (1987).
43. M. Marquart, J. Walter, J. Deisenhofer, W. Bode and R. Huber, *Acta Crystallogr., Sect. B*, **39**, 480 (1983).
44. S. Remington, J. Wiegand and R. Huber, *J. Mol. Biol.*, **158**, 111 (1982).
45. G. Fermi, M. F. Perutz, B. Shaanan and R. Fourme, *J. Mol. Biol.*, **175**, 159 (1984).
46. A. G. Amit, R. A. Mariuzca, S. E. Phillips and R. J. Poljak, *Science*, **233**, 747 (1986).
47. K. Suguna, E. A. Padlan, C. W. Smith, W. D. Carlson and D. R. Davies, *Proc. Natl. Acad. Sci. USA*, **84**, 7009 (1987).
48. K. Suguna, R. R. Bott, E. A. Padlan, E. Subramanian, S. Sheriff, G. H. Cohen and D. R. Davies, *J. Mol. Biol.*, **196**, 877 (1987).
49. P. Brick, T. N. Bhat and D. M. Blow, *J. Mol. Biol.*, **208**, 83 (1989).
50. J. Greer and B. L. Bush, *Proc. Natl. Acad. Sci. USA*, **75**, 303 (1978).
51. I. D. Kuntz, J. M. Blaney, S. J. Oatley, R. Langridge and T. E. Ferrin, *J. Mol. Biol.*, **161**, 269 (1982).
52. P. Zielenkiewicz and A. Rabczenko, *J. Theor. Biol.*, **111**, 17 (1984).
53. P. Zielenkiewicz and A. Rabczenko, *J. Theor. Biol.*, **116**, 607 (1985).
54. D. W. Fanning, J. A. Smith and G. D. Rose, *Biopolymers*, **25**, 863 (1986).

55. J. Novotny, M. Handschumacher, E. Haber, R. E. Brucoleri, W. B. Carlson, D. W. Fanning, J. A. Smith and G. D. Rose, *Proc. Natl. Acad. Sci. USA*, **83**, 226 (1986).
56. M. L. Connolly, *Biopolymers*, **25**, 1229 (1986).
57. R. L. Desjarlais, R. P. Sheridan, G. L. Siebel, J. S. Dixon, I. D. Kuntz and R. Venkataraghavan, *J. Med. Chem.*, **31**, 722 (1988).
58. Y. Chirgadze, N. Kurochkina and S. Nikonov, *Protein Eng.*, **3**, 105 (1989).
59. R. A. Lewis and P. M. Dean, *Proc. R. Soc. London, Ser. B*, **236**, 1411 (1989).
60. H. Wang, *J. Comput. Chem.*, **12**, 746 (1991).
61. F. Jiang and S. H. Kim, *J. Mol. Biol.*, **219**, 79 (1991).
62. B. K. Schoichet and I. D. Kuntz, *J. Mol. Biol.*, **221**, 327 (1991).
63. E. Katchalski-Katzir, I. Shariv, M. Eisenstein, A. A. Friesem, C. Aflalo and I. A. Vakser, *Proc. Natl. Acad. Sci. USA*, **89**, 2195 (1992).
64. M. Goldstein, in *Classical Mechanics*, Addison-Wesley, Reading, MA, 1980, p. 608.
65. R. C. Ladner, E. G. Heidner and M. F. Perutz, *J. Mol. Biol.*, **114**, 385 (1977).
66. K. Suguna, E. A. Padlan, C. W. Smith, W. D. Carlson and D. R. Davies, *Proc. Natl. Acad. Sci. USA*, **84**, 7009 (1987).
67. A. Wlodawer, J. Deisenhofer and R. Huber, *J. Mol. Biol.*, **193**, 145 (1987).
68. M. Laskowski Jr, *Biochem. Pharmacol.*, **29**, 2089 (1980).
69. M. W. Empie and M. Laskowski Jr, *Biochemistry*, **21**, 2274 (1982).
70. P. J. Carter, G. Winter, A. J. Wilkinson and A. R. Fersht, *Cell*, **38**, 835 (1984).
71. A. Horovitz, *J. Mol. Biol.*, **196**, 733 (1987).
72. A. Horovitz and A. R. Fersht, *J. Mol. Biol.*, **214**, 613 (1990).
73. G. K. Ackers and F. R. Smith, *Annu. Rev. Biochem.*, **54**, 597 (1985).
74. J. A. Wells, *Biochemistry*, **29**, 8509 (1990).
75. I. Schechter and A. Berger, *Biochem. Biophys. Res. Commun.*, **27**, 157 (1967).
76. A. Berger and I. Schechter, *Philos. Trans. R. Soc. London, Ser. B*, **257**, 249 (1970).
77. M. Laskowski Jr, M. Tashiro, M. W. Empie, S. J. Park, I. Kato, W. Ardelt and M. Wiczeorek, in *Proteinase Inhibitors: Medical and Biological Aspects* (eds N. Katunuma, H. Umezawa and H. Holzer), Japan Scientific Societies Press, Tokyo, 1983, p. 55.
78. W. Ardelt and M. Laskowski Jr, *J. Mol. Biol.*, **220**, 1041 (1991).
79. E. Weber, E. Papamokos, W. Bode, R. Huber, I. Kato and M. Laskowski Jr, *J. Mol. Biol.*, **149**, 109 (1981).
80. M. Fujinaga, R. J. Read, A. Sielecki, W. Ardelt, M. Laskowski Jr and M. N. G. James, *Proc. Natl. Acad. Sci. USA*, **79**, 4868 (1982).
81. S. M. Free Jr and J. W. Wilson, *J. Med. Chem.*, **7**, 395 (1964).
82. A. Horovitz and M. Rigbi, *J. Theor. Biol.*, **116**, 149 (1985).
83. J. E. Leffler and E. Grunwald, *Rates and Equilibria of Organic Reactions*, Wiley, New York, 1963.
84. A. R. Fersht, R. J. Leatherbarrow and T. N. C. Wells, *Nature*, **322**, 284 (1986).
85. J. N. Brønsted and K. Pedersen, *Z. Phys. Chem.*, **108**, 185 (1924).
86. A. Horovitz and R. D. Levine, *J. Theor. Biol.*, **141**, 557 (1989).
87. L. P. Hammett, *J. Am. Chem. Soc.*, **59**, 96 (1937).
88. E. R. Malinowski and D. G. Howery, *Factor Analyses in Chemistry*, Wiley, New York, 1980.
89. J. A. Wells, D. B. Powers, R. R. Bott, T. P. Graycar and D. A. Estell, *Proc. Natl. Acad. Sci. USA*, **84**, 1219 (1987).
90. M. D. Fothergill and A. R. Fersht, *Biochemistry*, **30**, 5157 (1991).
91. A. R. Fersht, R. J. Leatherbarrow and T. N. C. Wells, *Biochemistry*, **26**, 6030 (1987).
92. L. Serrano, A. Horovitz, B. Avron, M. Bycroft and A. R. Fersht, *Biochemistry*, **29**, 9343 (1990).
93. D. Sali, M. Bycroft and A. R. Fersht, *J. Mol. Biol.*, **220**, 779 (1991).

94. A. Horovitz, L. Serrano, B. Avron, M. Bycroft and A. R. Fersht, *J. Mol. Biol.*, **216**, 1031 (1990).
95. L. Serrano, M. Bycroft and A. R. Fersht, *J. Mol. Biol.*, **218**, 465 (1991).
96. R. Loewenthal, J. Sancho and A. R. Fersht, *J. Mol. Biol.*, **224**, 759 (1992).
97. J. Steyaert, A. F. Haikal, L. Wyns and P. Stanssens, *Biochemistry*, **30**, 8666 (1991).
98. A. Horovitz and A. R. Fersht, *J. Mol. Biol.*, **224**, 733 (1992).
99. A. Horovitz, L. Serrano and A. R. Fersht, *J. Mol. Biol.*, **219**, 5 (1991).

Chapter 3

New Biocatalysts via Chemical Modification

IAN M. BELL AND DONALD HILVERT

The Scripps Research Institute, La Jolla, CA, USA

1. INTRODUCTION

Emil Fischer's 'lock-and-key' analogy of enzymatic catalysis [1], which introduced the idea of enzyme–substrate complementarity, seems all the more remarkable when one considers that so little was known about the detailed nature of enzymes at that time. Their identity as proteins, for instance, was not firmly established for a number of decades after Fischer advocated his concept of a binding site which was tailored to recognize a specific substrate [2]. Despite such uncertainties, the mechanisms of enzymes were probed extensively by kinetic studies involving various substrates and their analogues, and the kinetic approach continues to be fundamental in modern enzymology. In the late twentieth century the tools of molecular biology, working in concert with such analytical techniques as X-ray crystallography and nuclear magnetic resonance spectroscopy, have provided a wealth of structural information about enzymes and proteins in general. The overall picture of enzyme mechanism, as assembled from all these sources, reveals the complexity and 'perfection' of these biocatalysts and simultaneously plants the desire in chemists to generate new enzyme-like catalysts.

Enzymes perform their chemistry with enviable rate enhancements and exquisite specificity, usually under very mild conditions of temperature and pH. These features have led to their increasing use in a variety of industrial and medical contexts, but such applications are limited by the enzymic activities which nature has provided. The challenge is to create new agents which exhibit the basic features of enzymes, but which may be generated for any chemical

process of interest. In addition to the practical benefits of such novel catalysts, this undertaking should yield a greater understanding of molecular recognition in biological systems, and of enzyme mechanism.

The development of general strategies for creating new protein-based catalysts has received a great deal of recent attention, in part because of the prevalence of protein catalysis in nature, although it should be noted that this is not the only possible approach. The chemical synthesis of model systems [3–5] and the production of RNA-based catalysts [6–8] are viable alternatives with great potential. With regard to protein biocatalysts, however, the most direct approach would involve the design of a suitable amino acid sequence that folds to give a catalytic binding site. Unfortunately, such a *de novo* design strategy is extremely difficult. Even without the additional considerations of catalysis, attempts to synthesize polypeptides with a given structure have met with only limited success [9]. In the few cases where the desired secondary structural elements have been successfully generated, the designed proteins have the characteristics of fairly mobile ‘molten globules’, in contrast to the typical close-packed, crystalline behavior of natural proteins [10,11]. At present, therefore, the design and synthesis of a protein which folds to give a native-like crystalline structure is a significant challenge, and the extra requirements of a catalytically functional binding site further complicate this task. Given our limited ability to construct functional proteins from their constituent amino acids, existing proteins, in particular enzymes and antibodies, provide valuable starting points for the preparation of new active sites. This chapter will focus on the chemical modification of such native proteins as an approach to novel biocatalysts.

2. GENERAL CHEMICAL MODIFICATION OF ENZYMES

Chemical modification of enzymes is by no means a new concept, and reports of such studies were made as far back as 1925 [12]. Until relatively recently, specific chemical alteration of functional groups was one of the only ways to address their mechanistic importance, and such protein modification studies have been thoroughly reviewed elsewhere [13]. In summary, there are many chemical agents which can modify a specific residue type in a protein, but not necessarily in a site-specific manner [14]. Oxidation, alkylation, acylation and other strategies have been employed, and although the modification methodologies were not always rational, the covalent alterations were often confined to a few residues in the enzyme which were then identified. The kinetic competence of the modified enzyme could then be correlated with the altered side chains and deductions made about the roles of those residues in catalysis. Studies of such modified enzymes with various substrates and substrate analogues have helped to ‘map’ their active sites and the conclusions reached have often been confirmed by later X-ray crystallographic studies. Changes in

substrate specificity, alteration of inhibition and activation profiles and the shifting of pH optima are some of the effects that have been realized by this approach. Furthermore, resistance to proteolysis has been achieved by derivatization of surface lysine residues with polyethylene glycol, which also confers nonimmunogenicity on the enzyme [15, 16]. These qualities greatly enhance the therapeutic potential of the derivatized catalyst since the proteolytic degradation of, and allergic reactions to, native enzymes have been major stumbling blocks in their acceptance as clinical drugs.

Thus, the use of broadly specific derivatization agents has allowed the participation of individual residues in catalysis to be investigated and the catalytic behavior of the enzymes manipulated, albeit often in an empirical manner. These modified enzymes contain chemically derivatized, nonnatural amino acid residues, and their altered specificity may usually be ascribed to the modified steric and/or electronic properties of these groups. A more elegant approach would be to modify an individual residue in a site-specific manner in order to rationally alter the catalytic function of the enzyme. In principle, such an approach could give rise to diverse activities since the amino acid in question could be changed into any unnatural residue or cofactor, but the earliest experiments in this area focused on the replacement of one naturally occurring amino acid with another.

3. CHEMICAL MUTAGENESIS

The substitution of one natural amino acid residue by another natural amino acid is now commonplace thanks to the powerful methodology of site-directed mutagenesis. However, before the advent of modern molecular biology, such replacements were achieved by chemical means, leading to the term 'chemical mutagenesis'. This approach was first developed by Neet and Koshland [17] and Polgár and Bender [18] for the conversion of the catalytic serine of subtilisin into a cysteine residue.

Serine proteases, such as subtilisin, contain an unusually reactive constellation of amino acid side chains at their active sites: the catalytic triad of Ser, His and Asp. In particular, the active site nucleophile (Ser) and general base (His) can be selectively modified with a wide range of reagents. In order to substitute the catalytic serine by cysteine, its side chain was first selectively sulfonylated with phenylmethanesulfonyl fluoride (PMSF). Displacement of the resulting sulfonate leaving group with thioacetate leads to an equivalent of the acylenzyme intermediate, which spontaneously deacylates to give the desired product (Figure 1). Thiolsubtilisin, with its synthetically formed active site, can be thought of as the first example of rational chemical modification to give a new biocatalyst, the first 'semisynthetic enzyme'.

The Ser to Cys mutation in thiolsubtilisin causes the enzyme to become a 'damaged' protease: it retains modest esterase activity but lacks the ability to

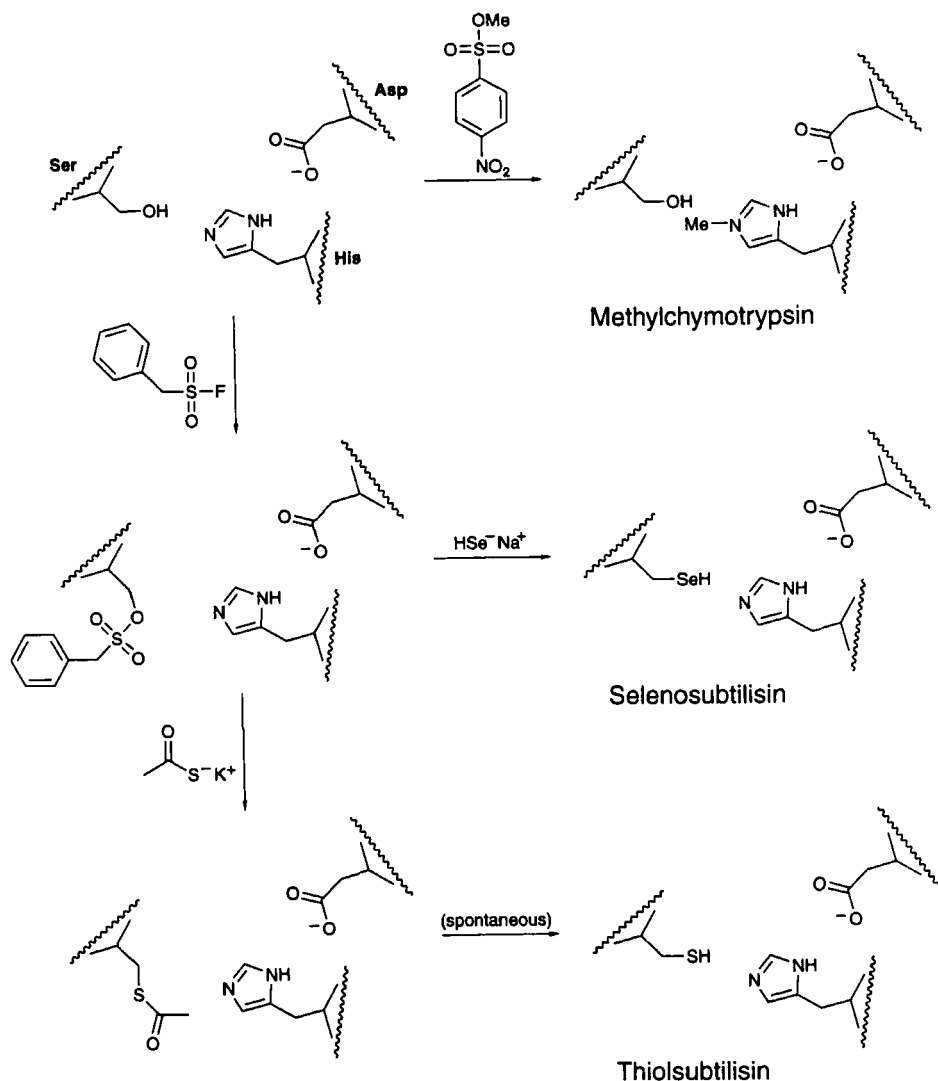


Figure 1 Active site modification of serine proteases to give ‘damaged’ proteases with potential peptide ligase activity

cleave amide bonds. A detailed kinetic study has shown that the introduction of cysteine at the active site significantly disfavors the formation of the acylenzyme intermediate [19, 20]. This species can be formed, however, from the reaction of the semisynthetic enzyme with activated esters, and it exhibits a marked preference for aminolysis over hydrolysis. The tendency to form amide bonds, coupled with the inability to cleave them, suggests that thiolsubtilisin would be useful as a peptide ligase. Such ligases would be invaluable for the

routine chemical synthesis of proteins from peptide fragments, since they potentially offer high specificity and stereoselectivity, minimal side reactions and little need for protecting groups. A number of proteases, including subtilisin, trypsin and papain, have been used successfully for peptide ligation. Under kinetic control, however, the enzyme-catalyzed peptide condensation can be hampered by the hydrolysis, instead of the desired aminolysis, of the transiently formed acylenzyme. Furthermore, because the reaction products contain amide bonds, they can be proteolyzed by the ligation catalyst. For the reasons stated above, thiolsubtilisin can circumvent these problems and it has been shown to be an effective peptide ligase by Kaiser and coworkers [21], producing a variety of peptides stereoselectively and in high yield. A second-generation thiolsubtilisin has been genetically engineered by Abrahmsén *et al.* [22]. A judicious mutation in the enzyme shifted the active site cysteine, partially relieving the steric crowding caused by the Ser to Cys substitution, to produce a significantly more potent ligase.

Analogous chemical mutagenesis methodology to that which produced thiolsubtilisin has also been applied to the mammalian serine protease trypsin (Ser replaced by Cys) [23] to give a similarly 'damaged' protease that is potentially a ligase with different substrate specificity to thiolsubtilisin. Interestingly, the converse chemical mutagenesis experiment with the thiol protease papain (Cys replaced by Ser) [24] destroyed all protease and esterase activity in the enzyme. Another example of a modification strategy which yielded a 'damaged' protease, while technically chemical modification rather than chemical mutagenesis, is the selective methylation of the active site histidine in α -chymotrypsin by methyl 4-nitrobenzenesulfonate [25] (Figure 1). The resulting methylchymotrypsin can cleave activated esters but is completely inert towards amides. Wong and coworkers [25] have exploited these properties to demonstrate efficient peptide ligation.

These examples reveal the viability of chemical mutagenesis as a strategy for producing catalysts with tailored function. However, a uniquely reactive residue in the template is required in order to allow for selective replacement, and this limits the strategy to a few residues in a small number of proteins. One advantage of chemical mutagenesis is that the yields of product are often very high, making it possible to isolate large amounts of semisynthetic enzyme in a very short time. Semisynthesis can also access mutations involving rare or unnatural amino acids, such as selenocysteine, which are difficult to produce in good yield by molecular biological techniques (see, however, [26–29]).

In this latter context, we have applied a similar methodology to that used to produce thiolsubtilisin in order to generate the first semisynthetic selenoenzyme, selenosubtilisin [19] (Figure 1). The incorporation of selenocysteine at the active site of subtilisin is a logical extension of the studies on thiolsubtilisin, and it seemed particularly attractive because of the extensive chemical utility of selenium. Furthermore, there is increasing interest in the biochemistry of selenium and natural selenoenzymes such as glycine reductase and glutathione peroxidase [30]. As expected, selenosubtilisin exhibits a similar 'damaged'

protease activity profile to its thiol analogue. However, the ratio of aminolysis to hydrolysis for the semisynthetic acylated intermediate of the selenoenzyme is an order of magnitude greater than that for thiolsubtilisin, making selenosubtilisin even more promising as a potential peptide ligase [19].

The ready availability of selenosubtilisin has facilitated its characterization. In contrast to the naturally occurring selenoenzyme glutathione peroxidase, for instance, the preparation of useful quantities of ^{77}Se -labeled selenosubtilisin is relatively straightforward, allowing characterization of various oxidation states of the semisynthetic enzyme by ^{77}Se NMR spectroscopy [31]. In addition, X-ray crystallography [32] and ^1H NMR spectroscopy [33] have provided valuable information about interactions between the selenium prosthetic group and the active site of its protein template.

As mentioned above, one reason for investigating a semisynthetic selenoenzyme is the unique chemistry of selenium. The element has access to several oxidation states and thus possesses interesting redox chemistry as is evidenced by the natural selenoenzymes. Glutathione peroxidase, for example, catalyzes the reduction of hydroperoxides by the thiol glutathione ($\gamma\text{-Glu-Cys-Gly}$), and is thought to be crucial in the prevention of oxidative damage in many tissues [34]. This redox reaction is highly efficient, with second-order rate constants approaching the diffusion limit, and the active site selenocysteine residue is known to be essential for the function of the enzyme [35]. In addition to its acyltransferase characteristics, selenosubtilisin can also mimic the activity of glutathione peroxidase, catalyzing the reduction of hydroperoxides by 3-carboxy-4-nitrobenzenethiol [36]. Detailed investigations into the nature of this catalysis [37, 38] have revealed that specific interactions between the selenocysteine moiety and active site residues, such as His64, modulate the activity of the semisynthetic enzyme. In fact, selenosubtilisin is at least 2000-fold more efficient at promoting the peroxidase reaction than the selenium side chain alone. Thus, both the synthetic prosthetic group and the protein template contribute to catalysis, making this truly a new enzyme. The kinetic studies on selenosubtilisin correlate well with crystallographic and spectroscopic data and they help to illuminate the mechanism of glutathione peroxidase itself.

4. MODIFICATION OF ENZYMES WITH COFACTORS

The peroxidase activity of selenosubtilisin requires the rare amino acid selenocysteine and its ability to perform redox chemistry. In this sense, the selenium moiety is acting like a cofactor. Indeed, the modification of proteins with cofactors is a general route to novel biocatalysts. Analogues of nicotinamide adenine dinucleotide (NAD^+), for instance, have been used to modify dehydrogenases covalently [39, 40]. In some cases, the modified dehydrogenases retained substantial redox capability even in the absence of exogenous NAD^+ , although the modifications did not cause much change in substrate specificity or in the type of chemistry catalyzed by the enzyme.

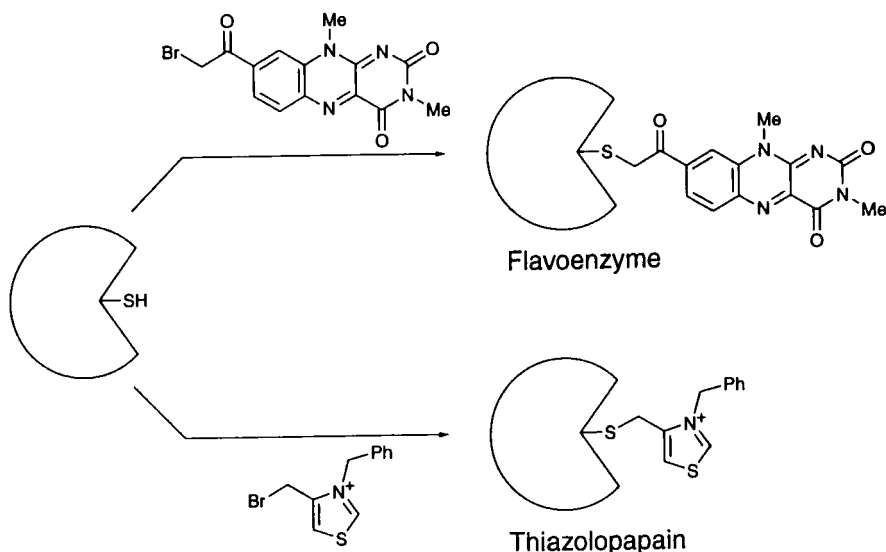


Figure 2 Modification of enzymes with cofactors

A more radical approach came from the laboratory of Kaiser, where a family of semisynthetic flavoenzymes were prepared by using 10-methylisoalloxazine derivatives to functionalize site specifically protein templates, including papain, glyceraldehyde-3-phosphate dehydrogenase (GAPDH) and hemoglobin [41]. The strategy relied upon the uniquely reactive cysteine residue in each of the proteins, which could be modified selectively with α -bromoketone-containing flavin analogues (Figure 2). These cofactors are attractive because of their intrinsic potential for such chemistry as oxidation of dihydronicotinamides and thiols, and electron transfer. It was expected that the semisynthetic enzymes would combine such redox activity with the substrate selectivity of their respective protein templates, and such behavior was indeed observed. Moreover, in some cases the artificial enzymes exhibited catalytic efficiencies which rival those of naturally occurring flavoenzymes.

For example, the protease papain has a preference for hydrophobic amino acids in its S_2 specificity site [2], which is reflected in the substrate selectivity of the semisynthetic flavopapain. The latter enzyme was found to catalyze the oxidation of *N*-alkyl-4-dihydronicotinamides and hydrophobic dithiols. Under aerobic conditions, molecular oxygen accepts electrons from the reduced flavin to allow turnover of the catalyst. In fact, the best substrate (*N*-hexyl-4-dihydronicotinamide) is turned over so efficiently that this aerobic reoxidation is the rate-determining step [42], and the nicotinamide is oxidized 1000 times faster by flavopapain than by the nonenzymatic model 8-acetyl-10-methylisoalloxazine [43]. In contrast, the best substrate for flavoGAPDH is the polar NADH molecule, in accord with the NAD^+ binding domain at the active site of the

protein template [44]. In this case, the specific recognition of NADH by the template gives rise to stereoselective transfer of hydride to the prosthetic flavin group.

The methodology developed by Kaiser has been extended to other proteins by Annan and Jordan [45], who covalently modified the active site cysteine of the thiamine-dependent enzyme pyruvate decarboxylase with 8 α -bromoacetyl-10-methylisoalloxazine. The semisynthetic flavoenzyme was shown to possess significant pyruvate oxidase activity. The flavin, in this case, interacts synergistically with the active site thiamine, predictably altering the partitioning of a key anionic intermediate. Suckling and Zhu [46] modified the catalytic cysteine in papain with 2-bromomethyl-*N*-benzylthiazolium bromide (Figure 2). The resulting thiazolopapain successfully mediated carbon-carbon bond formation with 6-oxoheptanal as the substrate to give predominantly two products derived from either intramolecular or intermolecular condensation of the substrate. Thiazolopapain thus provides an interesting counterpart to naturally occurring thiamine-dependent enzymes.

5. SITE-SPECIFIC NUCLEASES

Biology has been revolutionized by the discovery and utilization of restriction endonucleases and DNA ligases. These enzymes have given scientists the ability to 'cut and paste' fragments of DNA, immeasurably facilitating the manipulation and study of genetic material. The design of novel molecules which can cleave DNA in a site-specific manner will further advance DNA-cloning and DNA-mapping technologies. Similar agents for sequence-specific hydrolysis of RNA should also aid studies on RNA structure and function.

A number of research groups have investigated the possibility of generating such catalysts via chemical modification of existing proteins, and two basic approaches have developed. The first addresses one of the limitations of natural restriction endonucleases: that they have short DNA recognition sites (up to eight base pairs). These recognition sites therefore occur with relatively high frequency in random-sequence DNA, making the mapping of large DNA molecules, such as human chromosomes, problematic. Chen and Sigman [47] observed that the conversion of DNA-binding proteins, which have recognition sequences of approximately 20 base pairs, into endonucleases with the same recognition sites would circumvent this problem. To this end they covalently attached 1,10-phenanthroline moieties, which can chelate metal ions that may promote a nonspecific cleavage reaction, to the *trp* repressor from *E. coli* (Figure 3). The chemical modification procedure yielded a semisynthetic protein that contained four phenanthroline moieties in each subunit. Importantly, the derivatized DNA-binding protein retained its specificity, and was able to bind selectively to double-stranded DNA fragments that contained the *aroH* and *trpEDCBA* operator regions. Addition of copper(II) and a thiol reducing agent

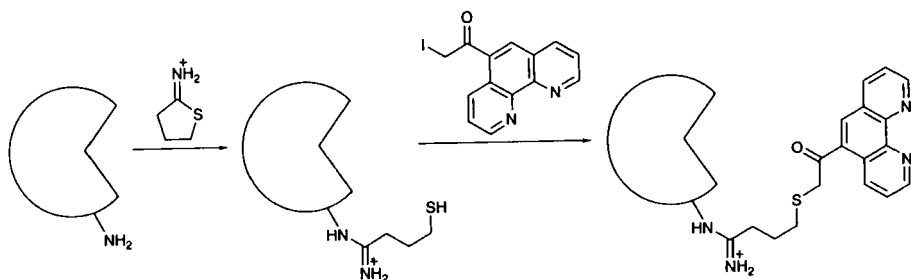


Figure 3 Attachment of a metal-chelating agent (a phenanthroline derivative) to a DNA-binding protein

caused highly selective cutting of the DNA backbone at the target sites. The cleavage is thought to be mediated by a nondiffusible, phenanthroline-bound 'copper-oxo' complex which can oxidatively attack proximal deoxyribose groups.

In principle, this strategy may be applied to other DNA- or RNA-binding proteins to create a family of semisynthetic nucleases. Ebright *et al.* [48] have applied the methodology to *E. coli* catabolite gene activator protein (CAP), selectively alkylating Cys178 with 5-iodoacetamido-1,10-phenanthroline. The modified CAP cut double-stranded DNA selectively at the center of the 22 base pair recognition site of CAP in a similar copper-mediated process to that described above. In an earlier study, Dervan and coworkers [49] attached an EDTA moiety to the N-terminus of a 52 residue peptide which was effectively the DNA-binding domain of Hin recombinase. The modified peptide bound in the minor groove of duplex DNA at a sequence corresponding to the Hin recombination site, and addition of iron(II) and a reducing agent initiated sequence-specific cleavage of the nucleotide chain. These results suggest that semisynthetic versions of such proteins should indeed be powerful tools for molecular biologists.

The second basic approach in this area has been developed by Schultz and coworkers, who have attached relatively nonspecific phosphodiesterases to oligonucleotides which may be expected to guide the nuclease enzyme to the complementary nucleotide sequence. The enzyme staphylococcal nuclease was engineered by oligonucleotide-directed mutagenesis to provide it with a suitable coupling site in the form of a cysteine residue (Figure 4). This allowed the protein to be linked, via a disulfide exchange reaction, to one of several synthetic oligonucleotides that contained a 3'-thiol [50]. The oligonucleotide was able to hybridize successfully with its target complementary sequence, and addition of calcium(II) permitted the semisynthetic enzyme to cleave single-stranded DNA [50] and RNA [51] adjacent to the oligonucleotide-binding site. During extended reaction times, it was found that autolysis of the coupled oligonucleotide occurred, leading to nonselective cleavage. However, second-generation semisynthetic nucleases have been genetically engineered, and these

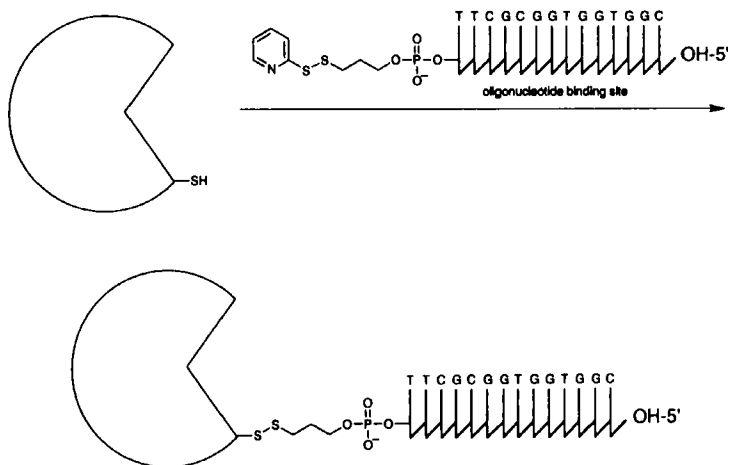


Figure 4 Derivatization of a nuclease with an oligonucleotide-binding site to give a site-specific DNA cleaving agent

have been shown to hydrolyze single-stranded DNA under more varied conditions than semisynthetic versions of the native enzyme [52]. The delivery of the modified nuclease to supercoiled, double-stranded DNA via D-loop formation and subsequent selective cleavage, has also been demonstrated [53]. These studies, and related work [54] with ribonuclease S, illustrate a potentially general route to semisynthetic nucleases capable of hydrolyzing DNA or RNA at any given target sequence.

6. OTHER SEMISYNTHETIC CATALYSTS

The design and synthesis of new biocatalysts is important, in part, because existing enzymes present us with a limited array of chemical activity: they only catalyze processes which nature has required. Similarly, the production of semisynthetic catalysts is limited by the availability of suitable, well-characterized protein templates. The development of catalytic antibodies circumvents this difficulty by harnessing the power of the mammalian immune system to provide specific, tailored binding pockets. By generating antibodies which bind a designed hapten, such as a transition state analogue, an array of novel biocatalysts has been realized [55–57]. However, the rate enhancements afforded by most catalytic antibodies are several orders of magnitude lower than those of corresponding enzymes, and this may reflect the difficulty in eliciting an antibody combining site with the appropriate constellation of general acids/bases, nucleophiles and cofactors essential for optimal catalysis.

Schultz and coworkers have addressed this issue directly by incorporating catalytic groups into antibody combining sites via chemical modification. Taking

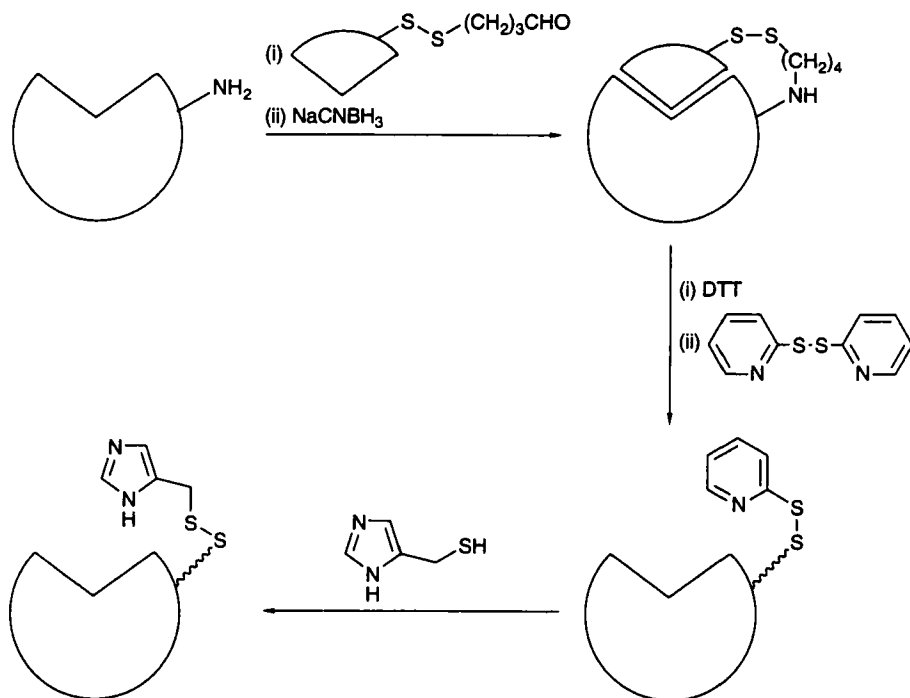


Figure 5 Introduction of a catalytic imidazole group into an antibody via affinity labeling

the 2,4-dinitrophenyl-binding antibody MOPC 315 as a protein scaffold, they modified its binding site with cleavable affinity labels as illustrated in Figure 5, and were able to introduce a thiol side chain, an imidazole moiety or a fluorescent reporter group [58, 59]. The imidazole-derivatized antibody catalyzed the hydrolysis of coumarin esters with rates that were approximately 1000 times the acceleration achieved with the simple model 4-methylimidazole [59]. Optimization of the tether which connects the catalytic group to the protein may lead to even more efficient catalysts. This approach is generally applicable to antibodies, since it does not presuppose any detailed structural knowledge of the binding site. In addition, the strategy could be readily extended to introduce a variety of catalytic cofactors.

Another approach to semisynthetic catalysts was recently reported by Shi and Griffin [60]. Using the relatively small glycopeptide antibacterial agent vancomycin as a template, they derivatized its C-terminus with alkyl, imidazole and amino functionalities. The antibiotic activity of vancomycin comes from its ability to bind to peptidoglycan precursors, preventing their incorporation into the bacterial cell wall. The chemically modified vancomycins retained the ability to bind a peptide analogue of these cell wall precursors and, moreover,

catalyzed the hydrolysis of a *p*-nitrophenyl carbamate group attached to the peptide analogue. Although the rate accelerations observed were small, they could be correlated with increases in positive charge on the modified C-terminus of vancomycin which presumably alter the microenvironment of the bound carbamate-peptide and facilitate carbamate hydrolysis.

7. FUTURE DIRECTIONS

Protein catalysts have great potential to contribute in diverse ways to research, industry and medicine. Both from an academic and practical standpoint, the creation of new biocatalysts is highly desirable. Chemical modification of proteins has developed from an often empirical methodology for manipulating enzyme specificity and mapping the crucial residues of active sites into a more exact science where rational alteration of enzymatic properties and activities is a reality. Specific covalent modification of active site residues has permitted the generation of a number of semisynthetic enzymes, and similar strategies have been successfully applied to other proteins, such as DNA-binding proteins and antibodies.

Chemical mutagenesis, the synthetic replacement of one amino acid residue by another, as elegantly demonstrated in the generation of thiolsubtilisin, is unfortunately limited to the small number of protein residues which may be selectively activated and altered. The advent of oligonucleotide-directed mutagenesis provides the ability to replace any natural amino acid residue with any other natural residue in a site-specific manner. The technique is precise and general, and has allowed the rational manipulation of enzyme properties such as substrate specificity [61,62]. Conversely, although selective chemical modification of proteins is, by definition, somewhat limited in the choice of residues which can be modified, it can be performed efficiently and rapidly on a large scale. Moreover, the incorporation of unnatural amino acids or cofactors is quite straightforward with chemical modification protocols, but difficult to accomplish in good yield with molecular biological techniques. Recent breakthroughs in the production of *in vitro* expression systems which allow the expansion of the genetic code to include unnatural amino acids are truly exciting, but have only produced very low yields of the desired mutant proteins and are confined to residues that are tolerated by the protein translation machinery [26–29]. Thus, recombinant technology has made possible the isolation, and subsequent characterization, of many native and mutant proteins but is restricted to a limited repertoire of amino acids. This can be contrasted with the direct chemical synthesis of proteins from their constituent residues, in which essentially any amino acids may be used readily. A case in point is the chemical synthesis of D-HIV protease, composed entirely of D-amino acids and the enantiomer of naturally occurring HIV protease, by Kent and coworkers [63]. This heroic effort demonstrates the feasibility of chemically synthesizing a 100 residue

polypeptide chain. Another report from the same laboratory describes the use of chemical methods to produce proteins with novel backbone linkages [64], which is also not possible by genetic methodology. As the chemical route to proteins becomes more practicable, so does the incorporation of an unnatural cofactor or label at virtually any position in the protein. Gray, Imperiali and colleagues [65] recently described the semisynthesis of cytochrome *c* mutants that contained bipyridyl alanine residues. They used solid-phase peptide synthesis to produce fragments containing residues 66–104, with the unnatural amino acids at position 72. These ‘mutant’ peptides were then attached to the peptide fragment 1–65, which was obtained from cyanogen bromide cleavage of the natural protein. One of the resulting cytochrome mutants was able to chelate a ruthenium ion via the bipyridyl side chain, and it demonstrated enhanced electron transfer between this ruthenium and the iron-containing heme embedded in the cytochrome. The extent to which peptide ligases, such as the ‘damaged’ subtilisin variants mentioned above, will contribute to the chemical synthesis of such proteins remains to be seen, but it might be expected that engineered ligases could accept a variety of unnatural amino acids or peptides as substrates.

The different strengths and weaknesses of chemical and genetic modifications of proteins lead to the conclusion that both approaches are worthwhile in the

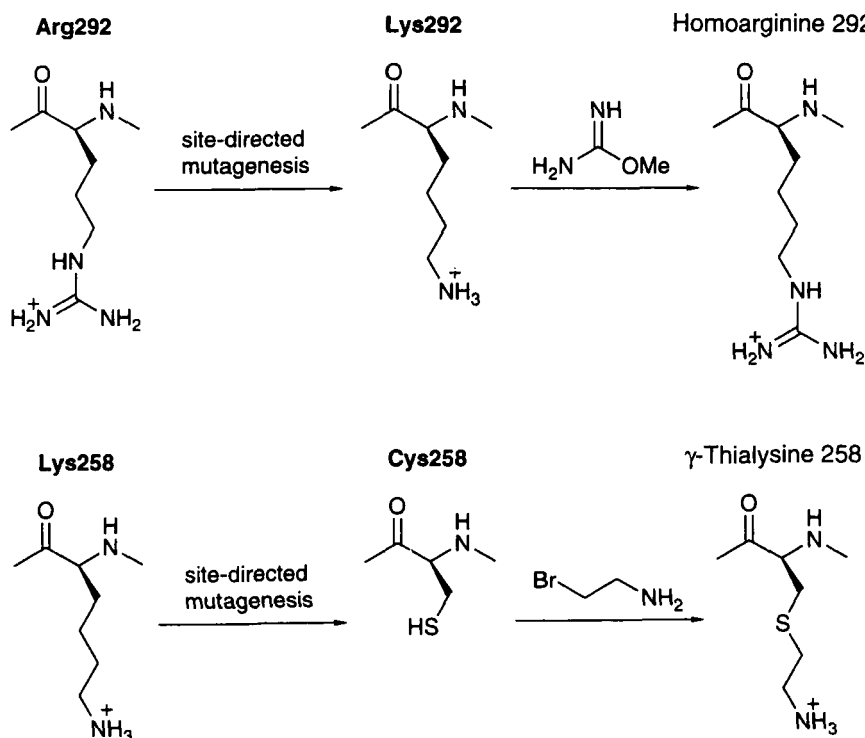


Figure 6 A combination of genetic and chemical modifications can produce subtle changes in active site side chains

pursuit of novel biocatalysts. Perhaps the last word should go to Kirsch and coworkers, who have utilized a combination of genetic and chemical methodologies in site-directed mutagenesis studies on aspartate aminotransferase. They realized that subtle changes in the side chain were often not possible for many residues because of the limited choice of replacements available in the other 19 amino acids. Replacement of a lysine residue, for example, must lead to a radical change of functional group since no other α -amino acid has a primary amine side chain. Consequently, they have employed ingenious combinations of site-directed mutagenesis and chemical modification to provide more subtle substitutions (Figure 6). The essential residue Lys258, for instance, was replaced by a Cys via genetic means and this thiol was selectively alkylated with 2-bromoethylamine to give a primary-amine-bearing side chain [66]. In another example, Arg292 was mutated to a Lys, which was then chemically guanidinated to give a homoarginine residue, in which the functionality is unchanged from Arg but the side chain is one methylene group longer [67]. Such subtle mutations of amino acid side chains have enabled Kirsch and coworkers to delicately probe the catalytic mechanism of aspartate aminotransferase. It is likely that similar fusions of genetic and chemical modification strategies will prove to be invaluable in advancing our understanding of enzyme mechanism, and in the creation of novel biocatalysts.

8. ACKNOWLEDGMENTS

This work was supported in part by NSF Grant CHE-9306267 and an SERC/NATO Postdoctoral Fellowship to Ian M. Bell.

9. REFERENCES

1. E. Fischer, *Ber. Dtsch. Chem. Ges.*, **27**, 2985 (1894).
2. A. Fersht, *Enzyme Structure and Mechanism*, 2nd edn, Freeman, New York, 1985.
3. J. M. Lehn, *Science*, **227**, 849 (1985).
4. D. J. Cram, *Science*, **240**, 760 (1988).
5. R. Breslow, *Ciba Found. Symp.*, **158**, 115 (1991).
6. T. R. Cech, *Annu. Rev. Biochem.*, **59**, 543 (1990).
7. A. A. Beaudry and G. F. Joyce, *Science*, **257**, 635 (1992).
8. D. P. Bartel and J. W. Szostak, *Science*, **261**, 1411 (1993).
9. S. F. Betz, D. P. Raleigh and W. F. DeGrado, *Curr. Opin. Struct. Biol.*, **3**, 601 (1993).
10. M. H. Hecht, J. S. Richardson, D. C. Richardson and R. C. Ogden, *Science*, **249**, 884 (1990).
11. D. P. Raleigh and W. F. DeGrado, *J. Am. Chem. Soc.*, **114**, 10 079 (1992).
12. L. Hugouenq and J. Loiseleur, *C. R. Hebd. Seances Acad. Sci.*, **181**, 149 (1925).
13. E. T. Kaiser, D. S. Lawrence and S. E. Rokita, *Annu. Rev. Biochem.*, **54**, 565 (1985).
14. R. L. Lundblad, *Chemical Reagents for Protein Modification*, 2nd edn, CRC Press, Boca Raton, FL, 1991.

15. Y. Kamisaki, H. Wada, T. Yagura, A. Matsushima and Y. Inada, *J. Pharmacol. Exp. Ther.*, **216**, 410 (1981).
16. A. Abuchowski, *Cancer Biochem. Biophys.*, **7**, 175 (1984).
17. K. E. Neet and D. E. Koshland Jr, *Proc. Natl. Acad. Sci. USA*, **56**, 1606 (1966).
18. L. Polgár and M. L. Bender, *J. Am. Chem. Soc.*, **88**, 3153 (1966).
19. Z.-P. Wu and D. Hilvert, *J. Am. Chem. Soc.*, **111**, 4513 (1989).
20. M. Philipp and M. L. Bender, *Mol. Cell. Biochem.*, **51**, 5 (1983).
21. T. Nakatsuka, T. Sasaki and E. T. Kaiser, *J. Am. Chem. Soc.*, **109**, 3808 (1987).
22. L. Abrahamsén, *et al.*, *Biochemistry*, **30**, 4151 (1991).
23. H. Yokosawa, S. Ojima and S.-I. Ishii, *J. Biochem.*, **82**, 869 (1977).
24. P. I. Clark and G. Lowe, *Eur. J. Biochem.*, **84**, 293 (1978).
25. J. B. West, *et al.* *J. Am. Chem. Soc.*, **110**, 3709 (1988).
26. C. J. Noren, S. J. Anthony-Cahill, M. C. Griffith and P. G. Schultz, *Science*, **244**, 182 (1989).
27. J. D. Bain, C. G. Glabe, T. A. Dix, A. R. Chamberlin and E. S. Diala, *J. Am. Chem. Soc.*, **111**, 8013 (1989).
28. J. D. Bain, C. Switzer, A. R. Chamberlin and S. A. Benner, *Nature*, **356**, 537 (1992).
29. H.-H. Chung, D. R. Benson and P. G. Schultz, *Science*, **259**, 806 (1993).
30. T. C. Stadtman, *Annu. Rev. Biochem.*, **59**, 111 (1990).
31. K. L. House, R. B. Dunlap, J. D. Odom, Z.-P. Wu and D. Hilvert, *J. Am. Chem. Soc.*, **114**, 8573 (1992).
32. R. Syed, Z.-P. Wu, J. M. Hogle and D. Hilvert, *Biochemistry*, **32**, 6157 (1993).
33. K. L. House, A. R. Garber, R. B. Dunlap, J. D. Odom and D. Hilvert, *Biochemistry*, **32**, 3468 (1993).
34. L. Flohé, in *Glutathione: Chemical, Biochemical, and Medical Aspects* (eds D. Dolphin, R. Poulson and O. Avamovic), Wiley, New York, 1989, p. 644.
35. J. W. Forstrom, J. J. Zakowski and A. L. Tappel, *Biochemistry*, **17**, 2639 (1978).
36. Z.-P. Wu and D. Hilvert, *J. Am. Chem. Soc.*, **112**, 5647 (1990).
37. I. M. Bell, M. L. Fisher, Z.-P. Wu and D. Hilvert, *Biochemistry*, **32**, 3754 (1993).
38. I. M. Bell and D. Hilvert, *Biochemistry*, **32**, 13 969 (1993).
39. H. Jornvall, C. Woenckhaus and G. Johnscher, *Eur. J. Biochem.*, **53**, 71 (1975).
40. M.-O. Mansson, P.-O. Larsson and K. Mosbach, *Eur. J. Biochem.*, **86**, 455 (1978).
41. D. Hilvert and E. T. Kaiser, *Biotechnol. Genet. Eng. Rev.*, **5**, 297 (1987).
42. C. Radziejewski, D. P. Ballou and E. T. Kaiser, *J. Am. Chem. Soc.*, **107**, 3352 (1985).
43. J. T. Slama, C. Radziejewski, S. R. Oruganti and E. T. Kaiser, *J. Am. Chem. Soc.*, **107**, 6778 (1984).
44. D. Hilvert, Y. Hatanaka and E. T. Kaiser, *J. Am. Chem. Soc.*, **110**, 682 (1988).
45. N. Annan and F. Jordan, *J. Am. Chem. Soc.*, **112**, 3222 (1990).
46. C. J. Suckling and L.-M. Zhu, *Bioorg. Med. Chem. Lett.*, **3**, 531 (1993).
47. C. H. B. Chen and D. S. Sigman, *Science*, **237**, 1197 (1987).
48. R. H. Ebright, Y. W. Ebright, P. S. Pendergrast and A. Gunasekera, *Proc. Natl. Acad. Sci. USA*, **87**, 2882 (1990).
49. J. P. Sluka, S. J. Horvath, M. F. Bruist, M. I. Simon and P. B. Dervan, *Science*, **238**, 1129 (1987).
50. D. R. Corey and P. G. Schultz, *Science*, **238**, 1401 (1987).
51. R. N. Zuckermann, D. R. Corey and P. G. Schultz, *J. Am. Chem. Soc.*, **110**, 1614 (1988).
52. D. R. Corey, D. Pei and P. G. Schultz, *Biochemistry*, **28**, 8277 (1989).
53. D. R. Corey, D. Pei and P. G. Schultz, *J. Am. Chem. Soc.*, **111**, 8523 (1989).
54. R. N. Zuckermann and P. G. Schultz, *J. Am. Chem. Soc.*, **110**, 6592 (1988).
55. P. G. Schultz and R. A. Lerner, *Acc. Chem. Res.*, **26**, 391 (1993).
56. J. D. Stewart, L. J. Liotta and S. J. Benkovic, *Acc. Chem. Res.*, **26**, 396 (1993).
57. D. Hilvert, *Acc. Chem. Res.*, **26**, 552 (1993).

58. S. J. Pollack, G. R. Nakayama and P. G. Schultz, *Science*, **242**, 1038 (1988).
59. S. J. Pollack and P. G. Schultz, *J. Am. Chem. Soc.*, **111**, 1929 (1989).
60. Z. Shi and J. H. Griffin, *J. Am. Chem. Soc.*, **115**, 6482 (1993).
61. J. A. Wells, B. C. Cunningham, T. P. Graycar and D. A. Estell, *Proc. Natl. Acad. Sci. USA*, **84**, 5167 (1987).
62. H. M. Wilks, *et al.*, *Science*, **242**, 1541 (1988).
63. R. C. deLisle Milton, S. C. F. Milton and S. B. H. Kent, *Science*, **256**, 1445 (1992).
64. M. Schnölzer and S. B. H. Kent, *Science*, **256**, 221 (1992).
65. D. S. Wuttke, H. B. Gray, S. L. Fisher and B. Imperiali, *J. Am. Chem. Soc.*, **115**, 8455 (1993).
66. A. Planas and J. F. Kirsch, *Biochemistry*, **30**, 8268 (1991).
67. P. W. White and J. F. Kirsch, *J. Am. Chem. Soc.*, **114**, 3567 (1992).

Chapter 4

Oligonucleotides: Superspecific Ligands for Targeting Nucleic Acids and Proteins and Development of Molecular Devices

V. V. VLASSOV

Institute of Bioorganic Chemistry, Novosibirsk, Russia

1. INTRODUCTION

Four decades have passed since the discovery by Watson and Crick of the principle of complementary interaction of nucleic acids [1]. According to this principle, antiparallel single-stranded nucleic acids can form specific complexes if in their sequences cytosine residues in one strand are located opposite to guanine residues in the other strand and thymine (or uracyl) residues are located opposite to adenosine residues in the other strand. This allows each nucleoside in one strand to form hydrogen bonds with the corresponding complementary residue in the other strand. Being formed by multipoint cooperative hydrogen bonding of a great number of functional groups, the complementary complexes of nucleic acids are extraordinarily stable and specific. They are fundamental to the events that define life: storage, processing and transmission of genetic information.

Since 1953 the advances in understanding of the mechanisms of specific interactions of nucleic acids, which provide the physicochemical basis for the most important biological processes such as multiplication, expression and reparation of genetic information, have been steady and impressive. One of the

important results of these studies was the recognition of the enormous targeting potential of oligonucleotides, which are fragments of nucleic acids long enough to retain the capability of specific interactions with nucleic acids and proteins [2–5]. These molecules of moderate size represent a unique class of specific ligands and potentially biologically active compounds primarily because of the simplicity of the rational design of oligonucleotides for targeting any arbitrary nucleic acid. Indeed, the rational design of linear oligomers built of a small number of monomers allows the targeting of any arbitrary nucleotide sequence according to the Watson–Crick recognition scheme, following the lock-and-key principle formulated by Emil Fischer. Recent studies have demonstrated that oligonucleotide analogues and derivatives capable of specific recognition of specific sequences in double-stranded nucleic acids can be designed. In this case again the recognition principle is given by the direct lock-and-key scheme which operates by formation of multiple hydrogen bonds along the contact surface of an oligonucleotide and a nucleic acid.

Recent studies have shown that the recognition potential of oligonucleotides is far from being limited by the possibility of forming complementary complexes with single-stranded nucleic acids and double-stranded nucleic acids. A great number of enzymes operating with nucleic acids and effectors of gene expression have been discovered and it has been shown that the proteins can form specific complexes with oligonucleotides containing the appropriate nucleotide sequences. The development of the molecular selection approach [6, 7] has opened up the possibility of generating oligonucleotide structures capable of specific binding to various proteins and even organic molecules which apparently have nothing to do with the molecular interactions of nucleic acids in nature.

The unique specificity of oligonucleotides as targeting structures, the possibility of rational design of oligonucleotide-based recognition elements and advances in oligonucleotide synthesis have resulted in the development of a number of key techniques for molecular biology and genetic engineering. The known principles of recognition of nucleic acids and the possibility of using reagents provided by nature, namely enzymes, have allowed nucleic acid chemists to develop simple methods of detection and almost unlimited possibilities of multiplication for the constructions they have designed, making for rapid progress in research. The enormous specificity of the complementary interactions of oligonucleotides has provided the possibility of designing various systems for the detection, isolation and sequencing of nucleic acids and the manipulation of DNA and RNA.

In 1967 Grineva *et al.* [8] described the first reactive derivatives of oligonucleotides designed for sequence-specific chemical modification of nucleic acids. This work laid the foundation for the development of a new field, namely that of affinity reagents targeted to nucleic acids for research purposes and for effecting the expression of specific genes. Since the first experiments with reactive derivatives of oligonucleotides [8, 9] were reported, a few important studies have initiated new stages of development in the field of the design and

investigation of oligonucleotide derivatives. Antiviral tests have been performed by Stephenson and Zamecnik, who first demonstrated the possibility of controlling virus proliferation using complementary oligonucleotides [10]. Miller *et al.* [11] have pioneered in the design of nonionic analogues of oligonucleotides in the hope of developing nuclease-resistant compounds capable of entering eukaryotic cells by diffusion through the cell membrane. It has been shown that oligonucleotide derivatives targeted to specific nucleotide sequences interact specifically with nucleic acids containing these sequences and affect the expression of the corresponding genes [2–5]. Therefore, targeting of specific mRNAs with complementary oligonucleotides (antisense oligonucleotides) and with complementary RNA produced by artificial constructs introduced into cells (antisense RNA) is used by molecular biologists to elucidate the biological role of specific genes. Oligonucleotide derivatives and analogues are receiving growing attention as gene-targeted, biologically active compounds and as potential therapeutics capable of inactivating foreign genomes of infectious agents and modulating gene expression [2–5, 12–15].

It is the purpose of this chapter to review the design and major applications of oligonucleotide derivatives targeted to nucleic acids and proteins, with an emphasis on the recognition problems and applications of the reactive derivatives used for the chemical modification and inactivation of specific biopolymers.

2. STRUCTURE OF NUCLEIC ACIDS

Of the known nucleic acid structures, the double-stranded complexes and the triple-stranded complexes play a major role in the folding and interactions of nucleic acids.

Nucleic acid duplexes are formed between two polynucleotide strands in the double-stranded nucleic acids. They are also formed between complementary sequences in single-stranded nucleic acids to give hairpin structures, which consist of two bound complementary stretches of nucleotides (stems) connected by a single-stranded nucleotide sequence. Intramolecular complementary interactions result in the formation of the secondary structure of the nucleic acid which arranges itself further to form a three-dimensional folded conformation (tertiary structure), allowing maximal hydrogen bonding between elements of the secondary structure.

Natural double-stranded DNAs and RNAs are regular, right-handed helices formed by two antiparallel polynucleotide chains. Heterocyclic bases in the nucleoside residues of the nucleic acids adopt the *anti*-conformation around the glycosidic bond and form hydrogen bonds with the residues in the complementary polynucleotide strand (Figures 1 and 2) [16]. In aqueous solutions, double-stranded DNA and RNA form helices of B-type conformation with 10 base pairs per turn and A-type conformation with about 11–12 base pairs per turn, respectively. The conformation of the nucleoside sugar is C_3' -*endo* for the

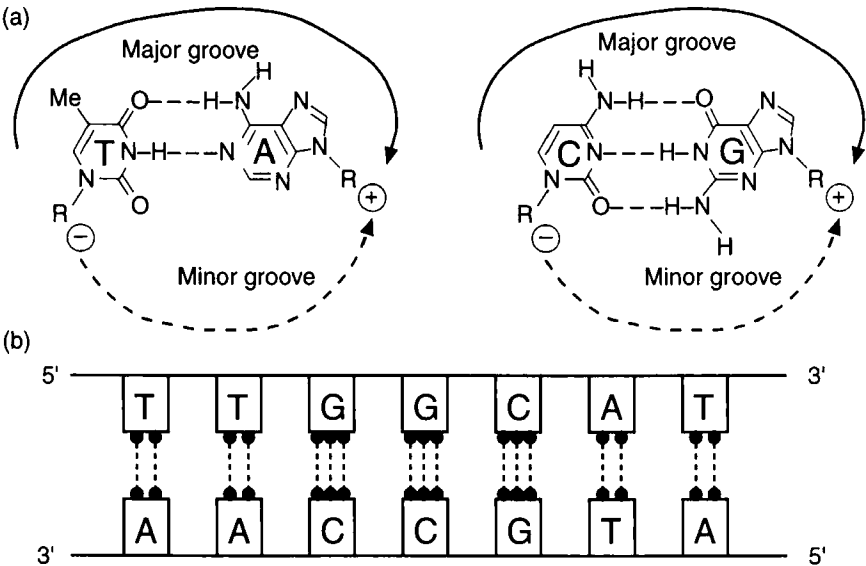


Figure 1 (a) Watson-Crick base pairs TA and CG, where R is a sugar residue. Pluses and minuses indicate the directions of the polynucleotide strands. (b) Schematic diagram of a complementary double-stranded complex formed by two nucleotide sequences

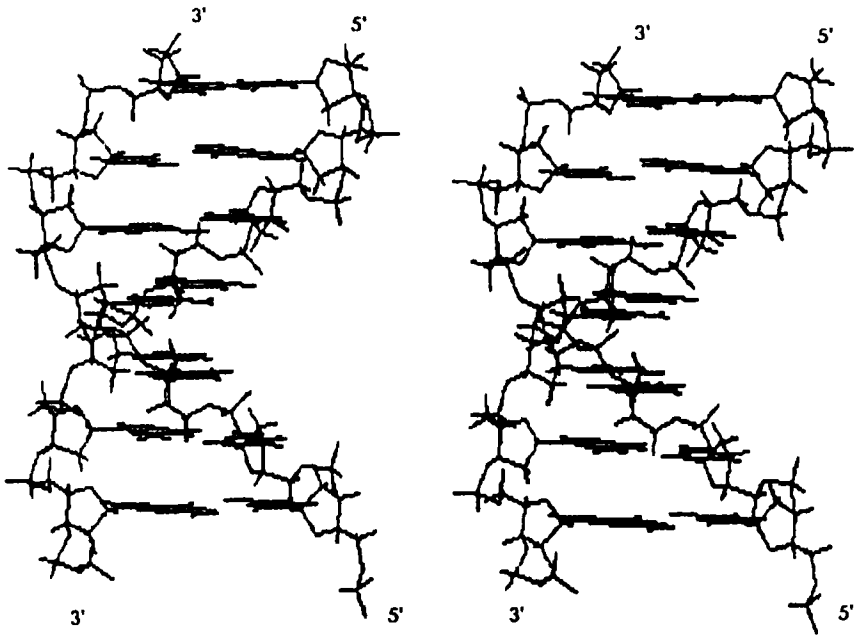


Figure 2 A stereoview of a duplex (B-form) of two deoxyribooligonucleotides, namely pTGTTTGG (top, 5'-terminus) and pCCAAACA (top, 3'-terminus)

A-type conformation and C_2' -*endo* for the B-type conformation. For the RNA duplexes only the A-type conformation is realized, while the DNA conformation can vary in different conditions to form a number of B-type-like conformations [16]. The complementary base pairs are stacked on each other along the nucleic acid helix. Phosphates are located at the outer surface of the helix. There are two grooves in the nucleic acid duplex where some groups of the heterocyclic bases are available to the external molecules. In the RNA duplex, the base pairs are displaced away from the axis of the helix toward the minor groove, and therefore the major groove is deep and narrow and the minor groove is shallow and wide. In the heteroduplexes formed by RNA and DNA, the hybrid helix conformation is similar to the A-type conformation.

Alternating purine–pyrimidine sequences in DNA can form a left-handed Z-conformation when the DNA is negatively supercoiled and in some salt conditions. In the Z-DNA, deoxyriboguanosine residues have the *syn*-glycosidic bond orientation of the base. The Z-form can also be formed by alternating GCGC sequences in double-stranded RNA. Z-DNA structures have been detected in cellular DNA and may be important for the regulation of gene expression.

When a homopurine sequence is present in one strand of the double-stranded nucleic acid and correspondingly a homopyrimidine complementary sequence is in the other strand, a third nucleotide strand can bind into the major groove of this duplex to give a triplex structure [17, 18]. There are two major types of triplexes, with the third strand being either homopyrimidine or homopurine. In the first case, pyrimidine bases of the third strand form so-called Hoogsteen base pairs [19] with the bases of the purine strand of the duplex to give triplets of bases T–TA and C–CG (Figure 3). Experimental studies have shown that the binding occurs according to the schemes shown in Figures 3(a) and 3(b). An *anti*-glycosidic bond orientation in the residues of the third strand is needed for triplex formation, and therefore the third strand binds in parallel to the purine strand of the duplex. The cytosine residues of the third strand must be protonated to form a hydrogen bond between N7(G) and N3(C), and therefore the complexes containing C–GC triplets are formed in slightly acidic conditions [17]. The triplets T–AT and C–GC are isomorphous, which allows formation of mixed sequence triplexes. The triple-stranded structures can be formed by both DNA and RNA. In the RNA triple-stranded complexes, the structure of the constituent duplex is of the A-type. In the DNA triple-stranded complexes, the duplex conformation depends on the sequence, and can have a conformation similar to the B-form or A-form.

When the third strand in the triplex structure is composed of purines, nearly isomorphous triplets A–AT(U) and G–GC can be formed (Figure 4). The triple-stranded complexes formed by purine oligonucleotides are stable in physiological conditions. In the cases investigated so far, the purine residues of the third strand can adopt the *anti*-conformation of the glycosidic bond, allowing the reversed Hoogsteen-type bonding with Watson–Crick base pairs when the strand is

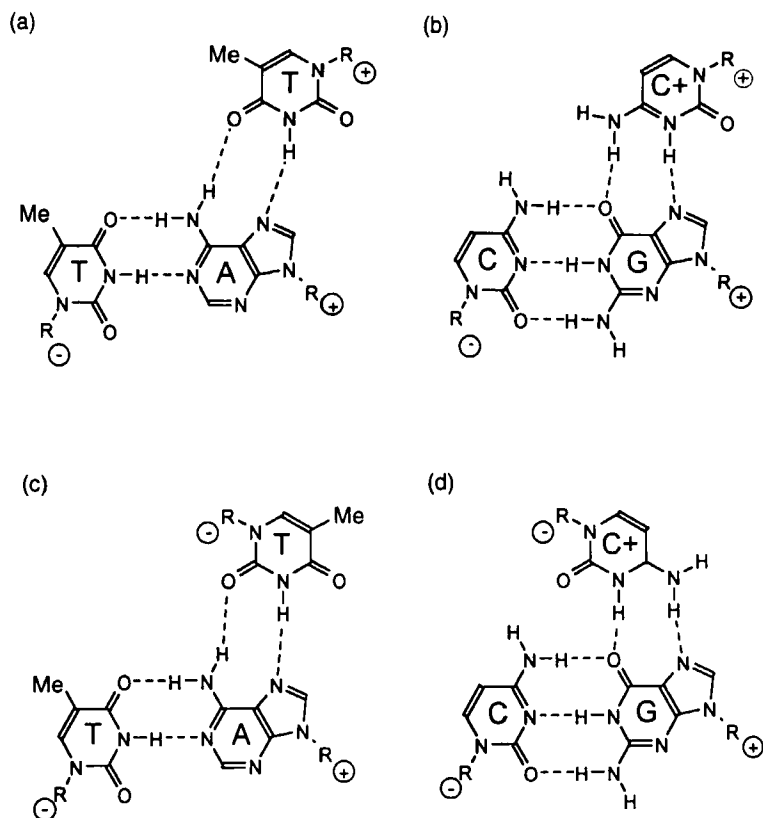


Figure 3 Base triplets which can be formed by pyrimidines of the third DNA strand and base pairs of the double-stranded DNA, where R is a sugar residue and pluses and minuses indicate the directions of the polynucleotide strands. (a, b) The pyrimidine base of the third strand is bound to the purine of the duplex by the Hoogsteen pairing. The orientations of the glycosidic bonds in the double-stranded DNA and in the third strand are *anti*. (c, d) The pyrimidine base of the third strand is bound to the purine of the duplex by the reversed Hoogsteen pairing. Nucleosides in all the three strands have the *anti*-conformation of the glycosidic bond

antiparallel to the duplex purine strand (Figures 4a and 4b). In theory, the Pu–PuPy triplex in the third strand can assume the *syn*-conformation of the glycosidic bond, which allows the formation of Hoogsteen-type hydrogen bonding with the Watson–Crick base pair (Figures 4c and 4d). A parallel orientation of the third strand to the duplex purine strand is required in this case.

It has been found that duplexes with a strong preference for purines in one strand can bind to oligonucleotides consisting of guanosine and thymidine residues [20]. Guanosines in the third strand are believed to form triplets, as shown in Figure 4(a), and thymidine residues interact with adenosines located

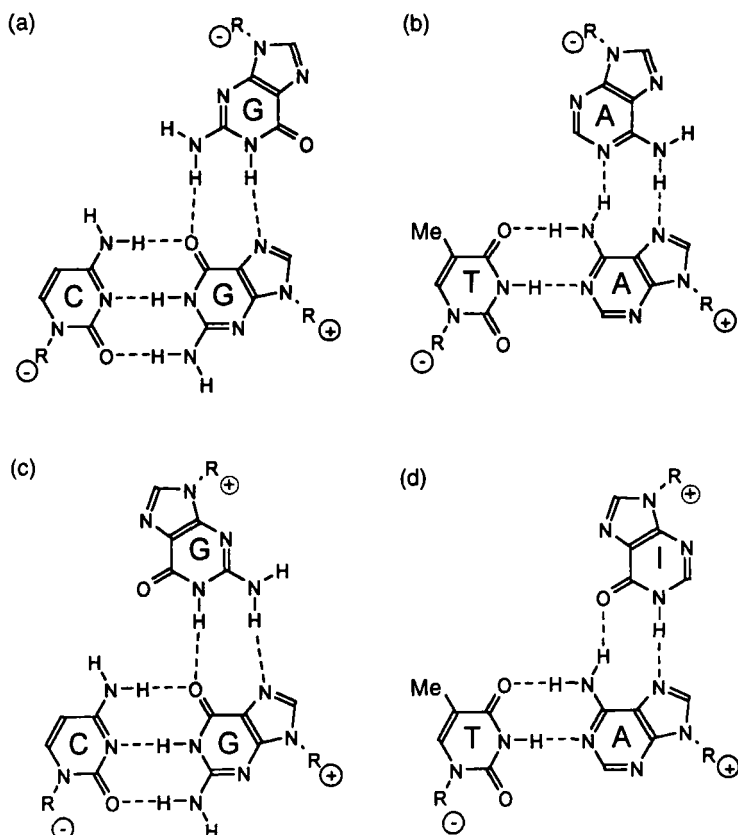


Figure 4 Base triplets which can be formed by purines of the third strand and base pairs of the double-stranded DNA, where R is a sugar residue and pluses and minuses indicate the directions of the polynucleotide strands. (a, b) The purine base of the third strand is bound to the purine of the duplex by the reversed Hoogsteen pairing. The orientation of the glycosidic bonds in the nucleosides of all three polynucleotide strands is *anti*. (c) Guanosine of the third strand is bound to guanosine of the duplex by the Hoogsteen pairing. (d) Adenosine cannot interact with the TA base pair by the Hoogsteen pairing. To recognize this base pair, inosine can be used. The nucleosides in the third strand have the *syn*-conformation of the glycosidic bond and the nucleosides of the duplex have the *anti*-conformation of the glycosidic bond in (c) and (d)

in either of the DNA strands. Details of the structures of different triple-stranded complexes have not yet been elucidated.

It has been found that homopurine-homopyrimidine mirror repeat DNA sequences in plasmids can form intramolecular triple-stranded structures (H-DNA) when the DNA is under negative supercoiling stress and/or an acidic pH [21–23]. The process includes disruption of the Watson–Crick base pairing and restructuring followed by formation of an intramolecular triplex (Figure 5).

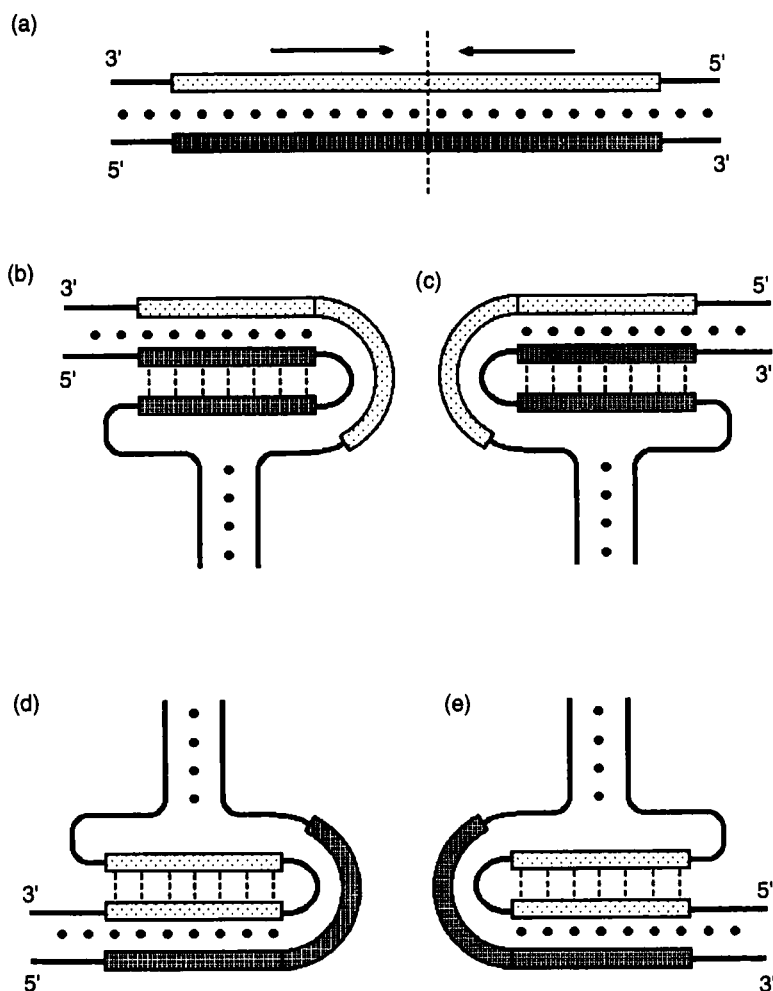


Figure 5 Intramolecular triplexes formed by a DNA sequence with homopurine-homopyrimidine mirror repeat. (a) Mirror image sequences are indicated by arrows; the vertical line shows the pseudosymmetry axis. H-DNA is formed when the pyrimidine lower strand folds back to form a triple-stranded structure with the left (b) or right (c) part of the target sequence. Half of the upper purine strand remains unstructured. H*-DNA is formed when the purine upper strand folds back to form a triplex with the left (d) or right (e) half of the pyrimidine sequence. Half of the lower purine strand remains unstructured

In mildly acidic conditions, an H-DNA structure consisting of C-GC and T-AT triads is formed by the pyrimidine strand with half of the purine strand, and the free purine strand forms a loop. In the presence of bivalent cations, an H*-DNA structure consisting of G-GC and A-AT base triads is formed by half of the pyrimidine strand, and the free pyrimidine strand remains unstructured. The homopurine-homopyrimidine sequences are often found in

regulatory regions of genes. The high stability of the triple-stranded complexes in physiological conditions, the interaction of specific cellular proteins with the triple-stranded complexes and the inability of DNA polymerases to read effectively through the triple helices preformed in a template DNA all give evidence that this structure could be of biological significance, perhaps for the regulation of gene expression.

Systematic investigations of the stability of triple-stranded complexes formed by various compositions of RNA and DNA strands have shown that the most stable complexes are formed by DNA duplexes and single-stranded homopyrimidine RNA sequences [24]. This finding allows one to speculate about the existence of regulatory RNA molecules capable of binding to homopurine-homopyrimidine DNA sequences.

Recently, a few other types of secondary structures for DNA and RNA have been discovered. Tetrameric DNA structures with exceptionally high stability have been shown to be formed by $G_n(A)T_m$ [25] sequences found overhanging at the ends of eukaryotic chromosomes and by sequences of $(GT)_n$ [26], which have been found in many introns and regulatory sites in genes of high organisms. The structures are formed by association of guanosine residues into square planar units, namely G-quartets, with each guanosine serving as both hydrogen bond acceptor and donor in Hoogsteen base pairs with two neighboring guanosines in a cyclic arrangement. Successive layers of the G-quartets stack on top of each other and form the structure. The quadruplexes can be formed by all-parallel and antiparallel orientations of the four DNA strands forming the structure. The formation of the structure exhibits a remarkable specificity with respect to the cation (potassium), which occupies the central cavity formed by the layers of the guanosine tetrads. The stability of the structure suggests its involvement in natural mechanisms. In supercoiled DNA, some sequences can form an electric structure including both H-DNA and a quadruplex structure [23]. Circular, covalently closed DNA and DNA loops with fixed ends in chromatin can be under a negative supercoiling stress when transcription occurs, and this promotes the formation of the above-mentioned DNA forms such as Z-DNA, H-DNA and some other structures with open, single-stranded DNA regions in physiological conditions [27–29].

It should be mentioned that there is an enzymatically mediated mechanism for the formation of triple-stranded complexes between double-stranded DNA and single-stranded oligo- and polynucleotides that are different from the complexes described above. The complexes are formed between DNA and an oligonucleotide or a polynucleotide which is complementary to one of the DNA strands. Formation of this type of complex occurs naturally within homologous genomes in the course of an important process, such as genetic recombination and produces heteroduplex joints that contain one strand from each of the parent molecules. Formation of the complex is assisted by special cellular proteins, namely recombination enzymes, and the exact mechanism of the binding between the polynucleotides has not yet been elucidated [30].

A number of unusual single-stranded and multistranded DNA and RNA structures have been shown to be formed in acidic conditions in the presence of different cations [31, 32]. For example, a four-stranded structure formed by antiparallel interlocking of protonated cytosine–cytosine parallel-stranded duplexes ('i-motif') is produced at acidic pH [32]. In this unusual structure, mutual intercalation of parallel-stranded C–C⁺ duplexes occurs to form a tetrad motif. It has been found that in certain conditions parallel-stranded DNA helices can be formed by some nucleotide sequences. This unusual structure is established by reverse Watson–Crick base pairing of dAdT with hydrogen bonds between N6H–O2 and N1–HN3 [33]. The biological role of these structures and their potential applications remain to be investigated.

The considered interactions between nucleic acid strands represent the basis for targeting nucleic acids with oligonucleotides. Watson–Crick binding allows targeting of single-stranded regions in RNA and open sequences in locally unwound DNA. The native double-stranded DNA can be targeted by oligonucleotides at homopurine–homopyrimidine sequences through triple-stranded complex formation.

3. OLIGONUCLEOTIDE DERIVATIVES FOR TARGETING NUCLEIC ACIDS

To achieve fully the goals of using oligonucleotides as potentially biologically active compounds and recognition elements in various techniques designed for manipulating nucleic acids and proteins, it is necessary to develop approaches for the conjugation of oligonucleotides to different compounds and to design modified versions of oligonucleotides which retain the ability to form specific complexes with the target nucleic acids. In this section, the main types of oligonucleotide analogues and the approaches for derivatization of oligonucleotides are overviewed. For a detailed consideration of the techniques and an explicit list of publications, the reader is referred to a few excellent recent reviews [3–5, 13–15].

3.1 Oligonucleotide Analogues

Oligonucleoside methylphosphonates and alkyl phosphotriesters of oligonucleotides are used as nonionic analogues of oligonucleotides [11, 34]. In these compounds, the negatively charged phosphate oxygen of the internucleotide bond is replaced by electroneutral functions (Figure 6a). The nonionic oligomers have been synthesized with either 2'-deoxyribonucleosides or 2'-*O*-methylribonucleosides. The compounds are resistant to nucleases and penetrate efficiently into mammalian cells. The lack of electrostatic repulsion between the oligonucleotide backbone and the complementary nucleic acid strand facilitates their association at low salt concentrations. Although the compounds are quite

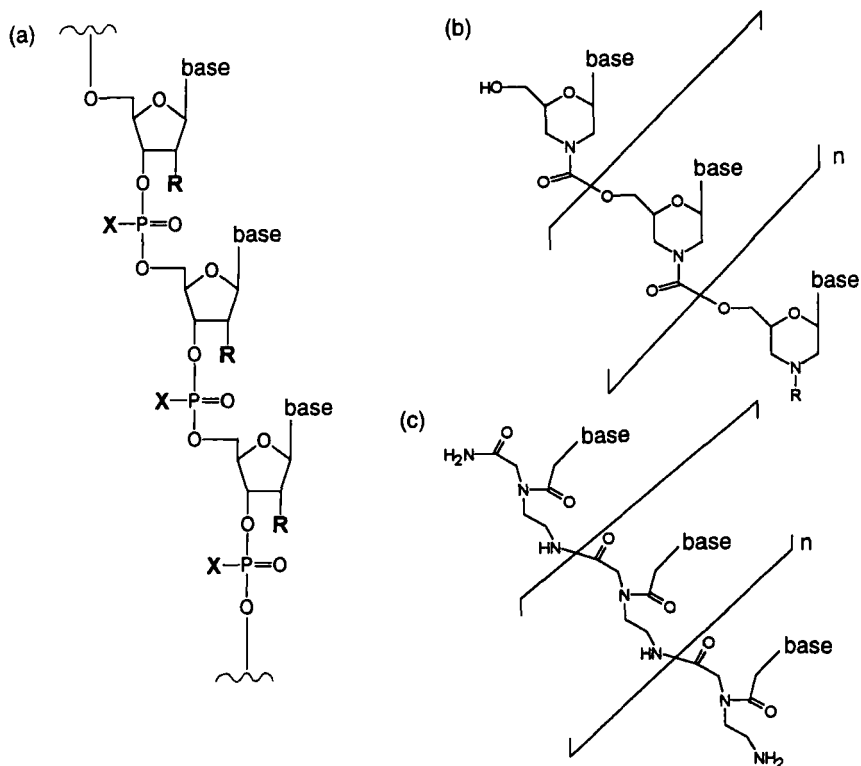


Figure 6 (a) Oligonucleotide analogues with modified phosphates and ribose: (X = O, natural phosphodiester oligonucleotide) X = Me, oligonucleotide methylphosphonate; X = EtO, oligonucleotide ethyl phosphotriester; X = S, phosphorothioate oligonucleotide analogue; X = R₂N, phosphoramidate; R = MeO, 2'-O-methylriboooligonucleotide. (b, c) Oligonucleotide analogues with a nonnatural backbone: (b) a morpholino-oligonucleotide and (c) an oligonucleotide with a polyamide backbone

hydrophobic, they can be dissolved in aqueous buffers to give concentrations as high as 0.5 mM. Oligodeoxypurine and oligodeoxypyrimidine methylphosphonates can bind to the corresponding homopurine-homopyrimidine sequences in double-stranded DNA to form triple-stranded complexes [34, 35]. Oligonucleoside methylphosphonates are studied extensively and are considered as promising compounds for *in vivo* applications.

The methylphosphonate linkage is chiral and exists in either the (*R*)_P- or (*S*)_P-configuration at phosphorus. Therefore, the oligomers synthesized by traditional methods represent mixtures of 2ⁿ diastereomers, where *n* is the number of chiral phosphorus atoms. These mixed diastereomers bind to complementary nucleotide sequences, and at low salt concentrations the stability of the complexes formed is similar to that of the complexes formed by the phosphodiester oligonucleotides. At high salt concentrations, the mixed complexes are less stable than similar complexes of phosphodiester oligonucleotides.

Experiments with oligonucleotides built of individual diastereomers have shown that the (*R*)_p-isomers form duplexes and triplexes with DNA of higher stability than the corresponding complexes with the (*S*)_p-isomers. Therefore, attempts have been made to develop synthetic approaches for preparing oligonucleotides with stereochemically defined internucleotide methylphosphonate bonds.

Phosphorothioate oligonucleotide analogues [36, 37] (Figure 6a) are closest to the natural nucleic acids. In these analogues, one of the nonbridging phosphate oxygen atoms in the internucleotide linkage is replaced by a sulfur atom, with a negative charge located mainly on sulfur. The compounds are nuclease resistant. They are taken up by mammalian cells with reasonable efficiency similar to that observed for the natural oligonucleotides. Phosphorothioate oligonucleotides have received much attention because of their antiviral activity. The internucleotide bond in these analogues is chiral; however, it affects their binding properties only slightly. To avoid the chirality problem associated with the presence of the asymmetric phosphorothioate center, phosphorodithioate oligonucleotide derivatives have been prepared [39]. The stability of the complementary complexes formed by the analogues decreases in the order unmodified oligonucleotides > phosphorothioates > phosphorodithioates.

Oligonucleotides containing various phosphoramidate linkages have been synthesized and shown to form complementary complexes with nucleic acids. Phosphoramidates derived from primary amines produce the more stable complexes compared to secondary amine phosphoramidates [40]. The presence of a few phosphoramidate linkages in mixed phosphodiester–phosphoramidate oligonucleotide derivatives decreases the stability of the complementary complexes only slightly. These analogues offer a great opportunity for structural variations through the use of various substituents on the amine structure.

Achiral oligonucleotide analogues have been synthesized in which the oxygens in the –OPO–ester group are substituted by –NH–, –CH– or –S– functions [2–5]. The introduction of a single 3'-OP(O)(O)NH-5' link into the backbone of the oligomer does not affect the hybridization properties of the oligonucleotide. A few phosphoramidate linkages considerably decrease the ability of an oligonucleotide to form complementary complexes.

The phosphodiester internucleotide connection has been replaced with a conformationally restricted acetal linkage [41]. It was expected that restricting the conformational freedom of the unbound oligonucleotide to resemble the bound conformation would increase the stability of the complementary complex of the oligonucleotide. Indeed, it has been shown that pyrimidine oligonucleotides with alternating phosphodiester and riboacetal linkages display a considerable increase in affinity for a double-stranded DNA target as compared to a control all-phosphodiester oligonucleotide.

Oligonucleotide analogues have been prepared in which the phosphodiester bonds are partially replaced by electroneutral 3'-CH₂NH(Me)OCH₂-5' linkages [42], and it has been shown that partial substitution of the natural

linkages by artificial ones does not affect the stability of the complementary complexes of the oligonucleotides.

Of the oligonucleotide analogues with modified sugars, the 2'-*O*-alkylated oligonucleotides are the most widely used [43]. The stability of the complementary complexes formed by these analogues with RNA is higher than that of the complexes formed by the parent oligonucleotides. The stability of the double-stranded complexes formed by the oligonucleotide analogues decreases in the order 2'-fluoro > 2'-*O*-methyl > 2'-*O*-propyl > 2'-*O*-pentyl, 2'-deoxy [44]. The analogues can also form triplexes with the corresponding sequences in double-stranded DNA.

Oligonucleotide analogues have been built from nucleotides with an anomeric α -configuration of the glycosidic bond (α -oligonucleotides) [45]. These compounds are resistant to nucleases. α -Oligonucleotides bind to the complementary sequences in single-stranded nucleic acids in a parallel orientation. α -Pyrimidine oligomers can bind to the corresponding homopurine-homopyrimidine sequences in DNA to form triple-stranded complexes in which the oligonucleotide strand is antiparallel to the DNA purine strand [46]. The stability of triplexes formed by DNA and α -anomeric oligomers is reduced as compared to the normal β -anomeric oligonucleotides.

Polyamide oligomers, termed peptide nucleic acids (PNAs) [47], have been synthesized in which the entire sugar phosphate backbone is replaced by *N*-ethylaminoglycine residues bearing heterocyclic bases (Figure 6c). The analogues form very stable complexes with complementary sequences in single-stranded RNA and DNA and can form triple-stranded complexes with homopurine single-stranded DNA (two PNA strands and one DNA strand). The unrestrained polyamide backbone of the PNA permits binding of the oligomers to the nucleic acid target in orientation-independent manner. Interestingly, the specificity of the double-stranded complex formation between PNA and single-stranded nucleic acids is as high as that determined for the natural nucleic acids [48]. Since the oligomers are electroneutral, they bind to nucleic acids efficiently in the absence of salts, where natural complementary complexes are destabilized. Therefore, peptide oligonucleotide analogues are capable of invading duplex DNA at a specific target site at low salt concentrations.

A number of artificial and noncanonical nucleoside bases have been introduced into oligonucleotides to modulate the stability of the complexes and the specificity of recognition in different conditions. Although synthetic approaches to introducing substituents in all positions of heterocyclic bases have been developed, only a few of them can be modified without interfering in the essential hydrogen bonding needed for complementary interactions. In this section only those modifications which do not interfere with the recognition properties of the monomers are considered. A comprehensive listing of analogues of pyrimidines and purines that have been incorporated into oligonucleotides can be found elsewhere [49].

Modifications at the C5-position or C6-position of pyrimidines remain in the major groove of the duplex and do not affect the formation of complementary complexes. Even bulky groups introduced at C5 are tolerated, and this is one of the best sites for tethering various functionalities to oligonucleotides. Another site for the conjugation of different groups is the 4-amino group of cytidine, which may be subjected to transamination with different amines. Linkers with amino groups have been introduced at the N4-position of cytosine, allowing the subsequent linking of various functionalities such as lanthanide chelates, biotin and various fluorescent labels and haptens for immunodetection. Reactive groups have also been introduced by functionalization of the C8-position of adenosine.

Modifications of the heterocyclic bases have been used to enhance the stability of oligonucleotide binding. Thus, C5-halogenated and C5-propyne analogues of pyrimidines form more stable complementary Watson–Crick base pairs than the parent unmodified pyrimidines. Oligonucleotides containing 2-aminoadenosine show increased strength of hybridization because the 2-amino group of the base creates a third hydrogen bond when paired with thymidine. These residues have been incorporated into oligonucleotides to increase the stability of the complementary complexes [50, 51].

To reduce the oligonucleotide net charge, pyrimidine 5- ω -aminohexyl residues have been introduced into the oligomers to neutralize the phosphate negative charge [52]. The resultant zwitterionic oligonucleotides bind more strongly to the complementary sequences than unmodified oligomers of the same sequence and the stability of the complexes is relatively insensitive to changes in ionic strength.

Considerable efforts have been made to design nucleosides with modified and artificial bases for the development of oligonucleotides that can target arbitrary sequences in dsDNA under physiological conditions. Replacement of cytosine with 5-methylcytosine in pyrimidine oligonucleotides reduces the pH effect on the stability of the triplexes [53]. To avoid the pH dependence of the triplex formation completely, cytosine was replaced by synthetic analogues that do not require protonation in order to form Hoogsteen hydrogen bonds with guanine. It has been shown that 7,8-dihydro-8-oxoadenine [54], pseudisocytidine [55] and 3-methyl-5-amino-1*H*-pyrazolo[4,3-*d*]pyrimidin-7-one [56] can replace cytosine and form triplexes which are stable at physiological pH.

Attempts have been made to design synthetic heterocycles that can bind specifically to different base pairs in double-stranded DNA. 1-(2-Deoxy- β -D-ribofuranosyl)-4-(3-benzamidophenyl)imidazole (D3) has been introduced into pyrimidine oligonucleotides to recognize T and C pyrimidine residues in both CG and TA base pairs of double-stranded DNA (the pyrimidine residues to be recognized were in the purine strand) [57]. The stabilities of the base triplets were found to decrease in the order D3–TA, D3–CG > D3–AT > D3–GC. 2-Deoxynebularine has been introduced into oligonucleotides to recognize both C and A residues in the Watson–Crick CG and AT base pairs, respectively [58].

7-Deaza-2'-deoxyxanthosine (X) has been used as a recognition analogue of thymidine which forms X-AT triplets [59].

3.2 Attachment of Functional Groups to Oligonucleotides

A number of approaches for derivatization of oligonucleotides by tethering various functionalities have been developed [4, 33, 34, 60, 61]. Conjugates have been designed to endow oligonucleotides with some new physical and chemical properties while retaining their ability to base pair. A great number of reactive groups have been conjugated to oligonucleotides to produce affinity reagents targeted to nucleic acids (see Section 6). Fluorescent and chemiluminescent groups are conjugated to oligonucleotides to prepare physical probes. Various groups have been tethered to oligonucleotides to modulate the specificity of their interactions with proteins and complementary nucleic acids, to target them to specific receptors and to protect them against nucleolytic enzymes in order to enhance the efficiency of antisense oligonucleotides *in vivo*.

Various functionalities can be built into the oligonucleotides directly in the course of the oligonucleotide synthesis using synthons carrying the necessary modifications at specific steps of the oligonucleotide elongation. During the synthesis, molecules other than nucleotides can be introduced into the growing oligonucleotide chain in a specific position. These artificial monomers can be equipped with the desired functionalities.

Other ways of introducing functionalities are based on chemical modification of some groups present naturally in oligonucleotides or modification of unique, reactive linker groups introduced into the oligomers at the desired sites in the course of their synthesis. The naturally occurring groups which can be modified in oligonucleotides include the amino groups of the bases, hydroxy groups of sugars and phosphate groups. Artificially introduced reactive groups include aliphatic amino groups and thiols, although there are some other possibilities. The amino groups can be used as a target for selective acylation with *N*-hydroxy-succinimide esters or anhydrides. Thiols can be alkylated in mild conditions by various electrophilic reagents. The reactive groups can be attached to oligonucleotides via convenient linkers, allowing regulation of the distance between the oligonucleotide and the reactive group to provide conformational freedom of the conjugated functionalities when needed.

Terminal phosphates or thiophosphates, which can be introduced into oligonucleotides chemically or enzymatically, can be used for the attachment of various bulky groups without affecting the hybridization properties of the oligonucleotides. The phosphate group can be activated, allowing the phosphorylation of primary amino groups in various molecules. Using this method, many groups and linkers with unique reactive groups have been conjugated to the terminal phosphates of oligonucleotides [61]. The thiophosphate can be alkylated with various compounds containing 2-chloro-ethylamino or iodoacetamido groups. Aliphatic amino groups can be tethered

to the terminal phosphate using polymethylenediamines as linker structures. Acylation of these groups can be used to attach various reporter and reactive groups to oligonucleotide termini.

The 2',3'-*cis*-diol group at the end of ribooligonucleotides and mixed deoxyribo-ribooligonucleotides is a unique site allowing the attachment of molecules with the benzaldehyde moiety by the formation of a cyclic acetal. Using this reaction, the alkylating agent *p*-(*N*-2-chloroethyl-*N*-methylamino)benzaldehyde has been attached to oligonucleotides [8].

The 3'-terminal ribonucleoside can be oxidized by NaIO_4 to give dialdehydes. The latter can be reacted with primary amino groups to form Schiff bases, which can be reduced with cyanoborohydride into morpholine structures that are stable under physiological conditions. An aldehyde group can be generated also at the 5'-end of oligonucleotides by periodate oxidation of a vicinal diol introduced during synthesis. A few approaches for conjugation to internucleoside phosphates have been developed. The most versatile method takes advantage of easy conjugation through phosphoramidates.

The length of the spacer between the functional group and the oligonucleotide can be important for providing optimal interaction between the oligonucleotide conjugate and the target. Thus, conjugated intercalating moieties should have a certain freedom for optimal interaction with the nucleic acid helix. The conjugated reactive groups should be capable of reaching potentially reactive sites of the target. For bulky groups such as fluorescent reporter groups or biotin, long linkers are needed to minimize the interference of these functions with hybridization of the oligonucleotide. Usually, no problems with oligonucleotide hybridization appear if the linker structure is long enough and the groups are conjugated to oligonucleotide termini or to heterocyclic bases at positions which are not involved in hydrogen bonding. Thus, even large protein molecules can be conjugated to oligonucleotides to form constructions capable of recognizing complementary sequences. Oligonucleotides have been conjugated to staphylococcal nuclease and to the catalytically active fragment of ribonuclease A (RNase S) in order to produce sequence-specific nucleic-acid-cleaving constructions. To accomplish the conjugation, Cys116 was substituted for Lys116 in staphylococcal nuclease by means of oligonucleotide-directed mutagenesis. An oligonucleotide carrying a 3'-terminal thiol was attached to this fragment via a disulfide bridge [62]. A similar approach has been used for conjugation to oligonucleotides of RNase S [63].

4. TARGETING OF SPECIFIC SEQUENCES IN RNA AND DNA WITH OLIGONUCLEOTIDE DERIVATIVES

Targeting of nucleic acids can be achieved by using oligonucleotides capable of recognizing specific single-stranded nucleotide sequences or the sequences arranged in more complex structures as described in Section 2. For efficient

targeting an appropriate target site in the nucleic acid should be selected by taking into account thermodynamic and kinetic parameters.

4.1 Kinetics and Thermodynamics of the Oligonucleotide–Polynucleotide Interaction

Single-stranded targets for oligonucleotides are RNAs which have complex and dynamic structures with open single-stranded sequences (loops) and double-stranded regions. The affinity of oligonucleotides for different sequences of RNA depends on their involvement in the folding of the molecule. The affinity of oligonucleotides for their complements is usually assessed using spectrophotometric techniques because complementary complex formation results in a decrease in absorbance. A value useful for determining the stability of oligonucleotide complexes in certain conditions is the melting temperature of the complex T_m , which is the temperature at which half of the molecules are duplex and half are single stranded. T_m depends on the structures of the interacting complementary sequences and their concentrations.

The stability and kinetics of formation of double-stranded and triple-stranded complexes of oligonucleotides are affected by the lengths and folding patterns of the interacting nucleotide sequences, the pH of the solution and the nature of the cations present. The stability of nucleic acid duplexes increases with increasing length of the oligonucleotide and increasing GC content of the interacting complementary sequences, in the latter case because of the higher stability of the GC base pair compared to the AT (AU) pair. The duplex stability is affected to some extent by the sequence of the duplex through the stacking interactions of the neighboring heterocyclic bases. The stability of heteroduplexes formed by specific RNA and DNA sequences may vary to a considerable extent. Thus oligo(dA)–oligo(rU) complexes are much less stable than oligo(dT)–oligo(dA) complexes because of the tendency of oligo(dA) clusters to assume the B-type conformation, whereas oligo(rU) can adopt only the A-type conformation. The mispaired bases or extra nucleotides in the central part of the interacting sequences cause a considerable destabilization of the duplexes. By contrast, terminal mispaired bases cause only a minor destabilization [64, 65].

Binding of short oligonucleotides [66–69] to unstructured complementary sequences does not depend on the oligonucleotide length or sequence and is determined entirely by the structure of the target. Typical association rate constants for oligonucleotides interacting with unstructured single-stranded targets vary in the range 10^6 – $10^7 \text{ M}^{-1} \text{ s}^{-1}$. Binding to folded nucleic acids occurs much less rapidly. Dissociation rates depend on the oligonucleotide sequence and length and are relatively unaffected by the target structure. All the practically important natural RNA targets form compact structures through intramolecular complementary interactions. Usually more than 50% of the RNA sequences are involved in the formation of the intramolecular double-stranded

regions. Secondary folding of nucleic acids can hide the sequences needed for oligonucleotide binding and interfere with the process for both kinetic and thermodynamic reasons [66–69]. Kinetically, the target involved in intramolecular base pairing may not dissociate long enough for the oligonucleotide to bind to it. Therefore folding of the target decreases the association rate. If the oligonucleotide and the target sequence can form intramolecular hairpins, the equilibrium constant of duplex formation is reduced by the equilibrium constant for hairpin formation [68]. The tendency of oligonucleotides to stem loop structures can be evaluated approximately as the sum of the free energies necessary to disrupt the secondary structure and hybridize the loop to the complementary oligonucleotide. Association rate constants for hybridization to structured regions range between 10 and $10^7 \text{ M}^{-1} \text{ s}^{-1}$ [67–69]. The result is that sequences in the loops of the nucleic acid structure bind oligonucleotides as strongly as the single-stranded oligonucleotides, while binding to stem sequences is much weaker or completely arrested. Binding of oligonucleotides to the loops depends considerably on the position of the complementary sequence in the loop. Oligonucleotides targeted to the 5'-end of the loop can bind to the target even more tightly than to an unfolded single strand, while oligonucleotides targeted to the 3'-side of the loop bind much less tightly to the loop than to a similar unfolded target. This effect can be explained in terms of the structure of the loop, which has stacked residues on the 5'-side that favor oligonucleotide binding and interfere with binding to the 3'-side, which requires restructuring of the loop [66].

From the above-mentioned values for the association rate constants it may be concluded that at micromolar concentrations oligonucleotides bind to complementary nucleic acids fairly rapidly. At subnanomolar concentrations, which are likely to be achieved when the compounds are introduced into an organism, the complex formation will require a few hours for oligonucleotides which bind to targets with a rate constant of $10^6 \text{ M}^{-1} \text{ s}^{-1}$ for the freely accessible target. If the target is structured and binding occurs more slowly, higher oligonucleotide concentrations are required to achieve hybridization in a reasonable time [66].

Most physicochemical studies with the triple-stranded complexes have been performed with pyrimidine oligonucleotides and the corresponding polypurine–polypyrimidine DNA sequences. The rates of formation and dissociation of the triple-stranded complexes are two orders of magnitude slower than those of the double-stranded complexes. In the case of the pyrimidine oligonucleotides containing cytidine residues, the process depends on the pH of the medium. It takes many hours to reach equilibrium in the process of triple-stranded complex formation under physiological conditions and at submicromolar concentrations of oligomers [70]. The triplex stability depends strongly on salt concentration. In the absence of magnesium ions and polyamines, the triplex melting temperature is linearly dependent on sodium ion concentration in the range 0.2–1 M. An increase in NaCl concentration from 0.02 to 0.3 M leads

to a considerable increase in the association rate constant. The dissociation rate constant is not affected by changes in NaCl concentration. The triple-stranded DNA with a third pyrimidine strand is stabilized considerably by magnesium ions and by the polyamines spermine and spermidine [71]. The presence of Mg^{2+} ions decreases the dissociation rate constant and considerably increases the association rate constant (by a factor of 1500 in 40 mM $MgCl_2$). Mismatches in the interacting oligomers affect considerably the dissociation rate constant. The stability of the hybrid triplexes formed by different combinations of RNA and DNA strands varies considerably for different combinations of the constituents. For increasing stability of the complexes, RNA is favored on both pyrimidine strands, whereas DNA is favored in the purine strand. It has been noticed that relatively unstable duplexes form more stable triplexes and vice versa [24]. In the purine motif triple-stranded complexes, A-AT triplets are more stable than G-GC triplets and Pu-PuPy triplex formation is independent of pH. Pu-PuPy base triplets are greatly stabilized by bivalent cations, especially Mn^{2+} , Cd^{2+} , Co^{2+} , Ni^{2+} and Zn^{2+} [22, 23].

4.2 Specificity of the Interaction of Oligonucleotide Derivatives with Nucleic Acids

When targeting a specific nucleotide sequence in cellular nucleic acids, scanning of databases is necessary to ensure that there are no matches in nontarget nucleic acids. The specificity of the site-specific recognition of nucleic acids by oligonucleotides depends on the relative stabilities of the perfect complex and the imperfect complexes formed at sites similar to the target [68, 72, 73]. Specific effects can be caused if most of the target sites are hybridized and the occupancy of nontarget sites is small.

Ideally, the oligonucleotide should interact only with the target nucleic acid. Statistically, a 17 nucleotides long sequence should be selected to achieve unique site targeting in a genome of the size of the human genome. If one considers the nonrandom nature of natural nucleic acid sequences and that the target is one of the mRNAs which represent a small portion of the genomic sequences, then one can assume that oligonucleotides containing as few as 11 nucleotides can be used as specific targeting structures [13]. Considering the length of the oligonucleotide address, it can be concluded that for specific targeting some optimal length of the oligonucleotide is required. On the one hand, the longer the sequence, the higher the probability of targeting a unique sequence within the nucleic acid mixture. On the other hand, even if only one true target exists in the system, other nucleic acids may still be affected by a long oligonucleotide because it is possible that it will bind to sequences of partial complementarity which will pair to only a portion of the oligonucleotide or form complexes with mismatches. These complexes will be less stable than the perfect complexes but they may be numerous enough to produce some side effects. Thermodynamic calculations [74] show that maximum specificity of oligonucleotide binding

occurs at the melting temperature, when half of the target is hybridized to the complementary oligonucleotide. Therefore, in the case of oligonucleotide probes used *in vitro*, high specificity is achieved by performing hybridization under 'stringent' conditions, i.e. at a temperature close to T_m , where mismatched complexes are dissociated. *In vivo* hybridization conditions cannot be controlled and only the oligonucleotide length and modifications can be varied to optimize the specificity.

Different types of imperfect complexes formed by oligonucleotides and their targets (Figure 7) are very different with respect to stability. The penalty for a single pyrimidine bulge in duplex RNA and DNA ranges from 2.8 to 3.5 kcal mol⁻¹. Nonspecificity because of mismatches in triplexes incurs essentially the same penalty [64]. The relative stability of triplexes decreases in the order perfect > single bulge > single mismatch > double mismatch. An abasic site in the Hoogsteen-pairing strand destabilizes the triplex as much as a mismatch. Mismatches at the termini of oligonucleotides cause only minor destabilization. It has been concluded that nonspecificity of oligonucleotide interactions arises primarily through the binding at the ends of the complexes and not because of the imperfect pairing in the center of the complex [64, 65].

The kinetics of interaction can modulate considerably the specificity of the effects of oligonucleotides when the binding results in some fast, irreversible modifications. Thus, results of experiments with ribozymes have shown that lengthening of the recognition sequence of the ribozyme does not increase the specificity of cleavage [75]. Although mismatched targets dissociated faster than the correct targets, the cleavage was too fast compared to the dissociation rate and the correct and incorrect targets were not distinguished well. Therefore, even incorrectly bound ribozymes cleaved the RNA before they dissociated, which reduced the specificity. Similar effects can be expected in all cases where the limiting event is binding and the inactivating event is fast: in the case of antisense oligonucleotides where the inactivation mechanism involves fast cleavage of the target RNA with the enzymes ribonuclease H and ribonuclease L, and in the case of derivatives of oligonucleotides which react with the targets rapidly when the complex is formed.

A few approaches have been proposed for modulation of the specificity and strength of the oligonucleotide–nucleic acid interaction. Conjugation of oligonucleotides to groups capable of noncovalent binding to nucleic acids, e.g. intercalating groups, renders the compounds capable of tighter binding to complementary sequences than the parent oligonucleotides. Intercalating groups such as polyaromatic agents, acridine and phenazinium and ethidium dyes (Figure 8) [13] form stable complexes with double-stranded nucleic acids. Being attached to an oligonucleotide, they increase the stability of the complementary complexes formed by the oligomer. Conjugation to these groups results in a T_m increase equivalent to elongation of the oligonucleotide by two or three nucleotide units. Conjugation to intercalating groups is a useful approach when a short nucleotide sequence needs to be targeted for some reason. It should be

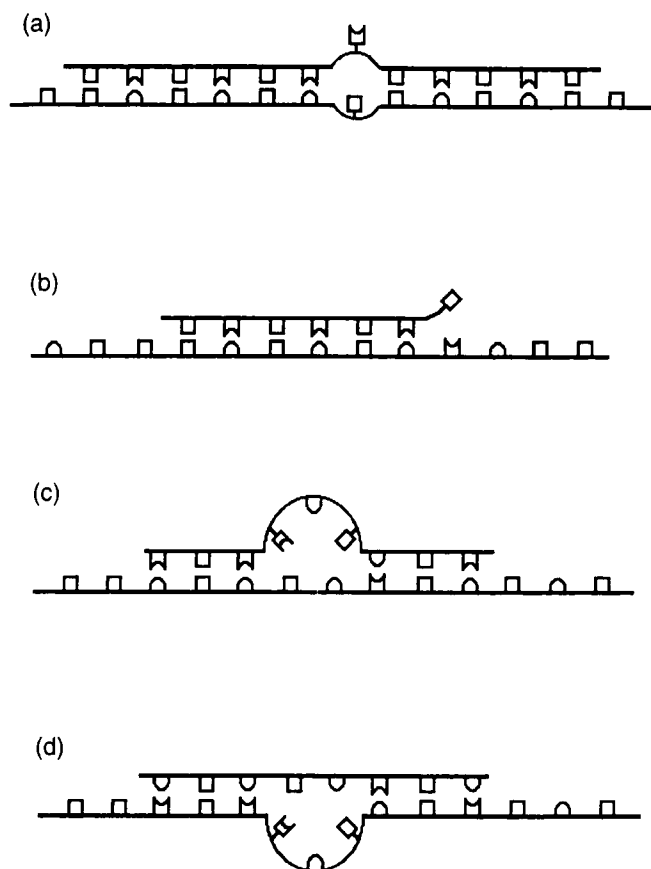


Figure 7 Imperfect complexes of oligonucleotides: (a) with an internal mismatch; (b) with a noncomplementary base at the terminus; and (c, d) with bulges in the oligonucleotide strand and in the target strand, respectively

noted that the stabilization by intercalating groups results from a sequence-independent interaction of the groups with the nucleic acid. This means that the specificity of targeting for an oligonucleotide with an intercalating group will be lower as compared to that of a longer, unmodified oligonucleotide having the same binding energy.

The specificity of targeting can be increased by using tandems of oligonucleotides (Figure 9) [76, 77]. It is known that oligonucleotides complementary to adjacent sequences in a nucleic acid bind to the target cooperatively through stacking interactions between the terminal bases of the oligonucleotides. These interactions provide additional energy which is equivalent to elongation of the oligonucleotide by two or three nucleotide residues. A similar phenomenon has been observed for oligonucleotides binding to double-stranded DNA [78].

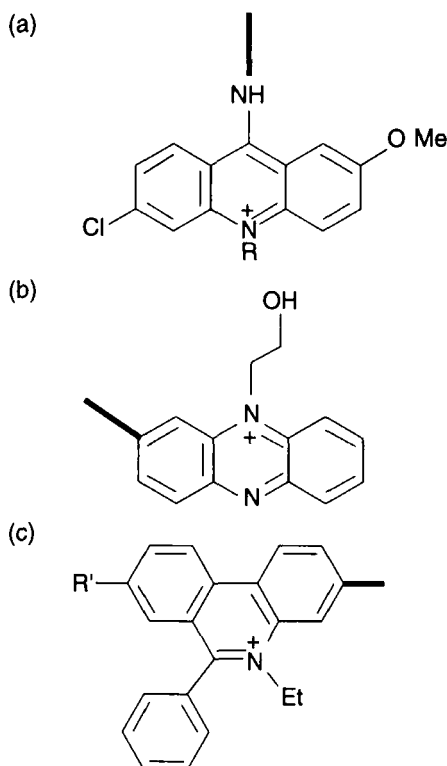


Figure 8 Intercalating groups which stabilize complementary complexes of oligonucleotides. The thick lines indicate sites of connection to linkers

In this case, the binding may also contribute to stabilization of the DNA conformation needed for the formation of the triple-stranded complex. When one of the oligonucleotides is equipped with an intercalating group, the end-to-end oligonucleotide interaction is strengthened and brings additional stability to the complexes [76]. It should be noted that the stabilizing effect achieved using oligonucleotide tandems is sequence specific because it is provided by a specific interaction of the neighboring oligonucleotide. The cooperative effect of tandem oligonucleotide binding has been employed to modulate the specificity of reaction for alkylating derivatives of oligonucleotides with single-stranded nucleic acids [77, 78]. This effect has also been used to modulate the efficiency of ribozyme action [79].

A strong, noncovalent binding of adjacent oligonucleotides is achieved when their terminal overhanging sequences can form a complementary complex (Figure 9). Because of this interaction, the two oligonucleotides bind to the target sequence as tightly as a single, long oligonucleotide of equivalent length [80]. The so-formed double-stranded zipper can be additionally stabilized by antibiotics capable of specific binding to sequences included in the zipper [81].

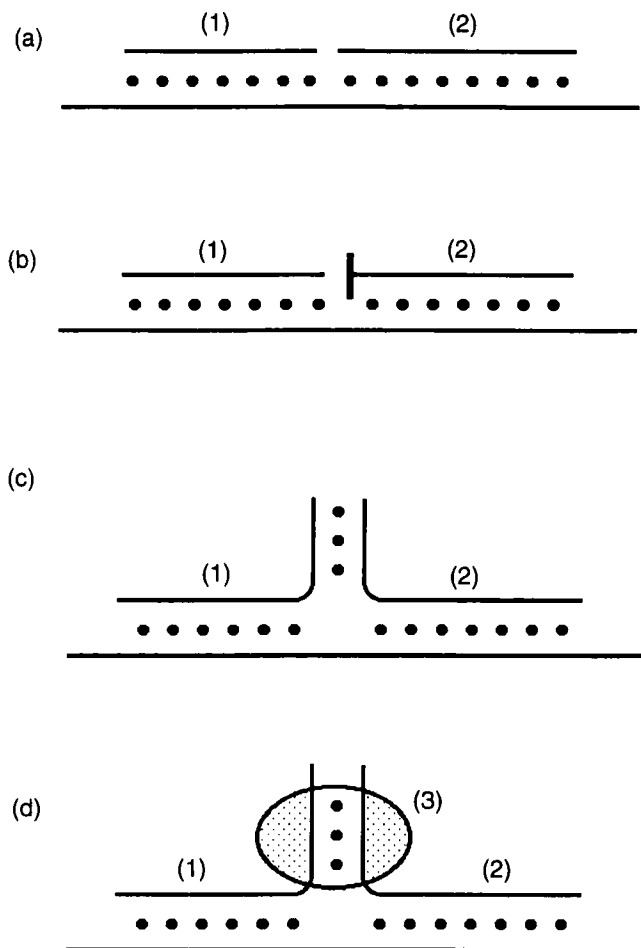


Figure 9 Schematic diagrams of cooperative oligonucleotide binding to target nucleic acids: (a) an oligonucleotide tandem which binds to juxtaposed nucleotide sequences cooperatively through end-to-end stacking interactions; (b) an oligonucleotide tandem stabilized by conjugation of one of the oligonucleotides to an intercalating group (shown by the thick vertical line); (c) an oligonucleotide tandem with complementary overhanging sequences; and (d) an oligonucleotide tandem with complementary overhanging sequences stabilized additionally by echinomycin (3) interacting with an AT-rich sequence in the helix formed by the overhangs

4.3 Selection of Target Sites and Oligonucleotide Addresses

Optimization of the target sites starts with analysis of the structure of the target to find the most accessible sequences where oligonucleotide binding would result in the formation of the maximal total number of binding interactions in the complex, including base pairs between the oligonucleotide and target and new base pairs in the rearranged target.

The currently available methods do not allow an accurate prediction of the secondary structure of large RNAs because of the great number of alternative secondary structures of the molecules that have similar free energy. Another complication is that in the cell, nucleic acids are always associated with proteins, and this also affects the secondary structure and availability of certain regions of the RNA to oligonucleotides. However, stable hairpins can usually be identified. Determination of the accessibility of specific sequences in nucleic acids to chemical and enzymatic probes allows us to choose between the structures and to find open regions. Juxtaposed residues of RNA in the folded structure can be determined using crosslinking with bifunctional chemical reagents. It is important that the chemical modification experiments can be performed *in vivo*; they should also provide information about the native conformation of the nucleic acid which is assumed by the RNA in the presence of proteins.

Special approaches are used to target some elements of RNA structure. In the heavily structured targets, only small loops are available for oligonucleotide binding. Therefore, the oligonucleotide analogues and derivatives with the highest affinity are used in this case, e.g. oligonucleotides with conjugated intercalating groups or oligonucleotide analogues that form complexes with enhanced stability. The possibility of forming triple-stranded complexes using oligonucleotides and elements of the RNA secondary structure has been exploited to give tight oligonucleotide complexes with specific RNA targets. Oligonucleotide clamps have been designed [82] which are covalently linked oligonucleotides capable of forming triple-stranded complexes with single-stranded RNA. One of the oligomers forms the Watson–Crick base pair with the target while the second oligomer engages in Hoogsteen base pairing (Figure 10a). The constructs form very tight complexes with RNA. The oligonucleotides can be connected by a synthetic linker or by a short nucleotide sequence. Intercalating groups can be placed at the oligonucleotide termini to stabilize the assembled triple-helix structure [82]. A strategy for targeting some hairpin structures in RNA has been developed [83]. The idea was to use oligomers containing two domains: one complementary to the single-stranded purine sequence at the bottom of the hairpin, and the second domain designed to form a triplex with both the hybridized first domain and the homopurine–homopyrimidine stem of the native RNA hairpin (Figure 10b). Apparently, the use of selection methods (Section 5.5) opens up to the possibility of developing oligonucleotides to target any element of nucleic acid structure.

Double-stranded DNA is relatively rigid and no steric complications appear when targeting any of the double-stranded sequences. The problem is that at present only the homopurine–homopyrimidine sequences of the DNA can be targeted. Attempts have been made to overcome the restrictive requirement for an all-pyrimidine sequence in one DNA strand. One of the approaches is to use two or three oligopyrimidine domains having appropriate 3'-3' or 5'-5' junctions between them to target juxtaposed, short oligopurine–oligopyrimidine stretches

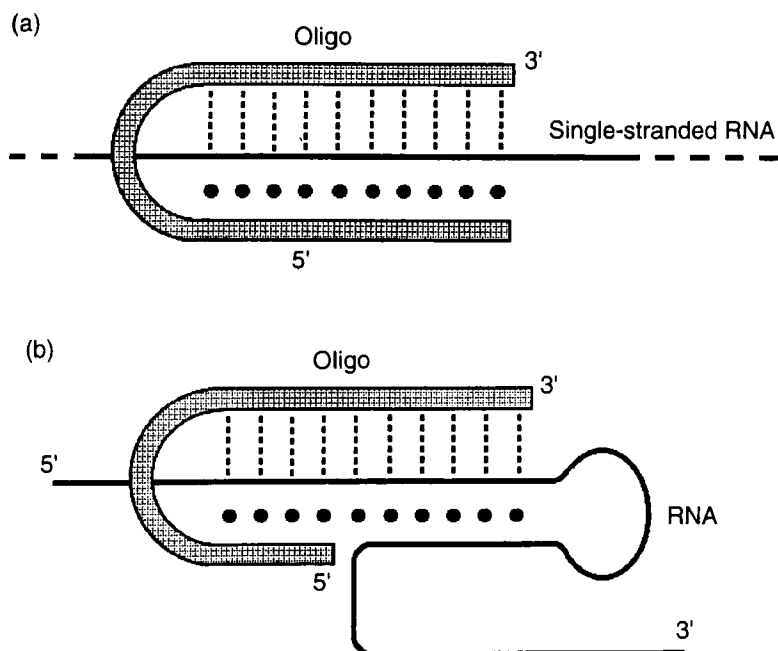


Figure 10 Targeting of single-stranded RNA by triple-stranded structure formation: (a) targeting of a homopurine sequence with a pyrimidine oligonucleotide construction which forms a duplex with the target sequence and folds back to form a triple-stranded structure [82]; and (b) targeting of the homopurine-homopyrimidine stem of a hairpin structure adjacent to a homopurine sequence. The pyrimidine oligonucleotide forms a duplex with the homopurine sequence and folds back to form a triple-stranded complex with the duplex and the RNA stem

in DNA (Figure 11) [84]. These linked domains are able to form cooperative triplexes with double-stranded DNA fragments. This multidomain approach may be developed further for the more general design of structures recognizing double-helix DNA by triple-helix formation. Neighboring oligopurine and oligopyrimidine sequences have also been targeted by oligonucleotides with normal phosphodiester linkages composed of two consecutive purine and pyrimidine blocks [85]. In the complex, both purine and pyrimidine recognition motifs are used for recognition of the blocks by both the DNA strands.

5. OLIGONUCLEOTIDE-BASED TECHNIQUES AND MOLECULAR DEVICES

Since oligonucleotides have become molecules that can be manipulated easily and reliably, they have found a variety of applications in different techniques and molecular devices where they are used as recognition elements. Although most of the current applications of oligonucleotides take advantage of the

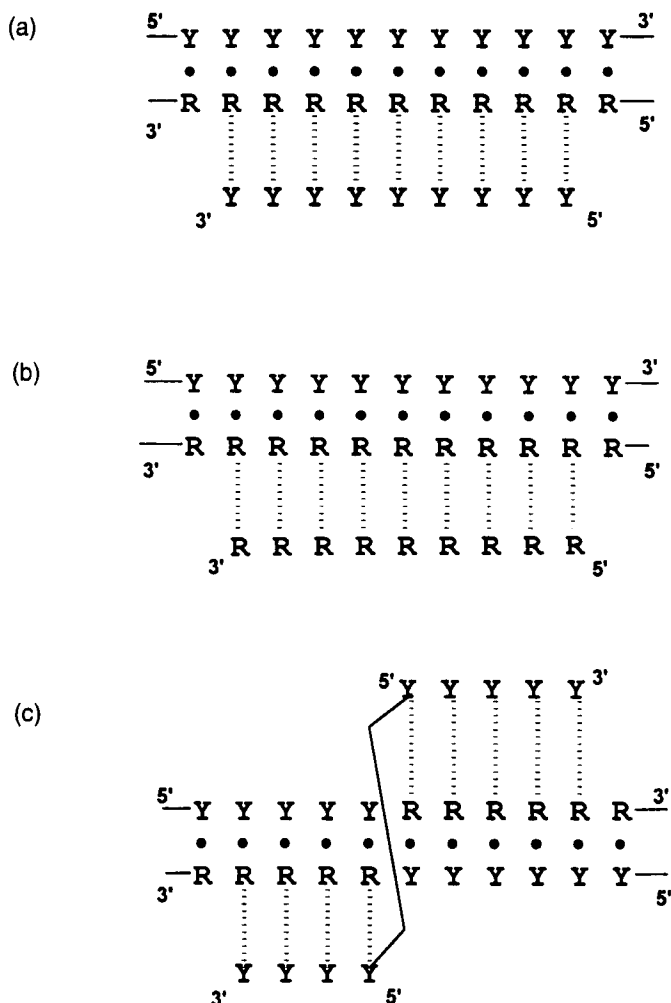


Figure 11 Targeting of homopurine-homopyrimidine sequences in double-stranded DNA: (a, b) a pyrimidine oligonucleotide and a purine oligonucleotide, respectively, bound to the homopurine-homopyrimidine sequence of DNA; and (c) an example of the targeting of two short, juxtaposed homopurine-homopyrimidine sequences by two covalently linked pyrimidine oligonucleotides

Watson-Crick base pairing, new and more general binding properties of the compounds are starting to be exploited.

5.1 Hybridization Probes and Ligands for Manipulation with Nucleic Acids and Proteins

Oligonucleotides tagged with various labels (radiolabeled groups, fluorescent groups or haptens recognized by specific antibodies) have become indispensable

tools for the identification of specific nucleic acids in research and as diagnostic probes in clinical studies [86]. Sorbents with immobilized oligonucleotides provide a simple and efficient means of isolation for nucleic acids containing specific nucleotide sequences. Sorbents containing single-stranded and double-stranded oligonucleotide sequences recognized by specific enzymes are used for affinity chromatography isolation of these proteins from cell lysates. A procedure for the isolation of specific DNAs has been developed that uses triple-helix formation [87]. In this procedure, specific DNA is captured by a biotin-conjugated oligonucleotide via intermolecular triplex formation. The complex is bound to streptavidin-coated magnetic beads. The bound DNA is recovered by elution with a mildly alkaline buffer that destabilizes the triple-stranded complexes. A similar technique is used for the isolation of specific RNAs [88].

Oligonucleotides are used as specific tags for labeling other macromolecules and constructions, e.g. microbeads carrying specific peptides or organic molecules [89, 90]. The idea is to use oligonucleotides as strings of information for tagging members of the chemical repertoire encoded by the oligonucleotides. When screening for molecules with specific binding properties, each molecule of the chemically synthesized entity is linked to a particular oligonucleotide sequence constructed in parallel. When a molecule is partitioned from the repertoire, the oligonucleotide tag can be used to identify the structure of the ligand.

Oligonucleotides have found a variety of applications in genome research. They are used to amplify nucleic acids, to single base changes in DNA and to tag specific sequences and identify specific DNA fragments with particular sequence motifs in genomic libraries. Triple-helix formation can enable highly specific enzymatic DNA cleavage, and it is anticipated that this method will be useful in genome mapping. A straightforward approach to the development of artificial constructions for cleaving DNA at specific sequences consists in the design of oligonucleotides that possess DNA-cleaving groups (see Section 6). However, it is possible to use oligonucleotides to enhance dramatically the specificity of cleavage by natural enzymes. A method has been developed [91] that allows large DNA to be cut at single sites using enzymes that normally attack multiple sites in the DNA. In this method, an oligonucleotide that binds to the DNA sequence for *EcoRI* restriction methylation is constructed. After binding of the oligonucleotide to the DNA, *EcoRI* methylase is used to methylate all restriction enzyme sites except that shielded by the oligonucleotide. The methylase and the oligonucleotide are then removed, and the DNA is cut quantitatively at the unique, unmethylated target site by the restriction endonuclease.

5.2 Oligonucleotides in the Sequencing of Nucleic Acids

Oligonucleotides play key roles in various approaches to DNA sequencing. Thus, in the currently used automated methods for DNA sequencing, oligonucleotides

are used as primers for producing sets of DNA fragments by enzymatic termination reactions that are resolved further by denaturing polyacrylamide gel electrophoresis to read the sequence. New sequencing methodologies are even more dependent on oligonucleotides and make use of the specific features of oligonucleotide hybridization. An efficient method for sequencing large DNA is that known as sequencing by primer walking. According to this method, a primer within a segment of known sequence is used to extend the sequence into an unknown region. The sequence thus determined is in turn used to select a primer to extend the sequence further. The process is repeated until the entire sequence of the nucleic acid is determined. The method uses the same template many times in a progressive manner, leading to the determination of the entire sequence of large DNA without subcloning. The disadvantage of this method lies in the need to synthesize a new oligonucleotide primer for each of the sequencing reactions. A solution to this problem has been found using the effect of cooperative oligonucleotide binding described in Section 4.2. It was proposed to use 'modular primers', which consist of short oligonucleotides (hexamers or pentamers) stacked on top of each other upon tandem annealing to a DNA template [92, 93]. Such complexes of oligonucleotides are found to prime DNA-sequencing reactions uniquely in conditions where individual short oligomers cannot prime the reaction. This permits the composition of constructions that prime sequencing from arbitrary sites having a complete library of short oligonucleotides (1024 for pentamers or 4096 for hexamers) from which specific sequences can be selected for assembling longer modular primers. Such oligonucleotide libraries eliminate the need for synthesis of new primers for each walking step and allow rapid and economical sequencing using an automated strategy.

A new strategy for efficient large-scale sequencing which relies entirely upon hybridization with oligonucleotides has been proposed [94–96]. According to the most advanced format, unknown DNA is hybridized to an array of support-bound oligonucleotides (sequencing chip). If octamers are used as the recognition elements, the chip should contain 66 536 elements. The hybridization pattern identifies the corresponding sequences in an unknown DNA, and because the linear DNA sequence is an assembly of overlapping shorter sequences, the complete structure can be elucidated from the binding data. Preliminary tests of this technique show that it could become the primary method for genomic sequencing except for fragments formed by multiple repeated sequences. The hybridization method offers the possibility of simultaneously analyzing thousands of samples, and clearly this approach is an ideal method for diagnostic applications such as the detection of infectious agents, analysis of mutations, HLA typing and genetic identification.

5.3 Oligonucleotide-Directed Mutagenesis

Oligonucleotide-directed site-specific mutagenesis is widely used to introduce mutations into DNA [97]. The method takes advantage of the possibility of

annealing partially complementary oligonucleotides to specific DNA sequences to be mutated. The oligonucleotides may contain a noncomplementary nucleotide, an excess of nucleotides or too few nucleotides. If the oligonucleotides are long enough, they form imperfect complexes with the target in which mismatches are tolerated. After the complex is formed, enzymatic elongation of the bound oligonucleotide *in vitro* or *in vivo* yields a DNA in which the new synthesized strand contains the oligonucleotide. After multiplication of the heteroduplex DNA, the modified strand gives rise to successive generations of mutated DNAs. In this way site-specific substitutions of nucleotide residues in DNA can be accomplished, and insertions and deletions of a few nucleotides can be achieved.

5.4 The Polymerase Chain Reaction

The polymerase chain reaction (PCR) is a rapid and sensitive technique for the detection of specific nucleotide sequences and manipulation of nucleic acids which has found a great number of applications ranging from molecular biology (nucleic acid labeling, quantitation, detection, cDNA cloning, mutagenesis) through forensic science and the diagnosis of genetic disorders to the detection of infectious agents [98–101]. The PCR is a method which can rapidly amplify a selected nucleotide sequence up to 10^6 times. Extremely small amounts of nucleic acids originating from only a few cells or even a single cell are required for the PCR, and the nucleic acid need not be purified. The technique can be directly used for DNA amplification; the amplification of RNA is also possible provided that the RNA is first converted into complementary DNA using reverse transcription and a suitable oligodeoxyribonucleotide primer.

The sequence to be amplified must be flanked by known DNA sequences (Figure 12). These sequences are used to bind two synthetic primers that are complementary to the opposite strands of the DNA duplex at each of the ends to be amplified. After binding to the DNA, the primers can be extended by a DNA polymerase, which results in synthesis of DNA across the region between the primers and duplication of the DNA in which the primers used are covalently incorporated into the 5'-end of the PCR product. Therefore the extension products also can bind the primers. This allows successive cycles of amplification which continue to double the number of DNA molecules synthesized in the previous cycle, and the fragment is amplified exponentially by a factor of 20^n (where n is the number of amplification cycles accomplished).

The PCR is used to detect and quantify mRNA sequences in cells and to copy and multiply known and unknown genes. When the primers are designed so as to contain restriction endonuclease cleavage sites, the amplification products can be cleaved at these sites and ligated at a suitable vector for multiplication of the DNA sequences of interest.

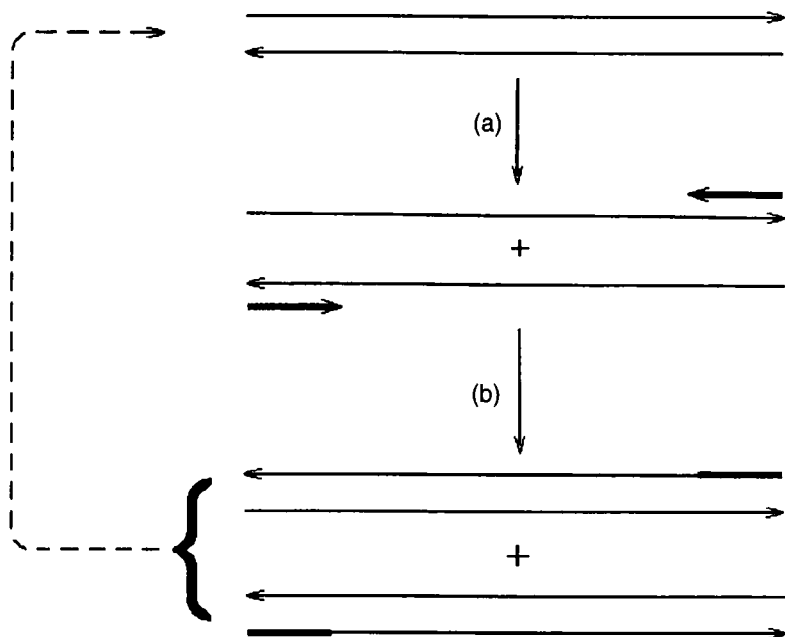


Figure 12 General scheme for the polymerase chain reaction (PCR): (a) a double-stranded DNA is denatured by heating to 90 °C and oligonucleotide primers are annealed to the DNA strands; and (b) the primers are extended at 70 °C with thermostable DNA polymerase to produce two double-stranded DNAs. The synthesized DNA can be used for further repeated copying (by repeating steps a and b)

5.5 Aptamers

A straightforward synthetic approach to the development of molecules with specific properties involves the rational design of chemical constructions that fold properly and present their functional groups in space in the geometry needed for creation of the active structure. This approach requires an exact knowledge of the active structure and considerable efforts in synthesis. Recently, an alternative approach for the development of oligonucleotides with some specific properties has been developed [6, 7, 102, 103]. This iterative approach is based on the repeated selection from large pools of randomly generated biomolecules of a few that present the desired properties. These rare variants are then multiplied. This cycle of selection and amplification can be repeated as many times as necessary to isolate a preparation of molecules composed entirely of the active species which can then be cloned and characterized. The selected ligand-binding nucleic acids are referred to as aptamers, from the Latin *aptus* (to fit). The approach is also known as SELEX, an acronym for systematic evolution of ligands by exponential enrichment.

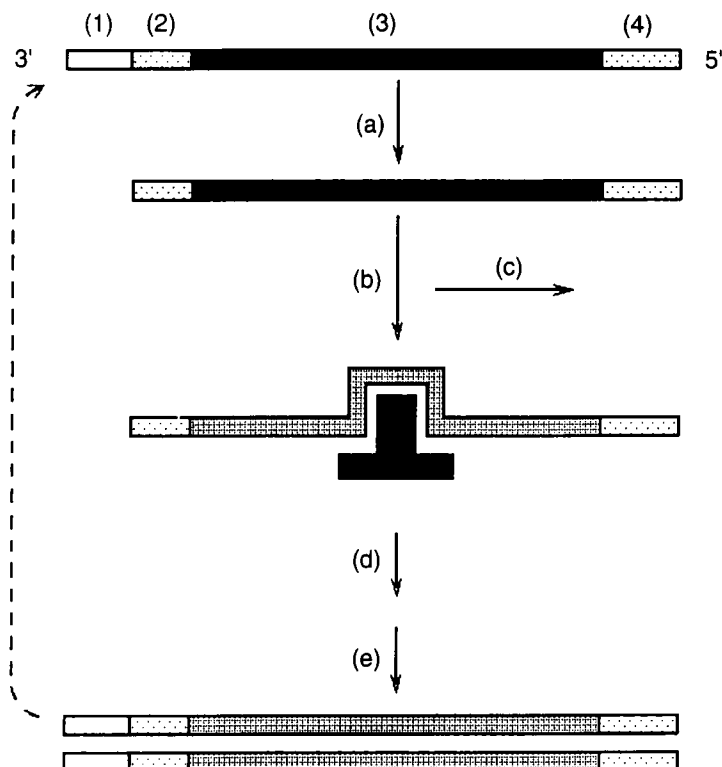


Figure 13 Principle of selection for RNA capable of a specific interaction (aptamer). The members of the DNA library possess a sequence including a T7 promoter needed for transcription (1) and a randomized sequence (3) flanked by definite sequences for binding oligonucleotide primers (2 and 4). (a) Transcription with T7-phage RNA polymerase yields a library of RNA sequences. (b) In the fractionation step, the RNA molecules capable of binding to a specific ligand are separated from the rest of the RNAs, which are discarded (c). (d) The complex is dissociated and reverse transcription is used to synthesize DNA copies of the isolated RNAs. (e) The DNA copies are amplified by the PCR using oligonucleotide primers, thus restoring the flanking sequences needed for the DNA construction to be transcribed. The synthesized duplex DNA provides an enriched library which can be repeatedly subjected to the entire selection procedure (steps a to e) until one or a few molecules are isolated which bind to the ligand with the highest affinity

A simple application of the method is in the development of oligonucleotides capable of specific interactions. The DNA sequence is chemically synthesized as a number of sequence-degenerate pools flanked by oligonucleotides of known structure (Figure 13). This is easily done by introducing a mixture of four synthons at each coupling step in the course of assembling the sequence to be made random. The complexities of the random pools used are as high as 10^4 , which means that in the pool there is an enormous number of molecules with

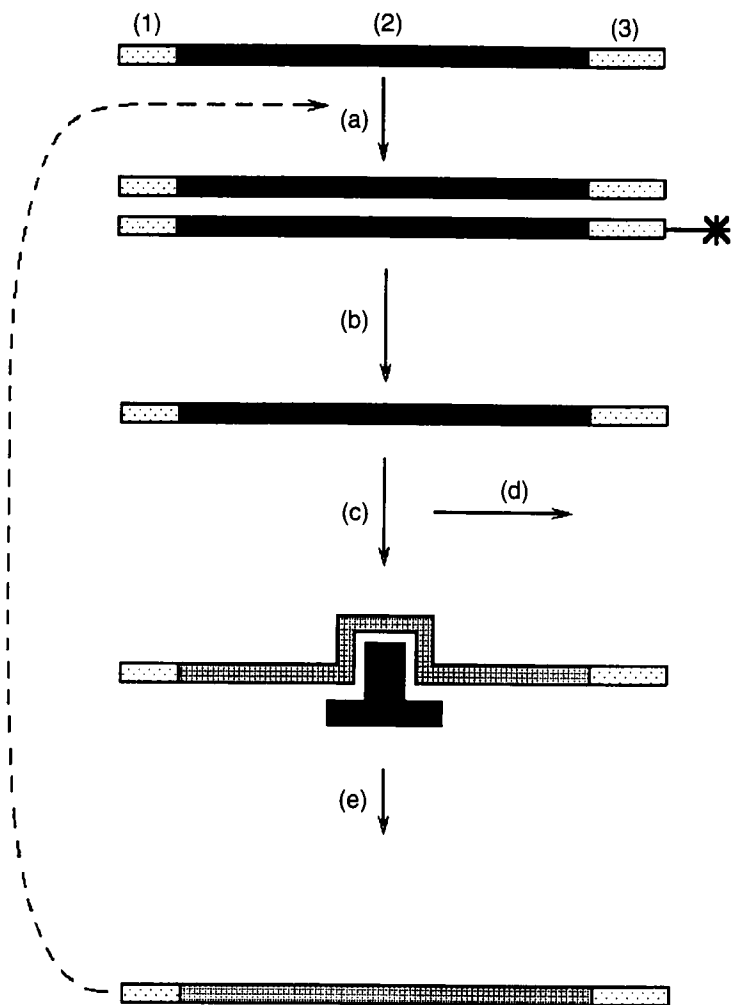


Figure 14 Principle of selection for DNA sequences capable of interacting with a specific protein. Members of the DNA library possess a central randomized sequence (2) flanked by the primer-binding sites (1 and 3). (a) Amplification of the DNA using one of the primers conjugated to biotin (shown by an asterisk) leads to the production of DNA with one labeled strand. (b) The DNA is denatured and the biotinylated strand is removed by absorption on an avidin-agarose column. (c) In the selection step, the single-stranded DNA is applied onto an affinity column with reversibly immobilized protein. The column is washed out to remove unbound DNA. (d) The protein is detached and washed out with the bound DNA from the column. (e) The protein is removed by phenol extraction and the DNA is recovered. The DNA is subjected to further cycles of amplification and selection for isolation of the sequence which binds the protein with the highest affinity

different structures, some of which may display the desired property. The target molecule is used as a binding structure to select oligonucleotides that can bind it because of the features of their tertiary structures encoded in their sequences. The selected pool of molecules is amplified by the PCR using primers complementary to the terminal flanking sequences of the oligomers. The produced population is enriched in molecules capable of binding to the target. Further rounds of selection allow full isolation of the sequences capable of tight binding. Similar techniques can be used for the selection of DNA oligonucleotides capable of some specific interactions (Figure 14) [104]. The key to the method is the use of cycles of selection. Most purification procedures are limited to a one-step purification factor of 100–1000 because of nonspecific absorption of nonspecific molecules. Repetition of cycles of purification and amplification increases the total purification factor multiplicatively. Five cycles are estimated to be sufficient to achieve an enrichment of 10^{10} – 10^{15} .

The great advantage of the approach is that it allows fast and simple selection of the tight-binding molecules from an enormous number of variants. RNA aptamers capable of specific binding to various organic dyes, amino acids, ATP, and other small molecules and proteins have been characterized [102, 103]. It has been found that nucleic acids which contain specific high-affinity binding sites for small ligands are present at a frequency of roughly one in 10^{10} – 10^{11} in pools of random-sequence RNA molecules. Naturally, the selection technology has been used to explore the recognition of double-stranded DNA by RNA. The selection of RNA molecules capable of binding to an oligopurine DNA sequence led to identification of the optimal RNA structure and established some mismatches in the RNA structure which are tolerated in the triple-stranded complex [105]. This approach will lead to the development of specific oligonucleotides for targeting any structural element in RNA and DNA molecules.

The described technology is based on the selection of molecules capable of some specific binding. However, any method that creates an enrichment in the population of molecules displaying some definite property can be used in the scheme shown in Figures 13 and 14 for fishing out molecules with that special property. The possibilities in this approach for the selection of molecules capable of specific interactions and reactions are limited only by the inventiveness of the researchers. The selection approach has been used to isolate active tRNAPhe variants from randomized libraries for the identification of RNA structures capable of undergoing intramolecular self-cleavage [106] and to develop ribozymes with unusual substrate specificity [107]. Apparently, aptamer technology provides new possibilities in the design of ligands for bioseparations, inhibition of enzymes and other biological activities. Using the selection approach, a DNA oligonucleotide with the sequence GGTGGTGTGGTTGG which binds and inhibits the human protease thrombin was developed [104]. It has been shown that the molecule inhibits the thrombin-catalyzed formation of fibrin clots in human plasma, suggesting its use as an anticoagulant.

Interestingly, the established structure of the oligonucleotide contains the guanosine quadruplexes mentioned in Section 2. The oligonucleotide forms a highly compact and unimolecular symmetrical structure consisting of two tetrads of guanosine bases and three loops connected by two TT loops and a TGT loop [108].

It should be noted that the described molecular selection technique allows the testing of more nucleic acid variants than are contained in the initially synthesized oligonucleotide pool. In the course of amplification, the sequences are randomly mutated because of the nonabsolute specificity of the nucleic acid polymerases. The mutation rate can be affected deliberately by varying the experimental conditions. The mutated molecules take part in further enrichment cycles and are amplified if mutations result in improvement of the assayed property. Therefore one can consider the iterative selection procedure also as a molecular evolution process in which mutations occur spontaneously in the course of multiplication of the oligonucleotides and evolutionary pressure is provided at the selection step.

An iterative selection procedure has been developed which allows the screening of oligonucleotide libraries without using enzymes [109]. This method allows the use of oligonucleotide analogues which cannot be multiplied by the PCR technique and extends the possibilities of the selection approach to the development of any synthetic oligomers built of a few standard monomers. The method was first tested with oligonucleotide analogues in the form of 2'-*O*-methylated ribooligonucleotides. Four libraries of *n*-mer oligonucleotides were synthesized which were built of a random sequence except for one position, where members of each family had a residue specific to that family. Binding of the families to the target was compared and the family displaying the most favorable properties was identified. At this point it was assumed that the specific residue at the assayed position in the best family should be present in molecules capable of the interaction. Then four new families of sequences were synthesized in which the oligonucleotides of all the families had identical optimal residues in the position identified at the preceding step. Different family-specific nucleotides were placed in another definite position. Comparison of the affinities of the families for the target allowed identification of the second nucleotide required for optimal binding. Repeating the synthesis of four families and the selection step finally allowed identification of optimal sequence for binding. The method was tested in experiments on the design of oligonucleotides for optimal targeting of the RNA loop in the activated *H-ras* oncogene mRNA. The investigated 2'-*O*-methylated oligonucleotide analogue library had 262 144 sequences in four pools of 65 536. As expected, the optimal target was found in the center of the loop. Furthermore, a library of 65 536 individual 8-mers in four pools (16 348 each) was screened for activity against human herpes simplex virus type 1 in cell culture. In this case, the selection criterion was the ability to suppress virus multiplication. The maximal activity was found to be associated with oligonucleotides containing tetraguanylate sequences. Interestingly, the best oligonucleotides can be expected to form guanosine quarters (see Section 2).

6. AFFINITY MODIFICATION OF NUCLEIC ACIDS AND PROTEINS WITH REACTIVE DERIVATIVES OF OLIGONUCLEOTIDES

6.1 Reactive Groups for Oligonucleotide Derivatives

A great number of affinity reagent derivatives of oligonucleotides have been synthesized and used for mapping nucleic-acid-binding centers of enzymes and nucleoprotein complexes [4, 12]. A major challenge confronting chemists developing reactive derivatives of oligonucleotides is the design of groups which will efficiently react with the target nucleic acid in complex systems *in vitro* and *in vivo* and will not affect nontarget biopolymers. In this section only the most advantageous groups which are prospective candidates for accomplishing efficient and specific modifications of biopolymers and for designing compounds for *in vivo* applications will be considered.

Alkylating groups [110] form covalent bonds with nucleic acids. The reactivity of nucleophilic centers of nucleic acids toward alkylating groups decreases in the order N7 atoms of purine bases > N1 and N3 atoms of adenine > N3 atom of cytosine > internucleotide phosphates. In alkaline medium the N3 atoms of uracil and thymine are alkylated easily. Alkylation of heterocyclic bases, especially guanine and adenine, results in weakening of the glycoside bonds. Therefore the polynucleotide chain may be selectively cleaved at the alkylation sites by treatment with piperidine. Analysis of the resulting fragments allows identification and quantitation of the modified residues.

Alkylation can proceed according to the S_N2 mechanism via a transition state formed by both reactants or according to the S_N1 mechanism where the alkylating agent is converted to a short-lived reactive intermediate in a monomolecular process and the intermediate reacts further with nucleophiles. In the case of the S_N2 process, the reaction rate is proportional to the concentration of nucleophiles in solution and obeys second-order kinetics. In the case of the S_N1 mechanism, conversion to the short-lived reactive intermediate is the rate-limiting step. Therefore the rate of reagent consumption is described by first-order kinetics and is independent of the concentration of nucleophiles. The S_N1 mechanism is typical of aromatic 2-chloroethylamines (Figure 15). Alkylating groups of both S_N1 -type (aromatic 2-chloroethylamines) and S_N2 -type (COCH_2Hal , ethyleneimine group) have been conjugated to oligonucleotides (Figure 16). Oligonucleotides bearing these groups can be used for sequence-specific modification of single-stranded and double-stranded DNA. The groups can be tethered to the terminal phosphates and ribose of oligonucleotides, to the 5-position of a deoxyuridine analogue, and to the 2-position of a deoxyadenine analogue [4, 111–114].

For accomplishing reaction in *in vivo* conditions, the groups reacting according to the S_N1 mechanism are more advantageous. In the cell numerous nucleophilic compounds are present and the groups reacting according to the

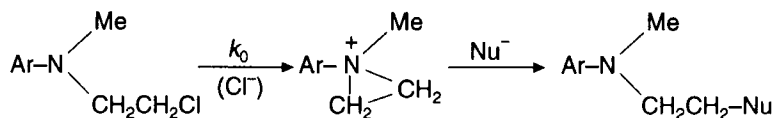


Figure 15 Alkylation with aromatic 2-chloroethylamines proceeds via formation of an intermediate ethyleneimmonium cation. The formation of this species is the rate-limiting step of the process

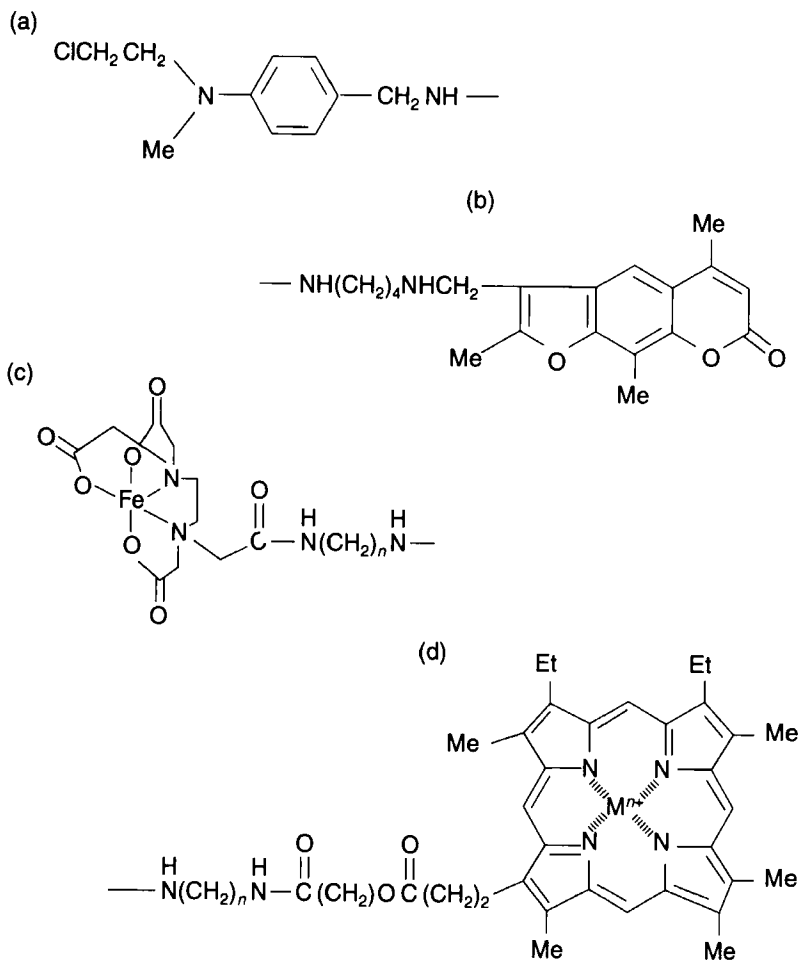


Figure 16 Reactive groups for oligonucleotide derivatives: (a) an aromatic 2-chloroethylamino group; (b) a psoralene group; (c) an EDTA-Fe group; and (d) a porphyrin group

S_N2 mechanism can be inactivated rapidly because of side reactions with these nontarget compounds. By contrast, the compounds reacting according to the S_N1 mechanism, e.g. aromatic 2-chloroethylamines, are inert before the formation of the reactive species occurs and can survive while finding their way to the cellular targets. This factor explains the considerable anticancer activity of aromatic 2-chloroethylamines such as chlorambucyl and sarcolysin.

Oligonucleotide derivatives have been synthesized in which the alkylating group is inactive and may be activated through special treatment. Thus the aromatic 2-chloroethylamino group in *N,N,N'*-tri(β -chloroethyl)-*N'*-(*p*-formylphenyl)propylene-1,3-diamine is relatively inert because of the electron-accepting aldehyde substituent in the benzene ring. The group is activated by the reduction of the aldehyde group to the alcohol group [115]. A few other alkylating groups activatable by mild chemical treatments have been developed [116, 117]. These groups can be triggered to react when needed, e.g. after annealing of a long oligonucleotide carrier to the target nucleic acid under stringent conditions.

Reactions are easily controlled in the case of the photoreactive oligonucleotide derivatives. Oligonucleotides with aromatic azido groups are widely used to label biopolymers interacting with nucleic acids [118]. Many of the UV-absorbing groups conjugated to oligonucleotides are activatable by irradiation and can bring about chemical modifications of the nucleic acids which bind to oligonucleotides equipped with these groups [4, 14, 15]. For the sequence-specific chemical modification of nucleic acids, the psoralene oligonucleotide derivatives seem to be the most promising. These representatives of the furocoumarin family [119] contain two photoactivatable double bonds: C3=C4 in the pyrone ring and C4'=C5' in the furane ring. Upon illumination with UV light (365 nm) they react with nucleic acids to yield monoadducts with pyrimidine bases (thymine is the most reactive) and diadducts with both strands, forming crosslinks [119–121]. Psoralene intercalates into the DNA helix in one of two conceivable orientations. After the photoactivation, either the 4',5'-furan double bond or the 3,4-pyrone double bond of psoralene undergoes an addition reaction with the 5,6-double bond of a pyrimidine to form a monoadduct. The photoadduct formed with the 3,4-pyrone bond of psoralene cannot be activated for further photochemical reactions with the nucleic acid. The product formed with participation of the 4',5'-furan double bond can take part in a second photoreaction with thymine in the other DNA strand to produce a crosslink. It should be mentioned that irradiation with short-wavelength UV light results in cleavage of the bonds between the psoralene residue and the nucleic acid (photoreversion of crosslinks).

The crosslinking groups can provide efficient modification of the targets in some controlled, optimal conditions; however, in an organism one can envisage a variety of potential side reactions which will decrease the efficiency of these groups. Another disadvantage of the crosslinking groups is that they form immunogenic conjugates with proteins [122] which will cause problems in the case of therapeutic applications. Therefore attempts have been made to develop

groups capable of cleaving nucleic acids. Equipment of oligonucleotides with such groups would yield compounds capable of destroying large numbers of nucleic acid targets without producing unwanted conjugates.

Reagents generating active oxygen species can cause efficient degradation of nucleic acids. Nucleic acids are degraded by the active oxygen species formed by chelated transition metals in the presence of either oxygen or hydrogen peroxide and some reducing agents. The possibility of developing efficient reactive groups for the oxidative destruction of nucleic acids is suggested by the fact that the anticancer antibiotic bleomycin, which has a chelated iron(II) ion that produces active oxygen species, is an extremely potent reagent for cleaving DNA *in vitro* and *in vivo*. In an attempt to develop small groups that produce active oxygen species, model chelating systems such as EDTA-Fe, 1,10-phenanthroline-copper and porphyrins have been tested and conjugated to oligonucleotides [4, 13–15, 53]. In the case of the EDTA-Fe complex, the DNA-damaging species are hydroxyl radicals. Therefore reaction of nucleic acids with oligonucleotide derivatives bearing these groups results in a diffuse modification pattern, with eight to 12 nucleotides of the target affected. For this reason, and also because of the self-degradation of the reactive group, the modification yields are usually below 10%. It has been shown that OH radicals react with both the sugar moiety and nucleic acid bases. However, the destruction of nucleic acids occurs primarily because of the damage to the sugar phosphate backbone. Oxidative degradation of DNA results in the formation of 3'-fragments with a 5'-phosphate terminus and 5'-fragments with either a 3'-phosphoglycolate or a 3'-phosphate terminus. In the other groups mentioned above, the active oxygen species cannot diffuse. Oligonucleotides with the copper-phenanthroline and porphyrin groups are more efficient and bring about localized damage to the target. More complex oligonucleotide conjugates with metal-chelating groups are under development. For example, the entire bleomycin molecule has been conjugated to an oligonucleotide and the construct was shown to cleave DNA in the vicinity of the target sequence [123].

Other potential nucleic-acid-cleaving groups are those compounds that can promote hydrolytic cleavage of the internucleotide bonds. Crosslinking of oligonucleotides to some nucleolytic enzymes has provided conjugates which retain the nuclease activity and cut the DNA near to the oligonucleotide-binding sites [62, 63]. Synthetic molecules that mimic the active centers of nucleolytic enzymes have been designed [124–127]. At present, the rational design of organic molecules of the size and complexity needed to mimic the entire active center of an enzyme and provide the corresponding arrangement of the residues characteristic of the active center is not possible. However, the active centers of nucleolytic enzymes contain only a few key groups involved directly in the cleavage of the phosphodiester backbone. The high turnover rate of natural enzymes is not needed for these enzyme mimics, and therefore one can try to design simplified versions of the active center where the residues have more conformational freedom and the active conformation can be realized by chance.

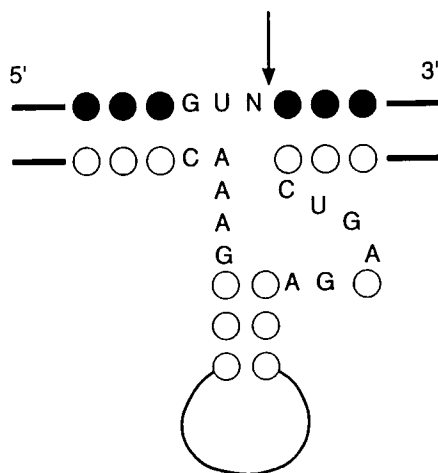


Figure 17 The hammerhead ribozyme: ribonucleotides important for the activity are shown by letters and open circles indicate nucleotides of the ribozyme where modifications and substitutions are tolerated. The arrow indicates the site of cleavage in the RNA substrate (top nucleotide line), in which GUN is a conserved sequence where the cleavage occurs, and black circles indicate the nucleotides complementary to the nucleotides of the ribozyme

Synthetic molecules capable of noncovalent binding to RNA and bearing imidazole residues that imitate the histidine residues of the active center of ribonuclease A have been shown to degrade tRNA efficiently in physiological conditions [127]. The results of these studies give hope for the design of oligonucleotides equipped with catalytic nucleic-acid-cleaving groups.

In recent years RNA structures possessing self-cleaving activity have been discovered [128, 129]. The catalytic RNA domains represent relatively simple structures which can be exploited as reactive groups for oligonucleotide construction. The simplest of the structures is the so-called 'hammerhead' structure (Figure 17), which was found in plant viroids and virusoids. The hammerhead catalytic center contains a few conserved nucleotides which are folded in a compact structure and flanked by sequences complementary to those surrounding the cleavage site of the target RNA. Cleavage occurs at the 3'-position with respect to the GU sequence to yield a 2'-3' cyclic phosphate and a 5'-OH terminus at the hydrolysis site. Many modified and substituted analogues of the hammerhead ribozyme have been synthesized chemically and tested to identify the sequences and groups essential for catalysis. It was found that only a few ribonucleotide residues need to be conserved in the structure to maintain catalytic activity. The rest of the molecule can be substituted by nucleotide analogues and some synthetic linkers, which makes the structure more resistant to nucleases [128-131]. Three other RNA-cleaving RNA structures have been identified in the RNA of tobacco ringspot (hairpin ribozyme), in

hepatitis delta virus and in the RNA component of RNase P, an enzyme which processes the 5'-leader segments from precursors of tRNAs [128–131]. These structures are also considered as potential RNA-cleaving groups for antisense constructions.

6.2 Sequence-Specific Chemical Modification of Nucleic Acids

The specificity of the affinity reactions between oligonucleotide derivatives and target nucleic acids is determined by several factors. When the reaction occurs within a complementary complex, the modification pattern is determined by the conformational freedom of the reactive group and by the relative reactivities and positions of the residues in the target available to the reactive group. Thus psoralene can react only with pyrimidine residues and a specific mutual orientation of psoralene and the pyrimidine base is required. Therefore oligonucleotides bearing psoralene groups will crosslink to the target sequence only if it possesses a pyrimidine residue in an appropriate position available for the psoralene tethered to the bound oligonucleotide. Alkylating agents such as 2-chloroethylamines can react with several nucleophilic centers in nucleotides. The modification pattern in this case is determined by the nucleotide sequence available to the reagent. Single guanosine residues can be alkylated efficiently if they are located in a favorable position and flanked by less reactive pyrimidine nucleotides. A few residues are affected if the alkylating agent can reach a few reactive centers [132, 133]. Optimization of the linker structures connecting aromatic 2-chloroethylamino groups to oligonucleotides can be performed in order to achieve efficient modification of single-stranded and double-stranded nucleic acids [134, 135]. As already mentioned, when the reagent produces active species capable of migrating in solution (e.g. EDTA–Fe), long stretches of nucleotides are affected and a diffuse modification pattern is observed.

In the case of single-stranded nucleic acids, a nonspecific reaction can occur with nontarget residues because of the tertiary folding of the target molecule which brings a remote sequence close to the oligonucleotide-binding site. This effect has been observed in experiments with single-stranded DNA fragments [132] and ribosomal RNA [136].

Another level of specificity defines the modification of a target sequence within a large nucleic acid or the modification of a specific nucleic acid in comparison to nontarget nucleic acids. A nonspecific reaction of nucleic acids can occur with nonbound oligonucleotide derivatives, resulting in a random modification of the polynucleotide. The efficiency of such a bimolecular process is usually very low and it can be suppressed by choosing appropriate reaction conditions. The most important source of nonspecific reactions of oligonucleotide derivatives is the formation of imperfect complexes (see Section 4). *In vitro*, the annealing of long oligonucleotide probes with nucleic acids under stringent conditions gives very high specificity of recognition. However, *in vivo*, hybridization has to occur in a narrow window of physiological conditions, and the specificity of targeting

can be affected only by choosing an appropriate oligonucleotide length and sequence and by modulating the affinity of the oligomers for the targets by means of chemical modifications. In practice, 15–25 nucleotides long oligomers are used for targeting RNA and DNA. Under physiological conditions, oligonucleotides of this length can form imperfect complexes with partially complementary nucleotide sequences. The potential contribution to the modification yield of the nonspecific reactions of imperfect complexes has been studied on the basis of stability data for imperfect oligonucleotide complexes [65]. It was concluded that under the conditions of rapid exchange of the bound and free oligonucleotide derivatives, nonspecific reactions occur mainly because of the formation of complexes with terminal mismatches. Although the affinity of oligonucleotides for these imperfect binding sites is lower as compared to the true target site, the total yield of the nonspecific modification may be comparable with that of the specific reaction. Since oligonucleotides larger than 18-mers can form complexes even with internal mismatches in physiological conditions, it is clear that side reactions with these oligonucleotides may cause severe problems. Indeed, the consequences of the nonspecific interactions of oligonucleotides have been clearly demonstrated in experiments where specific RNAs were targeted with antisense oligodeoxyribonucleotides in the presence of ribonuclease H. It was found that oligonucleotides can cause efficient degradation of nontarget cellular mRNAs through nonspecific complex formation. This explains the failure of some studies with antisense oligonucleotides and ribozymes aimed at inactivation of specific RNAs and the reported sequence-independent effects of antisense oligonucleotides [137–139].

Investigation of the reactions of nucleic acids with various oligonucleotide derivatives have demonstrated that under optimal conditions the specific modification of large RNAs and DNAs can be accomplished with pinpoint accuracy. Very specific alkylation and cleavage of dsDNA by *N*-bromoacetyl and EDTA oligonucleotide derivatives have been reported. Modification with *N*-bromoacetyl oligonucleotide derivatives is efficient and specific enough to accomplish cleavage of both DNA strands at a specific site in a yeast chromosome 340 000 bp in size with 90% yield [140]. Conjugates of oligonucleotides with nucleases can cleave DNA and RNA within the target sites [62, 63]. A stretch of five nucleotides is attacked by the enzyme near the target sequence, apparently through the flexibility of the linker connecting the enzyme to the oligonucleotide. When the oligonucleotide–enzyme conjugate was used in an excess, the cleavage yield reached 90%. Staphylococcal nuclease linked to the 5'-terminus of a homopyrimidine oligonucleotide capable of binding to the pUC19 plasmid cleaved the plasmid into two fragments which could be used further in cloning procedures [141]. Such conjugates may provide useful tools for manipulating nucleic acids.

Modification with reactive derivatives of oligonucleotides has been used to elucidate the structure of the nucleic acid complex. The orientation of the third strand in different triple-stranded nucleic acid complexes can be determined using affinity modification with reactive derivatives of oligonucleotides [46, 142].

Reactive derivatives of oligonucleotides can be used to place a tag or plant a reactive group onto a particular base within a nucleic acid. The label can be used to investigate the spatial arrangement in nucleic acid complexes. Using affinity modification with an alkylating oligonucleotide derivative, a residue with an amino group can be introduced into a specific position in tRNA. The amino group can be acylated selectively to attach a linker with an arylazido photoreactive group. A tRNA derivative with such a photoreactive group at a specific position has been used for the fine mapping of the tRNA-binding center of bacterial ribosome [143].

Investigation of the kinetics of RNA cleavage by various constructions containing a ribozyme-cleaving group has revealed some important features of the reaction mentioned in Section 4.2. Elongation of the targeting sequences to give enhanced stability of the complex is expected to facilitate cleavage. However, the elongation prevents rapid turnover of the ribozyme and interferes with the catalytic process. A compromise must be reached to balance complex formation and cycling of the ribozyme after cleavage [75].

Modification with derivatives of pyrimidine oligonucleotides equipped with alkylating groups and EDTA-Fe groups has been used to investigate the kinetics and thermodynamics of double-stranded and triple-stranded complex formation [144–146]. Site-specific modification of the target nucleic acid can be performed for a short time to achieve small modification extents. This eliminates the effects of side reactions and gives quantitative results from which thermodynamic and kinetic characteristics can be calculated. From the modification data, the association constants for reactive derivatives of oligonucleotides with single sites in structured, single-stranded polynucleotides and large DNAs can be determined and the effects of various parameters on complex formation can be evaluated.

Cooperative interactions of oligonucleotides bound to adjacent sequences of the nucleic acid target (see Section 4.2) can be investigated using alkylating derivatives of oligonucleotides. The oligonucleotide effectors can bind the nucleic acid on both sides of the reagent and increase considerably the modification efficiency. Use of effectors enhances the selectivity of modification through elongation of the recognition sequence. Similar studies have been done with a ribozyme [79]. Oligonucleotides complementary to the sequence of the target RNA adjacent to the ribozyme-binding site facilitate the interaction of the ribozyme with its specific sequence in the target.

Reactive derivatives of oligonucleotides provide efficient tools for accomplishing targeted chemical mutagenesis in intact DNA. In principle, these compounds may be expected to cause specific mutagenesis *in vivo*. In one study, a long oligonucleotide modified with *N,N,N'*-tri(β -chloroethyl)-*N'*-(*p*-formylphenyl)propylene-1,3-diamine to introduce multiple latent alkylating groups (see Section 6.1) was hybridized with T7-phage DNA. The complex was treated with sodium borohydride to trigger alkylation of the target and the alkylated DNA was used for transfection of *E. coli* cells. It was found that mutations occurred predominantly in the target gene [147]. A triple-helix-forming oligonucleotide

linked to psoralene was used to introduce targeted mutations into an intact λ -phage genome. UV irradiation of the complex resulted in the appearance of mutations in the target gene with a frequency at least 100-times higher than for nontarget genes. Practically all the mutations were localized within the target nucleotide sequence [148].

In the case of *in vivo* reactions, one can expect that the modified DNA residues will be removed or demodified by cellular repair enzymes. In order to introduce a modification which will be hard to repair and which will cause efficient inactivation of the target gene, bifunctional oligonucleotide derivatives with two alkylating groups capable of crosslinking the two DNA strands have been developed [149].

6.3 Affinity Modification of Proteins

Reactive derivatives of oligonucleotides are used for the affinity labeling of biopolymers that are interacting with nucleic acids [118]. Affinity labeling with alkylating and photoreactive oligonucleotide derivatives has been used to identify nucleic-acid-binding proteins on the cell surface [150,151] and in the bloodstream [152]. Modification with reactive derivatives of oligonucleotides has been used for the labeling of ribosomal proteins within the tRNA- and mRNA-binding centers of ribosomes [153] and for the investigation of the DNA-binding subunits in RNA polymerases [154].

Investigations of the kinetics of affinity labeling in different reaction conditions have shed light on the molecular interactions in DNA polymerases. Inactivation of the enzymes by reactive platinating derivatives of oligonucleotides has been investigated in the presence of various analogues of templates and primers. From the protective effects caused by the analogues, the kinetic and thermodynamic characteristics of the interaction within the template site of the enzyme can be determined.

7. DERIVATIVES OF OLIGONUCLEOTIDES AS INHIBITORS OF NUCLEIC ACIDS AND PROTEINS

Oligonucleotide derivatives are receiving growing attention from biologists and pharmacologists. These powerful and specific inhibitors of translation and transcription are used to investigate the physiological role of specific genes, and attempts have been made to use these compounds for therapeutic purposes, namely for the suppression of specific genes of infectious agents and over-expressed, unwanted cellular genes, e.g. oncogenes. To be of therapeutic value, oligonucleotide derivatives should be nontoxic, nonmutagenic and nonimmunogenic. They should be stable in the organism and capable of finding their way into cells and cellular compartments to the target nucleic acids. Attempts have been made to design oligonucleotide analogues and conjugates which possess these properties.

7.1 Targets for Oligonucleotide Derivatives

Potential cytoplasmic targets for oligonucleotides are tRNAs, ribosomal RNA in both subunits of ribosomes and mRNAs. In the nucleus, the targets are small nuclear RNAs, mRNA precursors and DNA. In infected cells, the pathogen-specific nucleic acids are targeted.

Targeting of ribosomal RNAs will affect the expression of all the cellular genes (Figure 18). However, specific effects can be achieved at the cellular level. Some ribosomal RNA sequences are specific for bacteria and mycoplasmas and can be targeted for specific elimination of these cells. Thus, the Shine–Dalgarno sequence of mRNAs in bacteria and mycoplasmas interacts with the anti-Shine–Dalgarno sequence in ribosomal RNA in the course of translation initiation. It has been shown that methylphosphonate oligonucleotide analogues that are complementary to the anti-Shine–Dalgarno sequence can inhibit protein biosynthesis by bacterial ribosomes and suppress proliferation of these bacterial strains with cell walls permeable to the oligonucleotide analogues [155]. The oligonucleotide AGAGGAGGU, complementary to the 3'-end of 16S ribosomal RNA, inhibits formation of the 70S ribosomal initiation complex [156].

The primary cytoplasmic targets are mRNAs, which are single-stranded polynucleotides readily accessible to complementary oligonucleotides, and, in the case of infected cells, nucleic acids of infectious agents. The latter often have some unique sequences which are not found in the cellular messengers. For example, a universal 5'-terminal mini-exon sequence is present in all mRNA species of trypanosomes and leishmanias, and this is an ideal target for the suppression of all gene expression in the parasite [157].

The most important target for oligonucleotides in modulating the expression of specific cellular genes and inactivating integrated viral genomes is double-stranded DNA. DNA can be targeted either by triple-stranded complex formation or by strand invasion by oligonucleotide analogues capable of displacing one of the DNA strands.

Also, oligonucleotides can be used for targeting specific proteins capable of interacting with certain RNA and DNA sequences.

7.2 Experiments with Isolated Biopolymers

Binding of oligonucleotides to single-stranded regions of RNA interferes with the interactions between RNA and protein factors. An efficient arrest of the translation of different mRNAs can be achieved with oligonucleotides and various analogues that are complementary to the sequences around the initiating codon and the 5'-terminal cap region of the messenger [13]. The inhibition is caused by the physical interference of the hybridized oligonucleotide with the binding of the initiation factors and ribosomes (Figure 19). Oligonucleotides that are complementary to the coding regions of mRNA usually do not arrest translation because ribosomes can displace the oligonucleotides from the

messenger. However, targeting of these sequences of mRNAs is important for discrimination between mRNAs with similar regulatory sequences, e.g. mRNAs of mutated and normal protooncogenes. Translation arrest can be achieved in this case if the oligonucleotide derivative damages the messenger or triggers degradation of the mRNA by cellular ribonucleases. There are a few enzymatic activities which destroy RNA when it forms a complementary complex. Thus RNase H digests the RNA component of the hybrid RNA-DNA complex [158] and RNase L degrades complexes of RNA with oligonucleotides that contain conjugated 2'-5' oligoadenylate sequences [159]. Ribonuclease H seems to be an important factor for enhancing the biological activity of oligodeoxynucleotides and phosphorothioate oligonucleotide analogues. The enzyme does not hydrolyze the complexes formed by oligonucleotide methylphosphonates and α -oligonucleotides, and therefore these analogues arrest translation only by blocking the transcription initiation when they mask the cap site of the messenger. RNase H can cleave loose, imperfect complexes formed by oligonucleotides with homologous sequences in nontarget RNAs. Rapid exchange of oligonucleotides in these complexes can lead to efficient degradation of nontarget RNAs [137, 138]. This nonspecific effect can be partially reduced by using chimeric oligonucleotides with a few phosphodiester bonds in the central part of the molecule flanked by sequences with modified phosphates [160]. In complexes of these oligonucleotides, cleavage can occur only at those RNA sequences in contact with the oligonucleotide domain between the phosphodiester bonds, while the specificity of complex formation is provided by the whole chimeric sequence.

Oligonucleotide derivatives capable of covalent crosslinking to the target arrest translation irrespective of their binding site on the messenger. It has been shown that equipment of oligonucleotides with alkylating groups [161, 162] and photo-reactive psoralene groups [163] gives irreversible translation inhibitors which arrest translation of the messengers even when targeted to the coding region of mRNA.

Binding of oligonucleotides to the RNA template can arrest the enzymatic synthesis of the complementary DNA strand. The transcription arrest is particularly efficient in the case of reactive oligonucleotide derivatives that form crosslinks with the target [164, 165]. It has been shown that binding of pyrimidine and purine oligonucleotides to DNA by triple-stranded complex formation inhibits the interaction of DNA with the specific proteins recognizing the overlapping sites. Such proteins include restriction endonucleases, methylases, transcription factors and DNA polymerase [166, 167]. Oligonucleotide binding to eukaryotic promoters represses the transcription process efficiently. Reactive derivatives of oligonucleotides show enhanced inhibitory activity. An oligonucleotide with a psoralene group at its 5'-end showed an irreversible and specific inhibition of the transcription of the target gene after irradiation with UV light [168]. Pyrimidine oligonucleotides with alkylating groups targeted to DNA sequences within the coding region of the gene arrest transcription completely after crosslink formation [169].

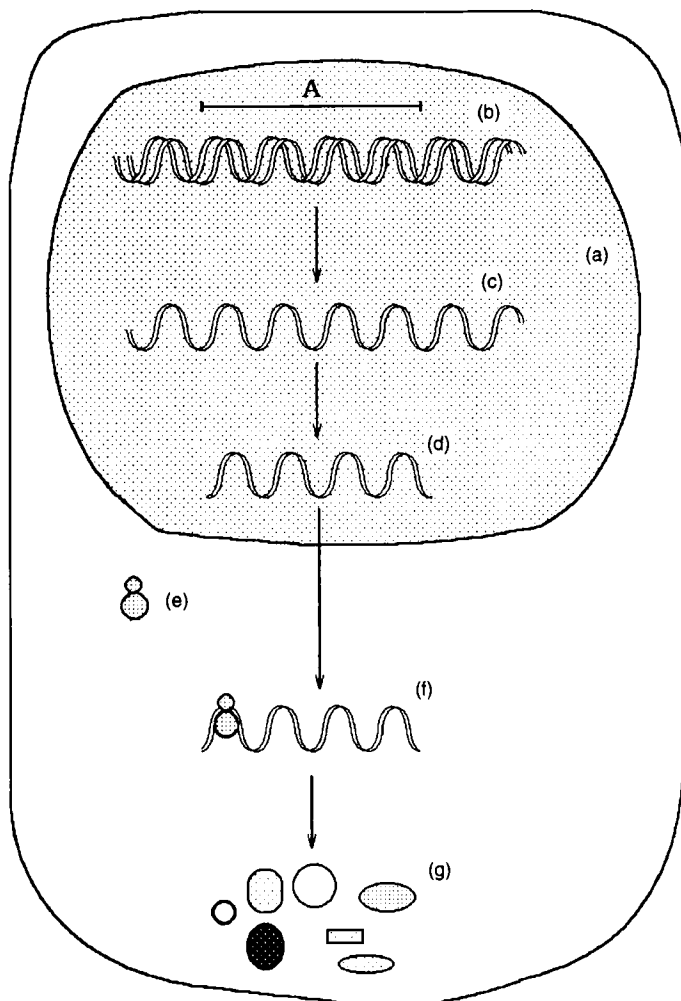


Figure 18 Schematic diagram of the information flow from DNA to protein in a eukaryotic cell and possible sites for oligonucleotide targeting. Shown are the nucleus (a) and DNA (b), where transcription from gene A leads to the formation of a specific RNA (c) which is processed to the mRNA precursor (d) and transferred to the cytoplasm where ribosomes (e) translate the messenger and build proteins from amino acids delivered by aminoacyl tRNAs (f). Different proteins (g) are synthesized by translation of the corresponding mRNAs. Oligonucleotides can be targeted to DNA by triplex formation or by hybridization to one of the DNA strands to arrest transcription of specific genes. Complementary oligonucleotides can bind to nascent RNAs that are interfering with the processing and transportation of mRNAs. Hybridization of oligonucleotides to mRNA can inhibit or completely arrest translation. The activity of ribosomes and tRNAs can be modulated by oligonucleotides that are complementary to accessible sites in the ribosomal RNAs and tRNAs. The activity of proteins involved in the splicing, transportation and translation of mRNAs can be affected by oligonucleotide inhibitors

Experiments with alkylating derivatives of oligonucleotides have revealed that these compounds can bind to some sequences of DNA which are naturally open in chromatin and in isolated cell nuclei [170]. These sequences can appear in the regions of DNA with high negative superhelicity, where the structure can open spontaneously. It has been shown that negative supercoiling of plasmids results in the opening of some sequences for oligonucleotide binding [171]. The development of oligonucleotide analogues which can form very tight complementary complexes and can displace one of the DNA strands [172–174] may allow easy targeting of any DNA sequence. It has been found that in certain conditions pyrimidine PNA oligomers can bind to DNA by strand invasion. As a result, a D-loop is formed in which the pyrimidine strand is free and the purine strand forms a very stable triple-stranded complex with two pyrimidine PNA molecules [172]. The D-loops induced by PNAs are sensitive to S1 nuclease, and if two PNA-binding sites on opposite DNA strands are located in close proximity, the DNA can be cleaved by S1 nuclease to produce sticky ends which can allow cloning of the fragments [174]. Binding of PNA oligomers to DNA inhibits the cleavage of DNA by restriction enzymes when PNA-binding sites overlap the restriction sites [173]. Transcription elongation is efficiently arrested at the PNA-binding sites.

Oligonucleotides can bind to nucleic-acid-related enzymes and positive and negative regulatory protein factors that interact with RNA and DNA [175]. Therefore they can serve as inhibitors and activators of specific gene expression. Apparent targets for oligonucleotide derivatives are polymerases of nucleic acids and regulatory proteins that recognize specific sequences in DNA. One of the proteins investigated so far is the ubiquitous factor Sp1, which stimulates transcription by Pol II from promoters that contain one or several hexanucleotide sequences such as GGGCGG. Double-stranded oligonucleotides containing the Sp1-binding sequence inhibit the transcription of the adenovirus E1b transcription unit, which is regulated by a single Sp1 site [176].

7.3 Experiments with Cellular Targets

Investigations of the natural uptake of oligonucleotide by cells [177–179] have shown that the compounds enter the endosomal compartment of the cell from where they escape to the cytoplasm with relatively poor efficiency (Figure 20). When the compounds reach the cytoplasm, they migrate rapidly to the nucleus [180]. The efficiency of oligonucleotide delivery into the cytoplasm can be greatly improved when the compounds are applied in complexes with cationic lipids. These lipids form complexes with nucleic acids that are capable of efficient binding to the negatively charged cell surface. Cationic liposomes enhance transfection efficiency dramatically; the most efficient of the lipids tested so far is lipospermin [181]. Use of cationic lipids reduces considerably the

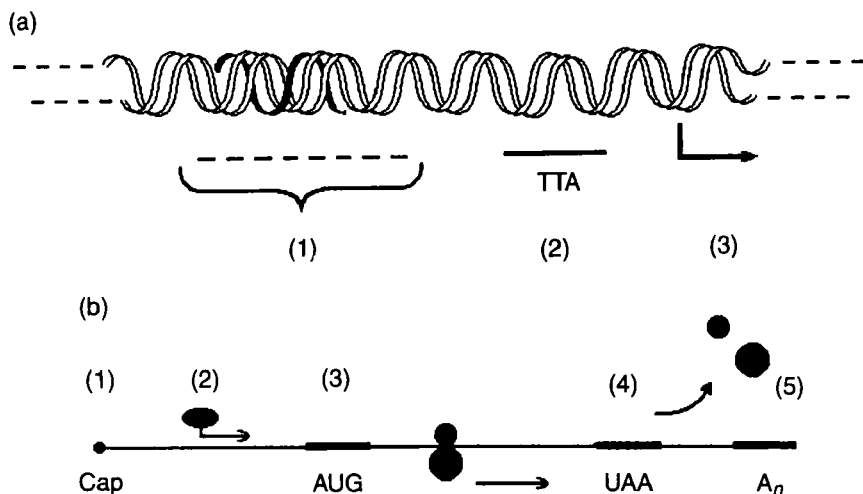


Figure 19 (a) Schematic illustration of the regulation of gene expression by oligonucleotides capable of triple-stranded complex formation. The oligonucleotide shown by a black band binds to a homopurine–homopyrimidine sequence (1) in the regulatory region upstream of the TTA box (2) in the vicinity of the transcription initiation site (3). If the bound oligonucleotide covers a sequence recognized by a protein regulator which activates or represses transcription, the process is inhibited or stimulated. If there is a homopurine–homopyrimidine sequence within the promoter region near the TTA box, the oligonucleotide can be targeted there to arrest transcription completely. (b) Schematic illustration of the translation process. A small ribosomal subunit (2) and protein initiation factors interact with the capped 5'-end of mRNA (1) and the active ribosome is assembled to start translation from the AUG codon (3). The process is terminated at the UAA codon (4), near to the terminal polyadenylated 3'-end of the messenger (5). The most efficient arrest of translation is achieved when the oligonucleotides hybridize to the 5'-end of the messenger, where the messenger has to have a biologically active conformation that is recognized by the initiation factors and ribosomes. Targeting to the initiation site of the messenger is only efficient when it results in cleavage of the messenger or the formation of crosslinks which cannot be overcome by ribosomes

concentration of oligonucleotide needed to achieve antisense inhibition of a specific cellular mRNA [50, 182]. Oligonucleotide uptake can be enhanced by conjugation to a lipophilic group that anchors the compound to the lipid cell wall. The most efficient lipophilic compound tested so far is cholesterol [183]. An important factor in increasing the cellular uptake of cholesterol-conjugated constructions in organisms and cell cultures in the presence of a complete growth medium is that they bind to plasma lipoproteins which are efficiently taken up by the cell through receptor-mediated endocytosis [184]. Conjugation of oligonucleotides to polycations capable of efficient binding to the negatively

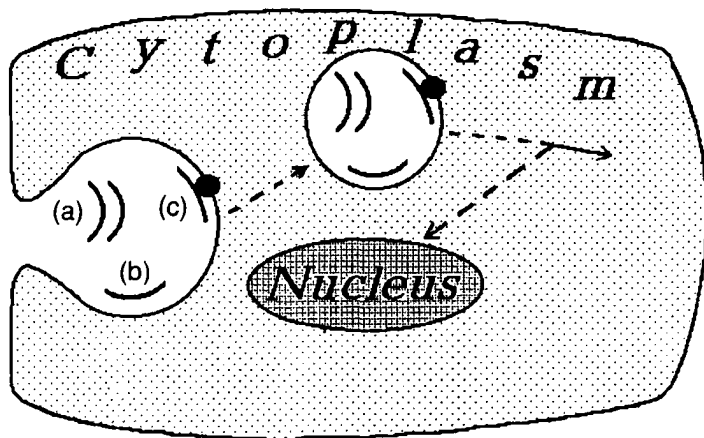


Figure 20 Schematic illustration of the interaction of oligonucleotides with a eukaryotic cell. Natural oligonucleotides and charged oligonucleotide analogues can enter cells as components of the solution in the process of liquid endocytosis (a). If the oligonucleotide derivative is absorbed at the cell surface (b), the uptake proceeds more efficiently (absorptive endocytosis). Very efficient uptake occurs if the derivatives interact with cell surface receptors (c). The oligonucleotides taken up are transferred inside the cell within membrane vesicle endosomes, from where the oligonucleotides can escape to the cytoplasm and further migrate into the nucleus

charged cell surface increases the efficiency of uptake. Most studies have been performed with the known polycationic drug carrier poly(L-lysine) [185]. In order to increase the uptake efficiency, oligonucleotides can be conjugated to compounds which bind to specific cellular receptors and are rapidly internalized by cells. Oligonucleotides can be conjugated to constructions bearing mannose and 6-phosphomannose which bind to cells that have specific sugar-binding membrane receptors such as lectins [186]. Modifications greatly enhance the uptake of oligonucleotides. Nonionic methylphosphonate and phosphotriester oligonucleotide analogues seem to be capable of passive diffusion through cellular membranes.

Potential chemotherapeutic applications require oligonucleotide analogues that are resistant to the nucleic-acid-degrading enzymes present in cells and in the bloodstream, of which the primary degrading activity is a 3'-exonuclease in serum. Considerable protection can be achieved by conjugation of different groups to the oligonucleotide at the 3'-terminus (capping) [187, 188]. Capping is not sufficient to prevent degradation of the compounds in those cells with high endonuclease activity. A radical solution to the problem consists in the

development of oligonucleotide analogues with modified internucleoside linkages. Ribose- and phosphate-modified oligonucleotide analogues such as α -oligonucleotides, phosphorothioate oligonucleotides and nonionic methylphosphonates and ethyl phosphotriesters of oligonucleotides are highly resistant to nucleolytic degradation.

When considering targeting cellular nucleic acids, one should take into account the fact that nucleic acids in the cell are always components of some nucleoprotein complexes. Thus, targeting of arbitrary sequences in DNA seems to be problematic because cellular DNA is a component of chromatin and many of the potential target sites in DNA can be expected to be masked by proteins. Some regions of cellular DNA are negatively supercoiled and transient opening of the DNA structure occurs in the course of transcription and translation; however, the possibility of targeting these sequences efficiently and the limitations of this approach remain to be investigated.

Incubation of eukaryotic cells with oligonucleotides that are complementary to specific mRNAs results in suppression of the synthesis of the corresponding proteins. The efficiency of the oligonucleotides correlates with their stability in cells and their ability to enter cells. Thus phosphorothioate analogues and methylphosphonate oligonucleotides require lower concentrations for producing the same inhibition effect as the normal deoxyribooligonucleotides.

Arrest of the biosynthesis of specific proteins has been used in the elucidation of their biochemical function. mRNAs that code for proteins taking part in cell proliferation are often targeted in order to develop approaches for controlling malignant cell growth. Oligonucleotides that are complementary to the mRNA of protooncogene *c-myc* enhance the differentiation of human promyelocytic leukemia HL 60 cells and inhibit the proliferation of these cells [189]. mRNAs of mutated protooncogenes and genes causing genetic disorders represent targets which differ from the normal mRNAs by a single nucleotide substitution. An example is the mRNA of the *ras* p21 protein which takes part in the cellular signal transduction pathway. A single mutation in either the twelfth or sixty-first codon of the cellular *ras* gene leads to the gene coding for a mutated protein which is directly involved in the development of the malignant cell phenotype. Methylphosphonate oligonucleotide analogues targeted to the mutated region of the *ras* mRNA overlapping the twelfth codon can cause 90% inhibition of the synthesis of the protein at a 50 μ M concentration [190]. Methylphosphonate oligonucleotides containing psoralene show high efficiency and specificity in inhibiting mutated *ras* mRNA translation [191].

The triple-helix approach has been used to arrest the function of certain cellular genes. A GT-motif oligonucleotide capable of binding to a sequence within the human *c-myc* promoter decreases selectively the steady-state concentration of *myc* mRNA [192]. An oligonucleotide targeted to a homopyrimidine-homopurine region in the promoter of the interleukin 2R α gene binds to the target *in vitro* and inhibits the synthesis of the corresponding mRNA in lymphocytes [193].

Double-stranded complexes of phosphodiester and phosphorothioate oligonucleotides representing binding sites for specific transcriptional regulatory proteins can bind to proteins in cells and interfere with the transcription of specific cellular genes. Oligonucleotides containing the κ B consensus or octamer sequences have been synthesized and they bind the nuclear factor (NF)- κ B or the octamer transcription factor, respectively [194]. These oligonucleotides inhibit the octamer-dependent activation of reporter genes in cultured cells.

The antiviral activity of oligonucleotides targeted to viral nucleic acids is well documented [195, 196]. Oligonucleotides inhibit virus proliferation by interfering with the life cycle of the virus in various ways. Besides the effects of oligonucleotides on the transcription and translation of virus-specific nucleic acids, oligonucleotides can prevent the processing and transportation of viral nucleic acids, inhibit viral enzymes, interfere with the packaging of virions and interfere with the virus-cell interaction. Inhibition of virus production in cell cultures is usually observed at micromolar concentrations of deoxyribo-oligonucleotides. Detailed studies of the effects of oligonucleotides on virus multiplication have been performed with HIV-1 [195].

Modifications that improve the ability of oligonucleotides to enter cells increase the antiviral potential of the compounds. Oligonucleotides conjugated to lipophilic groups such as cholesterol have demonstrated enhanced antiviral activity [197, 198]. However, cholesterol-conjugated control oligonucleotides having little or no base complementarity to the viral target are also effective [199]. Polylysine-conjugated oligonucleotides that are complementary to viral nucleic acids arrest the multiplication of the vesicular stomatitis virus and HIV-1 at concentrations much lower than the active concentrations determined for unmodified oligonucleotides [185]. Oligonucleotides conjugated to intercalating groups show enhanced antiviral activity as compared to the parent oligomers [200]. It should be noted that all the mentioned modifications protect the oligonucleotides against cellular nucleases, and this is an important factor that contributes to the enhancement of the antiviral effect.

In experiments with herpes simplex virus, HIV-1 and tick-borne encephalitis virus it has been found that the antiviral properties of oligonucleotides can be greatly improved by coupling them to photoreactive psoralene groups and alkylating groups [197, 201, 202].

Phosphorothioate oligonucleotides are promising antiviral agents [203]. However, it should be noted that considerable antiviral activity has been detected in experiments with oligonucleotides which apparently had no perfect complementary binding sites in virus RNAs. One explanation is that no oligonucleotide can be considered as a true control. One might expect the presence of nucleotide stretches in large viral RNAs which will form loose, imperfect complexes with oligonucleotides which can trigger degradation of the RNA by RNase H. Another source of the sequence-independent effect of the phosphorothioate oligonucleotides lies in their ability to inhibit DNA and RNA polymerases [203]. It has been found that the cellular receptor CD4, which is

the anchoring site for the human immunodeficiency virus HIV-1, interacts with oligonucleotides [204]. The ability of oligonucleotide analogues to bind to the CD4 receptor correlates with the anti-HIV activity of these compounds, suggesting that the anti-HIV effect is a result, at least in part, of their interfering with the virus–cellular CD4 interaction [205]. The possibility of inhibiting the transcription of HIV-1 in infected cells using oligonucleotides designed to form DNA triple helices within the virus promoter has been reported [206]. Acridine-linked oligothymidylates targeted to a homothymidine–homoadenosine region within the origin of replication of the SV 40 virus DNA inhibit virus multiplication in sensitive cells [207]. In animal tests it has been found that oligonucleotides administered by different routes distribute in all organs except for the brain and cause no acute toxicity at the concentrations needed to achieve biological effects. Oligonucleotides interact with immunoglobulins M and G in the bloodstream [152].

Oligonucleotides targeted to virus-specific nucleic acids have some curative activity in mice infected with herpes simplex and tick-borne encephalitis viruses [34, 208]. Subcutaneous perfusion of phosphodiester oligonucleotides targeted to the *N-myc* mRNA inhibits the growth of subcutaneous xenografts of *N-myc* that express neuroectodermal tumors [209]. These and other results of the first *in vivo* studies [210, 211] make one optimistic about the future therapeutic applications of oligonucleotides.

8. CONCLUSIONS

Oligonucleotides are versatile molecules that have a varied and expanding range of applications in various fields of research and development. At present, oligonucleotides are exceptionally well suited for the rational design of biologically active compounds and recognition elements needed for different molecular devices. The opportunities for development of oligonucleotides with novel properties have increased dramatically owing to the recent development of molecular selection approaches. This development has allowed oligonucleotides to challenge the monopoly enjoyed by antibodies as molecules for binding to others of different nature.

Oligonucleotide analogues and conjugates have the attributes necessary for pharmacological use. They provide an outstanding opportunity for rational drug design. Drug development using oligonucleotide-based technologies represents a revolutionary therapeutic strategy that directly targets and inhibits gene expression. The development of potential therapeutic applications of oligonucleotides faces a few problems which should be overcome by the design of appropriately modified oligonucleotide analogues and conjugates. It is still a challenge to chemists to design efficient, reactive oligonucleotide derivatives capable of efficient inactivation of target nucleic acids in organisms. Appropriate efforts will be required to develop chemical modifications or molecular carriers

for the delivery of oligonucleotides into cells. One can expect rapid development in this exciting research area.

9. REFERENCES

1. J. D. Watson and F. H. C. Crick, *Nature*, **171**, 737 (1953).
2. J. S. Cohen (ed.), *Oligodeoxynucleotides: Antisense Inhibitors of Gene Expression*, Macmillan Press, London, 1989.
3. E. Wickstrom (ed.) *Prospects for Antisense Nucleic Acids Therapy of Cancer and AIDS*, Wiley-Liss, New York, 1991.
4. D. G. Knorre, V. V. Vlassov, V. F. Zarytova, A. V. Lebedev and O. S. Fedorova, *Design and Targeted Reactions of Oligonucleotide Derivatives*, CRC Press, Boca Raton, FL, 1994.
5. S. T. Crooke and B. Lebleu (eds), *Antisense Research and Applications*, CRC Press, Boca Raton, FL, 1993.
6. A. D. Ellington and J. W. Szostak, *Nature*, **346**, 818 (1990).
7. C. Tuerk and L. Gold, *Science*, **249**, 505 (1990).
8. A. M. Belikova, V. F. Zarytova and N. I. Grineva, *Tetrahedron Lett.*, 3557 (1967).
9. J. Summerton, *J. Theor. Biol.*, **78**, 61 (1979).
10. M. L. Stephenson and P. C. Zamecnik, *Proc. Natl. Acad. Sci. USA*, **75**, 285 (1978).
11. P. S. Miller, L. Braiterman and P. O. P. Ts'o, *Biochemistry*, **16**, 1988 (1977).
12. D. G. Knorre and V. V. Vlassov, *Genetica*, **85**, 53 (1991).
13. C. Helene and J. J. Toulme, *Biochim. Biophys. Acta*, **1049**, 99 (1990).
14. J. Goodchild, *Bioconj. Chem.*, **1**, 165 (1990).
15. E. Uhlmann and A. Peyman, *Chem. Rev.*, **90**, 543 (1990).
16. W. Saenger, in *Principles of Nucleic Acid Structure*, Springer, New York, 1984.
17. P. B. Dervan, in *Oligonucleotides: Antisense Inhibitors of Gene Expression* (ed. J. S. Cohen), Macmillan Press, London, 1989, p. 197.
18. P. B. Beal and P. B. Dervan, *Science*, **251**, 1360 (1991).
19. K. Hoogsteen, *Acta Crystallogr.*, **16**, 907 (1963).
20. R. H. Durland, D. J. Kessler and M. Hogan, in *Prospects for Antisense Nucleic Acids Therapy of Cancer and AIDS* (ed. E. Wickstrom), Wiley-Liss, New York, 1991, p 219.
21. S. M. Mirkin, V. I. Lyamichev, K. N. Drushlyak, K. N. Dobrinin, S. A. Filippov and M. D. Frank-Kamenetskii, *Nature*, **330**, 495 (1987).
22. A. Dayn, G. M. Samadashwily and S. M. M. Mirkin, *Proc. Natl. Acad. Sci. USA*, **89**, 11 406 (1992).
23. O. N. Voloshin, A. G. Veselkov, B. P. Belotserkovskii, O. N. Danilevskaya, M. N. Pavlova, V. N. Dobrynin and M. D. Frank-Kamenetskii, *J. Biomol. Struct. Dyn.*, **9**, 643 (1992).
24. R. W. Roberts and D. M. Crothers, *Science*, **258**, 1463 (1992).
25. C. Kang, X. Zhang, R. Ratliff, R. Moyzis and A. Rich, *Nature*, **356**, 126 (1992).
26. O. F. Borisova, A. K. Shchyolkina, E. N. Timofeev and V. L. Florentiev, *FEBS Lett.*, **306**, 140 (1992).
27. E. Palacek, *Crit. Rev. Biochem. Mol. Biol.*, **26**, 151 (1991).
28. G. Yagil, *Crit. Rev. Biochem. Mol. Biol.*, **26**, 475 (1991).
29. A. I. H. Murchie, R. Bowater, F. Aboul-ela and M. J. Lilley, *Biochim. Biophys. Acta*, **1131**, 1 (1992).
30. R. D. Camerini-Otero and P. Hsieh, *Cell*, **73**, 217 (1993).
31. J.-L. Leroy, K. Gehring, A. Kettani and M. Gueron, *Biochemistry*, **32**, 6019 (1993).

32. K. Gehring, J.-L. Leroy and M. Gueron, *Nature*, **363**, 561 (1993).
33. H. Fritzsche, A. Akhebat, E. Taillandier, K. Rippe and T. M. Jovin, *Nucl. Acids Res.*, **21**, 5085 (1993).
34. P. S. Miller, P. O. P. Ts'o, R. I. Hogrefe, M. A. Reynolds and L. J. Arnold, in *Antisense Research and Applications* (eds S. T. Crooke and B. Lebleu), CRC Press, Boca Raton, FL, 1993, p. 189.
35. M. Matteucci, K.-Y. Lin, S. Butcher and S. Moulds, *J. Am. Chem. Soc.*, **113**, 7767 (1991).
36. M. H. Caruthers, in *Oligonucleotides: Antisense Inhibitors of Gene Expression* (ed. J. S. Cohen), Macmillan Press, London, 1989, p. 8.
37. G. Zon and T. G. Geiser, *Anti-Cancer Drug Des.*, **6**, 539 (1991).
38. E. P. Stirchak, J. E. Summerton and D. D. Weller, *Nucl. Acids Res.*, **17**, 6129 (1989).
39. M. H. Caruthers, G. Beaton, L. Cummins, D. Dellinger, D. Graff, Y.-X. Ma, W. S. Marshall, H. Sasmor, P. Shankland, J. V. Wu and E. K. Yan, *Nucleosides Nucleotides*, **10**, 47 (1991).
40. B. Froehler, P. Ng and M. Matteucci, *Nucl. Acids Res.*, **16**, 4831 (1988).
41. R. J. Jones, S. Swaminathan, J. F. Milligan, S. Wadwani, B. C. Froehler and M. D. Matteucci, *J. Am. Chem. Soc.*, **115**, 9816 (1993).
42. J.-J. Vasseur, F. Debart, Y. S. Sanghvi and P. D. Cook, *J. Am. Chem. Soc.*, **114**, 4006 (1992).
43. G. Paoletta, B. S. Sproat and A. I. Lamond, *EMBO J.*, **11**, 1913 (1992).
44. B. P. Monia, E. A. Lesnik, C. Gonzalez, W. F. Lima, D. McGee, C. J. Guinasso, A. M. Kawasaki, P. D. Cook and S. M. Freier, *J. Biol. Chem.*, **268**, 14 514 (1993).
45. B. Rayner, C. Malvy, J. Paoletti, B. Lebleu, C. Paoletti and J.-L. Imbach, in *Oligonucleotides: Antisense Inhibitors of Gene Expression* (ed. J. S. Cohen), Macmillan Press, London, 1989, p. 119.
46. J. S. Sun, C. Giovannangeli, J. C. Francois, R. Kurfurst, T. Montenay-Garestier, U. Asseline, T. Saison-Behmoaras N. T. Thuong and C. Helene, *Proc. Natl. Acad. Sci. USA*, **88**, 6023 (1991).
47. P. E. Nielsen, M. Egholm, R. H. Berg and O. Buchardt, in *Antisense Research and Applications* (eds S. T. Crooke and B. Lebleu), CRC Press, Boca Raton, FL, 1993, p. 363.
48. M. Egholm, O. Buchardt, L. Christensen, C. Behrens, S. M. Freier, D. A. Driver, R. H. Berg, S. K. Kim, B. Norden and P. E. Nielsen, *Nature*, **365**, 566 (1993).
49. Y. S. Sanghvi, in *Antisense Research and Applications* (eds S. T. Crooke and B. Lebleu), CRC Press, Boca Raton, FL, 1993, p. 273.
50. R. W. Wagner, M. D. Matteucci, J. G. Lewis, A. J. Gutierrez, C. Moulds and B. C. Froehler, *Science*, **260**, 1510 (1993).
51. G. M. Lamm, B. J. Blencowe, B. S. Sproat, A. M. Iribarren, U. Ryder and A. I. Lamond, *Nucl. Acids Res.*, **19**, 3193 (1991).
52. H. Hashimoto, M. G. Nelson and C. Switzer, *J. Am. Chem. Soc.*, **115**, 7128 (1993).
53. T. J. Povsic and P. B. Dervan, *J. Am. Chem. Soc.*, **111**, 3059 (1989).
54. M. C. Jetter and F. W. Hobbs, *Biochemistry*, **32**, 3249 (1993).
55. A. Ono, P. O. P. Ts'o and L.-S. Kan, *J. Am. Chem. Soc.*, **113**, 4032 (1991).
56. J. S. Koh and P. B. Dervan, *J. Am. Chem. Soc.*, **114**, 1470 (1992).
57. L. C. Griffin, L. L. Kiessling, P. A. Beal, P. Gillespie and P. B. Dervan, *J. Am. Chem. Soc.*, **114**, 7976 (1992).
58. H. U. Stilz and P. B. Dervan, *Biochemistry*, **32**, 2177 (1993).
59. J. F. Milligan, S. H. Krawczyk, S. Wadwani and M. D. Matteucci, *Nucl. Acids Res.*, **21**, 327 (1993).
60. M. Manoharan, in *Antisense Research and Applications* (eds S. T. Crooke and B. Lebleu), CRC Press, Boca Raton, FL, 1993, p. 303.

61. D. G. Knorre, V. V. Vlassov, V. F. Zarytova and A. V. Lebedev, *Sov. Sci. Rev. B Chem.*, **12**, 269 (1989).
62. D. R. Corey and P. G. Schultz, *Science*, **238**, 1401 (1987).
63. R. N. Zuckermann and P. G. Schultz, *J. Am. Chem. Soc.*, **110**, 6592 (1988).
64. R. W. Roberts and D. Crothers, *Proc. Natl. Acad. Sci. USA*, **88**, 9397 (1991).
65. M. P. Perelroyzen and A. V. Vologodskii, *Nucl. Acids Res.*, **16**, 4693 (1988).
66. S. M. Freier, in *Antisense Research and Applications* (eds S. T. Crooke and B. Lebleu), CRC Press, Boca Raton, FL, 1993, p. 67.
67. D. J. Ecker, in *Antisense Research and Applications* (eds S. T. Crooke and B. Lebleu), CRC Press, Boca Raton, FL, 1993, p. 387.
68. M. Chastain and I. Tinoco Jr, in *Antisense Research and Applications* (eds S. T. Crooke and B. Lebleu), CRC Press, Boca Raton, FL, 1993, p. 55.
69. J. R. Wyatt and I. Tinoco Jr, in *The RNA World*, Cold Spring Harbor Laboratory Press, 1993, p. 465.
70. M. Rougee, B. Faucon, J. L. Mergny, F. Barcelo, C. Giovannangeli, T. Montenay-Garestier and C. Helene, *Biochemistry*, **31**, 9269 (1992).
71. M. Cooney, G. Czernuszewicz, E. H. Postel, S. J. Flint and M. E. Hogan, *Science*, **241**, 456 (1988).
72. J.-L. Mergny, J.-S. Sun, M. Rougee, T. Montenay-Garestier, F. Barcelo, J. Chomilier and C. Helene, *Biochemistry*, **30**, 9791 (1991).
73. P. A. Beal and P. B. Dervan, *Nucl. Acids Res.*, **20**, 2773 (1992).
74. J. E. Hearst, *Annu. Rev. Phys. Chem.*, **39**, 291 (1988).
75. D. Herschlag, *Proc. Natl. Acad. Sci. USA*, **88**, 6921 (1991).
76. I. V. Kutyavin, M. A. Podyminogin, Yu. N. Bazhina, O. S. Fedorova, D. G. Knorre, A. S. Levina, S. V. Mamayev and V. F. Zarytova, *FEBS Lett.*, **238**, 35 (1988).
77. V. F. Zarytova, I. V. Kutyavin, S. V. Mamaev and M. A. Podyminogin, *Bioorg. Khim.*, **16**, 1653 (1990).
78. M. D. Distefano, J. A. Shin and P. B. Dervan, *J. Am. Chem. Soc.*, **113**, 5901 (1991).
79. J. Goodchild, *Nucl. Acids Res.*, **20**, 4607 (1992).
80. N. Colocci, M. D. Distefano and P. B. Dervan, *J. Am. Chem. Soc.*, **115**, 4468 (1993).
81. M. D. Distefano and P. B. Dervan, *J. Am. Chem. Soc.*, **114**, 11 006 (1992).
82. C. Giovannangeli, N. T. Thuong and C. Helene, *Proc. Natl. Acad. Sci. USA*, **90**, 10 013 (1993).
83. E. Brossalina and J.-J. Toulme, *J. Am. Chem. Soc.*, **115**, 796 (1993).
84. P. A. Beal and P. B. Dervan, *J. Am. Chem. Soc.*, **114**, 4976 (1992).
85. S. D. Jayasena and B. H. Johnston, *Biochemistry*, **32**, 2800 (1993).
86. U. English and D. H. Gauss, *Angew. Chem.*, **103**, 629 (1991).
87. T. Ito, C. L. Smith and C. R. Cantor, *Proc. Natl. Acad. Sci. USA*, **89**, 495 (1992).
88. H. O. Smith, K. Tabiti, G. Schaffner, D. Soldati, U. Albrecht and M. L. Birnstiel, *Proc. Natl. Acad. Sci. USA*, **88**, 9784 (1991).
89. J. Nielsen, S. Brenner and K. D. Janda, *J. Am. Chem. Soc.*, **115**, 9812 (1993).
90. M. C. Needels, D. G. Jones, E. H. Tate, G. L. Heinkel, L. M. Kochersperger, W. J. Dower, R. W. Barrett and M. A. Gallop, *Proc. Natl. Acad. Sci. USA*, **90**, 10 700 (1993).
91. S. A. Strobel and P. B. Dervan, *Nature*, **350**, 172 (1991).
92. L. E. Kotler, D. Zevin-Sonkin, I. A. Sobolev, A. D. Beskin and L. E. Ulanovsky, *Proc. Natl. Acad. Sci. USA*, **90**, 4241 (1993).
93. J. Kieleczawa, J. J. Dunn and F. W. Studier, *Science*, **258**, 1787 (1992).
94. K. R. Khrapko, V. P. Lysov, A. A. Khorlyn, V. V. Schick, V. L. Florentiev and A. D. Mirzabekov, *FEBS Lett.*, **256**, 118 (1989).
95. R. Drmanac, I. Labat, I. Brukner and R. Crkvenjakov, *Genomics*, **4**, 114 (1989).

96. R. Drmanac, S. Drmanac, Z. Strezoska, T. Paunesku, I. Labat, M. Zeremski, J. Snoddy, W. K. Funkhouser, B. Koop, L. Hood and R. Crkvenjakov, *Science*, **260**, 1649 (1993).
97. T. A. Kunkel, J. D. Roberts and R. A. Zakour, *Methods Enzymol.*, **154**, 367 (1987).
98. H. A. Erlich (ed.), *Principles and Applications for DNA Amplifications*, Stockton Press, New York, 1989.
99. M. Innis, D. Gelfand, J. Sninsky and T. White (eds), *PCR Protocols: A Guide to Methods and Applications*, Academic Press, New York, 1990.
100. N. Arnheim and C. H. Levenson, *Chem. Eng. News*, **68**, 36 (1990).
101. G. J. Nuovo, *PCR in situ Hybridization: Protocols and Applications*, Raven Press, New York, 1992.
102. J. W. Szostak and A. D. Ellington, in *The RNA World*, Cold Spring Harbor Laboratory Press, 1993, p. 511.
103. L. Gold, P. Allen, J. Binkley, D. Brown, D. Schneider, S. R. Eddy, C. Tuerk, L. Green, S. MacDougall and D. Tasset, in *The RNA World*, Cold Spring Harbor Laboratory Press, 1993, p. 497.
104. L. C. Bock, L. C. Griffin, J. A. Latham, E. H. Vermaas and J. J. Toole, *Nature*, **355**, 564 (1992).
105. D. Pei, H. D. Ulrich and P. G. Schultz, *Science*, **253**, 1408 (1991).
106. E. T. Peterson, J. Blank, M. Sprinzl and O. C. Uhlenbeck, *EMBO J.*, **12**, 2959 (1993).
107. A. A. Beaudry and G. F. Joyce, *Science*, **257**, 635 (1992).
108. K. Y. Wang, S. H. Krawczuk, N. Bischofberger, S. Swaminathan and P. H. Bolton, *Biochemistry*, **32**, 11 285 (1993).
109. D. J. Ecker, T. A. Vickers, R. Hanecak, V. Driver and K. Anderson, *Nucl. Acids Res.*, **21**, 1853 (1993).
110. W. C. J. Ross, *Biological Alkylating Agents*, Butterworths, London, 1962.
111. R. B. Meyer, J. C. Tabone, G. D. Hurst, T. M. Smith and H. Gamper, *J. Am. Chem. Soc.*, **111**, 8517 (1989).
112. T. J. Povsic and P. B. Dervan, *J. Am. Chem. Soc.*, **112**, 9428 (1990).
113. K. Kido, H. Inoue and E. Ohtsuka, *Nucl. Acids Res.*, **20**, 1339 (1992).
114. T. R. Webb and M. D. Matteucci, *J. Am. Chem. Soc.*, **108**, 2764 (1986).
115. R. I. Salganik, G. L. Dianov, E. B. Kokoza, L. P. Ovchinnikova, V. A. Kurbatov, A. A. Mustaev, A. A. Gall and G. V. Shishkin, *Mol. Biol. (Russia)*, **13**, 625 (1979).
116. J. Summerton and P. A. Bartlett, *J. Mol. Biol.*, **122**, 145 (1978).
117. B. L. Iverson and P. B. Dervan, *Proc. Natl. Acad. Sci. USA*, **85**, 4615 (1988).
118. D. G. Knorre and V. V. Vlassov, *Affinity Modification of Biopolymers*, CRC Press, Boca Raton, FL, 1989.
119. J. E. Hearst, *Annu. Rev. Biophys. Bioenerg.*, **10**, 69 (1981).
120. B. L. Lee, K. R. Blake and P. S. Miller, *Nucl. Acids Res.*, **16**, 10 681 (1988).
121. J. W. Tessman, S. Isaacs and J. E. Hearst, *Biochemistry*, **24**, 1669 (1985).
122. E. B. Brossalina, V. V. Vlassov and E. M. Ivanova, *Biokhimiya (Russia)*, **53**, 18 (1988).
123. D. S. Sergeev, V. F. Zarytova, S. V. Mamaev, T. S. Godovikova and V. V. Vlassov, *Antisense Res. Dev.*, **2**, 235 (1992).
124. R. Breslow, *Acc. Chem. Res.*, **24**, 317 (1991).
125. C.-H. Tung, Z. W. Wei, M. J. Leibowitz and S. Stein, *Proc. Natl. Acad. Sci. USA*, **89**, 7114 (1992).
126. J. Smith, K. Ariga and E. V. Anslyn, *J. Am. Chem. Soc.*, **115**, 362 (1993).
127. M. A. Podymingogin, V. V. Vlassov and R. Giege, *Nucl. Acids Res.*, **22**, 1 (1994).
128. O. C. Uhlenbeck, in *Antisense Research and Applications* (eds S. T. Croke and B. Lebleu), CRC Press, Boca Raton, FL, 1993, p. 83.

129. N. K. Tanner and M. Vasseur, in *Antisense Research and Applications* (eds S. T. Crooke and B. Lebleu), CRC Press, Boca Raton, FL, 1993, p. 415.
130. F. Benseler, D. Fu, J. Ludwig and L. W. McLaughlin, *J. Am. Chem. Soc.*, **115**, 8483 (1993).
131. S. M. Edgington, *Biotechnology*, **10**, 256 (1992).
132. V. V. Vlassov, V. F. Zarytova, I. V. Kutjavin, S. V. Mamaev and M. A. Podyminogin, *Nucl. Acids Res.*, **14**, 4065 (1986).
133. V. V. Vlassov, in *Antisense Research and Applications* (eds S. T. Crooke and B. Lebleu), CRC Press, Boca Raton, FL, 1993, p. 235.
134. N. V. Bulychev, Yu. N. Vorobjev, A. A. Gall, A. A. Koshkin and G. V. Shishkin, *Bioorg. Khim.*, **17**, 795 (1991).
135. E. B. Brossalina, E. M. Demchenko, V. V. Vlassov and S. V. Mamaev, *Antisense Res. Dev.*, **1**, 229 (1991).
136. M. A. Zenkova, G. G. Karpova, A. S. Levina, S. V. Mamaev and V. V. Soloviev, *Bioorg. Khim.*, **16**, 788 (1990).
137. T. M. Woolf, D. A. Melton and C. G. B. Jennings, *Proc. Natl. Acad. Sci. USA*, **89**, 7305 (1992).
138. A. Storey, D. Oates, L. Banks, L. Crawford and T. Crook, *Nucl. Acids Res.*, **19**, 4109 (1991).
139. D. Praseuth, A.-L. Guieysse, A. V. Itkes and C. Helene, *Antisense Res. Dev.*, **3**, 33 (1993).
140. T. J. Povsic, S. A. Strobel and P. B. Dervan, *J. Am. Chem. Soc.*, **114**, 5934 (1992).
141. D. Pei and P. G. Schultz, *J. Am. Chem. Soc.*, **113**, 9398 (1991).
142. E. Brossalina, E. Demchenko and V. V. Vlassov, *Nucl. Acids Res. Symp. Ser.*, **24**, 262 (1991).
143. D. M. Graifer, T. V. Venkstern, G. G. Karpova and I. A. Morozov, *Biopolymeri (Russia)*, **6**, 59 (1990).
144. D. G. Knorre and T. A. Chimitova, *FEBS Lett.*, **131**, 249 (1981).
145. V. V. Vlassov, D. G. Knorre, I. V. Kutjavin, S. V. Mamaev, L. M. Podust and O. S. Fedorova, *Bioorg. Khim.*, **13**, 1221 (1987).
146. S. F. Singleton and P. B. Dervan, *J. Am. Chem. Soc.*, **114**, 6957 (1992).
147. R. I. Salganik, G. L. Dianov, L. P. Ovchinnikova, E. N. Voronina, E. B. Kokoza and A. V. Mazin, *Proc. Natl. Acad. Sci. USA*, **77**, 2769 (1980).
148. P. A. Havre, E. J. Gunther, F. P. Gasparro and P. M. Glazer, *Proc. Natl. Acad. Sci. USA*, **90**, 7879 (1993).
149. I. V. Kutjavin, H. B. Gamper, A. A. Gall and R. B. Meyer, *J. Am. Chem. Soc.*, **115**, 9303 (1993).
150. L. A. Yakubov, E. A. Deeva, V. F. Zarytova, E. M. Ivanova, A. S. RYTE, L. V. Yurchenko and V. V. Vlassov, *Proc. Natl. Acad. Sci. USA*, **86**, 6454 (1989).
151. D. A. Geselowitz and L. M. Neckers, *Antisense Res. Dev.*, **2**, 17 (1992).
152. V. V. Vlassov, L. V. Pautova, E. Yu. Rykova and L. A. Yakubov, *Biokhimija (Russia)*, **58**, 1247 (1993).
153. D. A. Mundus, K. N. Bulygin, V. I. Yamkovoy, A. A. Malygin, M. N. Repkova, L. V. Vratskikh, A. G. Venjaminova, S. N. Vladimirov and G. G. Karpova, *Biochim. Biophys. Acta*, **1173**, 273 (1993).
154. T. I. Kolocheva, G. A. Nevinsky, V. A. Volchkova, A. S. Levina, V. V. Khomov and O. I. Lavrik, *FEBS Lett.*, **248**, 97 (1989).
155. K. Jayaraman, K. McParland, P. Miller and P. O. P. Ts'o, *Proc. Natl. Acad. Sci. USA*, **78**, 1537 (1981).
156. M. Trudel, J. Dondon, M. Grunberg-Manago, J. Finelli and R. H. Buckingham, *Biochimie*, **63**, 235 (1981).
157. P. Verspieren, A. W. C. A. Cornelissen, N. T. Thuong, C. Helene and J. J. Toulme, *Gene*, **61**, 307 (1987).

158. R. Y. Walder and J. A. Walder, *Proc. Natl. Acad. Sci. USA*, **85**, 5011 (1988).
159. P. F. Torrence, R. K. Maitra, K. Lesiak, S. Khamnei, A. Zhou and R. H. Silverman, *Proc. Natl. Acad. Sci. USA*, **90**, 300 (1993).
160. S. Agrawal, S. H. Mayrand, P. C. Zamecnik and T. Pederson, *Proc. Natl. Acad. Sci. USA*, **87**, 1401 (1990).
161. V. V. Vlassov, A. A. Godovikov, N. D. Kobets, A. S. RYTE, L. V. Yurchenko and A. G. Bukrinskaya, *Adv. Enz. Reg.*, **301** (1986).
162. V. V. Vlassov, A. A. Godovikov V. F. Zarytova, E. M. Ivanova and N. Yu. Nomokonova, *Mol. Biol. (Russia)*, **24**, 173 (1990).
163. J. M. Kean, A. Murakami, K. R. Blake, C. D. Cushman and P. S. Miller, *Biochemistry*, **27**, 9113 (1988).
164. V. V. Vlassov, V. F. Zarytova, I. V. Kutyavin and S. V. Mamaev, *FEBS Lett.*, **231**, 352 (1988).
165. B. C. F. Chu and L. E. Orgel, *Nucl. Acids Res.*, **17**, 4783 (1989).
166. L. J. Maher, P. Dervan and B. Wold, in *Antisense Research and Applications* (eds S. T. Crooke and B. Lebleu), CRC Press, Boca Raton, FL, 1993, p. 227.
167. J. M. Chubb and M. E. Hogan, *Trends Biotechnol.*, **10**, 132 (1992).
168. M. Grigoriev, D. Praseuth, A. L. Guieysse, P. Robin, N. T. Thuong, C. Helene and A. Harel-Bellan, *Proc. Natl. Acad. Sci. USA*, **90**, 3501 (1993).
169. S. L. Young, S. H. Krawczyk, M. D. Matteucci and J. J. Toole, *Proc. Natl. Acad. Sci. USA*, **88**, 10 023 (1991).
170. V. V. Vlassov, N. D. Kobetz, E. L. Chernolovskaya, S. G. Demidova, R. G. Borissov and E. M. Ivanova, *Mol. Biol. Rep.*, **14**, 11 (1990).
171. V. V. Vlassov, S. A. Gaidamakov, V. F. Zarytova, D. G. Knorre, A. S. Levina, A. A. Nikonova, L. M. Podust and O. S. Fedorova, *Gene*, **72**, 313 (1988).
172. D. Y. Cherny, B. P. Belotserkovskii, M. D. Frank-Kamenetskii, M. Egholm, O. Buchardt, R. H. Berg and P. E. Nielsen, *Proc. Natl. Acad. Sci. USA*, **90**, 1667 (1993).
173. P. E. Nielsen, M. Egholm, R. H. Berg and O. Buchardt, *Nucl. Acids Res.*, **21**, 197 (1993).
174. V. Demidov, M. D. Frank-Kamenetskii, M. Egholm, O. Buchardt and P. E. Nielsen, *Nucl. Acids Res.*, **21**, 2103 (1993).
175. C. Clusel, E. Ugarte, N. Enjolras, M. Vasseur and M. Blumenfeld, *Nucl. Acids Res.*, **21**, 3405 (1993).
176. H. Wu, J. S. Holcenberg, J. Tomich, J. Chen, P. A. Jones, S.-H. Huang and K. L. Calame, *Gene*, **89**, 203 (1990).
177. V. V. Vlassov and L. A. Yakubov, in *Prospects for Antisense Nucleic Acids Therapy of Cancer and AIDS* (ed. E. Wickstrom), Wiley-Liss, New York, 1991, p. 243.
178. S. Akhtar (ed.), *Delivery Systems for Antisense Oligonucleotide Therapeutics*, CRC Press, Boca Raton, FL, 1994.
179. S. Akhtar, Y. Shoji and R. L. Juliano, in *Gene Regulation Biology of Antisense RNA and DNA* (eds R. P. Erickson and J. G. Izant), Raven Press, New York, 1992.
180. J. P. Leonetti, N. Mechti, G. Degols, C. Gagnor and B. Lebleu, *Proc. Natl. Acad. Sci. USA*, **88**, 2701 (1991).
181. J.-P. Behr, B. Demeneix, J.-P. Loeffler and J. Perez-Mutul, *Proc. Natl. Acad. Sci. USA*, **86**, 6982 (1989).
182. C. F. Bennet, M.-Y. Chiang, H. Chan, J. E. Shoemaker and C. K. Mirabelli, *Mol. Pharmacol.*, **41**, 1023 (1992).
183. A. S. Boutorin, L. V. Guskova, E. M. Ivanova, N. D. Kobetz, V. F. Zarytova, A. S. RYTE, L. V. Yurchenko and V. V. Vlassov, *FEBS Lett.*, **254**, 129 (1989).
184. P. C. Smidt, T. L. Doan, S. Falco and T. J. C. van Berkel, *Nucl. Acids Res.*, **19**, 4695 (1991).
185. J.-P. Leonetti, G. Degols and B. Lebleu, *Bioconj. Chem.*, **1**, 149 (1990).

186. E. Bonfils, C. Depierreux, P. Midoux, N. T. Thuong, M. Monsigny and A. C. Roche, *Nucl. Acids Res.*, **20**, 4621 (1992).
187. J. G. Zengdegui, K. M. Vasques, J. H. Tinsley, D. J. Kessler and M. E. Hogan, *Nucl. Acids Res.*, **20**, 307 (1992).
188. J. F. R. Ortigao, H. Rosch, H. Selter, A. Frohlich, A. Lorenz, M. Montenarh and H. Seliger, *Antisense Res. Dev.*, **2**, 129 (1992).
189. T. A. Bacon and E. Wickstrom, *Oncogene Res.*, **6**, 21 (1991).
190. D. Brown, Z. Yu, P. Miller, K. Blake, C. Wei, H.-F. Kung, R. J. Black, P. O. P. Ts'o and E. H. Chang, *Oncogene Res.*, **4**, 243 (1989).
191. E. H. Chang, P. S. Miller, C. Cushman, K. Devadas, K. F. Pirollo, P. O. P. Ts'o and Z. P. Yu, *Biochemistry*, **30** (1991).
192. E. H. Postel, S. J. Flint, D. J. Kessler and M. E. Hogan, *Proc. Natl. Acad. Sci. USA*, **88**, 8227 (1991).
193. F. M. Orson, D. W. Thomas, W. M. McShan, D. J. Kessler and M. E. Hogan, *Nucl. Acids Res.*, **19**, 3435 (1991).
194. A. Bielinska, R. A. Shidvasani, L. Zhang and G. J. Nabel, *Science*, **250**, 997 (1990).
195. S. Agrawal, in *Prospects for Antisense Nucleic Acids Therapy of Cancer and AIDS* (ed. E. Wickstrom), Wiley-Liss, New York, 1991, p. 143.
196. C. A. Stein and Y.-C. Cheng, *Science*, **261**, 1004 (1993).
197. T. V. Abramova, V. M. Blinov, V. V. Vlassov, V. V. Gorn, V. F. Zarytova, E. M. Ivanova, D. A. Konevets, O. A. Plyasunova, A. G. Pokrovski, L. S. Sandakhchiev, F. P. Svinarchuk, V. P. Starostin and S. P. Chaplygina, *Nucleotides Nucleosides*, **10**, 419 (1991).
198. R. L. Letsinger, G. Zhang, D. K. Sun, T. Ikeuchi and P. S. Sarin, *Proc. Natl. Acad. Sci. USA*, **86**, 6553 (1989).
199. C. A. Stein, R. Pal, A. L. DeViko, G. Hoke, S. Mumbauer, O. Kinstler, M. G. Sarngadharan and R. L. Letsinger, *Biochemistry*, **30**, 2439 (1991).
200. A. Zerial, N. T. Thuong and C. Helene, *Nucl. Acids Res.*, **15**, 9909 (1987).
201. V. V. Pogodina, T. V. Frolova, T. V. Abramova, V. V. Vlassov, E. M. Ivanova, I. V. Kutiavin, A. G. Pletnev and L. A. Yakubov, *Dokl. Akad. Nauk SSSR*, **301**, 1257 (1988).
202. A. Jacob, G. Duval-Valentin, D. Ingrand, N. T. Thuong and C. Helene, *Eur. J. Biochem.*, **216**, 19 (1993).
203. A. M. Krieg, J. Tonkinson, S. Matson, Q. Zhao, M. Saxon, L.-M. Zhang, U. Bhanja, L. Yakubov and C. A. Stein, *Proc. Natl. Acad. Sci. USA*, **90**, 1048 (1993).
204. L. Yakubov, Z. Khaled, L. M. Zhang, A. Truneh, V. V. Vlassov and C. A. Stein, *J. Biol. Chem.*, **268**, 18 818 (1993).
205. W. Y. Gao, R. N. Hanes, M. A. Vazquez-Padua, C. A. Stein, J. S. Cohen and Y. C. Cheng, *Antimicrob. Agents Chemother.*, **34**, 808 (1990).
206. W. M. McShan, R. D. Rossen, A. H. Laughter, J. A. Trial, D. J. Kessler, J. G. Zengdegui, M. E. Hogan and F. M. Orson, *J. Biol. Chem.*, **267**, 5712 (1992).
207. F. Birg, D. Praseuth, A. Zerial, N. T. Thuong, U. Asseline, T. LeDoan and C. Helene, *Nucl. Acids Res.*, **18**, 2901 (1990).
208. V. V. Pogodina, T. V. Frolova, M. P. Frolova, T. V. Abramova, V. V. Vlassov, D. G. Knorre, A. G. Pletnev and L. A. Yakubov, *Dokl. Akad. Nauk SSSR*, **308**, 237 (1989).
209. L. Whitesell, A. Rosolen and L. M. Neckers, *Antisense Res. Dev.*, **1**, 343 (1991).
210. R. M. Burch and L. C. Mahan, *J. Clin. Invest.*, **88**, 1190 (1991).
211. M. Simons, E. R. Edelman, J.-L. DeKeyser, R. Langer and R. D. Rosenberg, *Nature*, **359**, 67 (1992).

Chapter 5

Macrocycles and Antibodies as Catalysts

D. B. SMITHRUD AND S. J. BENKOVIC

Pennsylvania State University, University Park, PA, USA

1. INTRODUCTION

Nature uses enzymes to catalyze a variety of biologically important reactions. Through their study, scientists have derived theories for how enzymes recognize their substrates and enhance product formation. Not to be outdone by nature, chemists have made attempts to design and produce their own catalysts based on a presumed understanding that first, enzymes provide a binding pocket where recognition of substrates depends on noncovalent interactions including ionic bonds, hydrogen bonds and van der Waals forces; and second, catalysis follows as a consequence of stronger binding of the transition state of the reaction relative to the binding of the substrate [1]. The product is then released (this is the rate-limiting step for many enzymes), freeing the enzyme to bind another substrate molecule and repeat the process. Therefore, if one was able to understand and consequently control the binding interactions between molecules, catalysts might be produced for many reactions.

The purpose of this article is to discuss the successes and present shortcomings of using macrocycles and antibodies as catalysts in a variety of reactions. We hope the informed reader may decide whether chemists have been indeed able to design and produce materials that are as efficient as the enzymes. Of course, the answer to this question depends on the definition of efficiency. In a recent article [2], we established a set of guidelines to determine if the catalytic rate, defined in terms of k_{cat} and K_{m} , is sufficient to produce the desired product in

high yield in the face of limitations imposed by competing reactions, and quantity of catalyst available (k_{cat} is the rate of product formation when the catalyst is saturated with substrate and K_{m} is the concentration of substrate that produces half of the maximal rate). For a single-substrate reaction as depicted in Figure 1, or for multisubstrate reactions, the amount of product formed can be simulated using values for k_{cat} , k_{uncat} , K_{m} and the concentrations of the components. Time courses were derived by varying both the $k_{\text{cat}}/k_{\text{uncat}}$ ratio and K_{m} using fixed concentrations of substrate (100 μM) and catalyst (1 μM) and setting a value for k_{uncat} ($1 \times 10^{-6} \text{ s}^{-1}$). The results suggested that effective catalysts should have $k_{\text{cat}}/k_{\text{uncat}}$ ratios $> 10^4$ and $K_{\text{m}} < 0.1 \text{ mM}$ for both types of reactions and not be subject to strong product inhibition. These values should be kept in mind when examining the characteristics of the catalysis for the reactions to be discussed.

2. CONVERTING MACROCYCLES INTO CATALYSTS

Host-guest chemistry has been a rapidly growing field since its inception with the development of crown ethers, which Pedersen [3] and others showed were capable of binding and transporting cations. Results from studies using crown ethers along with cyclodextrins, natural products composed of linked glucose units capable of binding organic compounds, inspired chemists to design and synthesize a wide range of host molecules such as spherands, cryptands and cyclophanes. Thermodynamic values for the complexation between hosts and a broad range of guest molecules have shown that their tight association derives from complementary interactions such as oppositely charged groups, hydrogen bond donors and acceptors, or van der Waals forces that stem from fluctuations in electron densities. Solvation and desolvation also play an important role in association, since their net balance is a major factor in the thermodynamic binding energy.

It seems reasonable to presume that catalysts might be produced by adding functional groups that are known to be catalytic residues in enzymes to the various binding hosts. In this regard, cyclodextrins were initially advantageous

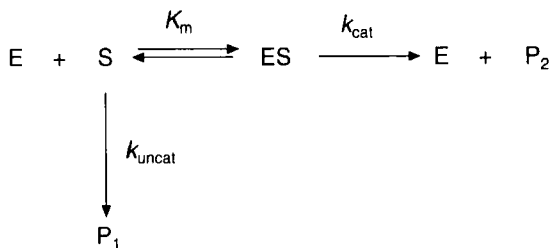


Figure 1

because their secondary hydroxy groups provided for an easy means of attachment. For example, Anslyn and Breslow [4] chose to add imidazole groups to β -cyclodextrin in an attempt to mimic RNases, which are capable of hydrolyzing the phosphate backbone of RNA. Ribonuclease A uses imidazoles from two histidines and an ammonium residue from lysine to cleave phosphodiester in a two-step process with a cyclic phosphate as an intermediate. The derivatized β -cyclodextrin (**1a**) (Table 1) catalyzed the opening of the cyclic phosphate of *t*-butylcatechol, requiring a 5:1 ratio of (**1a**) to phosphate to obtain saturation kinetics. The positioning of the imidazoles proved important, with an adjacent geometry being more effective than having the imidazoles placed on opposite sides of the cavity. Different orientations, however, did not alter the effect of pH. Studies performed on each cyclodextrin showed a bell-shaped pH profile for the rate of hydrolysis as a function of pH, suggesting that the imidazoles operated as acid–base catalysts. Furthermore, a proton inventory for the hydrolysis of the cyclic phosphate by (**1a**) found that two hydrogens were involved in phosphate cleaving.

Metals were combined with the cyclodextrins to enhance their catalytic ability. By attaching diethylenetriamine, (**1b**) was able to bind zinc, which in turn enhanced the cleavage of the cyclic phosphate of adenosine by a factor of five over the Zn^{2+} –triamine complex alone [5]. Metals have also been used to dimerize cyclodextrins [6a, 6b], which produced stronger-binding hosts and aligned the metal under the scissile bond. Dimerization gave extremely large association constants of the order of 10^{10} to 10^{11} for selective substrates, but the hosts were not catalytic. It should also be mentioned that strong binding can produce substantial product inhibition for substrates whose recognition elements do not appreciably change upon conversion to product.

An attempt was also made to mimic ATPases by using the polyammonium macrocycle (**2**) to catalyze the dephosphorylation of ATP, and, contrary to the cyclodextrins, it was more effective without added functional groups [7]. The negatively charged oxygen atoms of the phosphates were bound by the four positively charged amines (protonated at pH 7), and attack occurred by the free amine on the terminal phosphate. This macrocycle was capable of Michaelis–Menten kinetics, forming product at a rate of $0.064 \mu\text{M min}^{-1}$ with (**2**) at a concentration of $1 \mu\text{M}$.

A host capable of product turnover was developed by Rebek and coworkers [8a, 8b], who retooled their molecular clefts by adding converging carboxylic acids to form cavities that resemble the active sites of lysozyme and the aspartic proteases. By suspending the substrate over the aromatic amine, the amine serves either to donate a proton in the hemiacetal cleavage of glycoaldehyde dimer (**3b**) or to abstract a proton in the enolization of quinuclidinone (**3c**). For the dissociation of the glycoaldehyde dimer, 0.98 mM min^{-1} of the monomer was produced in the presence of $2.5 \mu\text{M}$ of (**3a**). As observed with (**1a**), the positioning of the attached groups in the catalyst proved important: the convergence of the carboxylic acids provided a more effective geometry, showing

Table 1

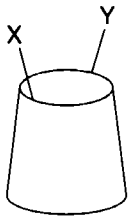
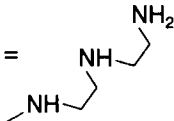
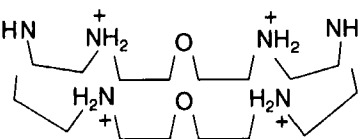
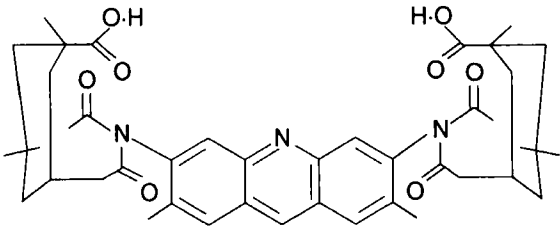
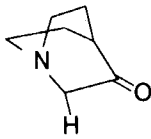
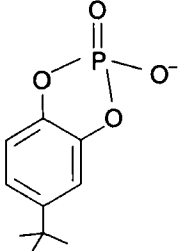
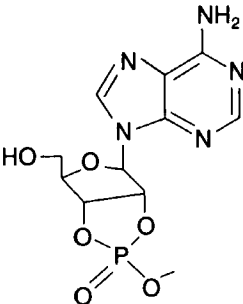
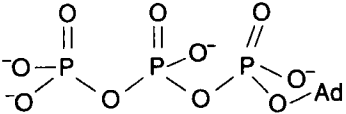
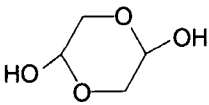
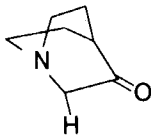
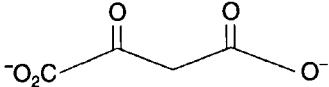
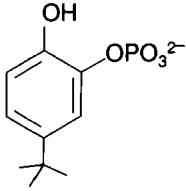
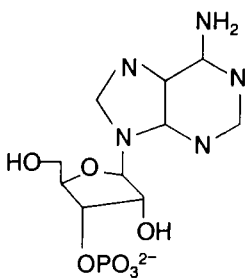
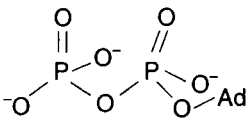
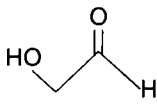
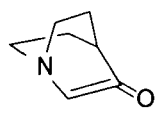
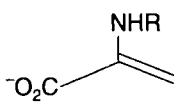
Catalyst	Substrate
 (1a): $X=Y=$  (1b): $X=OH, Y=$  (2)  (3a)  Cyclic peptide 	   (3b)  (3c)  

Table 1 (continued)

Product	$k_{\text{cat}}/k_{\text{uncat}}$	K_m (μM)
	—	—
	—	—
 (Ad=Adenosine)	256	100
	—	—
	833	6000
	423	8700

an increased rate of about 1000 compared to their divergence within the cleft. A more complete kinetic study was possible for the enolization of quinuclidinone because of its greater solubility. Michaelis–Menten kinetics were observed with about 5 mM s^{-1} of product being produced using a 2.5 mM solution of catalyst, which is a rate enhancement of 833 over background. More importantly, catalysis by these molecules showed turnover in both the cleavage of the hemiacetal and the enolization of quinuclidinone.

Recently, a macrocycle has been developed that is composed entirely of amino acids. Benner *et al.* [9] linked amino acids to form a cyclic host designed to catalyze the decarboxylation of oxaloacetate according to the mechanism provided in Figure 2. Although the catalyst was synthesized as a ring, its actual form in solution is an α -helix, which stems from the tertiary structure available from the polypeptides. Amino acids were chosen to be both catalytically active and to produce the α -helix required to align these residues in the correct geometry to form imine (5). An amino group at the N-terminus of the cyclic peptide was available to form the imine. Its activity at low pH was enhanced through the lowering of its $\text{p}K_a$ by interactions with the dipole moment of the α -helix. Positively charged amino acid side chains act to stabilize the carbolamine (4), the highest-energy intermediate. Michaelis–Menten catalysis was observed with this catalyst even though a well-defined binding pocket was not produced; the catalyst formed aggregates of helical bundles. Effective binding, as measured by K_m , of oxaloacetate was extremely weak, it being bound on the millimolar level, and the product was produced at a low rate of 0.07 mM min^{-1} using 0.2 mM of peptide.

3. DEVELOPMENT OF CATALYTIC ANTIBODIES

In the immune system, the role of an antibody is to bind a foreign molecule, thus triggering the removal of the complex and providing a means to purge the body of foreign substances. Therefore, antibodies have the potential to recognize

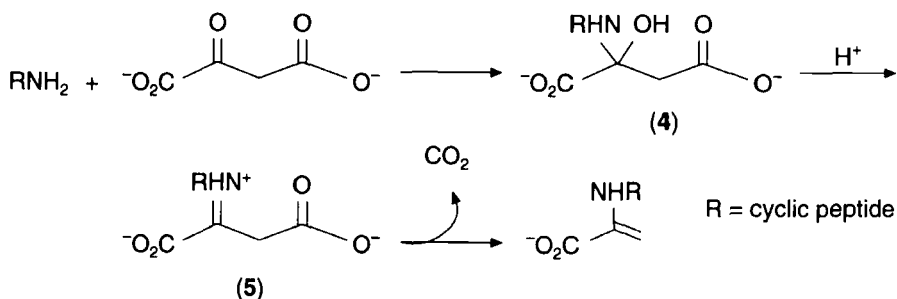


Figure 2 Proposed mechanism for the decarboxylation of oxaloacetate using a cyclic peptide as a catalyst

a large variety of compounds with high affinity, displaying binding constants of the order of 10^9 . If the responsible interactions were directed at transition state mimics, catalytic antibodies might be induced in response to a hapten that is a transition state analogue of a particular reaction [10]. Besides furnishing a direct route into designed protein catalysis, the production of antibodies gives a very large repertoire of antibodies, thus increasing the odds for finding a catalyst.

Some problems have been encountered in the genesis of the catalytic antibody field, specifically in the development of a strong immune response and a method to screen for catalytically active antibodies. Not all molecules will be recognized as a result of resemblance to self-compounds or because of their small size. To obtain an immune response against smaller haptens, they have to be conjugated to a carrier protein. Larger antigens such as proteins will be immunogenic if they are not natural to the animal. In a mature immune response, generally about 10^8 different antibodies are produced, which increases to 10^9 members upon somatic mutation. The screening method for monoclonal antibodies consists of determining whether the unique antibody produced by a given hybridoma cell binds the hapten, but there is no general direct method for screening a large number of cells for catalysis.

Structural information on antigen-antibody interactions is needed for successful hapten design, and this has been determined mainly through X-ray analysis. Crystal structures of both antibodies with and without bound antigens have been solved. Bound antigens include small organic molecules, proteins and steroids. The binding motif of antibodies is a β -barrel composed of six hypervariable loops, three from the heavy chain and three from the light chain. Differences in the amino acid compositions of the hypervariable loops give antibodies the ability to bind a large variety of compounds, but the basic binding motif is preserved. Furthermore, the flexibility of the loops also depends on their composition, although generally there appears to be only a small amount of structural change upon hapten binding. Therefore, it seems that antibody catalysis may not benefit from the induced fit mechanisms found with enzymes. The physical characteristics of the binding site also vary according to antigen size. Small haptens, in general, are embedded within pockets or clefts on the antibody surface, whereas long, polymeric antigens lie in elongated grooves. The maximum size of the binding site is somewhat controversial, but for proteins about 700–900 Å ($1 \text{ Å} = 0.1 \text{ nm}$) of protein surface area is buried, which is about seven amino acid residues [11, 12].

Initial studies revealed that catalytic antibodies do behave like enzymes in processing their substrates, not exactly like highly evolved enzymes but rather more primitive versions. In the following selected examples, the catalytic antibodies that were developed not only accelerated the desired reaction but also showed specificity for substrates that most closely resembled the hapten. The hapten itself acts as a competitive inhibitor to the reaction, a key test that the observed activity stems from the antibody. Haptens generally bind several

Table 2

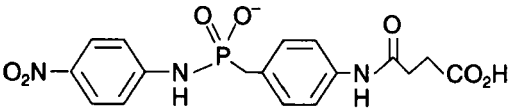
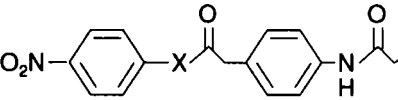
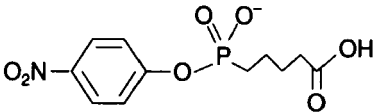
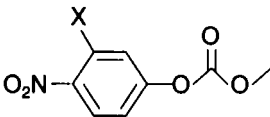
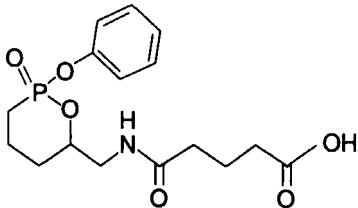
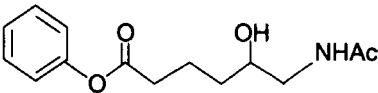
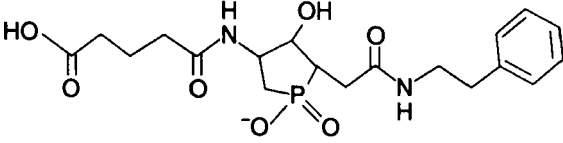
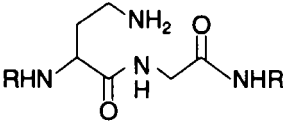
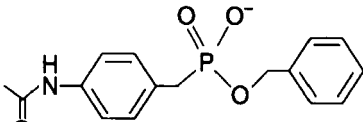
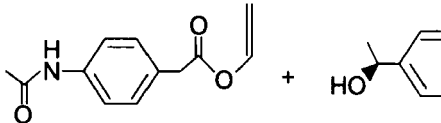
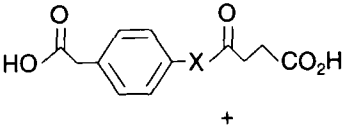
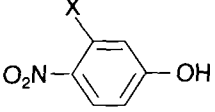
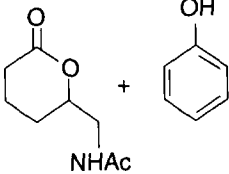
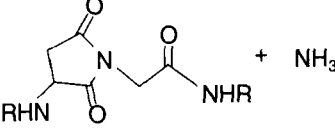
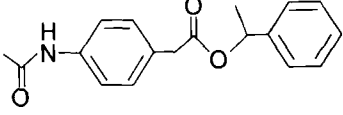
Hapten	Substrate(s)
 (6)	 (7a): X = O (7b): X = NH
 (8)	 (9a): X = H (9b): X = Me
 (10)	
 (11)	
 (12)	

Table 2 (continued)

Product(s)	$k_{\text{cat}}/k_{\text{uncat}}$	K_{m} (μM)
  	(7a): 2.7×10^4 (7b): 1.5×10^5 (9a): 810 (9b): 2300	53 377 660 7000
	167	76
	70	193
	$\geq 10^6$ (estimated)	3000 (ester) 7300 (alcohol)

orders of magnitude more tightly than the analogous substrates. The fact that hapten binding correlates with antibody turnover suggests that rate enhancements are derived from preferential binding of the transition state, in accord with the hypothesis explaining enzyme catalysis [1]. Finally, unlike many macrocycles, most catalytic antibodies give multiple turnovers, making them a practical means for obtaining enhanced product formation.

4. CATALYTIC ANTIBODIES SHOW ENZYME CHARACTERISTICS

Initially, several research groups chose the hydrolysis of various carbonyl compounds (mostly esters) that are cleaved by enzymes. Hapten design featured a phosphonate derivative because this moiety resembles the proposed tetrahedral transition state of hydrolysis. Furthermore, phosphonates have been shown to be effective 'transition state' inhibitors of proteolytic enzymes. One study [13] used phosphonate (**6**) (Table 2) as a transition state analogue to induce an antibody capable of catalyzing the hydrolysis of esters and anilides. Besides mimicking the proposed geometry of the hydrolytic transition state, the negatively charged oxygen atom of the phosphonate should produce complementary positively charged or hydrogen-bonding amino acid residues within the binding pocket. Substrates that bind to the antibody will experience this charge stabilization, thus further lowering the energy of the transition state for hydrolysis. The phenyl groups within the hapten increase its antigenicity, thus ensuring a strong immune response.

Hapten (**6**) led to the induction of the catalytic antibody NPN43C9. The rate of hydrolysis of the 4-nitrophenylester substrate (**7a**) by NPN43C9 is impressive, with a rate acceleration of 10^6 over background. This value is within a factor of about 500 of the catalytic rate for chymotrypsin with similar substrates. A pH-rate profile revealed a pH-independent step assigned to product release (4-nitrophenol) at $\text{pH} > 9$ and a step assigned to hydroxide attack at low pH. From kinetic investigation of a series of substrates hydrolyzed by NPN43C9, a multistep pathway with an acyl-bound intermediate was proposed (Figure 3) [14].

Evaluation of the rates for the individual steps of the mechanism for the hydrolysis of (**7a**) and (**7b**) revealed a striking difference between the way antibodies and enzymes process their substrates. When the rates are converted to free energies and plotted as a reaction coordinate diagram (Figure 4), it is evident that, unlike enzymes, the populations of the bound intermediates are not equivalent nor are the various transition states isoenergetic. NPN43C9, in particular, binds the phenolic product tightly, resulting in a reduction in the overall catalytic ability of the antibody. To date, tight binding of the product has been an inherent disability of antibodies.

Antibody catalysis of carbonate hydrolysis was also investigated concurrently [15]. Once again, the hapten (**8**) contains a phosphonate to represent the proposed

tetrahedral geometry of the carbonyl carbon in the transition state for carbonate hydrolysis. The rate acceleration was modest, with an enhancement of 810 over the background hydrolysis rate by hydroxide ion at pH 8.5. The investigation of the hydrolysis of various substrates by the antibody revealed that specificity for the substrate was accomplished by tighter binding of the substrate and not by a faster rate of hydrolysis. At low substrate concentrations, a greater amount of product was produced for the hydrolysis of methyl 4-nitrophenyl carbonate

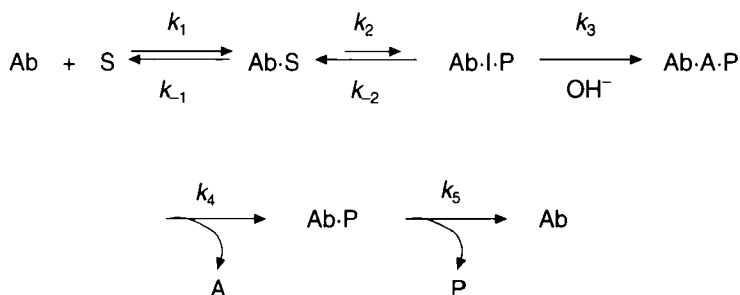


Figure 3 Proposed kinetic scheme for the hydrolysis of various substrates by NPN43C9, where Ab = NPN43C9, S = substrate, I = acyl intermediate, A = acidic product and P = phenolic product

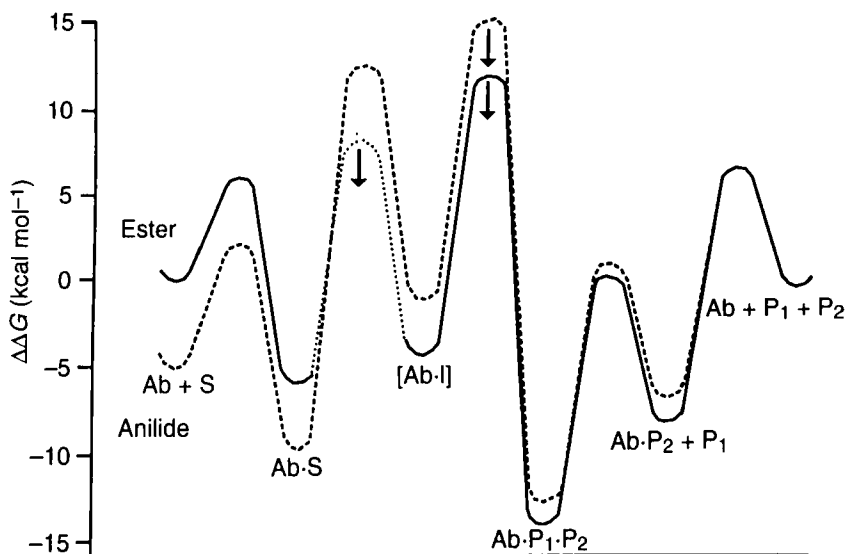
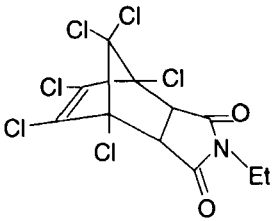
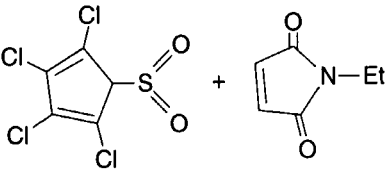
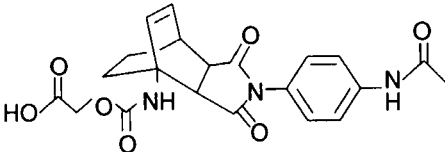
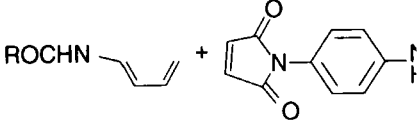
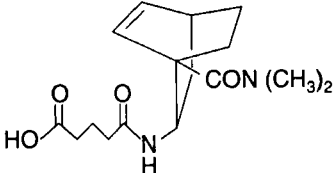
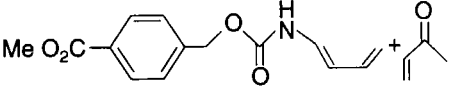
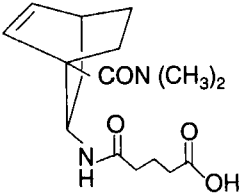
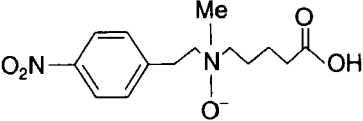
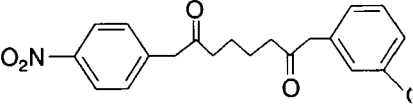


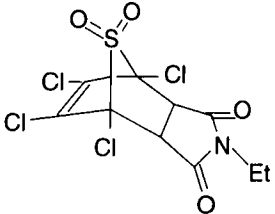
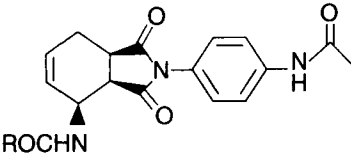
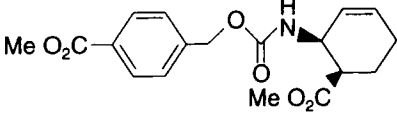
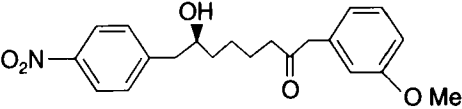
Figure 4 Free energy-reaction coordinate diagram for the antibody-catalyzed reaction depicted in Figure 3, where the ester and anilide are hydrolyzed at pH 7.0. The dashed lines are calculated values and the arrows indicate the maximal estimated values for the free energies of the transition states. Reproduced with permission from *Science*, **250**, 1135 (1990)

Table 3

Hapten	Substrate(s)
 (13)	
 (14)	 R = OCH₂CO₂H
 (15a)	
 (15b)	
 (16)	

^aConcentration of TCTD = 0.61 mM.^bConcentration of substrates = 100 μ M, antibody = 1 μ M.^cConcentration of NaBH₃CN = 50 mM.

Table 3 (continued)

Product(s)	$k_{\text{cat}}/k_{\text{uncat}}$	K_{m} (μM)
	108 ^a	2×10^4
	70 ^b	1130 (diene) 740 (dienophile)
	490 (<i>endo</i>) ^b 1810 (<i>exo</i>) ^b	960 (diene) 1700 (dienophile) 700 (diene) 7500 (dienophile)
	1760 ^c	52

(9a), which resembles hapten (8) closely, as a result of tighter binding than for methyl 3-methyl-4-nitrophenyl carbonate (9b). Even though (9b) had a faster hydrolysis rate, its poorer binding precluded efficient turnover by the antibody. Greater binding for (9a) is not surprising considering that these monoclonal antibodies were chosen for their ability to bind the hapten. Consequently, the observable characteristics for antibody catalysis are strongly influenced by the screening process.

The capability of catalyzing a lactonization reaction stereospecifically [16] revealed one advantage of antibodies compared to macrocyclic catalysts. As before, hapten (10) contained a phosphonate moiety used to represent a transition state that leads to the production of two different diastereomers along with their enantiomers. Antibodies were induced to only one diastereomer of (10), and a selected antibody gave a 170-fold rate acceleration compared to the background reaction. Moreover, this antibody catalyzed only the cyclization of one enantiomer of the racemate, providing an enantiomeric excess of 94%. Therefore, it was demonstrated that antibodies can provide a specific diastereomeric binding pocket, enabling the production of enantiomerically pure compounds.

Another advantage for antibodies is that they are naturally able to bind to proteins—a difficult union to achieve with synthetic catalysts—providing the motivation for developing catalysts capable of cleaving peptide bonds to inactivate proteins. One route [17a] entailed having an antibody cyclize a peptide to enhance self-hydrolysis. The appropriate binding pocket was constructed by raising antibodies against the substituted succinimide hapten (11). *N*-Acetyl-asparinglylglycyl(*N*-phenyl)amide was chosen as the substrate to represent an Asn–Gly bond. It was envisioned that the antibody would not only promote cyclization of the peptide but also catalyze hydrolysis of the succinimide intermediate to the aspartyl or the isoaspartyl products. Antibodies were found that hydrolyzed either the D-isomer, the L-isomer or both isomers of the starting peptide, and they were capable of catalyzing both steps of the reaction. Measurement of the individual rate constants [17b] suggested that the phosphonyl residue of (11) produces a more reactive site for cleavage in the binding pocket compared to the hydroxy group. As a result, product ratios depended upon the placement of the carbonyl of the peptide within the active site.

An indication of the potential of antibodies to rival enzymes in turnover efficiency was observed in antibody catalysis of transesterification in water [18]. The antibody, which was originally elicited to perform ester hydrolysis using the phosphonate (12) as the hapten, unexpectedly promoted the reaction of a benzylic ester with a variety of alcohols to form a second ester. Investigation of the steady-state reaction kinetics gave a series of parallel lines for the double-reciprocal plots of initial velocity against substrate concentration. This pattern is indicative of a ping-pong kinetic sequence. Additional studies with inhibitors supported this mechanism as follows: (1) the ester product gave noncompetitive

inhibition; (2) the second substrate bound by the antibody inhibited the reaction; and (3) for one set of reactants, the first product released was a competitive inhibitor of the second substrate. A covalent antibody intermediate, most likely an acyltyrosine, was directly observed in the presteady state of the reaction. An antibody capable of ping-pong kinetics is quite remarkable because it is a catalytic mechanism generally reserved for the highly advanced enzymes.

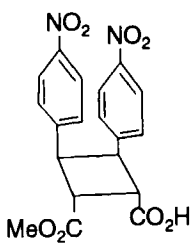
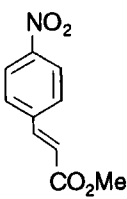
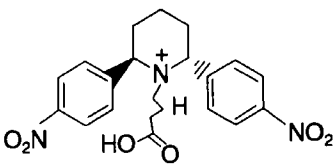
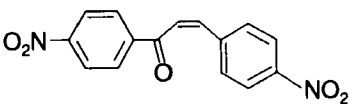
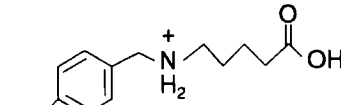
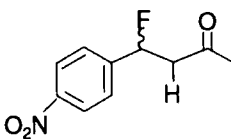
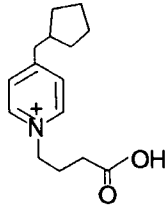
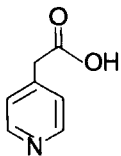
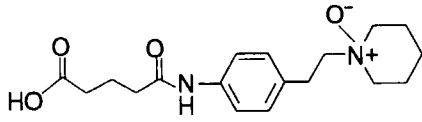
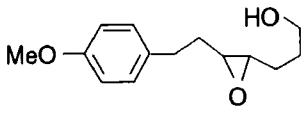
5. ANTIBODIES AS SYNTHETIC CATALYSTS

To compete with synthetic catalysts, antibodies must be able to control reaction rates and product formation for reactions that natural proteins usually cannot perform. A large variety of transition state analogues representing high-energy intermediates of synthetically important reactions have been synthesized and used as haptens [19]. The Diels–Alder reaction, providing regio- and stereochemical control of product formation, is an extremely important reaction for chemists in natural product synthesis. Since the rate-limiting step is an unfavorable entropic barrier, it has been an excellent choice for testing the catalytic potential of antibodies. Like enzymes, antibodies can act as entropic traps, catalyzing reactions by using binding energy to lower the translational and rotational activation entropies. Furthermore, this type of cycloaddition reaction does not necessarily require precise location of acid–base catalysts. One of the first successful ‘Diels–Alderases’ was found by Hilvert *et al.* [20], who modeled the reaction of tetrachlorothiophene dioxide (TCTD) with *N*-ethylmaleimide. Hapten (13) (Table 3) was designed to represent the intermediate, an unstable sulfoxide. Catalysis was envisioned to occur through lowering the energy of the transition state by aligning the reactants in the correct geometry. Furthermore, product inhibition would be avoided by the subsequent decomposition of the unstable intermediate to release sulfoxide and product from the antibody. The low solubility of tetrachlorothiophene dioxide prevented an accurate kinetic assay, but an effective molarity (concentration of reagents needed to give a rate enhancement equivalent to the catalyzed reaction) of at least 110 M of *N*-ethylmaleimide per binding site was measured.

Other ‘Diels–Alderases’ have also been produced. Braisted and Schultz [21] created an antibody which catalyzed the reaction between an activated diene and another maleimide derivative. Because under normal conditions the transition state proceeds through a boat conformation to produce a six-membered ring, the geometry of hapten (14) was based on a bicyclic ring. An additional benefit of this design results from the fact that the chair conformation is the favored ground state geometry of the product. Conversion to the chair conformation drives the product out of the binding pocket. The rate of product formation was $1.4 \mu\text{M s}^{-1}$ using $1 \mu\text{M}$ of the antibody.

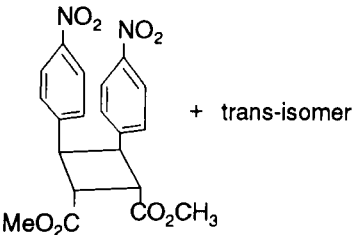
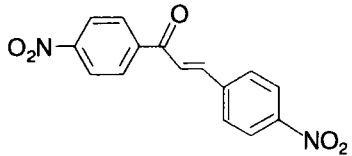
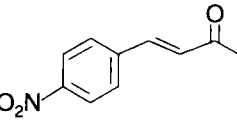
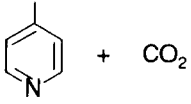
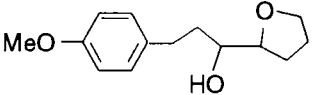
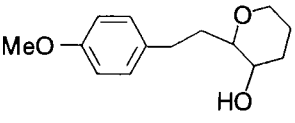
To be worthwhile to the synthetic community, ‘Diels–Alderases’ need to control both the regio- and stereochemistry of the product. Antibodies capable of

Table 4

Hapten	Substrate(s)
 (17)	
 (18)	
 (19)	
 (20)	
 (21)	

^aConcentration of substrates = 100 μ M, antibody = 1 μ M.

Table 4 (continued)

Product(s)	$k_{\text{cat}}/k_{\text{uncat}}$	K_{m} (μM)
	—	—
	1.5×10^4	220
	1700^a	182
	1.9×10^5	1.44×10^5
 + 	—	356

preferentially forming the disfavored product will be the most beneficial. To test antibodies for these requirements, the Diels–Alder reaction between *trans*-1-acylamino-1,3-butadiene and *N,N*-dimethylacrylamide, which has the potential of making eight different isomers, was chosen [22]. According to computational analysis, the *endo*-product is favored by $1.9 \text{ kcal mol}^{-1}$. Because this difference in free energy is expressed mainly in the transition state, this reaction is a prime candidate for having formation of the product being controlled by an antibody. Furthermore, 2 kcal mol^{-1} is well within the maximum energy of 12 kcal mol^{-1} available from binding (assuming a *K* value of 10^9).

Four out of the eight possible isomers, all having the same regiochemistry, are produced in solution, so antibodies were raised against haptens (**15a**) and (**15b**) that represent the *endo*- and *exo*-transition states, respectively. Calculations of the geometries of the transition state and the hapten suggested that their bond angles are in close agreement. The dihedral angle of the diene is within 0.4° of the dihedral angle of the alkene in the hapten. Furthermore, the attack angle of the diene with the dienophile (10.3°) is similar to the angle between the hapten bridgehead and the alkene moiety (19.7°). One major difference in the calculated geometries of the transition state and the hapten was in the bond lengths. The hapten covalent bonds were about $0.5\text{--}0.8 \text{ \AA}$ shorter than the forming bonds in the calculated transition state. It is not known whether this difference in geometry is crucial, but it was found that the antibodies that bound the hapten tightly did not bind the reactants strongly, with K_m values on the millimolar level. On the other hand, the antibodies were capable of producing a greater than 98% enantiomeric excess for both the *endo*- and *exo*-products, with effective molarities of 4.8 M and 18 M, respectively.

Control of product formation was also accomplished by an antibody in a more general type of synthetic reaction. The *N*-oxide-containing hapten (**16**) was used to produce an antibody capable of catalyzing regio- and stereospecific reduction of ketones [23]. In this study, a racemic amine oxide instead of a phosphonate was employed to represent the tetrahedral intermediate for the carbonyl reduction. Saturation kinetics were observed along with multiple turnovers, showing that the reducing agent NaBH_3CN did not inactivate the antibody. To test for regioselectivity, a diketone was used as the substrate and one antibody gave greater than 75:1 preference for reducing the ketone in the position nearer the *N*-oxide of the hapten. This antibody was also highly stereospecific, affording the (*S*)-alcohol with an enantiomeric excess of 96.0%; thus, both results indicate that chemistry occurred within the binding pocket.

6. ANTIBODIES AS TEMPLATES

The combination of strong binding with high specificity allows antibodies to control product formation using geometric constraints in their binding pockets.

An initial test for template control was a study whereby an antibody controlled the regio- and stereochemistry of the products formed in the photodimerization of methyl 4-nitro-*trans*-cinnamate [24]. Hapten (17) (Table 4) was chosen to induce an antibody capable of forming a cyclobutyl ring with the aromatic rings in a *cis*-conformation, which is a higher-energy conformation compared to the *trans*-compound. In the presence of an antibody specific for (17), not only was the desired compound produced upon dimerization, but its formation rate was also enhanced.

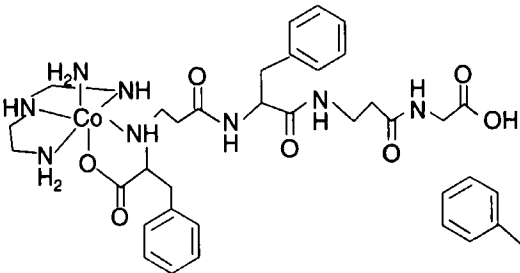
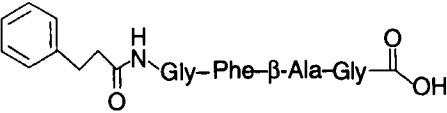
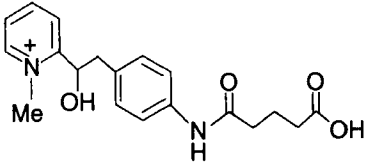
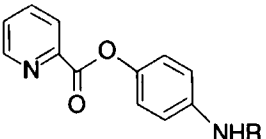
An extension of this approach of having an antibody control the formation of a specific isomer involved having an antibody interconvert the geometrical isomers of a compound [25]. To achieve this, the antibody was designed to have a nucleophilic residue in the binding pocket to assist in the rotation of the carbon-carbon double bond of an α,β -unsaturated ketone. Hapten (18) fulfilled the requirements with its *trans*-phenyl groups, which resemble a perpendicular rotational transition state, providing for a correctly shaped binding pocket. The probability of inducing a nucleophilic residue within the binding pocket was increased by introducing a positively charged amino group into the hapten at the appropriate location.

One antibody did catalyze the *cis-trans* isomerization with a respectable rate enhancement of 1.5×10^4 compared to the uncatalyzed reaction, and its characteristics were consistent with hapten design. The *trans*-hapten was bound more tightly than the *cis*-isomer, and generally antibodies which preferentially bound the *cis*-hapten were not catalytic. Both facts suggest that the isomerization does indeed proceed through a transition state with the aromatic rings passing through a *trans*-orientation. Support for the existence of a nucleophilic residue in the binding pocket was found by treating the antibody with diazoacetamide, which specifically blocks carboxy residues. With the addition of diazoacetamide, the antibody lost 85% of its catalytic activity.

This methodology of inducing catalytic groups in the binding pocket has been termed 'bait and switch'. The basic idea is to design a hapten containing a functional group (bait) that will induce an environment or a particular amino acid residue capable of performing catalysis of a substrate molecule containing a different residue (switch) in place of the bait. This technique furnishes a researcher some control in the design of binding pockets, an advantage held by synthetic catalysts.

A second application of this concept, which also utilized the electrostatic interactions that occur between antibodies and their haptens, was the induction of an antibody capable of β -elimination [26]. To induce a negatively charged residue, either an aspartate or glutamate in this case, near the hydrogen to be abstracted, an ammonium group was substituted for this hydrogen in hapten (19). The strongly immunogenic nitroaromatic group ensured that the substrate would be bound in the same geometry as the hapten. The formation of the α,β -unsaturated ketone occurred with a 1700-fold rate enhancement over the background rate. Support for a carboxylate being required in the binding pocket

Table 5

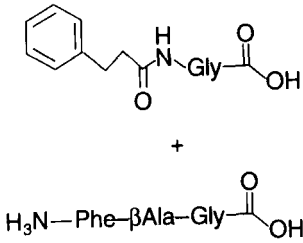
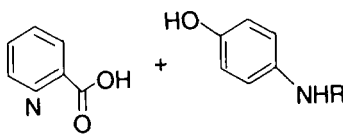
Hapten	Substrate(s)
 (22)	
 (23)	

was derived from the loss of 77% activity when diazoacetamide was incubated with the antibody.

As previously discussed, aromatic residues contained within haptens promote a strong immune response and can provide an anchoring point to align the substrate in the correct orientation. The corresponding pocket produced within an antibody provides a large amount of binding energy, and in one study this environment has been used to achieve catalysis. Decarboxylation of 4-pyridylacetic acid was enhanced by an antibody raised against hapten (20) containing a hydrophobic moiety in place of the acid group [27]. Success of this antibody depended strongly on the favorable binding energy provided by the aromatic ring to force the negatively charged carboxylate into a hydrophobic pocket. Once in the pocket, the ground state energy of the substrate was raised compared to the unbound substrate in the uncatalyzed reaction, resulting in a rate enhancement of 10^5 .

An attempt has been made to form a disfavored product in an intramolecular cyclization reaction [28]. It is known that the preferred geometry for cyclizations occurring from internal nucleophilic attack is a 180° alignment of the three atoms.

Table 5 (continued)

Product(s)	$k_{\text{cat}}/k_{\text{uncat}}$	K_m (μM)
	—	—
	1230	840

From Baldwin's rule, the opening of the epoxide of the substrate will produce the 5-*exo*-product over the 6-*endo*-product. To obtain the 6-*endo*-product, an antibody was raised against a hapten (**21**) composed of a six-membered ring containing an *N*-oxide moiety. The *N*-oxide represents the transition state geometry of the opening epoxide, and the negatively charged oxygen atom should produce a stabilizing amino acid residue in the binding pocket. In theory, when the substrate is bound by the antibody, this stabilization will be greater for the formation of the six-membered ring compared to the five-membered ring. The 6-*endo*-product was indeed formed, but at a very low rate of only $4.6 \mu\text{M min}^{-1}$, which apparently does not compete effectively with the formation of the 5-*exo*-product.

7. ANTIBODIES COMBINED WITH METALS

As discussed in the section on synthetic catalysts, metals can strongly activate bonds, making them a highly desirable addition to catalytic antibodies. Enzymes use metals, either ligated within active sites by amino acid residues or complexed

with porphyrins, to accelerate reaction rates. Antibodies capable of binding metals or porphyrins containing metals have been produced, and these complexes affect reaction rates. An antibody was constructed to perform peptide bond cleavage using a metal chelated to a tetramine complex [29]. The design of hapten (22) (Table 5) consisted of the desired peptide sequence linked to cobalt(III) coordinated to triethylenetetramine. It was envisioned that an antibody would bind the peptide, tetramine and metal (Zn^{2+}) in the correct geometry for amide hydrolysis. Results indicated that hydrolysis had occurred, but the slow rate precluded an accurate analysis. Antibodies capable of binding porphyrins provide yet another method for introducing metals. Metallation of a porphyrin has been catalyzed by an antibody raised against a distorted *N*-methylporphyrin [30]. This antibody not only acted as a ferrocyclase but also functioned as a peroxidase and reduced hydrogen peroxide.

It would be more convenient to have antibodies that bind metals without the addition of chelators. An antibody capable of metal binding has been produced by induction with hapten (23) that contains a pyridinium salt to represent the metal [31]. Since the reaction of interest was an ester hydrolysis, (23) also contained a hydroxy group in the position of the tetrahedral transition state. A flat pyridinium ring exposed the positive charge to the developing antibody to promote the formation of a partial coordination sphere within the antibody. One antibody did catalyze a hydrolysis reaction and required the presence of zinc. A metal-binding site has been constructed in the aforementioned antibody NPN43C9 [32a]. Through a double mutation via site-directed mutagenesis, two additional histidines were added to the antibody to give a total of three histidines near the binding pocket. By following a computer model, these histidines were placed in a manner that mimics the active site of carbonic anhydrase. The resulting antibody was a selective binder of zinc [32b].

8. CONCLUSION

This article has presented a brief overview of the different approaches available to convert compounds capable of binding into catalysts. For the macrocycles displaying Michaelis–Menten kinetics, turnover numbers are, in general, extremely low and for now the binding is too weak for these hosts to be effective catalysts according to the definition of efficiency given earlier. It is difficult to surmise whether efficient catalysts will emerge by simply adding functional groups to macrocyclic structures. This method generally will not work if efficient catalysis requires a precise alignment (within an angstrom) of the components within a catalyst–substrate complex—hence the fascination with combinatorial chemistry. Accordingly, the use of antibodies appears to be a better route because they are based on the same building blocks as enzymes and have a precisely sculptured active site molded to the hapten, thus providing the tuning necessary to position catalytic groups. However, the binding of ligands by antibodies

appears to have a finite limit (10^9) several orders of magnitude less than supposed for enzyme-transition state complexes. Moreover, this binding is not usually achieved through amino acid side chains that are capable of acid-base catalysis. To overcome this deficiency, catalytic groups have been either induced by creative hapten design or added through protein engineering. This may provide a means for antibodies to meet the requirements for being efficient catalysts. Therefore, the catalytic antibody and host-guest chemists face the same predicament. Both are trying to achieve the optimal alignment of catalytic groups within their own systems that enzymes have found through evolution.

9. REFERENCES

1. L. Pauling, *Am. Sci.*, **36**, 51 (1948).
2. J. D. Stewart and S. J. Benkovic, *Chem. Soc. Rev.*, 213 (1993).
3. C. J. Pedersen, *J. Am. Chem. Soc.*, **89**, 2495 (1967).
4. E. Anslyn and R. Breslow, *J. Am. Chem. Soc.*, **111**, 5972 (1989).
5. Y. Matsumoto, M. Komiyama, *J. Mol. Cat.*, **61**, 129 (1990).
6. (a) H.-J. Schneider and F. Xiao, *J. Chem. Soc., Perkin Trans. 2*, 387 (1992); (b) R. Breslow, *Isr. J. Chem.*, **32**, 23 (1991).
7. M. W. Hosseini, J.-M. Lehn, L. Maggiora, K. B. Mertes and M. P. Mertes, *J. Am. Chem. Soc.*, **109**, 537 (1987).
8. (a) J. Wolfe, D. Nemeth, A. Costero and J. Rebek Jr, *J. Am. Chem. Soc.*, **110**, 983 (1988); (b) K.-S. Jeong, T. Tjivikua, A. Muehldorf, G. Deslongchamps, M. Famulok and J. Rebek Jr, *J. Am. Chem. Soc.*, **113**, 201 (1991).
9. K. Johnsson, R. K. Allemann, H. Widmer, and S. A. Benner, *Nature*, **365**, 530 (1993).
10. W. Jencks, *Catalysis in Chemistry and Enzymology*, McGraw-Hill, New York, 1969.
11. D. R. Davies, E. A. Padlan and S. Sheriff, *Annu. Rev. Biochem.*, **59**, 49 (1990).
12. T. M. Fieser, J. A. Tainer, M. H. Geysen, R. A. Houghten and R. A. Lerner, *Proc. Natl. Acad. Sci. USA*, **84**, 8568 (1987).
13. K. D. Janda, D. Schloeder, S. J. Benkovic and R. A. Lerner, *Science*, **241**, 1188 (1988).
14. R. A. Gibbs, P. A. Benkovic, K. D. Janda, R. A. Lerner and S. J. Benkovic, *J. Am. Chem. Soc.*, **114**, 3528 (1992).
15. J. Jacobs and P. G. Schultz, *J. Am. Chem. Soc.*, **109**, 2174 (1987).
16. A. D. Napper, S. J. Benkovic, A. Tramontano and R. A. Lerner, *Science*, **237**, 1041 (1987).
17. (a) R. A. Gibbs, S. Taylor and S. J. Benkovic, *Science*, **258**, 803 (1992); (b) L. J. Liotta, P. A. Benkovic, G. P. Miller and S. J. Benkovic, *J. Am. Chem. Soc.*, **115**, 350 (1993).
18. P. Wirsching, J. A. Ashley, S. J. Benkovic, K. D. Janda and R. A. Lerner, *Science*, **252**, 680 (1991).
19. R. A. Lerner, S. J. Benkovic and P. G. Schultz, *Science*, **252**, 659 (1991).
20. D. Hilvert, K. W. Hill, K. D. Nared and M.-T. Auditor, *J. Am. Chem. Soc.*, **111**, 9261 (1989).
21. A. C. Braisted and P. G. Schultz, *J. Am. Chem. Soc.*, **112**, 7430 (1990).
22. V. E. Gouverneur, K. N. Houk, B. de Pascual-Teresa, B. Beno, K. D. Janda and R. A. Lerner, *Science*, **262**, 204 (1993).
23. L. C. Hsieh, S. Yonkovich, L. Kochersperger and P. G. Schultz, *Science*, **260**, 337 (1993).

24. A. Balan, B. P. Doctor, B. S. Green, M. Torten and H. Ziffer, *J. Chem. Soc., Chem. Commun.*, 106 (1988).
25. D. Y. Jackson and P. G. Schultz, *J. Am. Chem. Soc.*, **113**, 2319 (1991).
26. K. M. Shokat, C. J. Leumann, R. Sugawara and P. G. Schultz, *Nature*, **338**, 269 (1989).
27. J. A. Ashley, C.-H. L. Lo, G. P. McElhaney, P. Wirsching and K. D. Janda, *J. Am. Chem. Soc.*, **115**, 2515 (1993).
28. K. D. Janda, C. G. Shevlin and R. A. Lerner, *Science*, **259**, 490 (1993).
29. B. L. Iverson and R. A. Lerner, *Science*, **243**, 1184 (1989).
30. A. G. Cochran and P. G. Schultz, *Science*, **249**, 781 (1990).
31. W. S. Wade, J. A. Ashley, G. K. Jahangiri, G. McElhaney, K. D. Janda and R. A. Lerner, *J. Am. Chem. Soc.*, **115**, 4906 (1993).
32. (a) J. D. Stewart, V. A. Roberts, N. R. Thomas, E. D. Getzoff and S. J. Benkovic, *Biochemistry*, **33**, 1994 (1994) (b) J. D. Stewart, V. A. Roberts, M. W. Crowder, E. D. Getzoff and S. J. Benkovic, *J. Am. Chem. Soc.*, **116**, 415 (1994).

Chapter 6

Lock-and-Key Processes at Crystalline Interfaces: Relevance to the Spontaneous Generation of Chirality

*ISABELLE WEISSBUCH, RONIT POPOVITZ-BIRO,
LESLIE LEISEROWITZ AND MEIR LAHAV
The Weizmann Institute of Science, Rehovot, Israel*

1. INTRODUCTION

From the days of Pasteur, when he recognized that organic molecules associated with living matter are chiral, scientists have been puzzled as to the ancestry of homochirality in the living world. The bias for a single handedness in biopolymers from homochiral L- α -amino acids and D-sugars is still regarded as a most remarkable hallmark in nature.

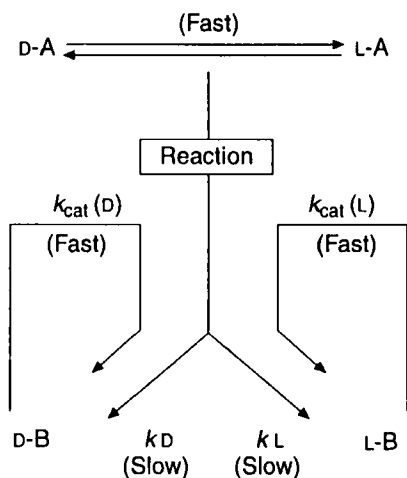
Theories for the origin of a single chirality in the biological world fall into two major categories—biotic and abiotic. The first suggests that the selection of one of the enantiomers took place at a later stage in the biological evolution of living matter. Till the present time such theories have been fundamentally speculative and imprecise as to the mechanistic details, and, most importantly, hardly amenable to experimental verification in any meaningful scientific sense [1]. The second scenario implies that chiral materials had been formed prior to the biopolymers. Such asymmetry could have emerged provided one can amplify a small fluctuation from the racemic state to magnitudes useful for biotic evolution. The small enantiomeric excess can be generated either through a determinate or a chance mechanism. The former must invoke the intrinsic chirality of the universe through various physical agents for producing

asymmetry, such as the weak nuclear interactions in which parity, i.e. space inversion, is not conserved [2, 3]. When these nonsymmetric forces operate on a prochiral or racemic mixture they may give rise to an energy difference between enantiomers [4, 5]. *Ab initio* calculations which take these forces into consideration suggest that L- α -amino acids, D-sugars and their polymers should be more stable than the corresponding unnatural enantiomers by a ratio of $1:10^{17}$. This difference is so minute that no single experiment has been performed so far where one could transform the unnatural enantiomers into the natural ones [6, 7].

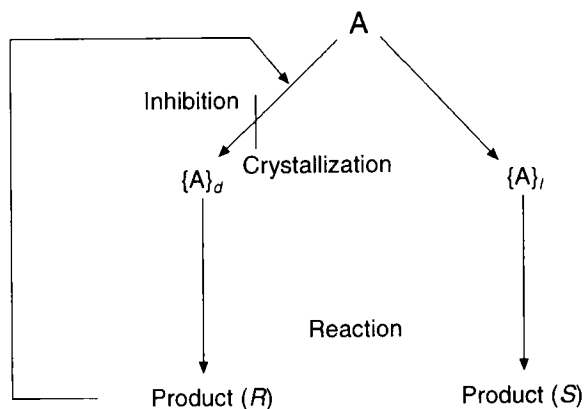
Mathematical models proposed by Frank [8] and later by Seelig [9] and Decker [10–13] have suggested that an efficient amplification of a small chance fluctuation from the racemic mixtures is feasible. In Frank's model a small enantiomeric excess can be efficiently amplified provided the system is designed such that one enantiomer acts as a catalyst for its own formation and as an inhibitor for the formation of the other enantiomer.

Calvin [14] proposed a general scheme for a rapidly racemizing material, where the minute excess of one of the enantiomers reacts, by chance, to yield product L-B (Scheme 1). The presence of B generates a catalyst which drives the reaction completely into the L-direction.

At present, it appears that none of these models rests on observation. Experimental systems which are the closest to these models involve the process of total asymmetric transformation, whereby nonchiral molecules or a rapidly racemizing system in solution crystallizes in a chiral arrangement. By controlled crystallization one can transform such a nonchiral system in solution into a solid of single chirality, as reported by Havinga [15, 16] for methylethylallylanilium iodide. A serious drawback of such systems as models for the spontaneous



Scheme 1



Scheme 2

generation of chirality is that the molecules will lose their chirality upon dissolution of the crystals. The loss of chirality can be avoided provided the crystallization is coupled to an asymmetric chemical reaction that yields a stable chiral product, where the asymmetry induced arises from the chiral environment of the crystal, as was first demonstrated by Penzien and Schmidt [17]. Green and Heller [18] proposed that for an amplification to take place, the reaction product formed in the first crystal should induce an autocatalytic crystallization of the same chiral phase. If, in addition, this product is an inhibitor of crystallization for the enantiomorphous phase, it will enhance the feedback process in the amplification of homochirality, as depicted in Scheme 2.

Several years ago a program was initiated for designing experiments to validate the process in Scheme 2. The materialization of this scheme demands an understanding of the ways in which molecules assemble into ordered nuclei and crystals and the stereochemical interactions of their growing surfaces with molecules of the environment. In the course of these studies on crystal growth we encountered several general aspects of molecular recognition and self-assembly which involve the following.

- (1) The role played by molecular lock-and-key mechanisms at crystal surfaces for habit modification in crystals and the reduction in symmetry of mixed crystals.
- (2) Self-organization of amphiphilic molecules at the air-liquid interface, forming transient nuclei of two-dimensional or three-dimensional crystallites *en route* to crystal completion.
- (3) Lock-and-key mechanisms at the crystal-solution interface involving the effect of solvent on crystal shape and growth.
- (4) Recognition of crystal embryos by molecular additives in the early stages of the nucleation process.
- (5) Probing the embryonic clusters at the molecular level with the aid of modern analytical tools.

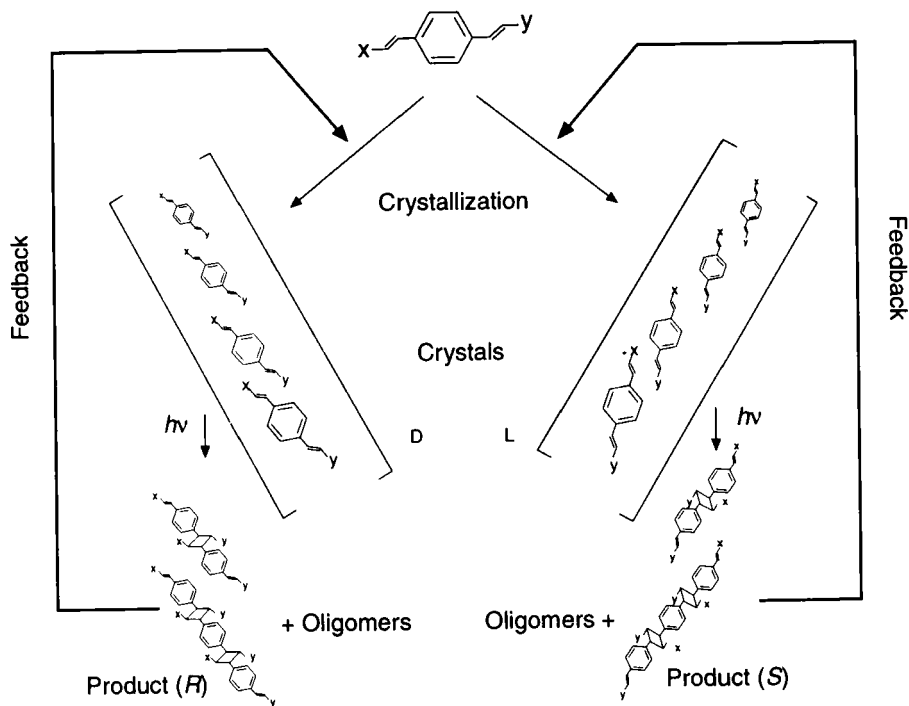
The generality of these aspects for molecular recognition and self-organization in a variety of different molecular systems will be discussed. The knowledge acquired from these approaches will be combined for the design of a cooperative system that fulfills the requirements of Scheme 2 and encompasses the spontaneous separation of racemic α -amino acids into chiral territories through the assistance of centrosymmetric crystals of α -glycine.

The diverse topics described here leading to the experiments on the chiral separation of α -amino acids are not given in chronological order, but rather in terms of a logical elucidation of the different elements involved therein.

2. GENERATION AND AMPLIFICATION OF CHIRALITY IN THE SOLID STATE

2.1 Nonchiral Molecules Crystallizing into Chiral Crystals

Crystallization of nonchiral molecules into chiral crystals starting from a single nucleus can induce the entire ensemble of molecules to assemble into crystals of one handedness. This process is generally termed 'total asymmetric transformation of the second type'. The most commonly known inorganic systems which might undergo such a process are quartz and sodium chlorate. Over the years, total asymmetric transformations have been encountered for several different types of organic molecules, such as the inclusion complexes of urea, tri-*o*-thymotide [19] and the high-temperature polymorph of binaphthyl [20]. In a given chiral single crystal of these systems all the molecules reside in a chiral environment and assume a chiral conformation. A topochemical reaction which takes place in the solid phase can transform the conformational chirality in the crystal into stable molecular chirality, thus generating in a closed system chiral materials enriched with one enantiomer [17]. However, in a large number of independent crystallization experiments, the precipitations of crystals of opposite handedness have identical probabilities. On the other hand, if a chiral product formed by chance in the first event can preserve and amplify the initial chirality by exerting control on fresh crystallization of the starting material, it may drive the entire system to precipitate preferentially into one of the two enantiomorphous phases. In order to validate this model, advantage was taken of the photopolymerization behavior of nonchiral dienes in the solid phase [21, 22] (Scheme 3). Following the empirical rules of molecular packing, several nonchiral phenyl diacrylates (Table 1) were found to be appropriate. In these solid-state reactions, the stereochemistry of the product is determined by the crystalline arrangement of and distances between the carbon-carbon double bonds of nearest molecular neighbors. Furthermore, high enantiomeric yields were anticipated in these



Scheme 3

Table 1 Structural formulas and space groups of achiral monomers packing in chiral crystal structures

Monomer	R ₁	R ₂	Space group
(1)	3-Pentyl	Methyl	<i>P</i> 2 ₁
(2)	3-Pentyl	Ethyl	<i>P</i> 1
(3)	3-Pentyl	<i>n</i> -Propyl	<i>P</i> 1
(4)	(<i>R,S</i>)- <i>s</i> -Butyl	Ethyl	<i>P</i> 1
(5)	(<i>R,S</i>)- <i>s</i> -Butyl	<i>n</i> -Propyl	<i>P</i> 1
(6)	Isopropyl/3-pentyl	Ethyl	<i>P</i> 1
(7)	<i>t</i> -Butyl	Ethyl	<i>P</i> 1

transformations since they occur exclusively in the crystalline lattice and cease to do so in disordered regions caused by melting or partial destruction of the crystal during the reaction.

Single crystals of homochiral phases were prepared either from the melt starting from a single nucleus or by slow evaporation from methanol solution in conjunction with autocatalytic seeding effects [23–26]. The best-suited compounds proved to be (1), (2) and (3) (Table 1). The single crystals contain only one of the two enantiomeric motifs depicted in Scheme 3. High enantiomeric yields of dimers, trimers and oligomers were obtained on irradiation.

Of particular interest are monomers (4) and (5). The racemic mixtures of these molecules crystallize in a chiral space group. In a single crystal all the chromophores reside in a chiral environment where the racemic *s*-butyl groups exert a minor chiral influence on the packing, being disordered in the lattice. Crystallization of a racemate under conditions close to equilibrium yields the two chiral crystalline phases *d* and *l* in equal amounts owing to the presence of a eutectic. However, fast crystallization yields a metastable state with unequal amounts of the two enantiomorphous phases. Irradiation of such a mixture yielded chiral products. However, it is not possible to predict the absolute configuration of the reaction product. Of paramount importance for the amplification of chirality in such systems is that on crystallization of an (*R*:*S*) mixture with an enantiomeric excess as low as 60:40, the overall system should precipitate in a homochiral single crystal. The chirality of the phase is determined by the enantiomer present in excess, yielding an asymmetric synthesis with quantitative enantiomeric yield.

In recent years several groups, in particular those of Scheffer [27], Toda [28] and Ohashi [29] have reported 'spontaneous' asymmetric synthesis for more than a dozen nonchiral or racemic materials. Therefore the generation of chirality by such a route is feasible.

The close stereochemical relationship between the rigid products and the crystalline matrix in which they were generated suggested that dimers, trimers and oligomers might influence the crystallization process by acting as nucleating centers for the crystalline phase in which they were generated. All seven monomers which appear in Table 1 were crystallized in the presence of enantiomerically pure dimers, trimers and oligomers generated photochemically from the same phase in a previous step. Unexpectedly, it was observed in all experiments carried out that a large excess was precipitated of the phase enantiomorphous to that in which the additive itself had been generated (Figure 1) [30, 31].

The stereochemical similarity between the molecular structures of the chiral cyclobutane products as additives and the crystal structure of the enantiomorph whose precipitation was inhibited was found to be of paramount importance. Other parameters, such as temperature, concentration and the nature of the medium of crystallization, had only a secondary effect on the enantiomeric excess

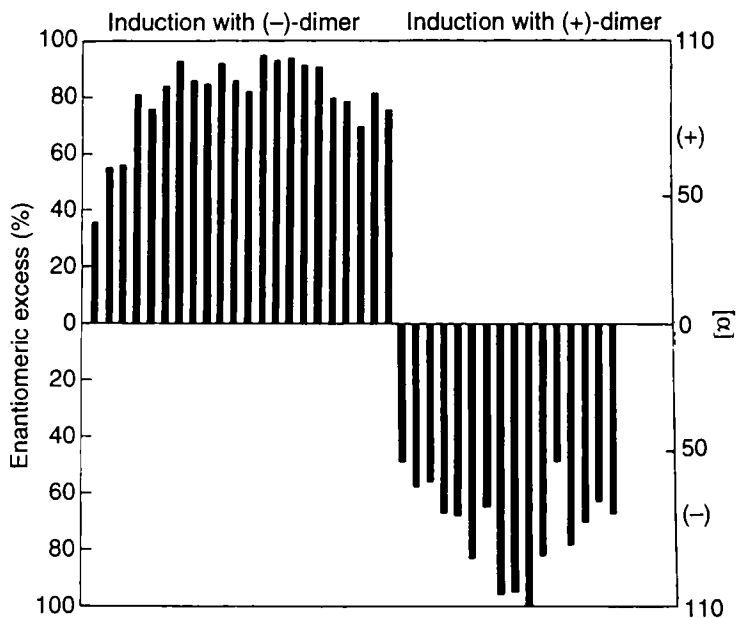


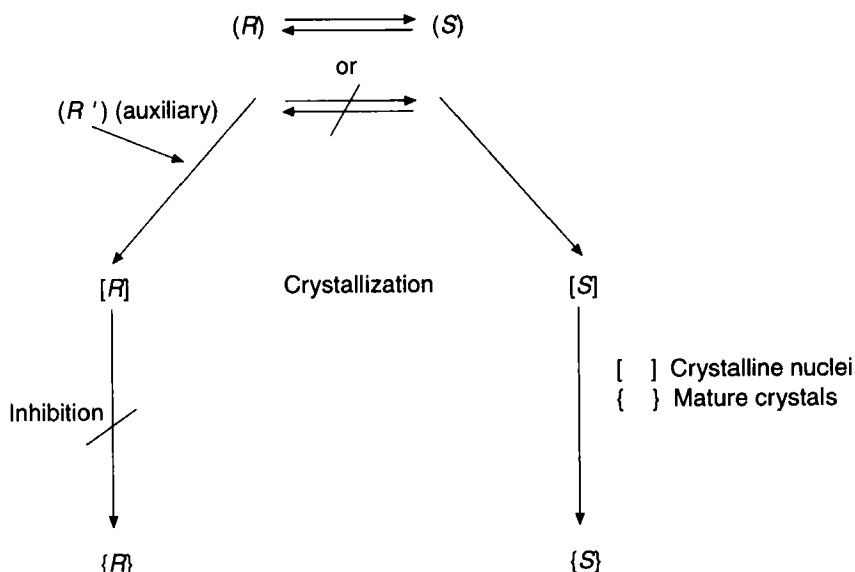
Figure 1 Asymmetric yields of induction for monomer (1) with the (+)-dimer and (–)-dimer of (4) (see Table 1) under various experimental conditions

of the crystalline mixture. On the basis of the above-proposed mechanism, a direct amplification cannot be accomplished in the systems under discussion. Nevertheless, since the observed effects are general, they should be applicable to other systems. Consequently, a clear understanding of this asymmetric induction was considered advantageous for the design of the required systems needed for the amplification process.

Further kinetic and stereochemical studies resulted in the formulation of a mechanism according to which the additive is enantioselectively adsorbed in small amounts at the surfaces of the growing crystals and, possibly, on crystalline nuclei of the same absolute configuration as the additive, causing a drastic decrease in the rate of growth of this phase. Such an effect induces a shift in the crystallization equilibrium towards the precipitation of the unaffected enantiomorphous phase, as illustrated in Scheme 4.

2.2 Role of Additives in Crystallization

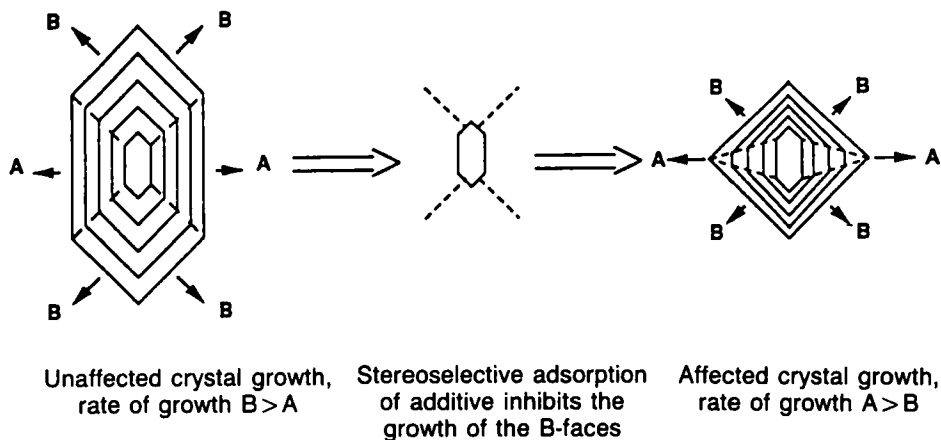
In order to obtain further insight into the stereospecific interactions between crystals and additives, morphological studies on crystals grown pure and in the presence of the molecular additives were performed systematically [32–37]. In general, the morphology of a crystal is defined by the relative rates of growth of the crystal in different directions. When growth perpendicular to a given face



Scheme 4

is specifically inhibited, the area of this face is expected to increase relative to the areas of the other faces of the same crystal, as depicted in Scheme 5. Differences in the relative surface areas of the various faces of pure and affected crystals allow, therefore, identification of the faces where adsorption took place and the crystallographic directions of growth involved. This morphological analysis was carried out on a variety of organic compounds crystallized in the presence of additives of molecular structures similar to those of the corresponding substrate molecules. A stereochemical correlation between the structures of the affected crystal surfaces and the molecular structure of the additive could be deduced for a large variety of systems [38]. It became possible to infer that the additive may be stereospecifically adsorbed only at those faces in which the part of the adsorbate that differs from the substrate emerges from the crystal. The adsorbed additive inhibits the regular deposition of oncoming layers of substrate molecules, impeding growth perpendicular to that face and leading to a concomitant relative increase in its surface area. Once this mechanism was established it became possible to exploit it in order to modify systematically the morphology of crystals by tailoring additives which bind at preselected faces and inhibit growth in a predictable manner.

The mechanism proposed in the above example is, in principle, applicable for the kinetic resolution of racemic mixtures, as will be discussed in Section 4.4.



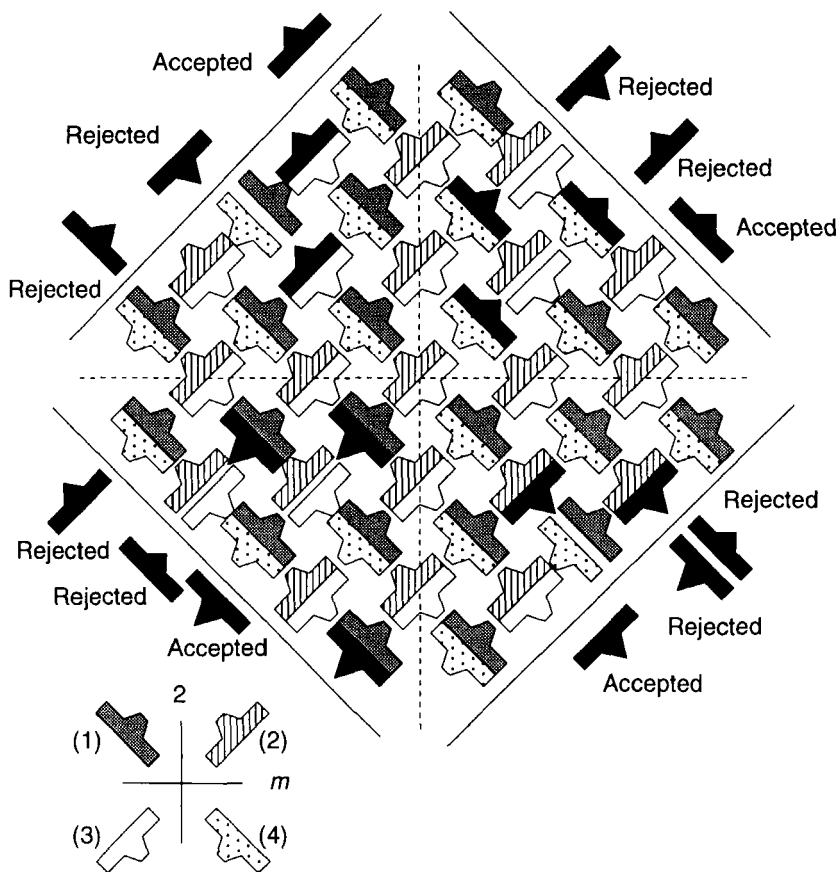
Scheme 5

2.3 Transformation of Nonchiral Crystal Hosts into Chiral Solid Solutions

In this section we demonstrate that the spontaneous generation of asymmetry is not confined to chiral crystals but can be extended to nonchiral crystals, which generally have among their symmetry elements glide planes, centers of inversion and possible combinations thereof leading to two-fold or two-fold screw symmetry.

Such crystals may be thought of as being composed of two chiral ensembles of opposite handedness, within each of which chiral molecular sites are related by translation and rotation symmetry elements. Were it possible to replace enantioselectively some of the molecules of only one of the two enantiomeric ensembles by additive molecules, then the original host crystal would have been transformed into a chiral solid solution. We demonstrated that such a process is possible via selective occlusion of the guest molecules through chiral crystal faces [39–43].

Crystals display surface arrangements that are a subset of the arrangement in the bulk. This characteristic may be expressed in terms of the two-dimensional (2D) symmetry relating molecules at a certain surface, which is generally lower than the three-dimensional (3D) symmetry within the bulk. For example, a surface structure cannot be centrosymmetric or have an axis of rotation or glide plane parallel to it. Thus, molecules which are related to each other by centers of inversion in the bulk are not so at the surface. Chiral faces which are related by a center of inversion or by mirror symmetry are defined as enantiotopic [44]. Additive molecules may be selectively adsorbed at a subset of surface sites on chiral faces and so occluded at a subset of the bulk symmetry-related sites. These aspects of crystal symmetry are illustrated schematically in Scheme 6 for a



Scheme 6

molecular arrangement of space group $P2_1/c$, and so in a motif of point group symmetry $2/m$, which is the most commonly observed in molecular crystals. (The c -glide is a combination of the mirror symmetry and half a translation along the c -axis; the two-fold screw 2_1 along the b -axis is a combination of a two-fold rotation and half a translation along b . The symbol $2/m$ specifies a two-fold axis perpendicular to a mirror plane m , the combination of which generates a center of inversion.) The packing arrangement shown in Scheme 6 describes the relative orientations of the molecules in the crystal.

In the arrangement shown the cross-hatched figures (1) are related to the vertical striped ones (2) by two-fold screw rotation, to the white figures (3) by glide symmetry and to the dotted ones (4) by centers of inversion. Note the corresponding symmetry operations of the point group $2/m$ in the inset. The crystal shown develops symmetry-related faces of the type $\{011\}$. Applying the mechanism of stereoselective

adsorption to the black additive molecules at the four symmetry-related {011} faces, the additives may occupy only one of the four symmetry-related molecular sites at each face. These sites at the four {011} faces are different and related to each other by the $2/m$ point symmetry of the crystal. The symmetry of each of the four sectors of the crystal is triclinic $P1$ since the additive molecules occluded therein are related by translation symmetry only. Thus, the guest molecules are anisotropically distributed within the grown crystal, preferentially occluded through different surface sites on the various faces, leading to a mixed crystal composed of four sectors coherently intertwined and of reduced symmetry.

We describe here two representative systems, (*E*)-cinnamamide–(*E*)-cinnamic acid and (*E*)-cinnamamide–(*E*)-2-(β -thienyl)acrylamide, in which an ‘absolute’ asymmetric photodimerization is a consequence of the concept of crystal symmetry lowering.

(a) Asymmetric synthesis in centrosymmetric host crystals: the system (E)-cinnamamide–(E)-cinnamic acid [44]

(*E*)-Cinnamamide packs in a centrosymmetric crystal in space group $P2_1/c$, containing four symmetry-related molecules. The crystals assume a prismatic morphology elongated in the *b*-direction (Figure 2).

The basic structural motif within the crystal consists of ribbons of centrosymmetric, hydrogen-bonded dimers interlinked by $N-H_a \cdots O=C$ bonds related by translation along the *b*-axis. These ribbons make plane-to-plane contact of 3.8 Å (1 Å = 0.1 nm) across carbon–carbon double bonds via centers of inversion along the *a*-axis. Each ribbon may be considered as being composed of two ‘heterochiral’ stacks related to each other by a center of inversion. These stacks are given the same labels (1, 2, 3 and 4) as the corresponding symmetry-related molecules in the bulk which adopt chiral conformations.

In order to understand how cinnamic acid additive molecules are adsorbed and occluded into the crystal along the symmetry *b*-axis, we describe the orientations of molecules 1–4 *vis-à-vis* the pertinent crystal surfaces. At the +*b*-end of the crystal the $N-H_a$ bonds emerge from the crystal for molecules in stacks 2 and 3, which are of the same handedness, whereas for molecules in homochiral stacks 1 and 4, which are enantiomeric to 2 and 3, the corresponding $N-H_a$ bonds point into the crystal. The opposite situation holds at the –*b*-end.

The (*E*)-cinnamic acid guest molecules will be attached preferentially to the +*b*-side of the crystal at the ends of stacks 2 and 3, forming hydrogen bonds between the carboxy groups and the $N-H_a$ bonds of the amides emerging at that side. Binding of the guest to stacks 1 and 4 at the same +*b*-side of the crystal would result in repulsion between the oxygen atoms of the carboxy and

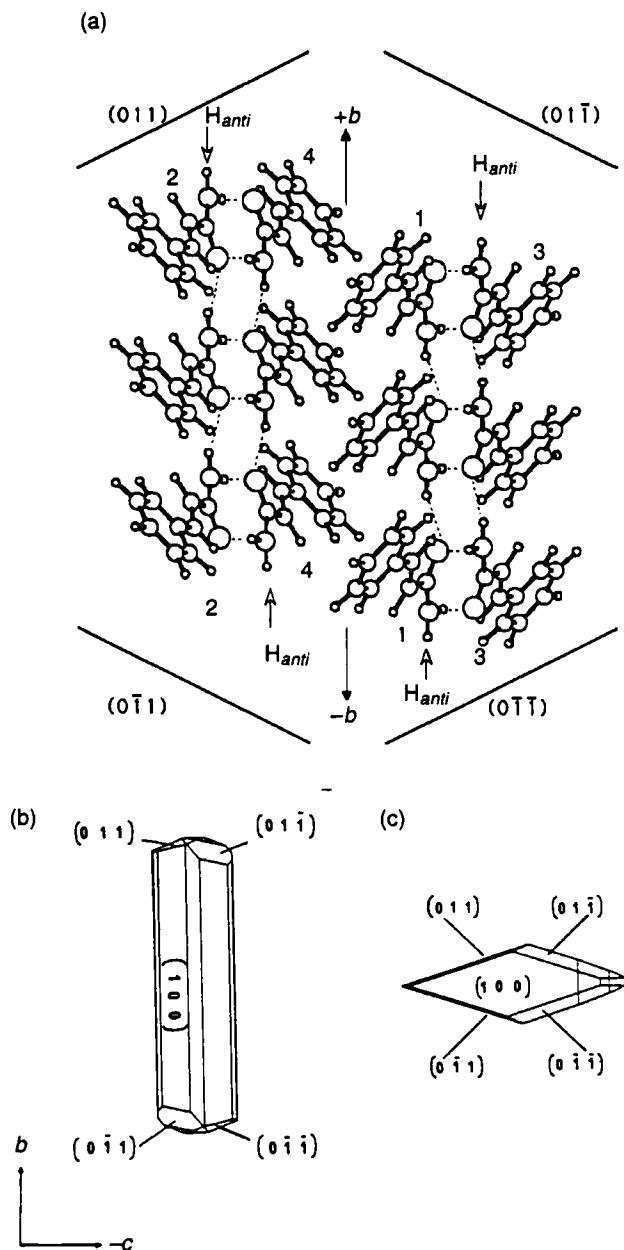


Figure 2 (a) Packing arrangement of *(E)*-cinnamamide viewed along the α -axis. The $[011]$ faces are shown and the four symmetry-related sites are denoted. (b) Crystal morphology of pure *(E)*-cinnamamide. (c) Crystal morphology of *(E)*-cinnamamide grown in the presence of *(E)*-cinnamic acid (Note that the $N-H_{anti}$ bond is designated in the text by $N-H_d$)

amide groups, and so the guest will not be favorably adsorbed at such sites. By symmetry, *E*-cinnamic acid is expected to bind easily and subsequently be occluded into the homochiral sites 1 and 4 from the $-b$ -side, but not at sites 2 and 3. The calculated intermolecular energy loss within the crystal bulk at a guest site compared to binding a host molecule at the same site is as high as 8 kcal mol^{-1} ($1 \text{ cal} = 4.2 \text{ J}$). Thus, in keeping with the above-discussed concepts, the mixed crystal of (*E*)-cinnamamide–(*E*)-cinnamic acid is not centrosymmetric, but rather composed of two enantiomorphous halves of symmetry $P2_1$.

It is well documented that (*E*)-cinnamamide photodimerizes in the solid phase to yield the centrosymmetric dimer truxillamide. Such a reaction takes place between the close-packed amide molecules of two 'enantiomeric' stacks across centers of inversion along the a -axis (Figure 3).

UV irradiation of solid solutions of (*E*)-cinnamamide–(*E*)-cinnamic acid yields, apart from the α -truxillamide homophotodimers, the two chiral amide–acid mixed dimers. Following the proposed mechanism of the selective occlusion of cinnamic acid we expect, as shown in Figure 3, the (*R,R,R,R*)-dimer and (*S,S,S,S*)-dimer to be formed at the $+b$ sector and $-b$ -sector of the crystal, respectively. Large mixed crystals of (*E*)-cinnamamide–(*E*)-cinnamic acid were grown and material was cleaved from the opposite b -poles and irradiated with UV light (Figure 4). The enantiomeric excess of the photodimers at the two poles of the crystal ranged from 40 to 60%, depending upon the temperature at which the crystal was grown. This asymmetric synthesis is also unique compared with conventional asymmetric syntheses insofar as the crystal structure permits a prediction of the chirality of the product molecules. This absolute assignment of structure is the outcome of the fact that in centrosymmetric crystal arrangements, the positions of the host and guest molecules are unambiguously determined given the mode of guest adsorption and occlusion into the crystal.

(b) The system (E)-cinnamamide–(E)-2-(β -thienyl)acrylamide [39, 45]

We may now query whether selective additive adsorption on the crystal surface followed by occlusion and a concomitant reduction in crystal symmetry will occur in systems where the difference in binding energy between the host and guest molecules at the surface will be lower than that of the host amide and guest acid described above. For example, in a solid solution bearing host phenyl and guest thienyl groups, it is well documented that in the crystalline state a phenyl ring can be easily replaced by a thienyl, although its atoms do not directly coincide with those of the phenyl group.

When a crystal of (*E*)-cinnamamide was grown in the presence of (*E*)-2-(β -thienyl)acrylamide, a change in morphology was observed. The four {011} faces were more developed, implying that the thienyl moieties were occluded primarily

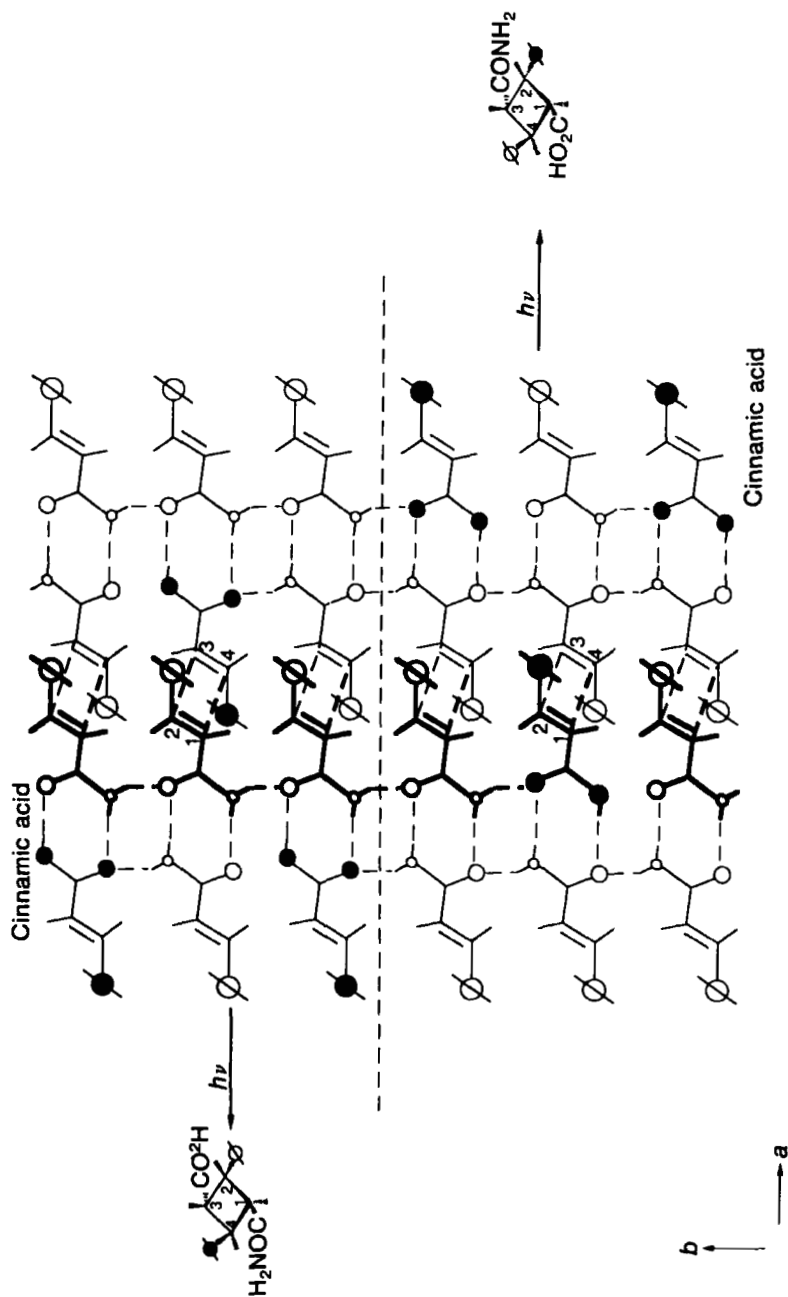


Figure 3 View along the *c*-axis of the packing arrangement of *(E)*-cinnamamide (open circles) containing occluded cinnamic acid (filled circles). The hydrogen-bonded dimers are related by translation along *a* and *b*. Photodimerization takes place between neighboring molecules from the stack in bold print and the underlying 'centrosymmetrically' related stack. Note that the horizontal dashed line separates the top from the bottom half of the crystal

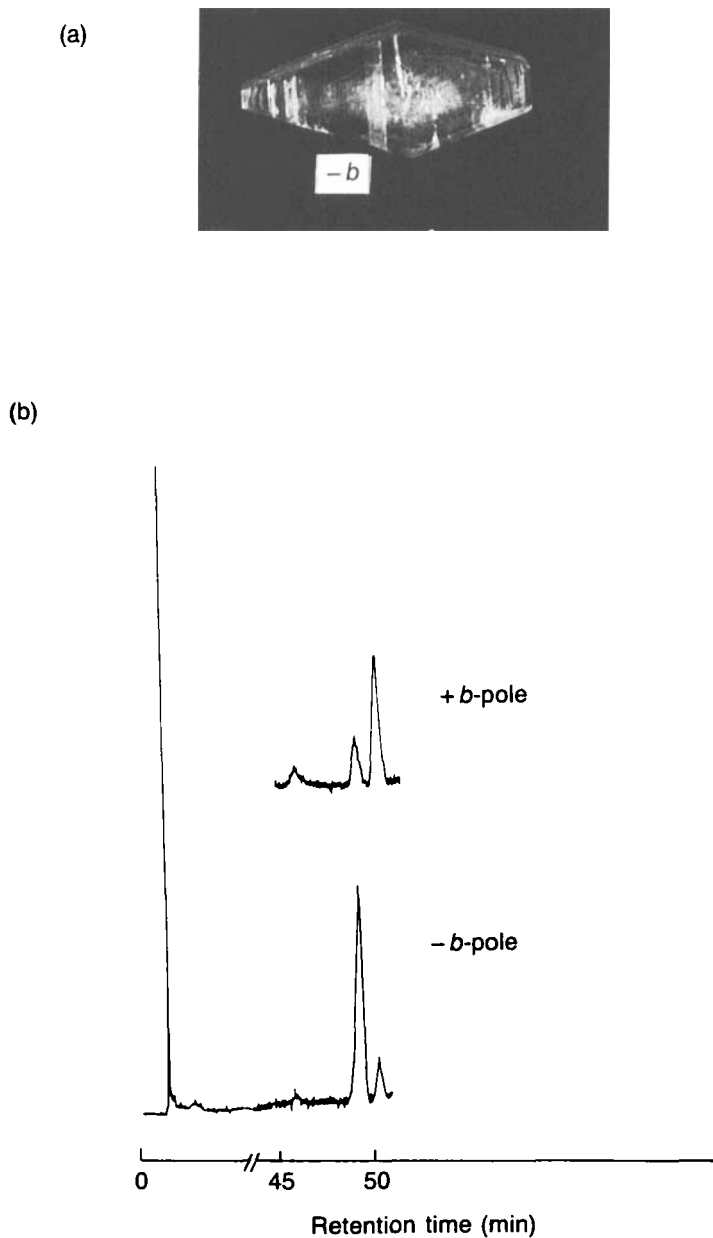


Figure 4 (a) Photograph of an (*E*)-cinnamamide crystal grown in the presence of (*E*)-cinnamic acid. The $-b$ -pole of the crystal is indicated. (b) Enantiomeric analysis by gas chromatography of the derivatives of the photodimerization products (shown in Figure 3) isolated from the $+b$ -pole and $-b$ -pole of the specimen crystal

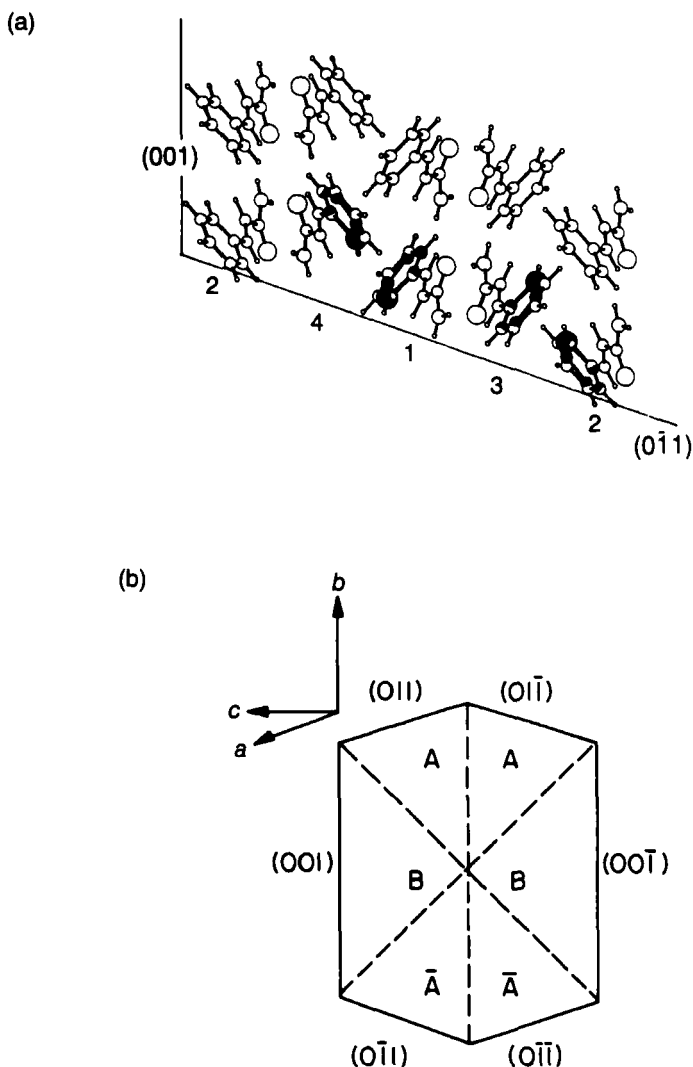


Figure 5 (a) Packing arrangement of (*E*)-cinnamamide showing the four different surface sites at the $(0\bar{1}1)$ face. The filled atoms are those of β -thienyl rings in the positions they would assume were they to replace cinnamamide molecules. (b) The morphology of a host-additive crystal of (*E*)-cinnamamide-(*E*)-2-(β -thienyl)acrylamide showing the six sectors of reduced crystal symmetry

through these faces. (*E*)-cinnamamide molecules at these faces display a herringbone arrangement involving interactions between aromatic C-H groups and the π -electron system of neighboring phenyl rings related by two-fold screw symmetry. To effect site-selective adsorption and occlusion, the herringbone contact involving the C-H groups and π -electron clouds was replaced by

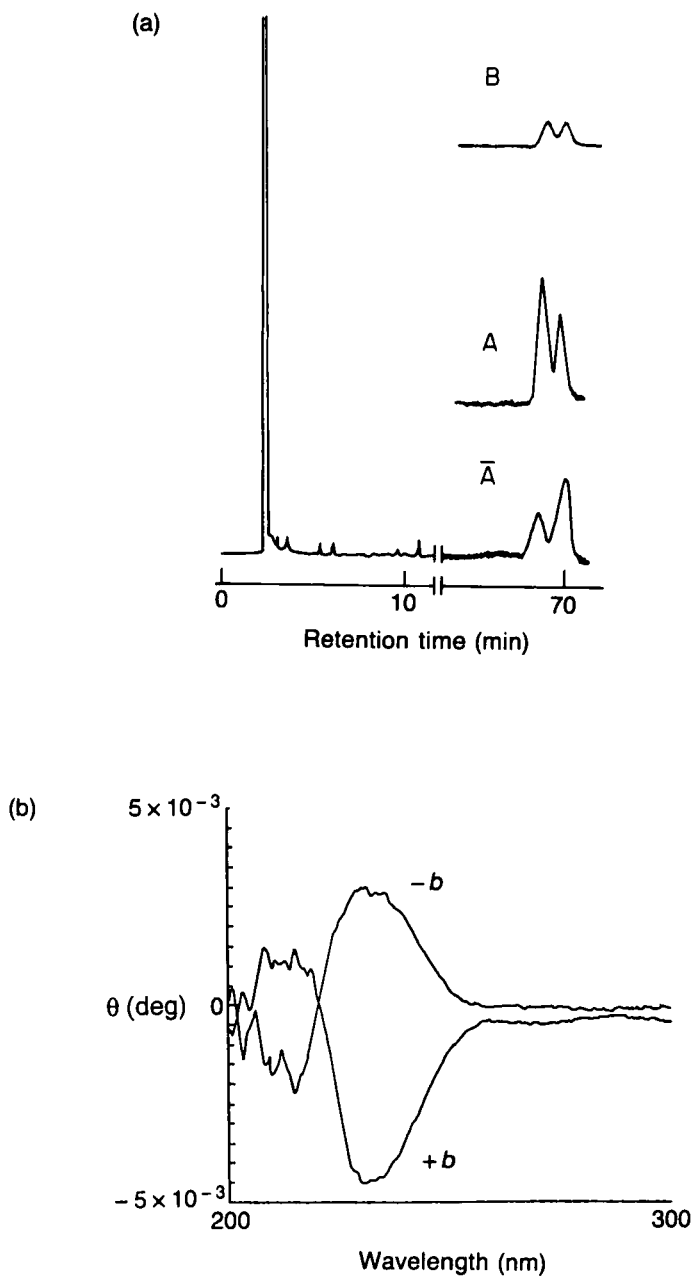


Figure 6 (a) Enantiomeric analysis by gas chromatography of the derivatized heterophotodimers of (*E*)-cinnamamide-(*E*)-2-(β -thienyl)acrylamide from crystal sectors, A, $\bar{A}$ and B (see Figure 5). (b) Circular dichroism spectra of the above photodimers obtained from crystal sectors A (+ *b*-pole) and $\bar{A}$ (− *b*-pole)

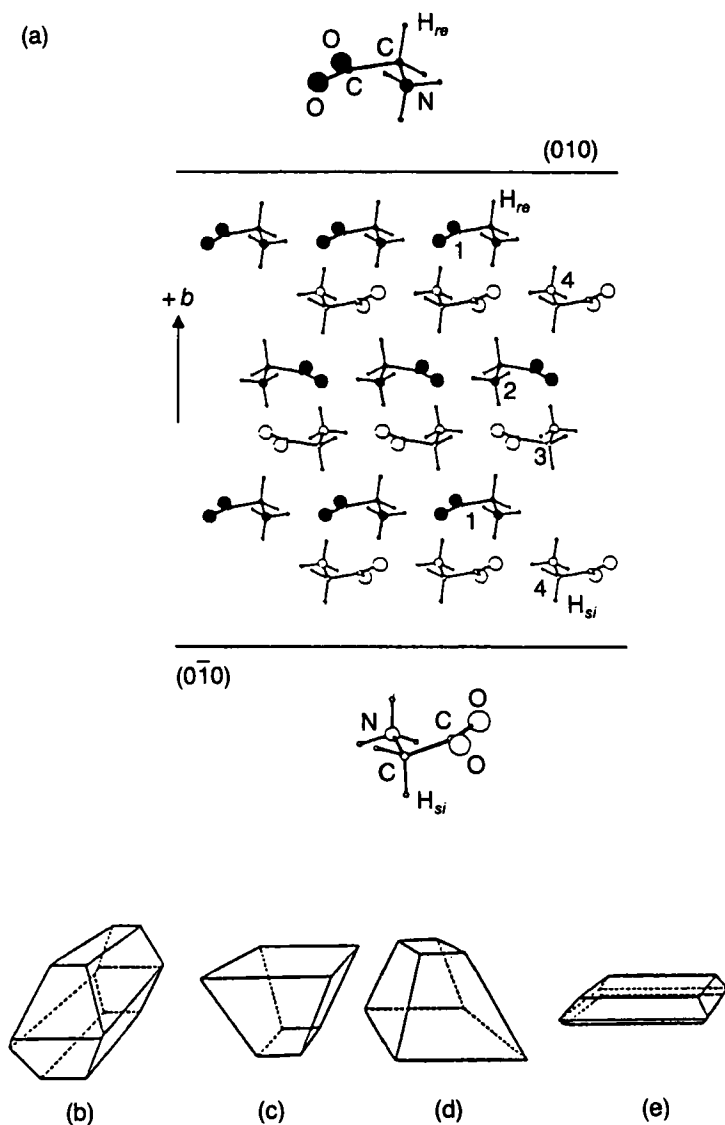


Figure 7 (a) Packing arrangement of α -glycine. Molecules 1 and 2 have their $C-H_{re}$ bonds directed along $+b$; molecules 3 and 4 have their equivalent $C-H_{si}$ bonds along $-b$. (b-e) Crystal morphologies of α -glycine: (b) pure α -glycine; and (c-e) crystals grown in the presence of (*R*)- α -amino acid, (*S*)- α -amino acid and (*R,S*)- α -amino acid additives, respectively

unfavorable contacts between sulfur lone pair electrons and the π -electron system. Following this line of thought, as shown in Figure 5(a), the guest (*E*)-2-(β -thienyl)acrylamide cannot easily occupy surface site 3 on the (0 $\bar{1}$ 1) face because the sulfur would point into the crystal bulk. At surface sites 2 and 4 the thienyl rings of the guests would lie almost parallel to the face and their relative ease of adsorption would depend only upon the tilt of the ring with respect to the face. The guest should be easily adsorbed at site 1 on the (0 $\bar{1}$ 1) face since the sulfur atom would emerge from the crystal surface.

Preferential occlusion at site 1 will yield the triclinic space group *P*1. Adsorption of (*E*)-2-(β -thienyl)acrylamide takes place also through the {001} faces, as revealed by the morphology change. Selective occlusion through these faces would lead to symmetry *P**c*, given that the guest should be adsorbed more easily at the (001) surface sites 1 and 2 than at 3 and 4. The mixed crystal should thus be divided into six sectors of reduced symmetry (Figure 5 b), with the structures of the sectors related to each other by the $2/m$ point symmetry of the host crystal; the two joined sectors of symmetry *P*1 of type A at one half of the crystal will be congruent and enantiomorphous to the two joined sectors of type $\bar{A}$ at the opposite half. The two *P**c* sectors (type B) will, by symmetry, have opposite senses of polarity along the *c*-axis.

We have demonstrated previously that irradiation with monochromatic light ($\lambda > 350$ nm) of a solid solution of (*E*)-cinnamamide-(*E*)-2-(β -thienyl)acrylamide grown from the melt leads to the formation of racemic mixtures of the

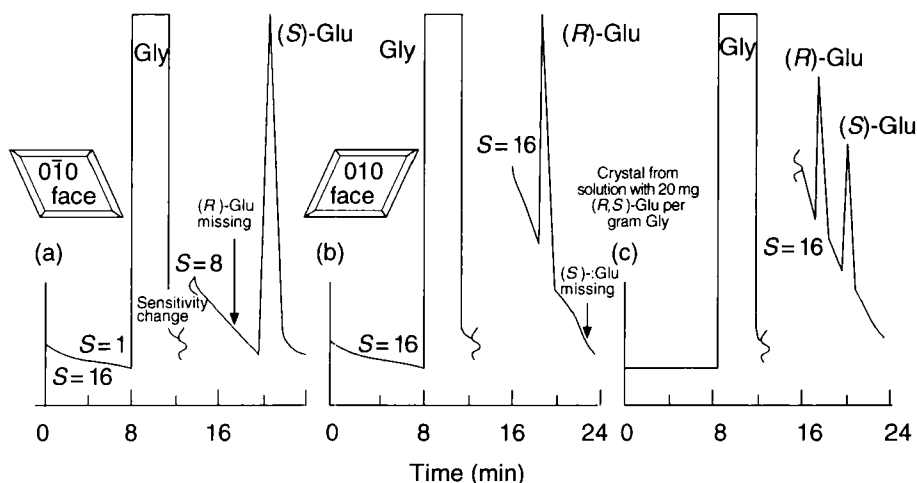


Figure 8 Enantiomeric distribution of occluded (*R,S*)-glutamic acid in a specimen platelike crystal of glycine determined by high-performance liquid chromatography (HPLC). (a) Only the (*S*)-glutamic acid additive is present in material shaved from the (0 $\bar{1}$ 0) face of the crystal. (b) Only the (*R*)-glutamic acid additive is present in material shaved from the (010) face. (c) Relative distribution of (*R,S*)-glutamic acid in the remaining whole crystal (*S* = sensitivity in arbitrary units)

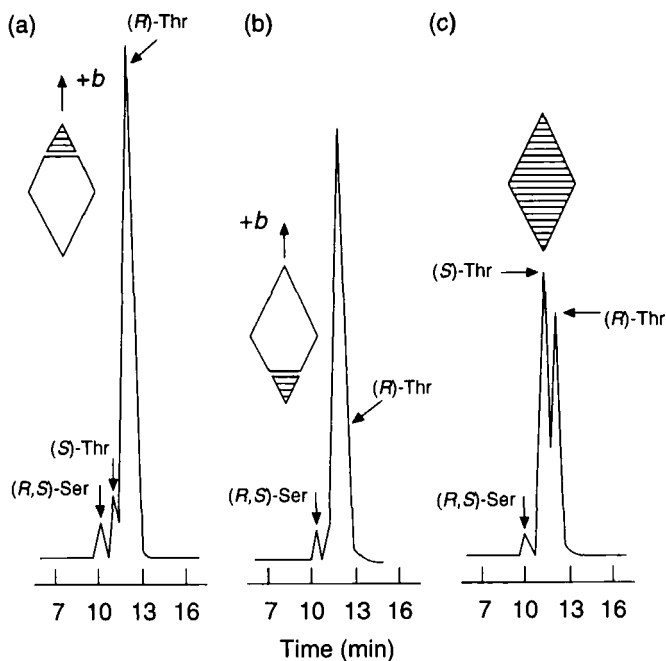


Figure 9 Enantiomeric analyses by HPLC of the occluded additive in (*R,S*)-serine crystals which were grown in the presence of racemic threonine. (a) Sample taken from the tip of the crystal at the +*b*-side. (b) Sample taken from the tip of the same crystal from the -*b*-side. (c) Analysis of the whole crystal. The small amount of unresolved serine in each case is residual from the previous separation of threonine from most of the serine on a cation exchange column

heterodimers [46]. In this mixed crystal there is no energy transfer from the host to the guest, so that the photoproduct distribution reflects directly the molecular occupancy of (*E*)-2-(β -thienyl)acrylamide inside the (*E*)-cinnamamide crystal. Consequently, on the basis of the arguments presented above, irradiation of sectors A and $\bar{A}$ of an (*E*)-cinnamamide crystal grown from solution and containing guest (*E*)-2-(β -thienyl)acrylamide should yield optically active dimers of opposite handedness. This expectation was confirmed experimentally by circular dichroism and gas chromatography analysis on a chiral column (Figure 6).

The absolute configurations of the two mixed dimers obtained from sectors A and $\bar{A}$ were assigned by NMR studies and found to be in agreement with the proposed mechanism. Furthermore, irradiation of the B-sector of the crystal yielded a racemic mixture. This model was independently confirmed by low-temperature X-ray and neutron diffraction analyses on different crystal sector

specimens, according to which the sectors A and $\bar{A}$ have reduced crystal symmetry $P1$ of opposite handedness and sector B has symmetry Pc .

(c) Generation of chirality in centrosymmetric crystals: the system α -glycine- α -amino acids [47, 48]

The principle of reduction in symmetry is general and was found to be applicable to a large variety of crystals. An important system, and one relevant to life processes, consists of the α -amino acids. Thus we focus on the centrosymmetric α -form of glycine grown from aqueous solution in the presence of α -amino acid additives.

Glycine, the one naturally occurring, nonchiral α -amino acid, is trimorphic (α , β , γ). The stable α -form appears in a monoclinic centrosymmetric

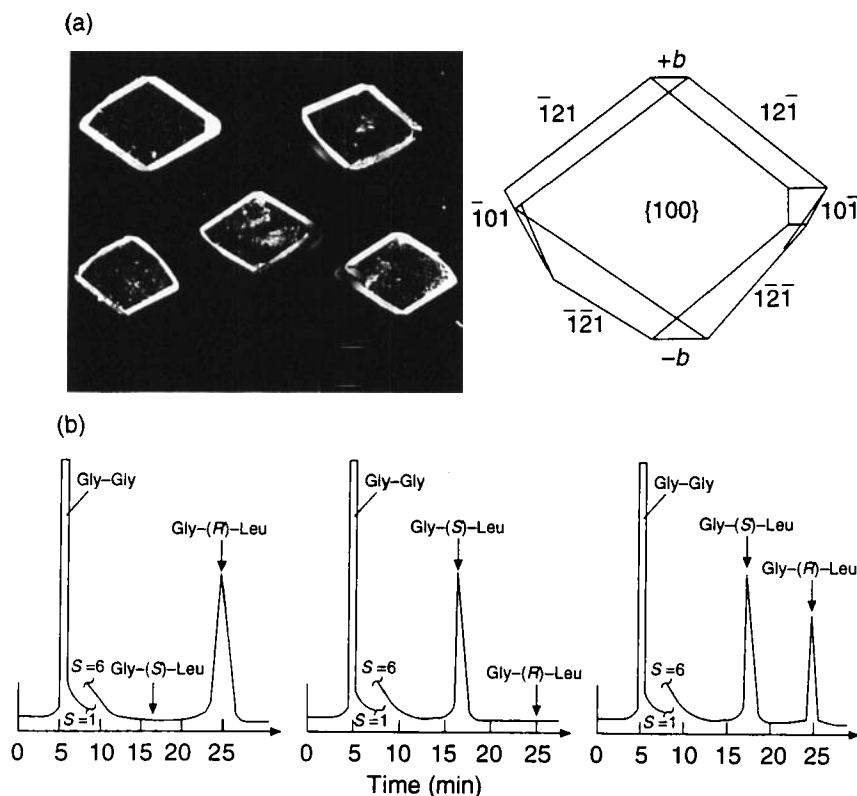


Figure 10 Gly-Gly crystals grown in the presence of racemic Gly-Leu. (a) Photograph and computer-drawn morphology. (b) Enantiomeric analyses by HPLC of samples taken from single crystals (left to right): sample taken from the $+b$ -end of the crystal, sample taken from the $-b$ -end of the crystal and sample from the whole crystal

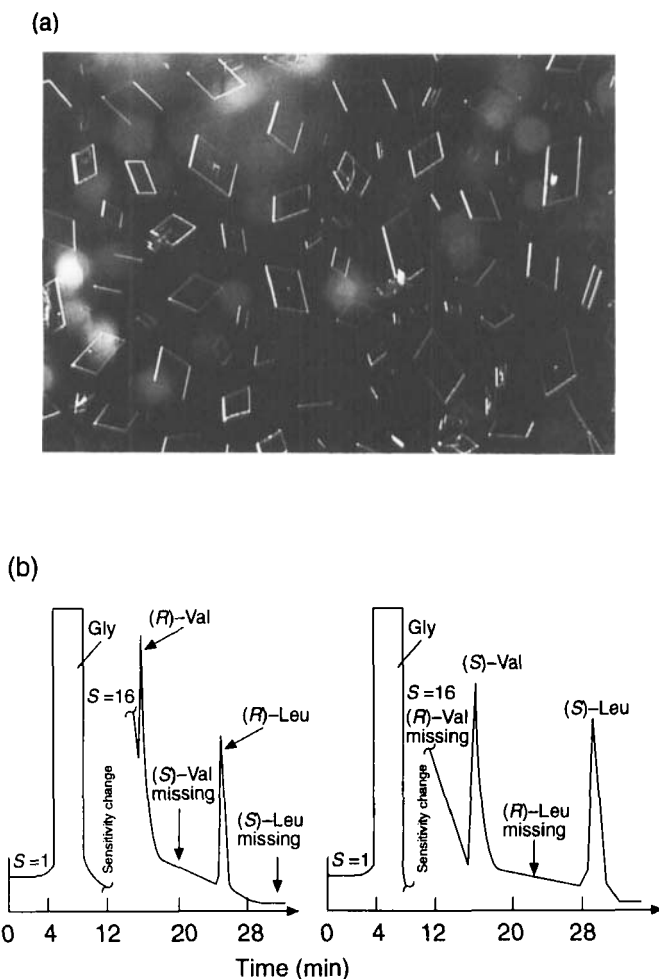
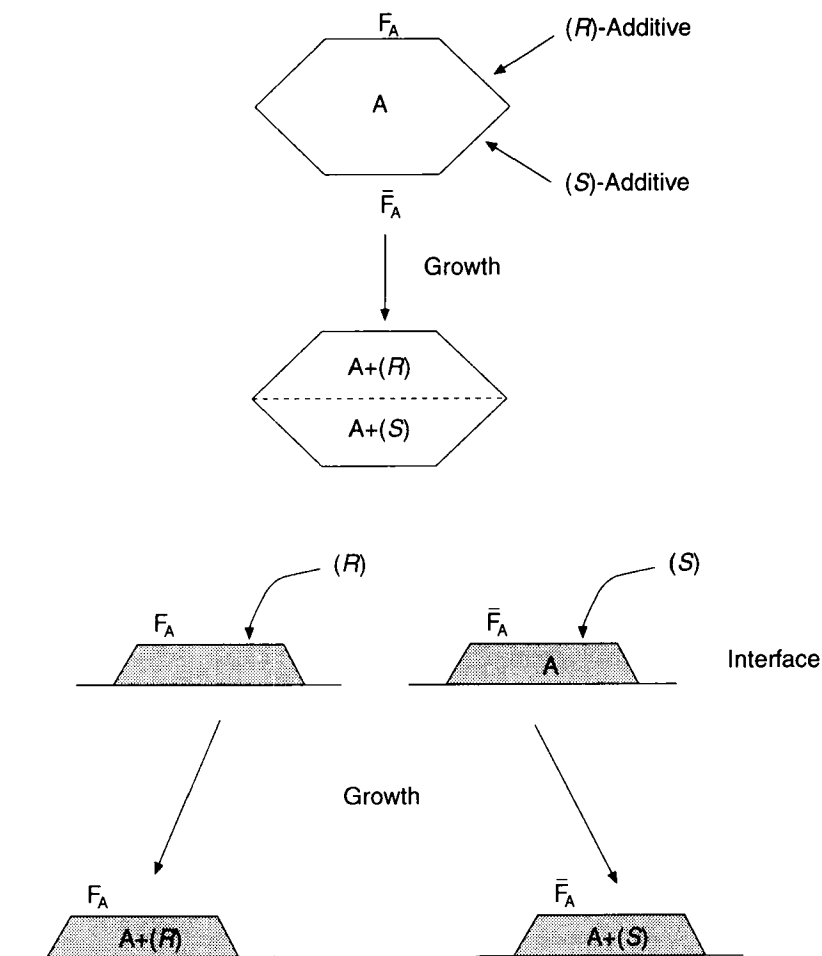


Figure 11 (a) Platelike crystals of α -glycine floating at the solution surface as grown from an aqueous solution containing (*R,S*)-leucine and (*R,S*)-valine. (b) Enantiomeric analyses by HPLC of single crystals of α -glycine: (left) crystals with their (010) face exposed to the air; and (right) crystals with their (010) face exposed to the air

arrangement of point symmetry $2/m$ [49]. The crystal faces relevant to the discussion are of the type {010}, as shown in Figure 7.

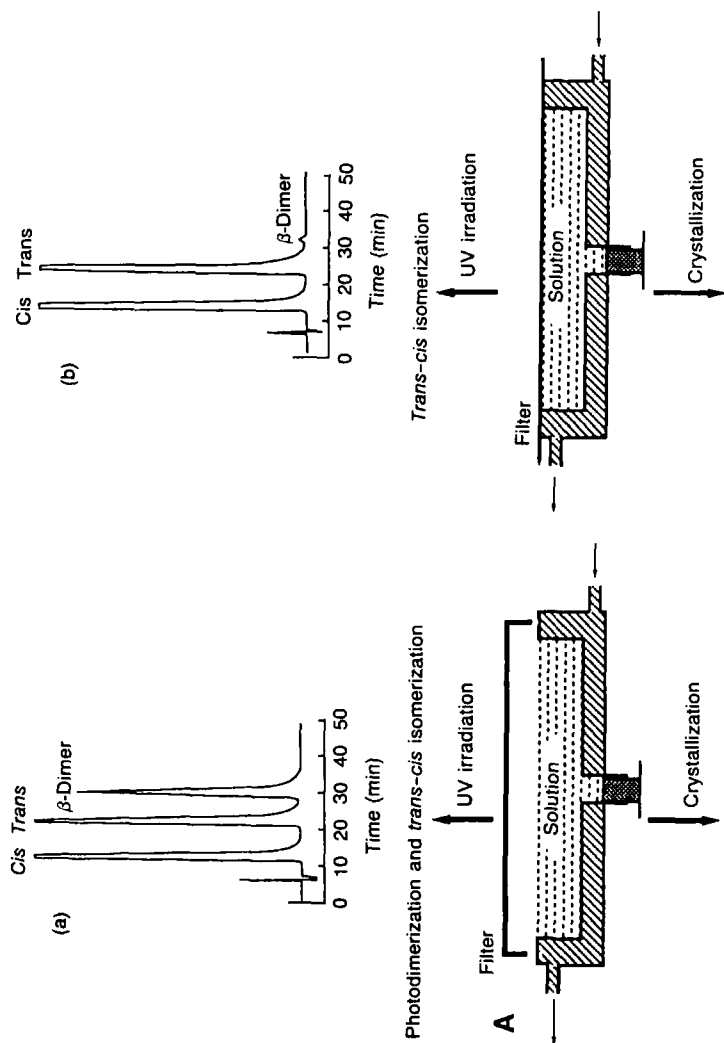
Of the four symmetry-related molecules (labeled 1, 2, 3 and 4 in Figure 7a), 1 and 2 are related by two-fold screw symmetry and are oriented such that their C-H_{re} bonds point in the +*b*-direction and so emerge from the (010) face.



Scheme 7

By crystal symmetry, the $C-H_{si}$ bonds of molecules 3 and 4, related to 1 and 2 by a center of inversion, point toward $-b$ and so emerge from the $(0\bar{1}0)$ face. Replacement of the $C-H_{re}$ and $C-H_{si}$ groups by α -amino acid side chains leads to molecules of (R) -configuration and (S) -configuration, respectively. Thus, only (R) -amino acid additives can substitute for a glycine molecule at sites 1 and 2, and only on face (010) , whereas only (S) -amino acids can be adsorbed at sites 3 and 4 on face $(0\bar{1}0)$. This constraint arises from the steric requirement that the additive molecule be recognized on a $\{010\}$ surface as a 'substrate' molecule, with the α -amino acid side chain emerging from the crystal surface.

α -Glycine crystallizes from water as bipyramids (Figure 7b). (R) - α -Amino acid additives induce the formation of pyramids (Figure 7c) with a (010) basal plane.



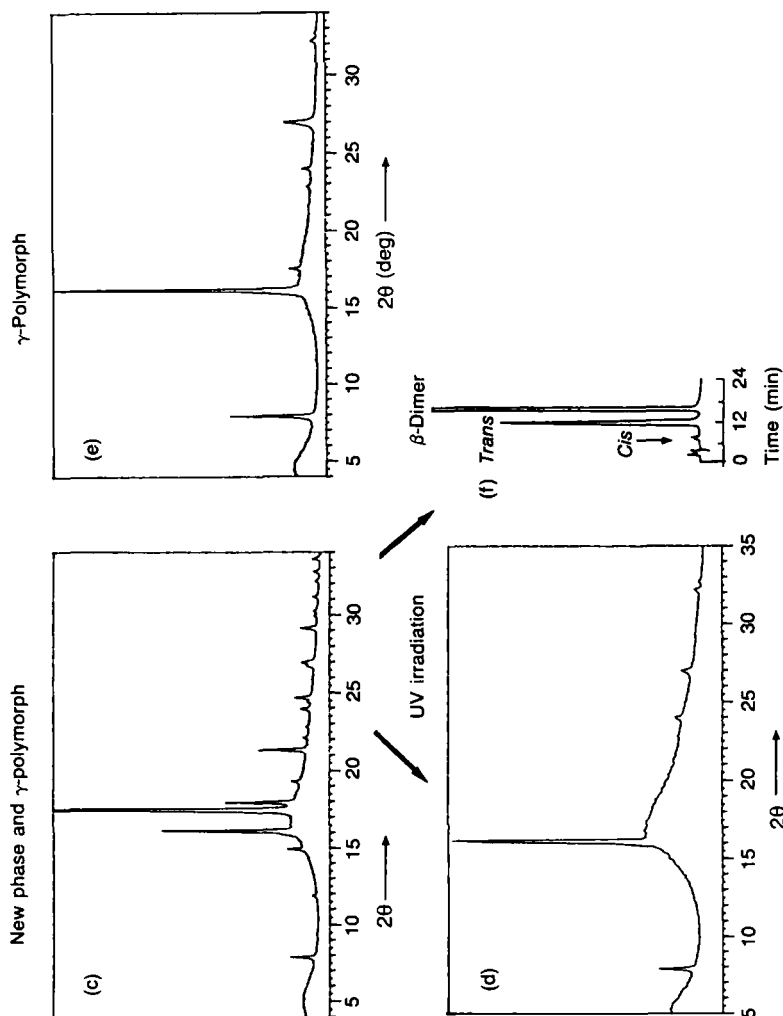


Figure 12 (a, b) Schematic drawings of the two experimental setups and the corresponding analyses by HPLC of the solutions irradiated in the two types of vessel: (a) sample from the open vessel; and (b) sample from the closed vessel. (c–e) X-Ray powder patterns of the crystalline materials: (c) sample precipitated in the open vessel; (d) sample as in (c) after solid-state UV irradiation; and (e) sample precipitated in the closed vessel. (f) Analysis by HPLC of the UV irradiation product from the sample obtained as in (c)

(*S*)- α -Amino acids induce the enantiomorphous morphology (Figure 7d). Racemic additives cause the formation of {010} plates (Figure 7e), because growth at both the $+b$ -side and $-b$ -side of the crystal is inhibited. These plates have been found, by enantiomeric HPLC analysis, to contain 0.02–0.2% racemic additive occluded in the crystal bulk, with the two enantiomers totally segregated in the two crystal sectors at the $+b$ -half and $-b$ -half (Figure 8).

As expected, the (*R*)-enantiomers populate the $+b$ -half of the crystal and the (*S*)-enantiomers the $-b$ -half. In terms of the arguments given above, the crystal symmetry of each half must be reduced from $P2_1/n$ to $P2_1$, and the two sectors are enantiomorphous. It is noteworthy that one may assign the absolute configuration of the resolved α -amino acid additives through the morphological changes they selectively induce at one set of enantiotopic faces of the centrosymmetric crystal.

Such a 'spontaneous' segregation of enantiomers inside a centrosymmetric crystal is general and has been demonstrated for other systems, such as the segregation of racemic threonine inside the crystals of (*R,S*)-serine [50] (Figure 9) and glycylleucine dipeptide inside crystals of glycylglycine [51] (Figure 10).

The ability to transform a single centrosymmetric host crystal into a mixed crystal of lower symmetry provides a new means of designing chiral matrices for the performance of spontaneous asymmetric transformations. Envisage a centrosymmetric single crystal grown at an interface, either floating on the solution or in contact with a glass surface such that one of the two enantiotopic

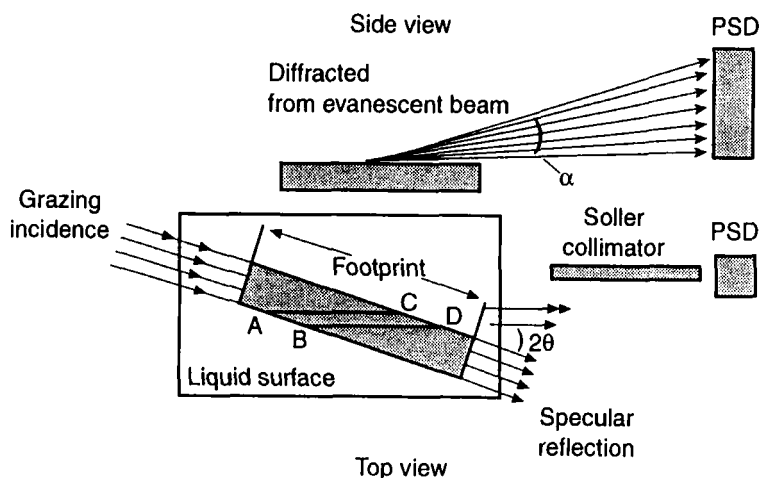


Figure 13 Side and top views of the grazing incidence X-ray diffraction geometry. The footprint of the grazing incident beam is indicated by the shaded area. The position-sensitive detector (PSD) has its axis along the vertical. Only the cross-beam area ABDC contributes to the measured scattering. The Soller collimator consists of thin vertical absorbing foils stacked together to define the horizontal resolution of the detector

faces is blocked. For example, imagine a single crystal of α -glycine with its $(0\bar{1}0)$ face blocked by an interface and the opposite enantiotopic (010) face exposed to a supersaturated glycine solution also containing a racemic mixture of other α -amino acids. Only the (*R*)-amino acids will be occluded through the (010) face into the growing crystal. By symmetry, were the $(0\bar{1}0)$ face exposed to the solution, the (*S*)-amino acids would be occluded. Such a process should result in a net resolution, one enantiomer being occluded into the single crystal so that the solution becomes enriched in the other. The net resolution of leucine and valine occluded into such floating, platelike single crystals of α -glycine, as shown in Figure 11, confirmed these expectations.

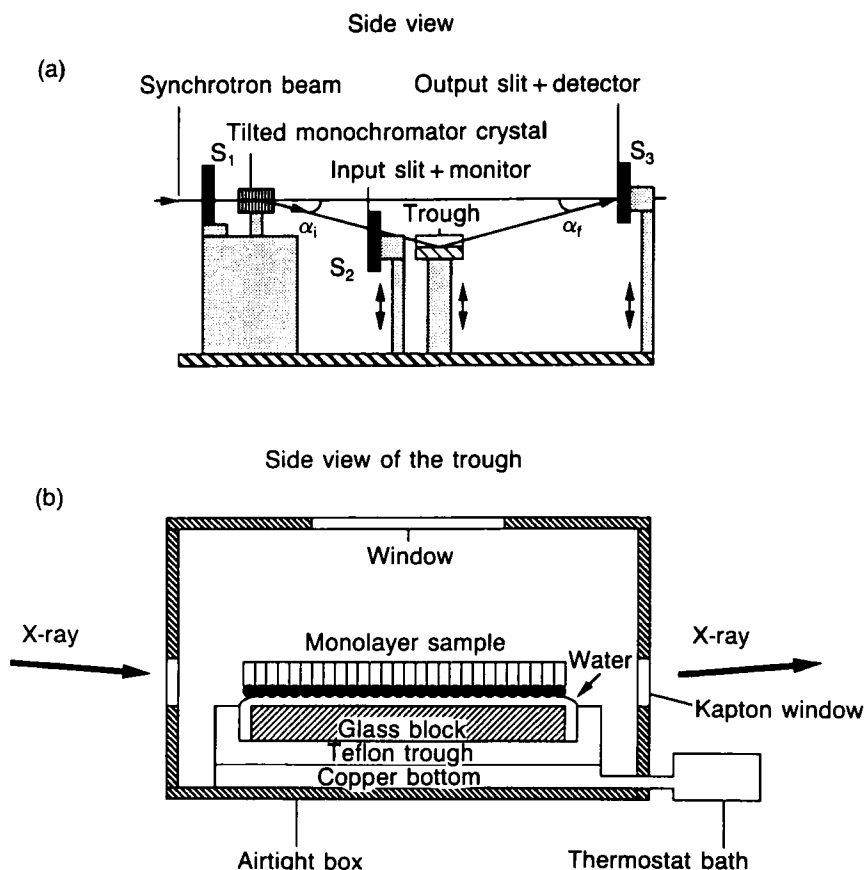


Figure 14 Example of an experimental setup for a liquid surface diffractometer. (a) Side view of the vertical scattering plane. The beam directions are defined by slits. The monochromatic beam is bent down towards the sample by tilting the monochromator crystal. The incident beam intensity is monitored after slit S_2 . (b) Blown up view of the trough showing the monolayer sample spread on a thin film of water. The glass block provides a thin liquid film (about 0.3 mm thick) and thus effectively reduces surface capillary waves

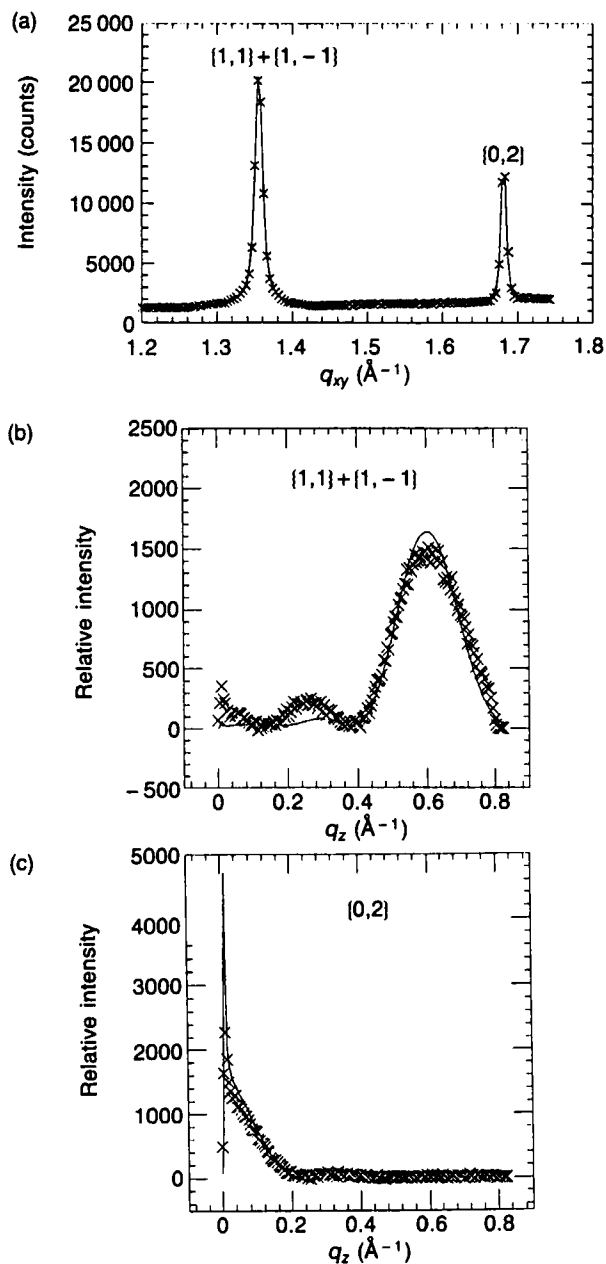


Figure 15 Grazing incidence X-ray diffraction (GID) measurement of an uncompressed monolayer of 4-(octadecyl)oxycinnamamide at 5 °C on water. (a) Bragg peaks along the horizontal scattering vector $q_{xy} (= 4 \pi \sin(\Theta/\lambda))$, where Θ is indicated in Figure 13). (b, c) Observed and calculated Bragg rod intensity profiles along the vertical scattering vector $q_z (= 2 \pi \sin(\alpha/\lambda))$, where α is indicated in Figure 13) for each of the two reflections

In the absence of an outside chiral influence, however, a glycine crystal grown from a racemic mixture of α -amino acids will be equally likely to float on its (010) or (0 $\bar{1}$ 0) face. Since the two enantiotopic {010} faces of glycine interact differently with a chiral environment, there might be a preference for the crystals to be attached to the interface with one of the enantiotopic faces. If such an asymmetric induction can be achieved, one might be in a position to amplify the optical activity generated in a single crystal by extending the model proposed in Scheme 4 to centrosymmetric crystals (Scheme 7).

We addressed the question as to whether there are processes which promote and, moreover, replicate a specific orientation of growing crystals floating at the air–solution interface. The first process should involve the design of a template for induced nucleation at the interface and, as presented in Section 3, involves the self-aggregation of amphiphilic molecules at the air–solution interface into two-dimensional crystalline arrays. The second process should cause an enantioselective inhibition of ‘wrongly’ oriented crystal nuclei at the interface. This mechanism is intimately linked to the structure and dynamics of the nuclei as influenced by the presence of additives, and shall be discussed in Section 4. Finally, in Section 5 an experiment is described in which these two processes operate in tandem.

3. SELF-AGGREGATION OF AMPHIPHILIC MOLECULES AT AIR–SOLUTION INTERFACES

We examine the self-aggregating properties of water-soluble amphiphiles at liquid surfaces in terms of their ability to induce 3D nucleation of appropriate cosolute molecules in the solution. These studies are complemented by the use of the corresponding water-insoluble amphiphiles for induced 3D nucleation, taking advantage of the fact that the 2D crystal structures of such self-aggregates can be determined by modern techniques.

3.1 Modeling the Crystal Nucleation of 4-Methoxycinnamic Acid [52, 53]

When amphiphilic molecules are dissolved in water, they are partitioned between the solution and the air–solution interface. The degree of coverage of the surface can be evaluated by surface tension measurements. Such studies on saturated aqueous solutions of 4-methoxycinnamic acid (MCA) demonstrate that the

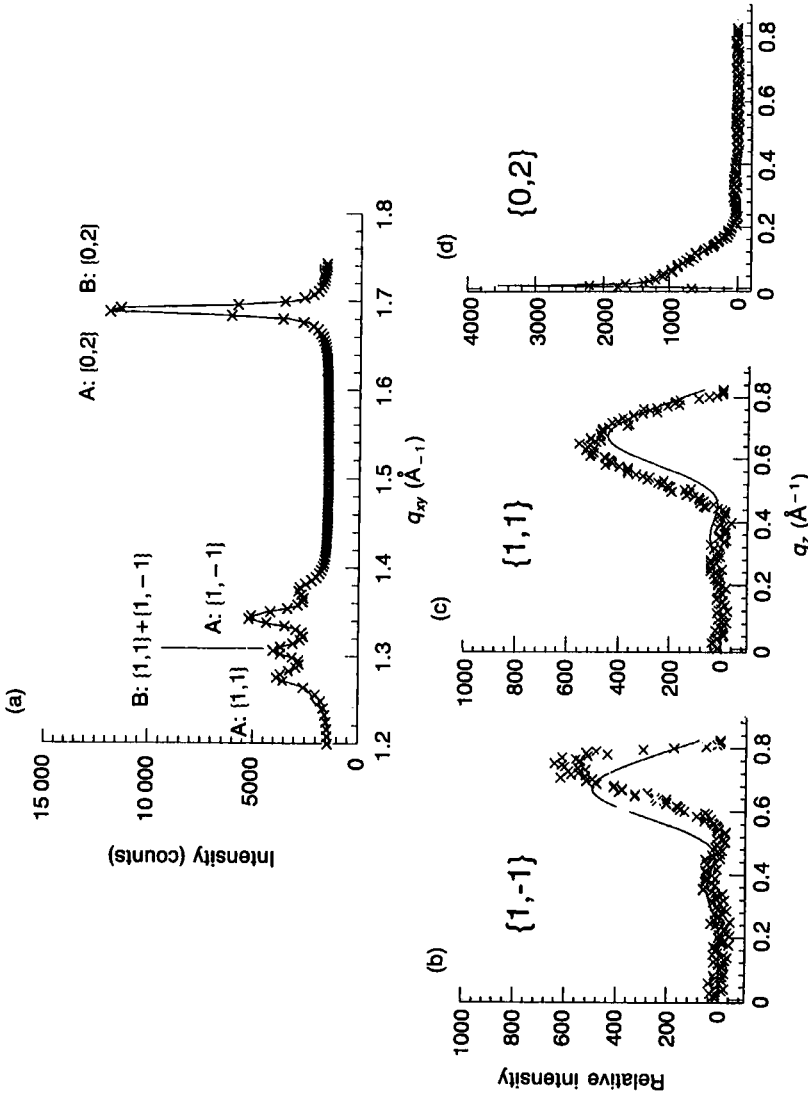


Figure 16 Grazing incidence X-ray diffraction measurement of an uncompressed monolayer of 4-(octadecyl)oxycinnamic acid at 5 °C on water. (a) The GID pattern showing two phases A and B. The B-phase is similar to the structure of the corresponding amide (see Figure 14). The A-phase contains three reflections, indicating a pseudorectangular unit cell. (b–d) Observed and calculated Bragg rod intensity profiles for each of the three reflections which belong to phase A

solution–air interface is covered by $\sim 30\%$ of the acid. This result was confirmed by second-harmonic generation spectroscopy, which did not provide, however, information as to how the molecules are organized at the molecular level.

Structural considerations, based on the packing modes of the cinnamic acids, suggested that MCA molecules should form, at the air–solution interface, close-packed stacks of ‘translationally’ related molecules separated by $4.0 \text{ \AA}$ and stabilized by aromatic ring interactions and Coulombic forces between the carboxy groups. This packing should influence the photochemical behavior of these molecules, as was established from the photoreactivity in the solid state. Thus, while all the cinnamic acids undergo *trans*–*cis* photoisomerization in solution, the MCA aggregates at the air–solution interface are expected to undergo a $2\pi + 2\pi$ photodimerization leading to the mirror symmetric cyclobutane 4,4'-dimethoxy- β -truxinic acid. Differentiation between the products formed in the bulk solution and at the interface was done by irradiating MCA solutions in two different experimental setups: closed glass vessels completely filled with solution and open glass vessels with a large solution–air interface (Figure 12). The closed vessel resulted in a *trans*/*cis* ratio of about 1:3 (Figure 12b), whereas the open vessel yielded, in addition, 5–20% of the mirror photodimer (Figure 12a), depending upon the surface area of the interface and the exposure time. The molecular organization at the surface could be gradually reduced by the addition of ethanol to the solution (35%), as made manifest by the formation of only traces of the photodimer.

A similar structural analysis of 4-methoxy-(*E*)-cinnamamide implies that it should form, at the interface, two-dimensional clusters within which the molecules form stacks interlinked by $\text{N-H} \cdots \text{O}=\text{C}$ hydrogen bonds. Because the hydrogen-bonding repeat distance is about $5 \text{ \AA}$ and the cross-sectional area of the molecule perpendicular to its long axis is about $25 \text{ \AA}^2$, in this arrangement the separation between the carbon–carbon double bonds of molecules in neighboring stacks is about $5 \text{ \AA}$, which is too large a distance to allow photodimerization. Indeed, irradiation of the amide resulted in only *trans*–*cis* photoisomerization, in keeping with the model.

In order to probe the role played, if any, by the ordered self-aggregates of MCA at the interface in promoting the formation of new polymorphic forms, this material was precipitated from aqueous solution, by cooling, in the two experimental setups. The open vessel yielded a new crystalline phase (Figure 12c) not previously reported, which precipitated together with the known phase which had been previously defined as the γ -form. UV irradiation of the mixed crystalline phases yielded the mirror photodimer. This dimer must arise from the new crystalline phase because the γ -form is light stable and the diffraction peaks associated with the new phase disappeared upon irradiation (Figure 12d). From the crystallization experiments performed in the closed vessels, the γ -phase precipitates either in pure form or with only small amounts of the new phase (Figure 12e).

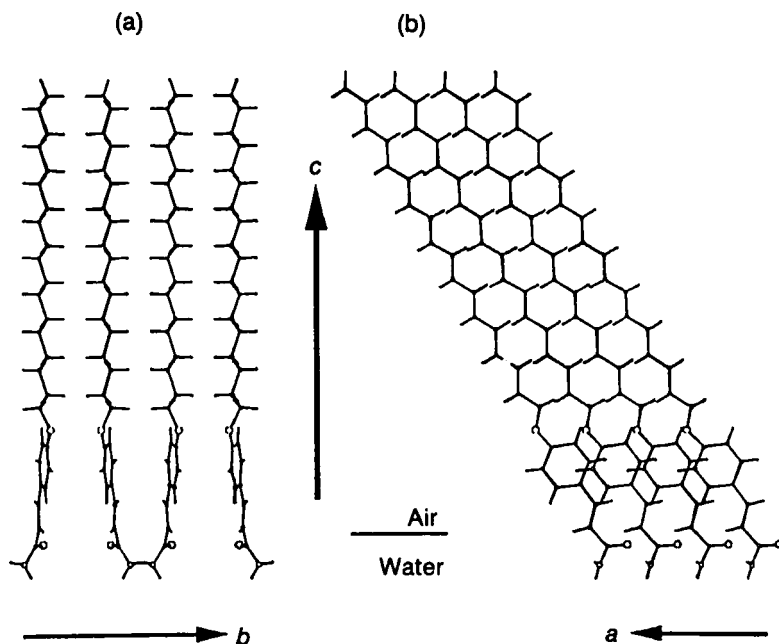


Figure 17 Packing arrangement of the uncompressed monolayer of 4-(octadecyl)oxycinnamide at 5 °C on water. (a, b) Views along the *a*-axis and *b*-axis, respectively, where the rectangular *ab* plane is parallel to the water surface

The molecular packing characteristics of 4-methoxycinnamic acid and cinnamamide on water are, at present, not amenable to direct elucidation by modern diffraction methods. Nevertheless, recent advances in surface X-ray diffraction have made feasible the structural elucidation of monolayers of water-insoluble amphiphiles on liquid surfaces by grazing incidence X-ray diffraction (GID) using synchrotron radiation [54, 55]. In this respect we took advantage of the photochemical behavior of monolayers of water-insoluble 4-alkoxycinnamic acid and cinnamamide amphiphiles ($\text{C}_{18}\text{H}_{37}\text{OC}_6\text{H}_4\text{CH}=\text{CHX}$, $\text{X}=\text{CO}_2\text{H}$, CONH_2) spread on the water surface. UV irradiation of the amphiphilic acid monolayer yields the mirror dimer, whereas the amide system only undergoes *trans*-*cis* isomerization in a manner akin to the 4-methoxy analogues described above. Thus we examined by GID the two-dimensional crystalline properties of uncompressed monolayers of 4-(octadecyl)oxycinnamic acid and 4-(octadecyl)oxycinnamamide on the water surface in the light of their photochemical behavior [56]. The GID measurements were made on the liquid surface diffractometer [54] shown schematically in Figures 13 and 14. Both monolayers form self-aggregated 2D crystalline clusters according to their GID patterns shown in Figures 15(a) and 16(a). The widths of the diffraction peaks indicate perfect 2D crystalline order ranging from 200 to 1000 Å. Since the

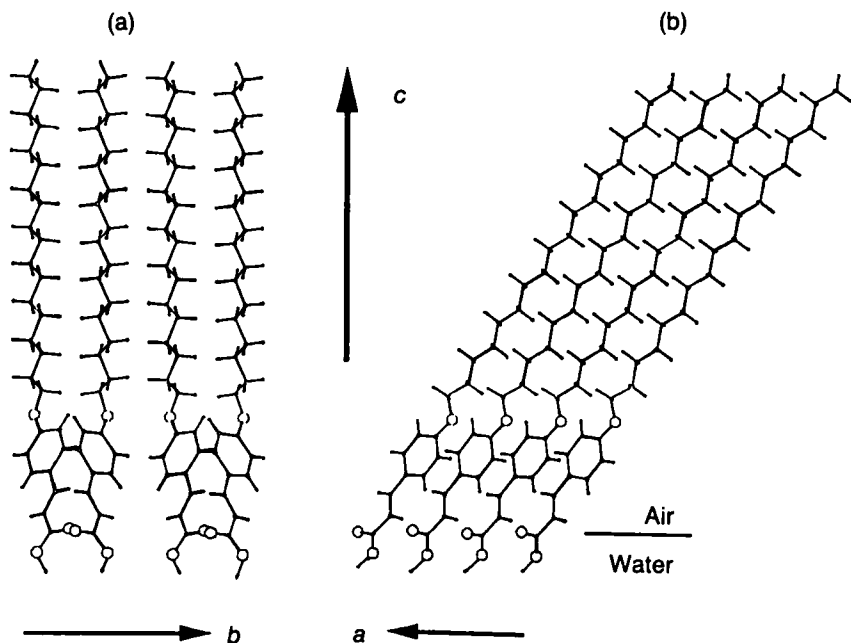


Figure 18 Packing arrangement of the A-phase of the uncompressed 4-(octadecyl)-oxycinnamic acid monolayer at 5 °C on water. (a, b) Views along the a -axis and b -axis, respectively

clusters on the water surface are crystalline in only two dimensions, the GID reflections are composed of Bragg rods extending in a direction perpendicular to the water surface (Figures 15b, 15c and 16b–16d), and so were measured by a position-sensitive detector. The X-ray intensity distribution along these Bragg rods provides information on the packing arrangement of the molecular chains (Figures 17 and 18). In the case of the amide system, the molecules are related by a glide symmetry separated by 5 Å and interlinked by $\text{N-H} \cdots \text{O}=\text{C}$ bonds. This 5 Å separation (Figure 17) precludes intermolecular photodimerization. The GID pattern of the acid shows the coexistence of two 2D crystalline phases. In contrast to the amide system, the packing structure of the major phase (Figure 18) shows a 4 Å separation between the $\text{C}=\text{C}$ moieties of neighboring glide-related cinnamic acid head-groups which are essentially parallel, in keeping with the formation of the mirror photodimer.

The crystallization experiments, photochemical behavior and direct structure determination of the 2D crystals by grazing incidence diffraction demonstrate the formation at the air–solution interface of ordered two-dimensional aggregates which play an important role as nuclei *en route* to 3D crystal precipitation.

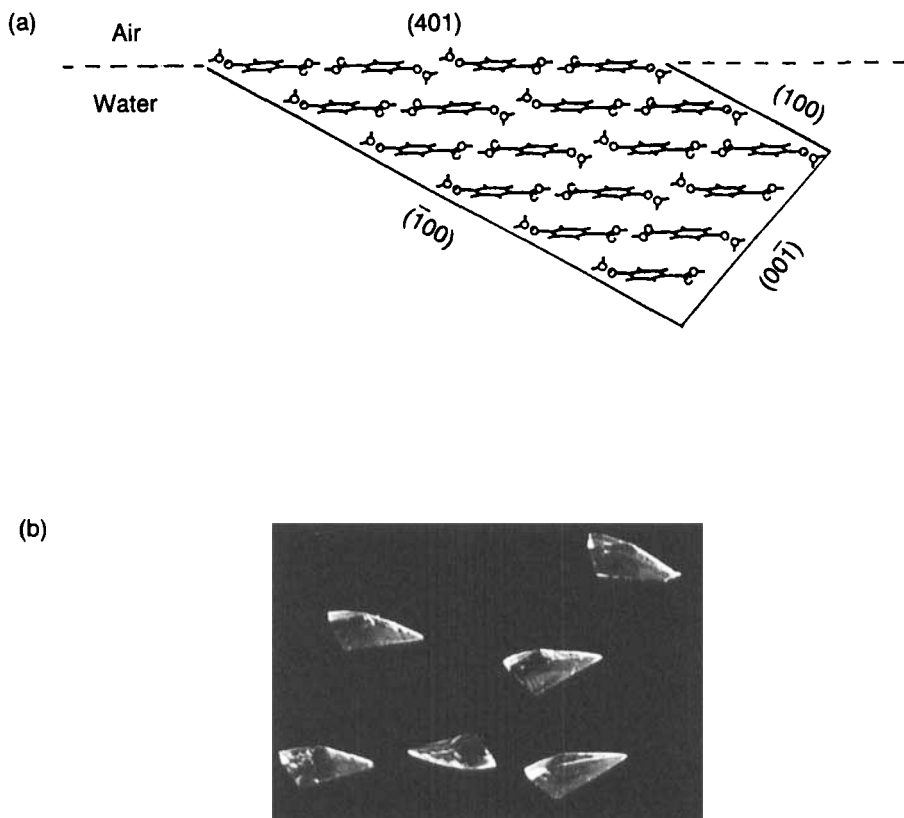


Figure 19 (a) Packing arrangement of a 4-hydroxybenzoic acid (HBA) crystal with the (401) face viewed 'edge on' as delineated by the crystal faces. (b) Photograph of HBA crystals nucleated by a monolayer of 4-(octadecyl)oxybenzoic acid amphiphile at the solution surface

3.2 Oriented Nucleation of 4-Hydroxybenzoic Acid Monohydrate [57, 58]

Self-assembled two-dimensional ordered aggregates of water-soluble hydrophobic molecules at the water surface can serve as a template for the oriented crystallization of suitable cosolute molecules at the solution surface. The cosolute under discussion is 4-hydroxybenzoic acid (HBA), which precipitates as a monohydrate from aqueous solution. The crystal is composed of hydrogen-bonded layers within which the molecules form cyclic hydrogen-bonded carboxy dimers interlinked by $\text{O}-\text{H}\cdots\text{O}$ hydrogen bonds between the phenolic OH groups and the water molecules. These neighboring layers, parallel to the (401) plane of the crystal, make interlayer contact to complete the crystal structure (Figure 19).

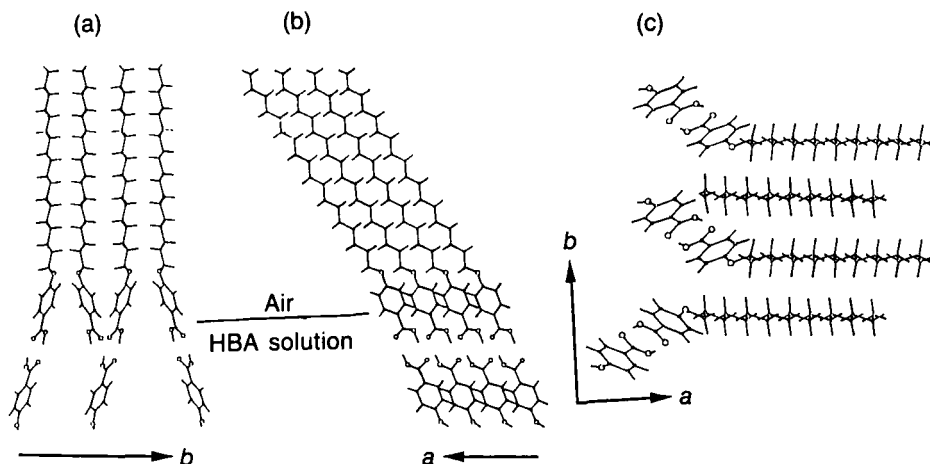


Figure 20 Packing arrangement of a 4-(octadecyl)oxybenzoic acid monolayer with bound HBA solute molecules: (a, b) views along the *a*-axis and *b*-axis, respectively; and (c) view perpendicular to the solution surface. For clarity, the head-group of one chain was removed

Interestingly, the (401) plane is not expressed as an observed face in the crystal morphology. But when HBA is crystallized in the presence of 4-methoxybenzoic acid, the crystals first appear floating on their (401) faces. The result suggests that the additive molecules, which are more hydrophobic than HBA, accumulate at the solution surface and so facilitate the formation of layered aggregates with HBA in a structure which mimics the (401) layer of HBA. This study was complemented by the use of water-insoluble 4-substituted benzoic acid amphiphiles ($C_nH_{2n+1}XC_6H_4CO_2H$, $X=O, NH$, $n=16, 18$) for the induced nucleation of HBA. The structures at the interface were monitored by surface pressure–area isotherms, optical second-harmonic generation (SHG) [57], specular X-ray reflectivity and GID using synchrotron radiation [58]. The GID measurements of uncompressed monolayers of $C_{18}H_{37}OC_6H_4CO_2H$ on pure water and on a solution containing 4-hydroxybenzoic acid showed the presence of 2D crystalline aggregates. The crystallinity of the monolayer on a solution containing 4-hydroxybenzoic acid was considerably enhanced. Analysis of the GID patterns showed that 40% of the solute molecules were bound to the amphiphile and tilted at an angle of 36° from the normal to the water surface (Figure 20). We propose the following mechanism to account for the oriented nucleation of HBA under the monolayer. In the (401) plane of the HBA crystal the molecules are related by a $6.4 \text{ \AA}$ translation along the *b*-axis, forming an array. Neighboring arrays are interlinked by hydrogen-bonded water molecules across a two-fold screw axis, where the separation between the phenolic oxygens is about $3 \text{ \AA}$ (Figure 21a). The monolayer molecules form ordered arrays separated by $6.2 \text{ \AA}$ along the *a*-axis (Figure 21b), which is similar to the $6.4 \text{ \AA}$

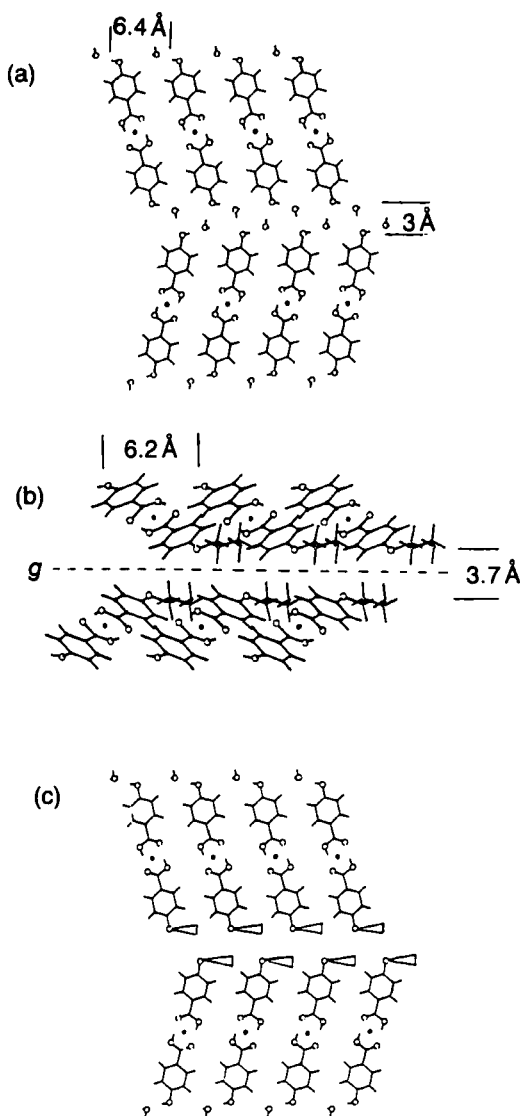
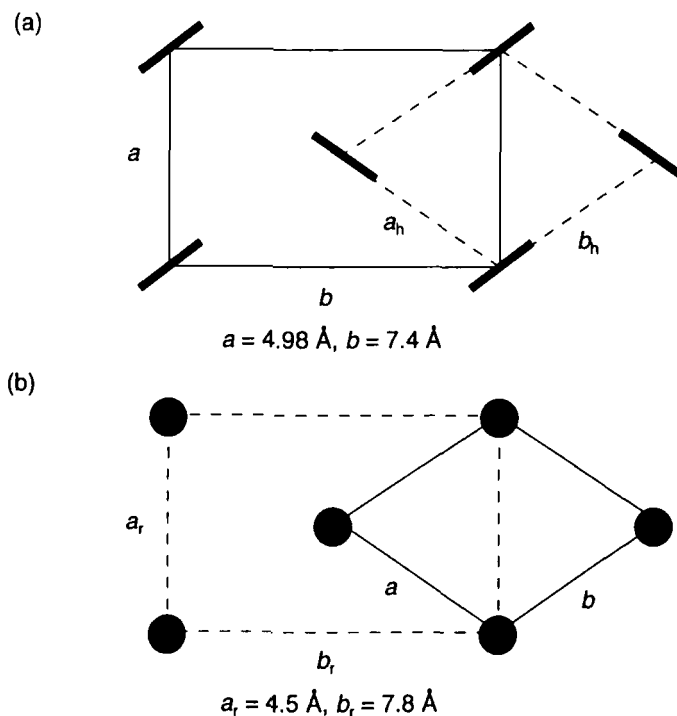


Figure 21 Proposed mechanism to explain the oriented nucleation of HBA crystals as induced by the 4-(octadecyl)oxybenzoic acid monolayer. (a) Packing arrangement of the HBA molecules in the monohydrate crystals viewed perpendicular to the (401) face. (b) Packing arrangement, as viewed perpendicular to the solution surface, of a double array of cyclic hydrogen-bonded dimers formed between molecules of the 4-(octadecyl)-oxybenzoic acid monolayer and the bound HBA solute. Each array contains three molecular units related by a $6.2 \text{ \AA}$ translation. For clarity, only part of each hydrocarbon chain is shown. (c) Proposed rearrangement of the double array of monolayer molecules to serve as a template for HBA crystal nucleation. Note the similarity to the corresponding double array in (a) and that the wedges represent hydrocarbon chains

repeat in HBA. Finally, two such neighboring amphiphilic molecular arrays with bound solute molecules, related by glide symmetry and separated by $3.7 \text{ \AA}$, can be regarded as an ordered double layer which resembles the corresponding double molecular array in the (401) plane of HBA. Clearly we must still assume a molecular reorientation of the cyclic hydrogen-bonded dimers of the amphiphilic head-groups and bound solutes (i.e. $-\text{C}_6\text{H}_4\text{CO}_2\text{H} + \text{HO}_2\text{CC}_6\text{H}_4\text{OH}$) to best mimic the (401) crystal layer at the nucleation stage and facilitate epitaxial nucleation (Figure 21).

3.3 Epitaxial Nucleation of Ice by Amphiphilic Alcohols [59, 60]

We now examine the template-induced nucleation of ice from supercooled water. Recent studies have demonstrated that the freezing temperature provides a very sensitive measure of the match between the template and the layer structure of ice. Water-soluble alcohols generally reduce the freezing point of water. On the other hand, in view of the self-aggregating properties of water-insoluble aliphatic amphiphiles on water and the layer-packing characteristics of crystalline, long-chain alkanes, it was expected that water-insoluble aliphatic alcohols would form 2D crystalline self-aggregates at the water surface to give



Scheme 8

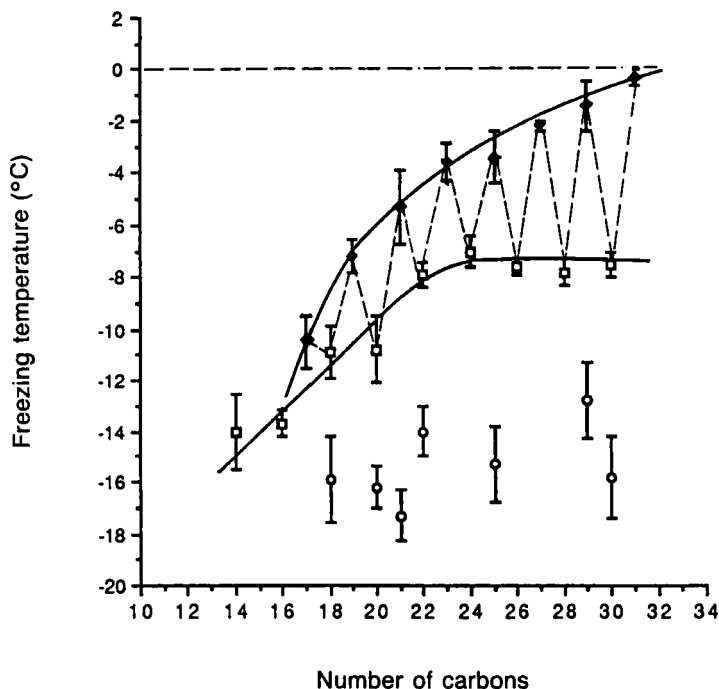


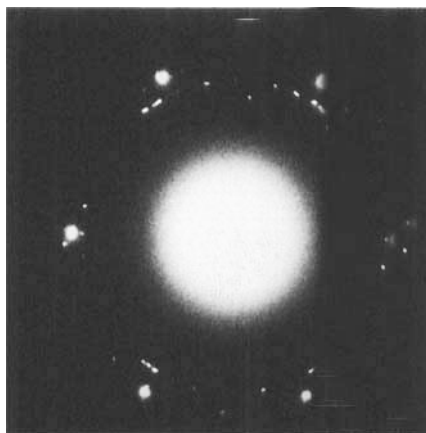
Figure 22 Freezing temperatures of supercooled water drops covered by monolayers of aliphatic alcohols $C_nH_{2n+1}OH$ ($\square$ C_n -even, $\blacklozenge$ C_n -odd) and carboxylic acids $C_{n-1}H_{2n-1}CO_2H$ ($\circ$) versus the number of carbon atoms in the chain

a 2D net of hydroxy groups which matches well the *ab* layer structure of hexagonal ice. Such an arrangement should induce the freezing of supercooled water. GID studies on some uncompressed monolayers of alcohols [61] in the series $C_nH_{2n+1}OH$ indeed demonstrate that such monolayers form highly crystalline arrangements with an *ab* net similar in dimension to the *ab* layer of hexagonal ice (Scheme 8).

The expected catalysis of ice nucleation by the alcohols was confirmed experimentally using the homologous series $C_nH_{2n+1}OH$ ($n=13-31$). The freezing point of ice is sensitive to the length and parity of the chain. The curve for the C_n -odd series increases asymptotically with chain length, reaching a maximum of -1°C for $n=31$. The C_n -even series behaves differently; the freezing point curve reaches a plateau of about -7.5°C for n in the upper range 22–30 (Figure 22).

The requirement for an extended lattice match between the nucleator and nucleant was made evident by a substantial reduction in freezing point when miscible amphiphiles $C_nH_{2n+1}OH$ and $C_mH_{2m+1}OH$ ($n=30$, $m=31$) were used. These mixed systems reduce the freezing point presumably because of OH site defects at the monolayer–water interface. On the other hand, immiscible

(a)



(b)

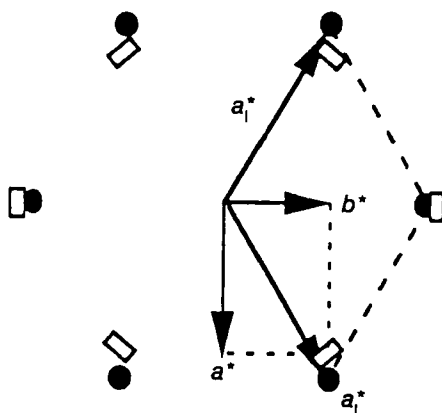


Figure 23 (a) Electron diffraction pattern from an essentially single two-dimensional crystallite of a $C_{31}H_{63}OH$ monolayer on a single crystal of hexagonal ice. (b) Schematic relation between the reciprocal a^*b^* lattices of ice (filled circles) and the monolayer (open rectangles)

alcohols such as hydrocarbons and fluorocarbons do not affect the ice-nucleating temperature, since the 2D nets of the alcohols are not perturbed. The epitaxial growth of ice under a $C_{31}H_{63}OH$ monolayer was demonstrated by an electron diffraction measurement [62] (Figure 23).

Table 2 Dimensions of the a -axis and b -axis of the rectangular unit cell, areas per molecule ($= ab/2$), projected areas per molecule and molecular tilt^a angles obtained from the grazing incidence X-ray diffraction measurements on a series of aliphatic alcohol ($C_nH_{2n+1}OH$, $n = 13-31$) uncompressed monolayers spread on a pure water subphase (70% coverage) at 5 °C

Alcohol	a (Å)	b (Å)	Surface area per molecule (Å ²)	Projected area per molecule (Å ²)	Tilt t_b (°)	Tilt t_a (°)
$C_{31}H_{63}OH$	4.98	7.45	18.6	18.4	7.0	1.5
$C_{30}H_{61}OH$	4.99	7.49	18.7	18.5	7.7	1.7
$C_{23}H_{47}OH$	5.00	7.56	18.9	18.6	9.5	1.2
$C_{20}H_{41}OH$	5.05	8.00	20.2	19.1	19.0	2.4
$C_{18}H_{37}OH$	4.99	8.19	20.4	19.2	19.4	2.9
$C_{16}H_{33}OH$	4.86	8.41	20.4	19.6	15.8	2.8
$C_{14}H_{29}OH$	4.84	8.47	20.5	19.8	14.8	3.0
$C_{13}H_{27}OH$	4.86	8.55	20.8	20.3	12.0	3.0

^a The parameters t_b and t_a refer to the tilt of the chain in the b -direction and a -direction, respectively. Total tilt $\approx (t_b^2 + t_a^2)^{1/2}$.

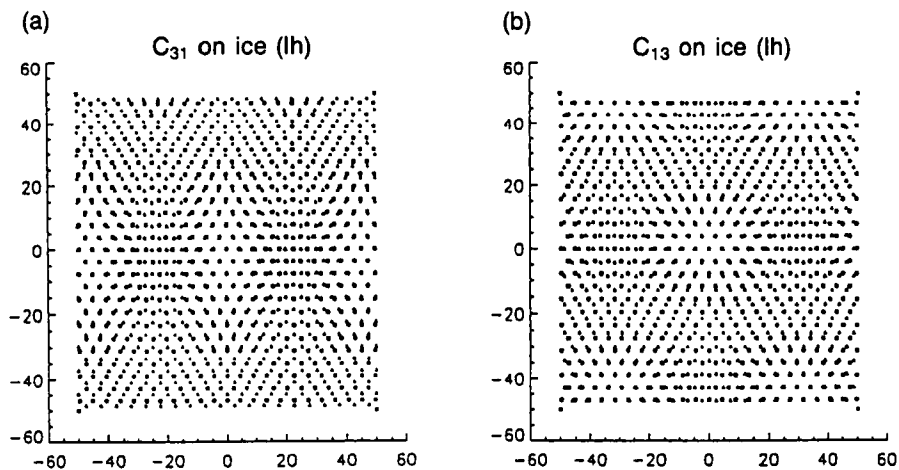


Figure 24 Diagram of the superimposed oxygen positions within the *ab* layer of hexagonal ice (triangles) and the oxygen positions in monolayers of (a) $C_{31}H_{63}OH$ and (b) $C_{13}H_{27}OH$ (squares)

Systematic GID studies of the monolayer series $C_nH_{2n+1}OH$ ($n = 13-31$) on water at 5°C were performed [63–65]. The unit cell data are listed in Table 2. Superposition of an *ab* lattice of an alcohol monolayer onto the *ab* lattice of hexagonal ice results in a moiré fringe pattern. A comparison of these patterns for the extreme cases of $C_{13}H_{27}OH$ and $C_{31}H_{63}OH$ (Figure 24) demonstrates that the number of well-overlapped oxygens is fewer for $n = 13$ than for $n = 31$, in keeping with their ice-nucleating behavior. The lattice fit between ice and the different monolayers in terms of the freezing curve (Figure 22) may be expressed somewhat differently, for example by a comparison between the area per H_2O molecule in the *ab* plane of ice ($= 17.5 \text{ \AA}^2$) and the area per molecule of the amphiphilic alcohol (Table 2); the fit improves with n increasing from 13 to 23, at which point the surface area per amphiphile levels off asymptotically to a value of $18.5 \text{ \AA}^2$ [64].

Monitoring by GID the structure of a $C_{31}H_{63}OH$ monolayer on water (Figure 25) as the latter was cooled to freezing just below 0°C indicated not only that the monolayer maintained its structural integrity, but that the average domain size of many of the ice crystals in their *ab* plane was as low as $25 \text{ \AA}$ [63]. These data provide some information on the maximum threshold size of ice nucleated by the alcohol monolayer just below 0°C . For example, if we assume that the ‘critical’ nucleus is a hemisphere in shape and of maximum diameter $25 \text{ \AA}$, then the number of H_2O molecules in such a domain is in the range 70–100.

The analysis of the grazing incidence diffraction data, however, does not clearly explain the differences in freezing points induced by the C_n -odd and

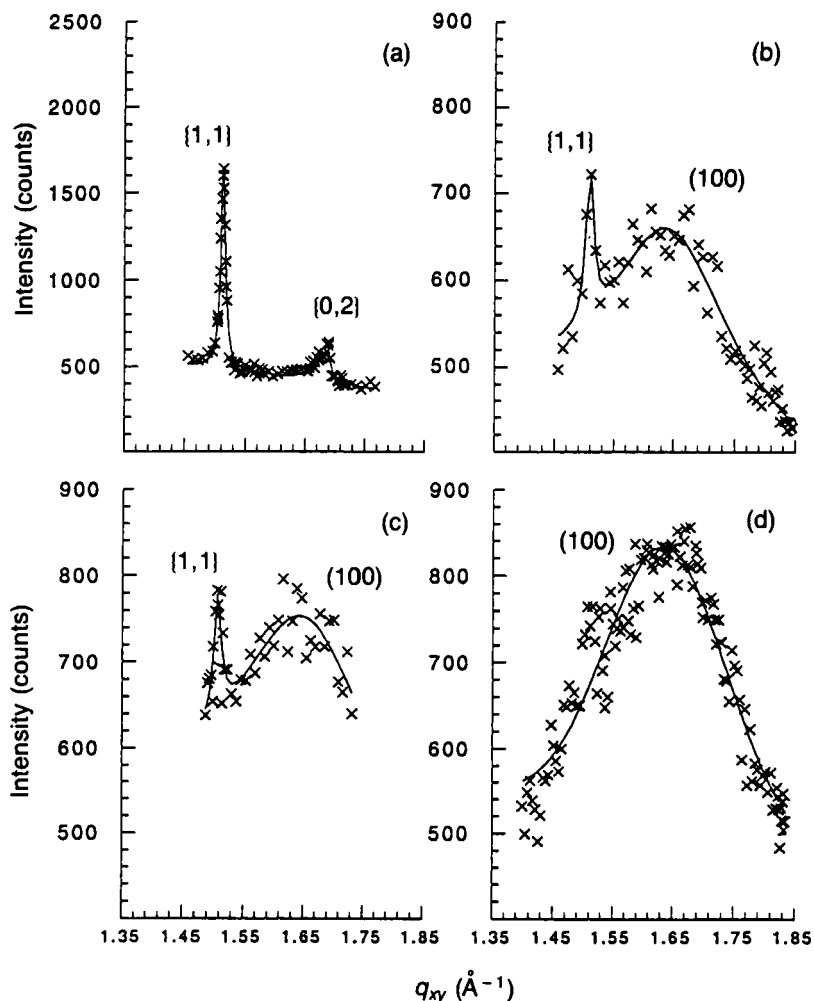
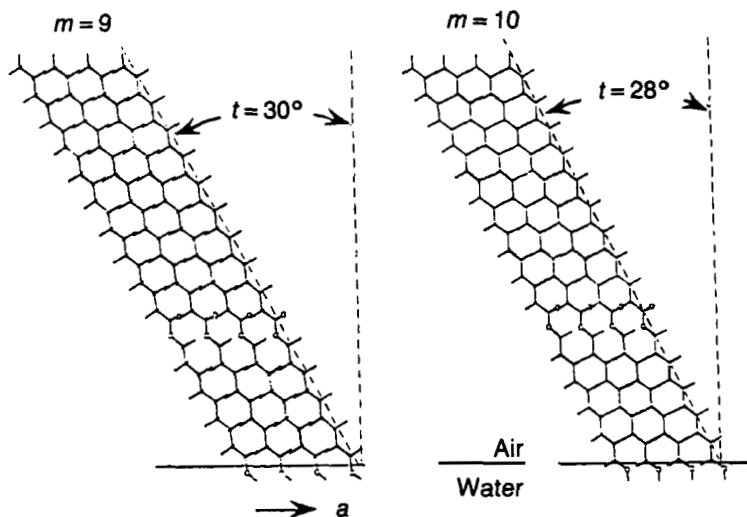
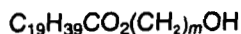


Figure 25 Grazing incidence X-ray diffraction measurements made on a monolayer of $\text{C}_{31}\text{H}_{63}\text{OH}$ over pure water cooled to freezing. (a) The two Bragg peaks $\{1,1\}$ and $\{0,2\}$ of the monolayer on water. The temperature of the water subphase was about 4°C . (b) First stage of ice crystal nucleation under the monolayer. The $\{1,1\}$ Bragg peak of the monolayer and the (100) Bragg peak of ice are visible. The temperature of the subphase was just below 0°C . (c) Same as part (b) but after a time interval of approximately 15 min. (d) After a further 15 min only the (100) Bragg peak of ice is visible

C_n -even series. For example, the GID patterns of the $\text{C}_{30}\text{H}_{61}\text{OH}$ and $\text{C}_{31}\text{H}_{63}\text{OH}$ monolayers are almost the same, and thus are their overall packing characteristics. Since the molecules are tilted by an angle of about 10° from the vertical, one could deduce from the ice-nucleating results that the hydrocarbon chain structures for $n=30$ and 31 must be very similar, and so the orientations of

(a)



(b)

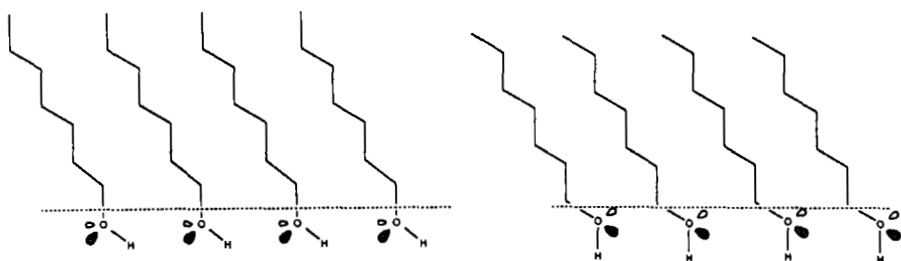


Figure 26 (a) Two-dimensional crystal structures of monolayers of the hydroxyalkyl esters $\text{C}_{19}\text{H}_{39}\text{CO}_2(\text{CH}_2)_m\text{OH}$ ($m=9$ and 10) on a water subphase, viewed along the b -axis. (b) Blowup of the hydroxy head-group orientations at the monolayer–water interface

their OH groups with respect to the water surface must be different. A clear indication of the role played by the orientation of OH groups was provided by the freezing point measurements and structure determinations for the series $\text{C}_n\text{H}_{2n+1}\text{CO}_2\text{C}_m\text{H}_{2m}\text{OH}$ ($n=19$, $m=9$, 10). The monolayer with $m=10$ is a better ice nucleator than that with $m=9$. The freezing point measurements were independent of the odd/even parity of n , in agreement with the model [60]. The structures of these monolayers are shown in Figure 26. The chain arrangements of the two structures are essentially the same; the difference lies

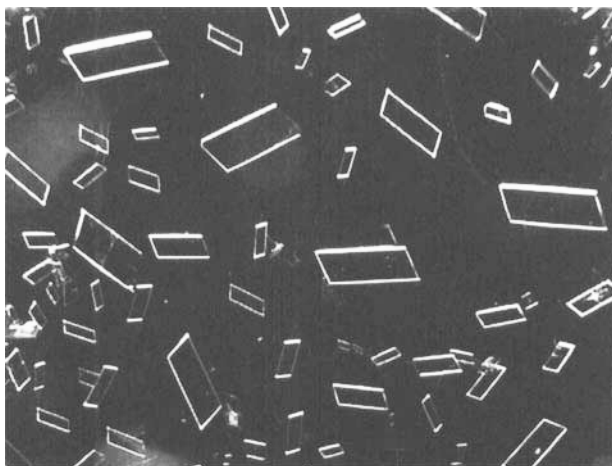
in their C–OH orientations. For $m=9$ the O–H bond and all the oxygen lone pair electrons are equally exposed to water; for $m=10$ the O–H bond points vertically into the water subphase.

These studies demonstrate that a stringent structural fit is required between the two-dimensional net of the alcohol OH groups and the two-dimensional surface of the hexagonal ice for efficient induced freezing to take place.

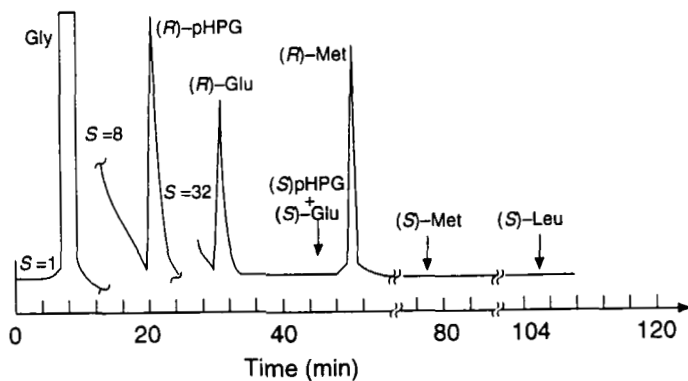
3.4 Enantioselective Nucleation of α -Glycine by Hydrophobic α -Amino Acids [47]

At this stage, we introduce the element of chirality in the process of induced nucleation at the interface. As described previously, when glycine crystals are grown from an aqueous solution in the presence of a racemic α -amino acid, they display a $\{010\}$ plate morphology. When the racemic α -amino acid additives are hydrophobic, the plates have a strong tendency to float at the air–solution interface. On simple symmetry grounds the probability is equal for a given crystal to float on its (010) or $(0\bar{1}0)$ face. Growth of α -glycine in the presence of small amounts of resolved amino acids bearing hydrophobic side chains, such as (*S*)-leucine, (*S*)-phenylalanine or (*S*)- α -aminooctanoic acid, yields floating crystals of pyramidal shape, always exposing their $(0\bar{1}0)$ basal face to the air. When crystals were grown in the added presence of equal amounts of racemic α -amino acids such as *p*-hydroxyphenylalanine, glutamic acid and methionine, all the resulting plates had their $(0\bar{1}0)$ faces exposed to air (Figure 27a), and so the crystals occluded only the (*R*)- α -amino acid additives according to the HPLC analysis (Figure 27b). The orientation of the crystal is also immediately apparent from the shapes of the $(0\bar{1}0)$ faces, which display two-dimensional chirality. These results indicate that the hydrophobic α -amino acids have a tendency to self-aggregate at the solution surface and induce $(0\bar{1}0)$ -oriented nucleation of α -glycine. The role played by the hydrophilic (*R*)- α -amino acids, subsequent to the induced orientation of glycine, was to inhibit growth of the floating crystals in the $+b$ -direction into the subphase, inducing formation of the (010) face.

The packing arrangements of a variety of racemic, hydrophobic α -amino acids show that these molecules, interlinked by hydrogen bonds, form centrosymmetric bilayers akin to those of α -glycine. Within each layer the molecules are related by translation symmetry and so are homochiral. In the event that the hydrophobic α -amino acids, as additives, form ordered, hydrogen-bonded, two-dimensional chiral clusters at the air–solution interface akin to the layer arrangements observed in their own crystals, such aggregates can form pseudocentrosymmetric hydrogen-bonded bilayers with glycine solute molecules. Following this line of thought, a bilayer at the air–solution interface composed of hydrophobic (*R*)- α -amino acids bound to glycine solute molecules would resemble the top (010) bilayer of an α -glycine face (Figure 7), and so enantioselectively trigger its nucleation. On this basis we explain why hydrophobic



(a)



(b)

Figure 27 (a) Floating α -glycine crystals grown in the presence of 1% (w/w of glycine) (S)-leucine and racemic *p*-hydroxyphenylglycine (0.5%), racemic glutamic acid (1%) and racemic methionine (0.5%). All the crystals float with their (010) faces exposed to air and exhibit the same enantiomorphous morphology. (b) HPLC enantiomeric analysis of *p*-hydroxyphenylglycine (pHPG), glutamic acid, methionine and leucine occluded inside the crystals of α -glycine

Table 3 Crystallization of α -glycine underneath floating Langmuir monolayers

Monolayer	Limiting area per molecule ($\text{\AA}^2$)	Degree of orientation (%)	Face exposed to monolayer	Crystallization rate ^a (h)
(S)-5- α -Cholestan-3 β -OCOCH ₂ CH(NH ₃ ⁺)CO ₂ ⁻	38	0		Slow
(R)-(X) ₂ NCOCH ₂ CH(CO ₂ H)SCH ₂ CH(NH ₃ ⁺)CO ₂ ⁻ [X = ME(CH ₂) ₁₄ CO ₂ CH ₂ CH ₂ -]	77			
(S)-CF ₃ (CF ₂) ₉ CH ₂ OCOC(CH ₂) ₃ CH(NH ₃ ⁺)CO ₂ ⁻	30	50-79	(010)	Fast (s)
(S)-Me(CH ₂) ₁₄ CONH(CH ₂) ₃ CH(NH ₃ ⁺)CO ₂ ⁻	27	88-93	(010)	Fast (s)
(R)-Me(CH ₂) ₁₇ OCO(CH ₂) ₂ CH(NH ₃ ⁺)CO ₂ ⁻	29	73-97	(010)	Fast (s)
(S)-Me(CH ₂) ₁₇ OCO(CH ₂) ₂ CH(NH ₃ ⁺)CO ₂ ⁻	29	73-93	(010)	Fast (s)
(R)-Me(CH ₂) ₁₇ OCO(CH ₂) ₂ CH(NH ₃ ⁺)CO ₂ ⁻	25	85-96	(010)	Fast (s)
(S)-Me(CH ₂) ₁₅ CH(NH ₃ ⁺)CO ₂ ⁻	25	78-94	(010)	Fast (s)
(R,S)-Me(CH ₂) ₁₅ CH(NH ₃ ⁺)CO ₂ ⁻	25	50	(010), (010)	Fast (s)
(S)-Me(CH ₂) ₁₇ NHCOCH ₂ CH(NH ₃ ⁺)CO ₂ ⁻	27	39-77	(010)	Fast (s)
(S)-Me(CH ₂) ₁₇ NHCO(CH ₂) ₂ CH(NH ₃ ⁺)CO ₂ ⁻	27	51-96	(010)	Fast (s)
(R)-Me(CH ₂) ₁₄ CONH(CH ₂) ₄ CH(NH ₃ ⁺)CO ₂ ⁻	27	>99	(010)	Fast (s)
(S)-Me(CH ₂) ₁₄ CONH(CH ₂) ₄ CH(NH ₃ ⁺)CO ₂ ⁻	27	>99	(010)	Fast (s)
(R)-Me(CH ₂) ₁₄ CONH(CH ₂) ₂ CH(NH ₃ ⁺)CO ₂ ⁻	27	>99	(010)	Fast (s)
(S)-Me(CH ₂) ₁₄ CONH(CH ₂) ₂ CH(NH ₃ ⁺)CO ₂ ⁻	27	>99	(010)	Fast (s)
(R)-Me(CH ₂) ₁₇ OCOCH ₂ CH(NH ₃ ⁺)CO ₂ ⁻	29	>99	(010)	Fast (s)
(S)-Me(CH ₂) ₁₇ OCOCH ₂ CH(NH ₃ ⁺)CO ₂ ⁻	29	>99	(010)	Fast (s)
(R)-Me(CH ₂) ₁₁ OCOCH ₂ CH(NH ₃ ⁺)CO ₂ ⁻	29	>99	(010)	Fast (s)

^aIn hours (h) or seconds (s).

(*R*)- α -amino acids and (*S*)- α -amino acids induce nucleation of glycine from the (010) and (0 $\bar{1}$ 0) faces, respectively. Further support for the proposed model was provided by a series of crystallization experiments on glycine grown in the presence of *t*-butylglycine and neopentylglycine. Owing to the bulky nature of their hydrocarbon chains, these α -amino acids cannot constitute a hydrogen-bonded layer as in α -glycine, and indeed crystallize in the form of hydrates which act as hydrogen-bonding filler molecules. Consequently, the additive cannot induce oriented nucleation of α -glycine [66].

More direct information on the structure of the nucleating domains was obtained from the use of water-insoluble α -amino acids bearing hydrophobic chains of different cross-sectional area [67–69]. The results on the oriented

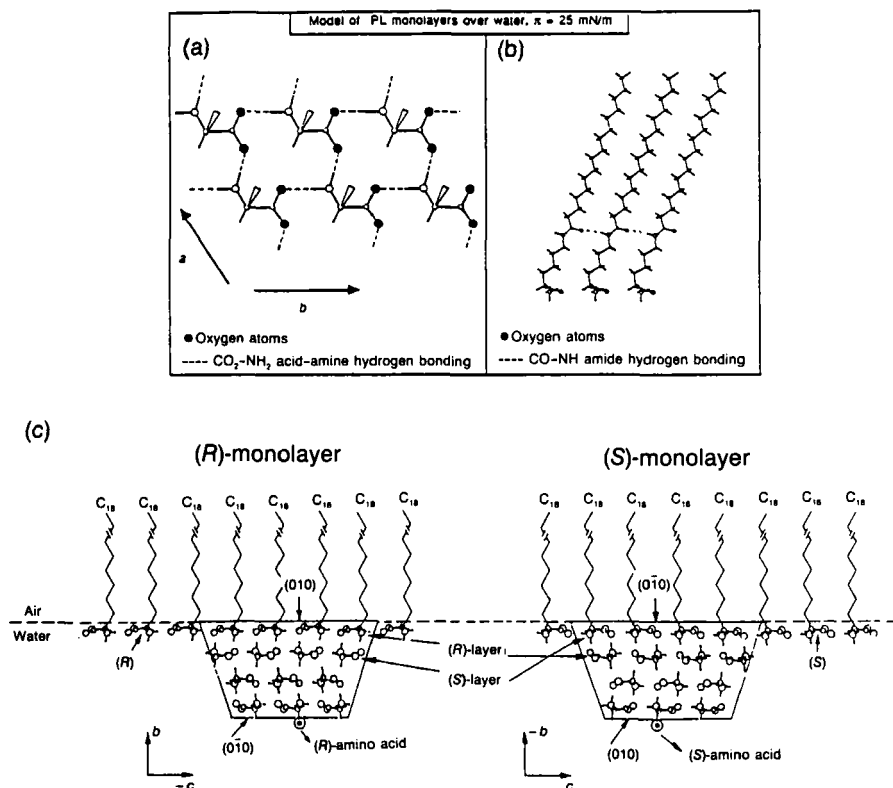


Figure 28 Model of the two-dimensional packing arrangement of a compressed (25 mN m^{-1}) (*R*)-palmitoyllysine monolayer over pure water. (a) View perpendicular to the monolayer plane; the hydrocarbon chains are represented by wedges. The $\text{NH}\cdots\text{O}$ (carboxylate) hydrogen bonds are shown by dashed lines. (b) View along the *a-b* axis showing the molecules in a row parallel to the *a+b* direction, interlinked by $\text{N-H}\cdots\text{O}=\text{C}$ (amide) hydrogen bonds. (c) Schematic representation of pyramidal crystals of glycine attached with the (010) crystal face to a compressed (*R*)- α -amino acid monolayer and with the (0 $\bar{1}$ 0) face to an (*S*)- α -amino acid monolayer, respectively

nucleation of α -glycine, shown in Table 3, are similar to those obtained using the water-soluble counterparts.

The monolayer structures on water of several of these water-insoluble amphiphiles were determined *in situ* by GID using synchrotron radiation. The packing arrangement of *N*^ε-palmitoyl-(*R*)-lysine [70] obtained from such an analysis is shown in Figure 28. The layer motif of the (*R*)- α -amino acid head-groups of the amphiphile is very similar to that of the α -glycine crystal at the (010) face. Schematic views of the (*R*)- α -amino acid monolayer bound to α -glycine at its (010) face and of the (*S*)-monolayer bound to the (0 $\bar{1}$ 0) glycine face are shown in Figure 28(c). Similar GID studies on α -amino acids bearing fluorocarbon chains demonstrate that the cross-sectional area of the fluorocarbon chain ($\sim 28 \text{ \AA}^2$) is somewhat larger than the area per molecule of α -glycine in its layer structure ($\sim 25 \text{ \AA}^2$), an observation in keeping with the poor orientating effect by such monolayers on induced floating glycine crystals [71, 72].

In this section we have shown that amphiphilic molecules form 2D crystalline self-aggregates at the water surface, by virtue of which they may induce oriented nucleation of solute molecules in the subphase through epitaxy. In the case of α -amino acid amphiphiles the oriented nucleation of α -glycine may be achieved enantiospecifically.

4. MOLECULAR RECOGNITION DURING CRYSTAL NUCLEATION

We now address the process of control on crystal nucleation, with the ultimate aim of enantioselective inhibition of crystal nucleation. Any crystallization process must cross a nucleation barrier. The classical theory of crystal nucleation describes the process of aggregation of molecules to form clusters in terms of a total free energy which has two components: the bulk free energy, which is negative, and the positive free energy associated with the interfacial energy between the cluster and its environment. At a certain size of the cluster, the two terms become equal and the total free energy is zero. Above this critical size, the total free energy being negative, the clusters should grow spontaneously into mature crystals. Below the critical size the clusters are not stable and may disintegrate. This treatment does not, however, take into account the structural and stereochemical aspects of the nucleus [73].

As discussed earlier, we have generally assumed that as the molecules associate in supersaturated solutions, they form embryos with structures resembling those of the yet to be grown crystals. In systems which display polymorphism, embryos of all phases are expected to be formed, each embryo resembling structurally the crystal into which it will eventually develop. Additive molecules may therefore be designed which can be targeted enantioselectively and stereospecifically at nuclei of undesired crystalline phases and induce the precipitation

of metastable crystals by a kinetically controlled process. Several examples will be described where this naïve hypothesis was instrumental for the precipitation of the desired phase.

In all such studies the solvent clearly plays an important role in determining the size and, for polymorphic forms, the structures of the nuclei. Thus we shall first discuss the role played by the solvent in crystal growth from a stereochemical viewpoint.

4.1 Lock-and-Key Processes in Solvent Effects on the Crystal Growth of Solvates

Since the solvent is an inevitable 'additive' present in the solution of a substrate undergoing crystallization, the role played by the solvent in crystal growth must be taken into consideration. The solvent, like a 'tailor-made' additive, interacts differently with the nonsymmetry-related faces of a growing crystal, and likewise with the facets of crystal nuclei.

This analogy is demonstrated here by several examples. To facilitate the transition from additive-surface to solvent-surface interactions, crystalline hydrates were chosen for analysis so that added solvents, such as methanol and higher alcohols, may be considered as 'tailored' solvent inhibitors [74–77]. Asparagine precipitates from water in the form of a monohydrate, exhibiting as many as 18 faces (Figure 29a). In the pure crystal structure (Figure 30) the hydrate molecules are oriented so that an O–H bond of each of the four symmetry-related molecules emerges from a {010} face, while the other O–H bond points towards the crystal interior. These water molecules at the two opposite {010} crystal surfaces can be replaced by alcohol molecules with the

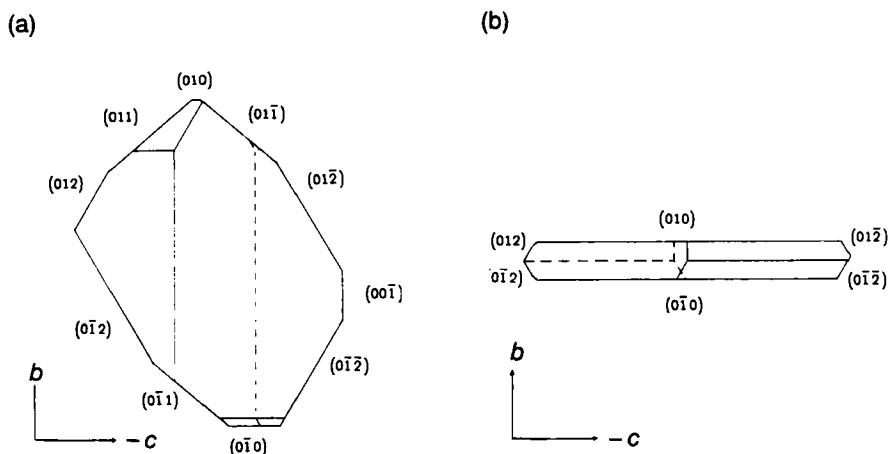


Figure 29 Crystal morphologies of (*S*)-asparagine monohydrate grown from aqueous solution (a) pure and (b) in the presence of (*S*)-aspartic acid or methanol

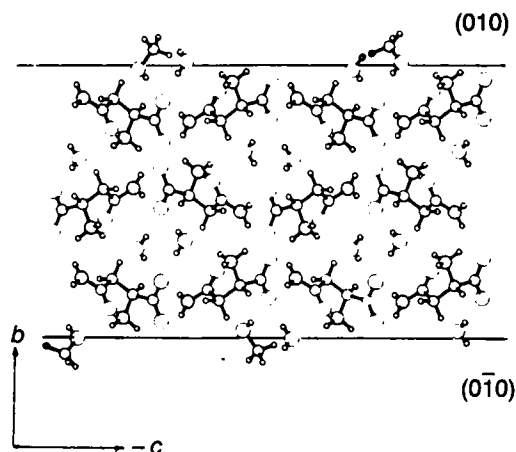


Figure 30 Packing arrangement of (*S*)-asparagine monohydrate viewed along the *a*-axis. The four adsorbed methanol molecules are on the two {010} faces

hydrocarbon moiety emerging from the crystal, and so impede growth of these faces. On the {011} faces only two out of the four water molecules can be replaced, suggesting that the morphological change would be less dramatic. Stereoselective adsorption was demonstrated both by {010} plate formation (Figure 29b) and by stereospecific etching of the {010} faces on dissolution of the asparagine crystals in aqueous solutions of asparagine containing alcohol additives.

A second example is α -rhamnose monohydrate, which crystallizes into a polar arrangement and exhibits hemihedral faces. Such a class of crystals is most appropriate since one may establish the effect of the solvent at the opposite ends of the polar axis. The packing arrangement is shown in Figure 31. The hydrate molecules are oriented with their O–H bonds pointing more towards the $+b$ -direction than to the $-b$ -direction of the crystal. Thus alcohol additives should induce changes from the regular crystal morphology more at the $+b$ -side of the crystal than at the $-b$ -side. Upon addition of methanol to the solution, a decrease in the rate of crystal growth in the $+b$ -direction as against the $-b$ -direction *vis-à-vis* that of the crystal grown in aqueous solution was observed, in agreement with the model (Figure 32). This simple experiment also permits an unambiguous assignment of the orientation of the rhamnose molecule in the crystal, and hence the sense of chirality of rhamnose, independent of the Bijvoet method of anomalous X-ray scattering.

4.2 A Relay Mechanism for the Crystal Growth of Racemic Alanine and γ -Glycine [75]

We now examine the role of solvent in the growth properties of polar crystals, focusing on the γ -form of glycine and (*R,S*)-alanine. These two crystals have

remarkably similar packing and morphological features (Figure 33), with an $(00\bar{1})$ face perpendicular to the polar c -axis at one end of the crystal and capped faces at the opposite end. The absolute orientation of the crystal structure *vis-à-vis* the polar axis was established for both γ -glycine and (R,S) -alanine by making use of tailor-made additives, according to which the CO_2^- groups are

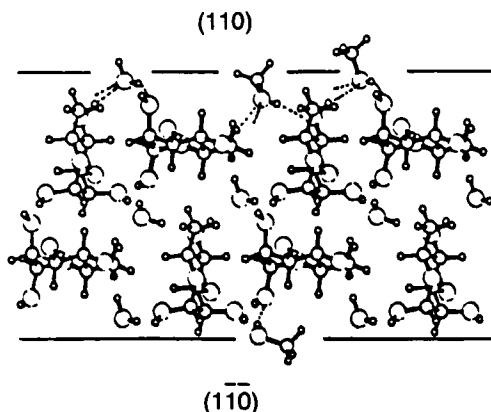


Figure 31 Packing arrangement of α -rhamnose monohydrate viewed down the a -axis. The two different hydrate water molecules oriented with their O-H bonds emerging from the (110) face are shown replaced by methanol. On this surface both water molecules, which are each tightly bound via three hydrogen bonds, may be replaced by methanol. On the $(\bar{1}\bar{1}0)$ face methanol can replace only one of the two water molecules and is loosely bound by only one hydrogen bond

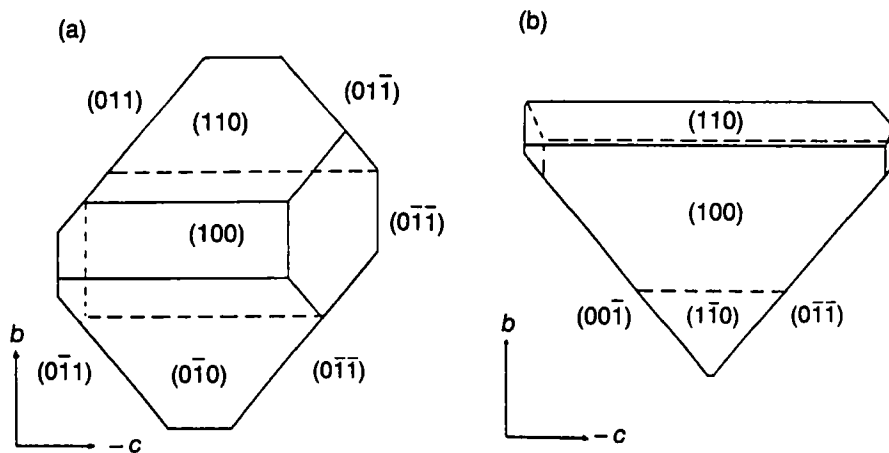


Figure 32 (a) Crystal morphology of α -rhamnose monohydrate grown from aqueous solution. (b) Crystal morphology obtained from a 9:1 mixture of methanol and water

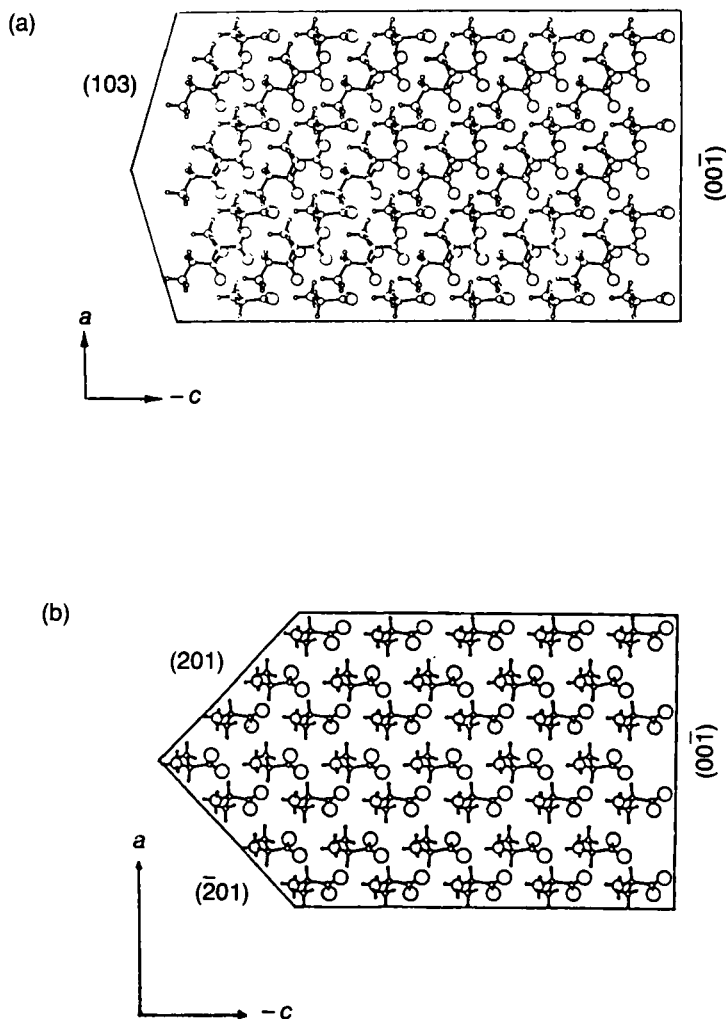


Figure 33 Packing arrangements of (a) γ -glycine and (b) (R,S) -alanine crystals, delineated by the developed faces. The capped faces which cut the $+c$ -end of the polar axis expose NH_3^+ groups and the flat $(00\bar{1})$ faces expose CO_2^- groups

exposed at the 'flat $-c$ -end' while the NH_3^+ groups are exposed at the 'capped $+c$ -end' [74]. Crystal growth and dissolution experiments also indicate that in aqueous solution the carboxylate $-c$ -ends of crystals of γ -glycine and (R,S) -alanine grow and dissolve faster than the amino $+c$ -ends.

The question remains as to which polar end of the crystal water may bind more tightly. We must also correlate the macroscopic growth phenomena with the solvent binding. The packing arrangements of both crystal structures reveal

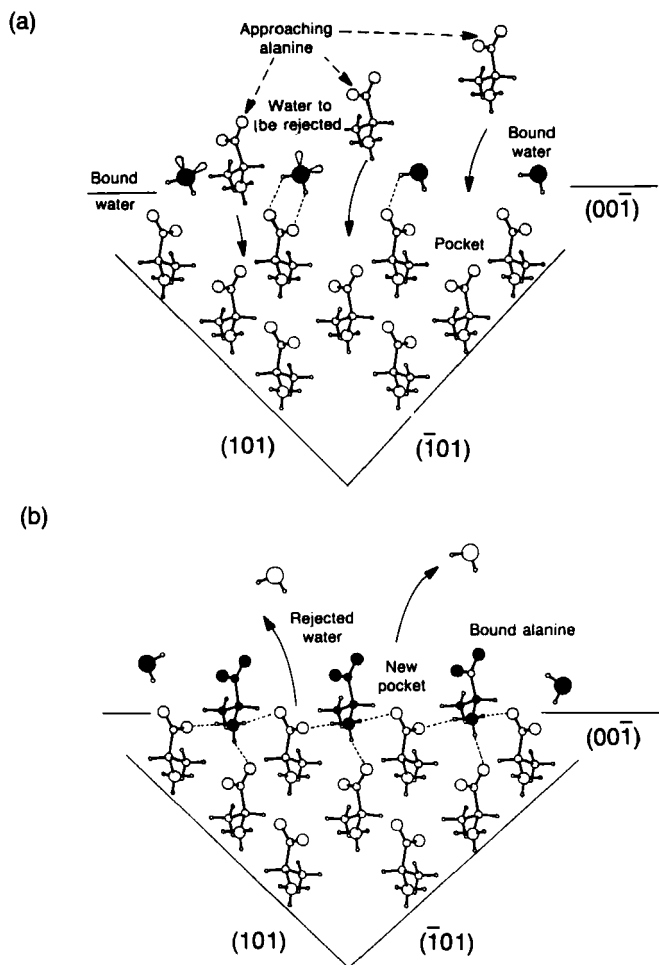


Figure 34 The (001) face of (*R,S*)-alanine during the crystal growth process. (a) Approaching solute alanine molecules are about to be bound within the pockets of this face. Also shown are water molecules bound to the outermost layer of this face. The pockets remain primarily unsolvated because lone pair–lone pair oxygen–oxygen repulsion inhibits the binding of water within them. (b) Alanine molecules are bound via three N–H···O hydrogen bonds. The previously bound water molecules are shown being rejected by O (water)···O (carboxylate) lone pair–lone pair repulsions

that their (001) carboxylate faces comprise regular pockets on a molecular level and can be regarded as corrugated in two dimensions. The size and shape of these pockets are suited to fit the NH_3^+ groups of the α -amino acid solute molecules. The NH_3^+ groups form hydrogen bonds with the oxygens of the CO_2^- groups inside the pockets (Figure 34). Replacement of an NH_3^+ group by an oriented water molecule will result in a weakly attractive or even a repulsive interaction

in the pocket since at least one N-H bond must be replaced by a lone pair electron lobe of the water oxygen atom, so leading to oxygen (lone pair)-oxygen (lone pair) repulsions. Presumably then, the water molecules will be bound to the carboxylate oxygens from above the pocket, leaving the pocket either unhydrated or populated with weakly bound water, and so easily accessible to approaching solute molecules. As glycine or alanine molecules dock into adjoining pockets, the CO_2^- groups of the newly added substrate molecules are within 3 Å of the water molecules bound on the outermost surface, leading to their expulsion and thereby generating a new unsolvated pocket on the crystal surface. This relay process of solvent water binding and expulsion helps growth and dissolution by both desolvating the surface and perpetuating the natural corrugation of the surface on a molecular level. On the other hand, the {001} and {201} faces of (*R,S*)-alanine at the opposite +*c*-end of the crystal, which expose NH_3^+ and Me groups, are relatively smooth and comprise molecules which are equally accessible for water binding. The corresponding three {10 $\bar{1}$ 3} faces of γ -glycine have a similar surface structure exposing NH_3^+ and CH_2 groups, and their molecular surfaces should be equally accessible to water. Experimental support in favor of this relay mechanism of growth was provided by adding to the water solution methanol, which can dock within the pockets through weak $\text{Me}\cdots\text{O}$ interactions and was found to impede growth of the (00 $\bar{1}$) faces.

4.3 Crystal Polymorphism: Polar versus Nonpolar Crystals

Polymorphic systems are appropriate for studying the effect of tailor-made inhibitors on crystal nucleation. Since we have already discussed the effect of solvent on the crystal growth of γ -glycine, we shall focus here on the polymorphic behavior of α -glycine and γ -glycine [78]. The stable α -polymorph precipitates from aqueous solution and is centrosymmetric, whereas the less stable γ -polymorph, which is obtained from acid or base solution, is polar and grows unidirectionally in aqueous solution along the polar axis at the (00 $\bar{1}$) face [74]. Let us assume that an aqueous solution of supersaturated glycine contains embryonic clusters with structures and 'morphologies' akin to the α -form and γ -form. In order to induce precipitation only of the γ -form it is necessary that the additive should bind at least to the four fast-growing {011} faces of α -glycine (Figure 35) and so inhibit its nucleation and, at the same time, bind to the γ -form primarily at its nongrowing capped end (Figure 33b) and so leave the crystal essentially unaffected.

We have already pointed out that the natural α -amino acid additives in racemic form bind specifically to the two {010} faces of α -glycine and retard their growth, resulting in the formation of plates instead of the usual bipyramids. But these additives do not prevent crystal formation, and so a search was made for additives which will inhibit growth of the four {011} faces of α -glycine in order to prevent its appearance. Hexafluorovaline was chosen for this purpose.

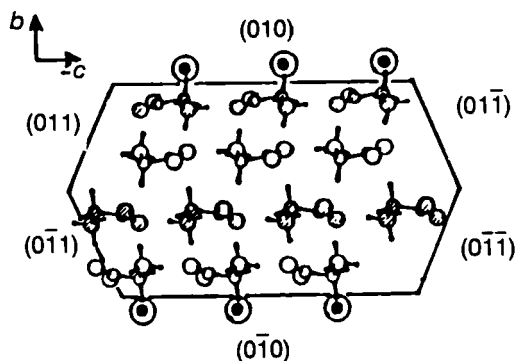


Figure 35 Packing arrangement of α -glycine viewed along the a -axis and delineated by its crystal faces

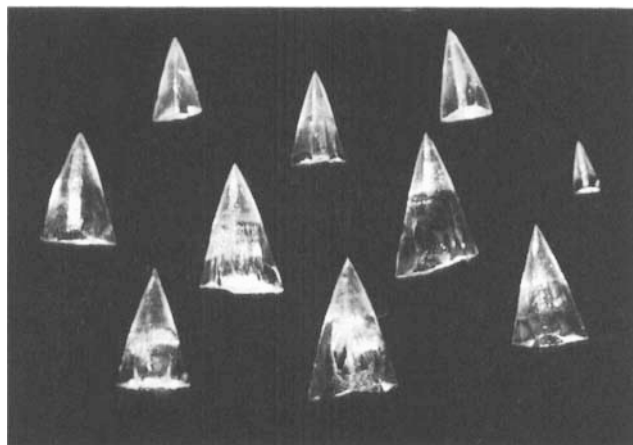


Figure 36 Photograph of γ -glycine crystals grown from an aqueous solution containing racemic hexafluorovaline (3% w/w of glycine)

It cannot be adsorbed at the $\{010\}$ faces because of steric repulsion imposed by the hexafluoroisopropyl group, but was found to attach strongly to the four $\{011\}$ glycine faces. Regarding the effect of this α -amino acid on γ -glycine, it can bind to the NH_3^+ groups at the nongrowing capped faces but not to the fast-growing $(00\bar{1})$ face where the carboxylate groups are exposed. Addition of 1% of hexafluorovaline causes precipitation of α -glycine where the $\{011\}$ faces are strongly affected, but 3% induces the expected precipitation of the γ -form (Figure 36). Other materials such as chloranil, acetic acid and ammonia, which also affect these faces, induce the precipitation of the γ -form.

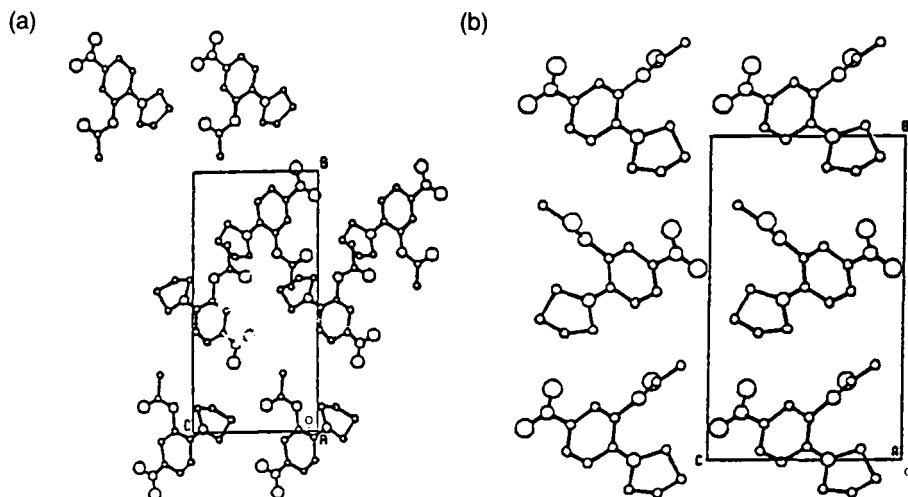


Figure 37 (a) Packing arrangement of the stable nonpolar α -form of PAN. (b) Packing arrangement of the metastable polar β -form of PAN

Another successful example involves the polar and nonpolar structures of *N*-(2-acetamido-4-nitrophenyl)pyrrolidine (PAN) [79]. The stable polymorph is nonpolar, whereas the metastable form is polar and displays optical second-harmonic generation (SHG). The crystal structures are depicted in Figure 37. Addition of a polymeric inhibitor which is the PAN derivative of poly(*N*-acryloyl-12-dodecanoic acid) blocks the growth of both sides of the nonpolar crystal but only one side of the metastable form. Indeed, addition of as little as 0.03% of the inhibitor induced crystallization only of the metastable polymorph of PAN.

4.4 'Tailor-Made' Inhibitors for the Resolution of Enantiomers [80, 81]

Another ramification of the selective inhibition of crystal nucleation is the kinetic resolution of enantiomers by crystallization. In a supersaturated solution of a racemate undergoing spontaneous resolution, we may envisage that the solution contains equal amounts of clusters of structures akin to those of the corresponding mature chiral phases. On this basis, soluble 'tailor-made' chiral polymers can be designed which may enantioselectively bind to, and consequently destroy, the corresponding homochiral clusters and so inhibit their growth into macroscopic crystals (Scheme 4). The enantiomeric crystalline phase will precipitate unperturbed. 'Tailor-made' additives have been successfully used for the resolution of a large class of conglomerates including threonine, glutamic acid hydrochloride, asparagine monohydrate, *p*-hydroxyphenylglycine, *p*-toluenesulfonate and 3-phenylhydracrylic acid.

Table 4 Typical resolution experiments on histidine hydrochloride monohydrate (3.2 g in 5 ml H₂O) in the presence of polymeric additives at 50 °C

Polymer	Polymer content (% w/w)	$[\alpha]_D^{25}$ ^a (°)	Enantiomeric excess (%)	Chemical yield (%)	Seed (0.5 mg)
None		-0.9	9.3 (R)	11	(R)-His·HCl·H ₂ O
None		+0.8	8.3 (R)	10.3	(R)-His·HCl·H ₂ O
(S)-PA-Phe ^b	1	-8.8	91.6 (R)	10	No seed
(S)-PA-Phe	1	-9.6	100 (R)	7	No seed
(S)-PA-Phe	1	-9.6	100 (R)	12	(R)-His·HCl·H ₂ O
(R)-PA-Phe	1	+9.6	100 (S)	12	(S)-His·HCl·H ₂ O
(S)-PAO-Tyr ^c	1	-9.6	100 (R)	11	(R)-His·HCl·H ₂ O
(R)-PAO-Tyr	1	+9.6	100 (S)	10	(S)-His·HCl·H ₂ O

^a C = 5 in 5 N HCl.^b PA-Phe = poly(*p*-acrylamidophenylalanine).^c PAO-Tyr = poly(*p*-acryloxytyrosine).

Another example involves the design of auxiliary molecules for the induced resolution of systems where a small energy gap exists between the racemic compound and the resolved enantiomorphs, illustrated here for racemic histidine hydrochloride [82]. This material precipitates at room temperature in the form of a stable racemic hydrochloride dihydrate. At temperatures above 45 °C the material crystallizes in the form of a conglomerate as histidine hydrochloride monohydrate. Chiral polymeric inhibitors were designed, based on an analysis of the packing arrangements of the two crystals, in order to inhibit growth of the stable racemic compound and one of the enantiomorphs. Addition of as little as 1% of the chiral inhibitor poly [*p*-acrylamido-(*S*)-phenylalanine] induced resolution of the racemic mixture with almost quantitative enantiomeric excess. Slight modification of the polymer, such as replacement of the *p*-amino-phenylalanine side chain by lysine, was sufficient to prevent the resolution (Table 4).

4.5 Induced Twinning of Racemic Alanine

A similar attempt to resolve the chiral and racemic forms of alanine whose crystal structures are very similar, where the latter is thermodynamically the more stable, resulted in an unusual twinned structure of the racemic form [83]. (*R,S*)-Alanine crystallizes in the polar space group *Pna*2₁ with the packing arrangement shown in Figure 33(b). The chiral form precipitates in the space group *P*2₁2₁2₁ with a packing arrangement very similar to that of (*R,S*)-alanine.

As discussed previously, crystals of (*R,S*)-alanine grown from pure aqueous solutions are [001] needles elongated in *c* and delineated by four {210} side faces, a (00 $\bar{1}$) 'flat face' and four 'capped faces' at the two opposite ends of the *c*-axis (Figures 33b and 38a). In aqueous solution, the crystals grow unidirectionally at the 'flat (00 $\bar{1}$) face'. Crystallization of racemic alanine in the presence of a

small amount of the optically pure α -amino acid threonine, or phenylalanine, say of (*S*)-configuration, yielded crystals which proved to be (*R,S*)-alanine assuming a propeller-like shape with the long axis of the propeller parallel to *c* (Figure 38b). These crystals exhibit an enantiomorphous morphological symmetry 222, which differs from the *mm*2 point symmetry of the pure crystal. On symmetry grounds we might have expected the affected crystals to exhibit morphological point symmetry 112, being grown in the presence of the chiral α -amino acid additive. One way to account for the 222 morphology is to assume that the two halves along the *c*-axis are twinned about the central *ab* plane of the crystal. This would also be in accordance with the fast crystal growth with equal rates at the two opposite ends of the propeller, which is only possible if both ends expose CO_2^- groups. Thus the two twinned parts of the crystal, which appear to be stitched together across the middle section of the crystal, are surprisingly attached at 'faces' each exposing the positively charged NH_3^+ groups. When (*R,S*)-alanine was grown in the presence of the α -amino acid additive of opposite handedness, in this case (*R*)-threonine or (*R*)-phenylalanine, the twinned crystals assumed the enantiomorphous morphology (Figure 39).

In order to explain this unusual twinning the following mechanism was proposed. Assume that supersaturated solutions of (*R,S*)-alanine contain a

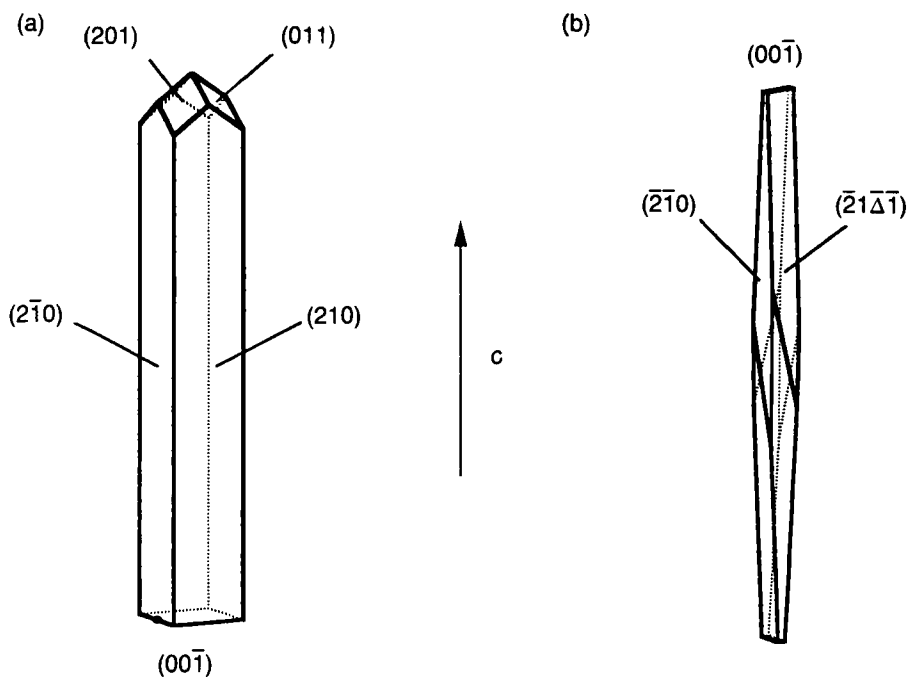


Figure 38 (a) Morphology of (*R,S*)-alanine crystals. (b) Morphology of (*R,S*)-alanine twin crystals grown in the presence of 1% (*S*)-threonine or (*S*)-phenylalanine

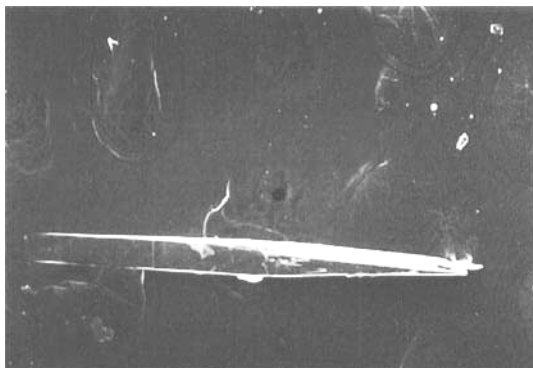
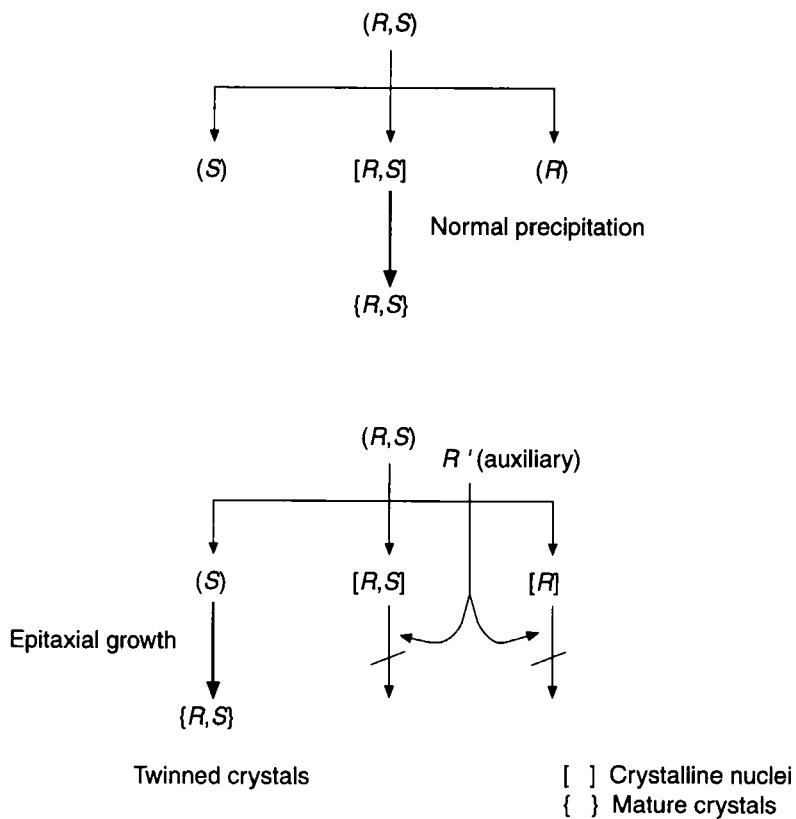


Figure 39 Scanning electron micrograph of a twinned crystal of (R,S) -alanine grown in the presence of (R) -phenylalanine



Scheme 9

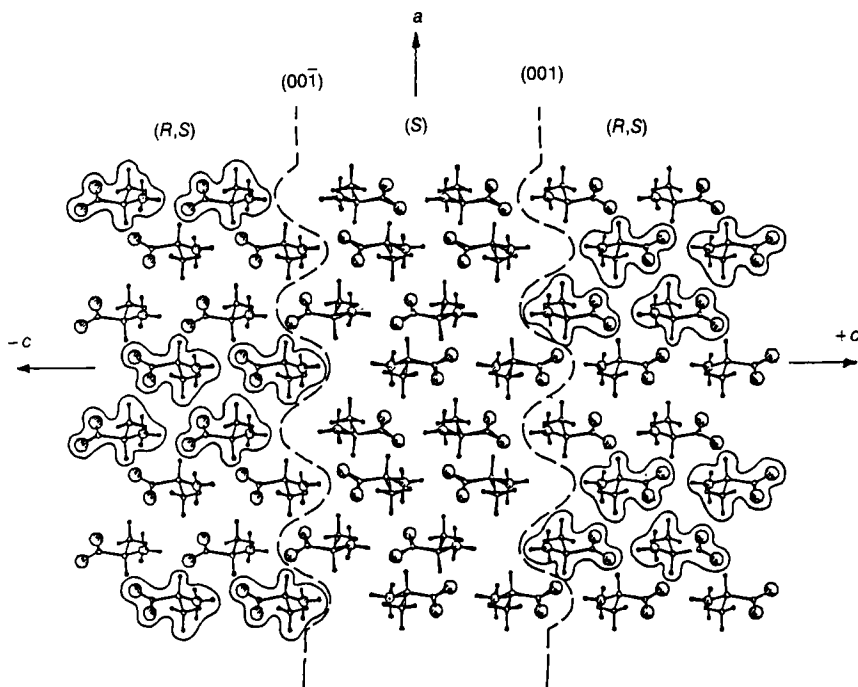


Figure 40 Proposed molecular arrangement of the 'nucleus' composed of a central (S)-'nucleus' onto whose (001) and (00 $\bar{1}$) 'faces' (R,S)-nuclei have been epitaxially grown. Note that the two (R,S)-'nuclei' are related by two-fold screw symmetry about the *a*-axis of (S)-alanine

variety of clusters which resemble crystals of the racemic (R,S)-form and resolved (R)-form and (S)-form of alanine. The (*R'*)- α -amino acid additive present in the solution should stereospecifically inhibit in the initial stages of crystal growth the formation of clusters of the (R,S)-form and (R)-form of alanine. The (*R'*)-additive should bind to a lesser extent, if at all, to the (S)-nuclei, and so should not prevent their growth (Scheme 9). In view of the structural similarity of (R,S)-alanine and (S)-alanine, and given the fact that the latter appears in space group $P2_12_12_1$, we may rationalize the formation of (R,S)-twins by assuming that the (S)-nuclei serve as templates for the epitaxial nucleation and growth of (R,S)-crystals at the opposite ends of (S)-alanine along its nonpolar *c*-axis, as shown in Figure 40. This model was supported by X-ray diffraction studies of a twinned crystal.

4.6 Monitoring the Effect of Auxiliaries on Organic 'Nanocrystals'

In the examples involving crystal nucleation described above, the structures and 'morphologies' of the nuclei were invoked from indirect methods. Recent GID studies of amphiphilic molecules at air-solution interfaces have proved that certain compounds, when spread on liquid surfaces, yield not the usually

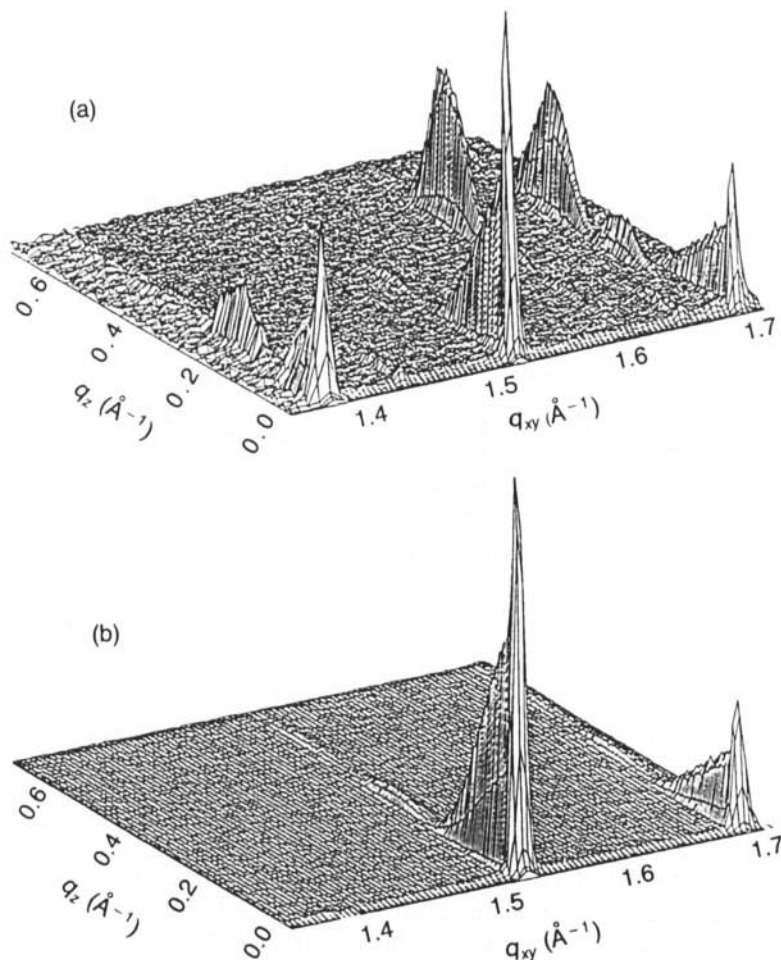


Figure 41 Two-dimensional intensity plots (along q_{xy} and q_z) of the GID patterns on a 70% formamide–water subphase: (a) pure arachidamide; and (b) arachidamide containing 5% 3-arachidamidopropionamide

observed monolayer films but rather multilayer crystallites [84]. For example, arachidamide ($\text{C}_{19}\text{H}_{39}\text{CONH}_2$) when spread on water self-aggregates into a 2D crystalline monolayer in which the molecules are tilted by 18° from the vertical. On the other hand, when spread on liquid formamide, the amphiphiles self-organize into dimorphic trilayers [85]. In one dimorph the molecules are aligned perpendicular to the liquid surface, whereas in the other the molecules are tilted by 17° from the normal to the liquid surface, as shown in Figure 42(a). On 70% formamide in water, the amphiphile organizes with the coexistence of the tilted trilayer and a monolayer phase where the molecules are aligned vertically. The GID pattern is shown in Figure 41(a).

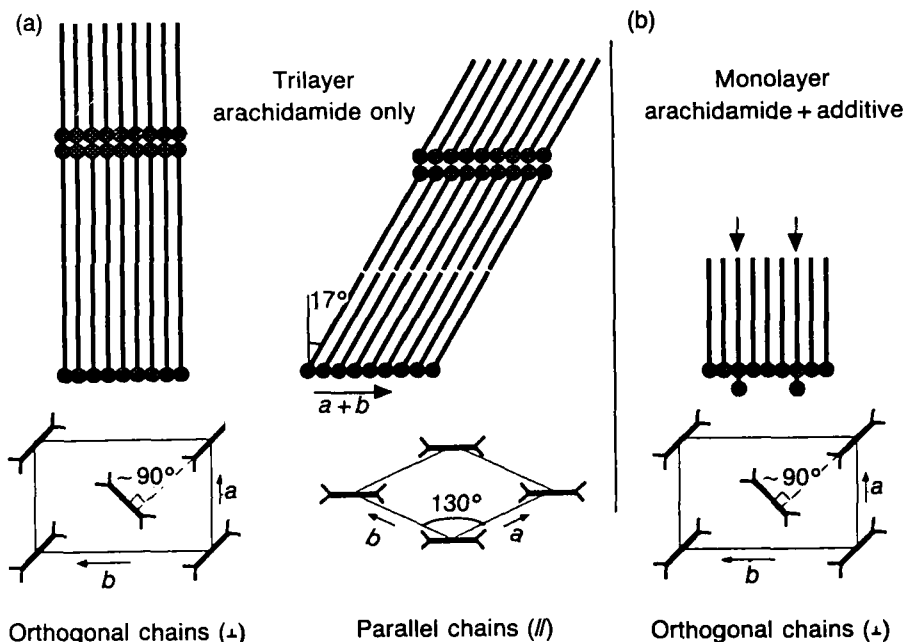


Figure 42 Schematic representation of the packing arrangements of the arachidamide molecule. (a) Trilayer phases with orthogonal and parallel packing arrangements. (b) Monolayer phase obtained by addition of the additive with the orthogonal packing arrangement. Note that in each case the bottom part represents the projection of the hydrocarbon chains onto a plane perpendicular to the long molecular axis and the top part represents the side view

Addition of 5–10% of the ‘tailor-made’ additive *N*-methyларachidamide or 3-arachidamidopropionamide ($\text{C}_{19}\text{H}_{39}\text{CONHC}_2\text{H}_4\text{CONH}_2$) inhibits the formation of the trilayers according to the GID pattern (Figure 41b), yielding monolayer crystallites in which the molecules are vertically aligned (Figure 42b).

The bolaamphiphiles $\text{HO}(\text{CH}_2)_n\text{OH}$ form such multilayers on the water surface, where the number of layers appears to be dependent on the conditions of preparation, temperature and chain length. For example, the bolaamphiphile $\text{HO}(\text{CH}_2)_{22}\text{OH}$ forms multilayers. Their morphology was imaged by cryo-transmission electron microscopy (cryo-TEM) on vitreous ice (Figures 43a and 43b) [86]. These multilayers were also deposited on smooth mica surfaces and their thickness and size determined by scanning force microscopy (SFM). Figure 43(a) shows such crystallites with an average thickness of four layers. The multilayer formation is inhibited upon addition of 10% of $\text{C}_{21}\text{H}_{43}\text{CO}_2\text{H}$ or $\text{C}_{24}\text{H}_{49}\text{OH}$ [87], resulting in monolayers (and some residual bilayers) as shown by SFM (Figures 44b–44d) and cryo-TEM (Figures 43c–43e).

The structures of the $\text{HO}(\text{CH}_2)_n\text{OH}$ multilayers were further examined by X-ray reflectivity and grazing incidence X-ray diffraction (GID). Analysis of the

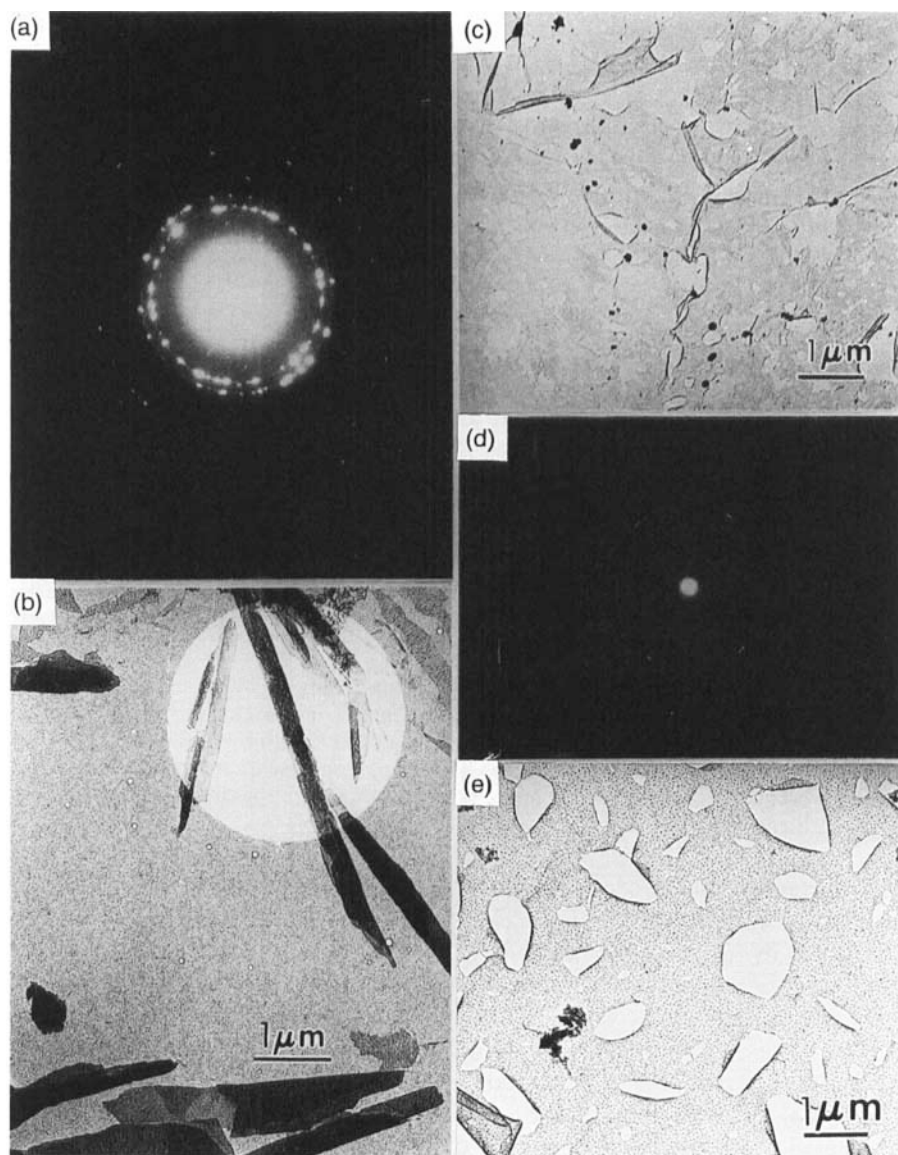


Figure 43 (a) Transmission electron diffraction pattern from crystallites of $\text{HO}(\text{CH}_2)_{22}\text{OH}$ amphiphile within the aperture shown in (b). (b) Bright-field image of crystallites of $\text{HO}(\text{CH}_2)_{22}\text{OH}$ deposited on collodion-coated grids. (c) Bright-field image of a film obtained from a (9:1) mixture of $\text{HO}(\text{CH}_2)_{22}\text{OH}$ and $\text{Me}(\text{CH}_2)_{23}\text{OH}$ deposited on a collodion-coated grid. (d) Electron diffraction pattern from a self-suspended monolayer film of (9:1) $\text{HO}(\text{CH}_2)_{22}\text{OH}$ - $\text{Me}(\text{CH}_2)_{20}\text{CO}_2\text{H}$ deposited on a bare grid and whose bright-field image is shown in (e). All images and diffraction patterns were obtained after partial or complete sublimation of the underlying vitreous ice

GID data yielded multilayer structures at almost atomic resolution. The existence of different crystalline phases was demonstrated by GID measurements (Figure 45a). For example, $\text{HO}(\text{CH}_2)_{30}\text{OH}$ is dimorphic: one phase (B) contains bilayers in which the molecules are highly tilted (Figures 45d and 46c), and the other phase (A) is a monolayer in which the molecules are vertically aligned (Figures 45b, 45c, 46a and 46b) [88]. It remains now to find conditions to control this polymorphic behavior, in a manner akin to 3D crystals, with the use of stereospecific ‘tailor-made’ additives. Studies are in progress for the inhibition of growth in multilayer crystallites of chiral mixtures and for the induction of chiral resolution with the aid of tailor-made additives.

All these studies are in agreement with the model, suggesting that the nuclei assume crystalline structures which can be stereospecifically recognized by the inhibitors in the environment. This knowledge will be used in the next section for the generation of chirality according to Scheme 2.

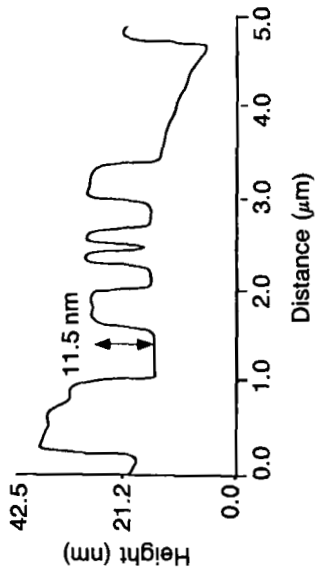
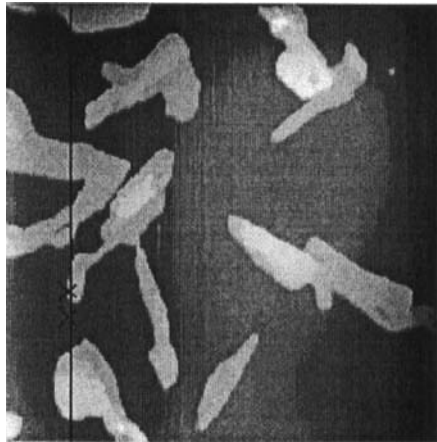
5. CHIRAL AMPLIFICATION IN THE SYSTEM GLYCINE-SOLUBLE α -AMINO ACIDS

So far we have described three different effects which are of general applicability in self-organizing systems:

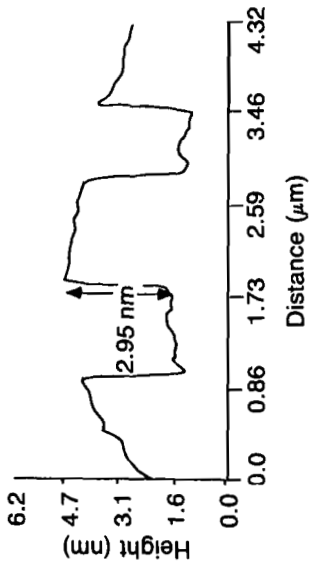
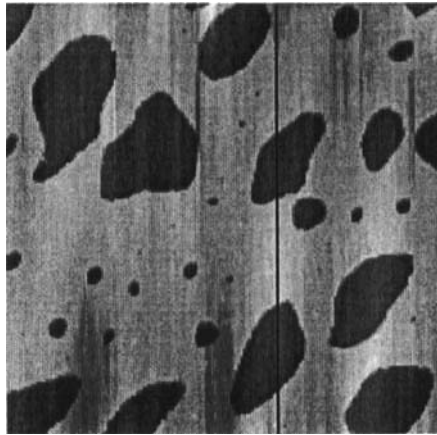
- (1) the transformation of centrosymmetric single crystals into chiral mixed crystals when grown with a chiral face attached to an interface and the opposite enantiotopic face exposed to a racemic solution;
- (2) the self-organization of amphiphilic molecules to form functional clusters as transient nuclei for the oriented nucleation of crystals; and
- (3) the enantioselective inhibition of crystal nucleation with the use of chiral additives.

These three diverse effects can now be combined to design a model system for the spontaneous generation and amplification of chirality. The system of α -glycine grown at the air-solution interface in the presence of racemic α -amino acids provides a feasible model since the centrosymmetric α -glycine single crystals are transformed into mixed chiral solid solutions (as described in Sections 2.3c and 3.2). Such crystals contain only one enantiomer selectively occluded through a {010} face exposed to the solution. Furthermore, hydrophobic α -amino acid additives can aggregate at the solution surface into ordered domains that may serve as templates for the oriented nucleation of α -glycine [89]. The third effect which comes into play is whether the oriented crystallization can be controlled not only by promoting the nucleation of α -glycine ‘correctly’ oriented towards the solution but also by enantioselective inhibition of α -glycine ‘wrongly’ oriented. A series of kinetic experiments were performed on the crystallization of glycine in the presence of a mixture of a racemic hydrophobic α -amino acid [1% w/w (*R,S*)-leucine] with varying amounts of resolved hydrophilic α -amino acids such as (*S*)-alanine, (*S*)-serine and (*S*)-histidine

(a)



(b)



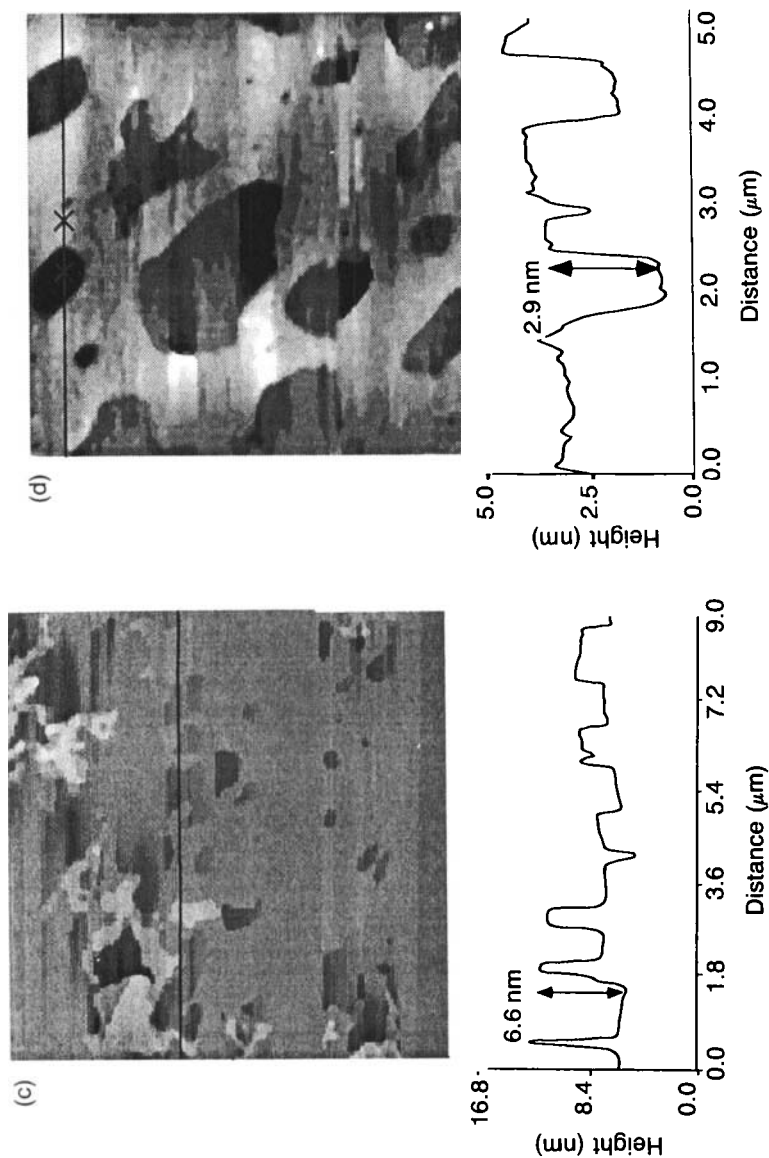


Figure 44 SFM images and the corresponding height profiles of the $\text{HO}(\text{CH}_2)_{22}\text{OH}$ amphiphile deposited from the water subphase on mica. (a) Multilayer crystallites of the pure amphiphile. (b) Monolayer film created by addition of 10% $\text{Me}(\text{CH}_2)_{22}\text{OH}$. (c) Taken from the same sample as in (b), in a different region, showing the existence of some bilayers in addition to the monolayer. (d) Monolayer film created by addition of 10% $\text{Me}(\text{CH}_2)_{20}\text{CO}_2\text{H}$

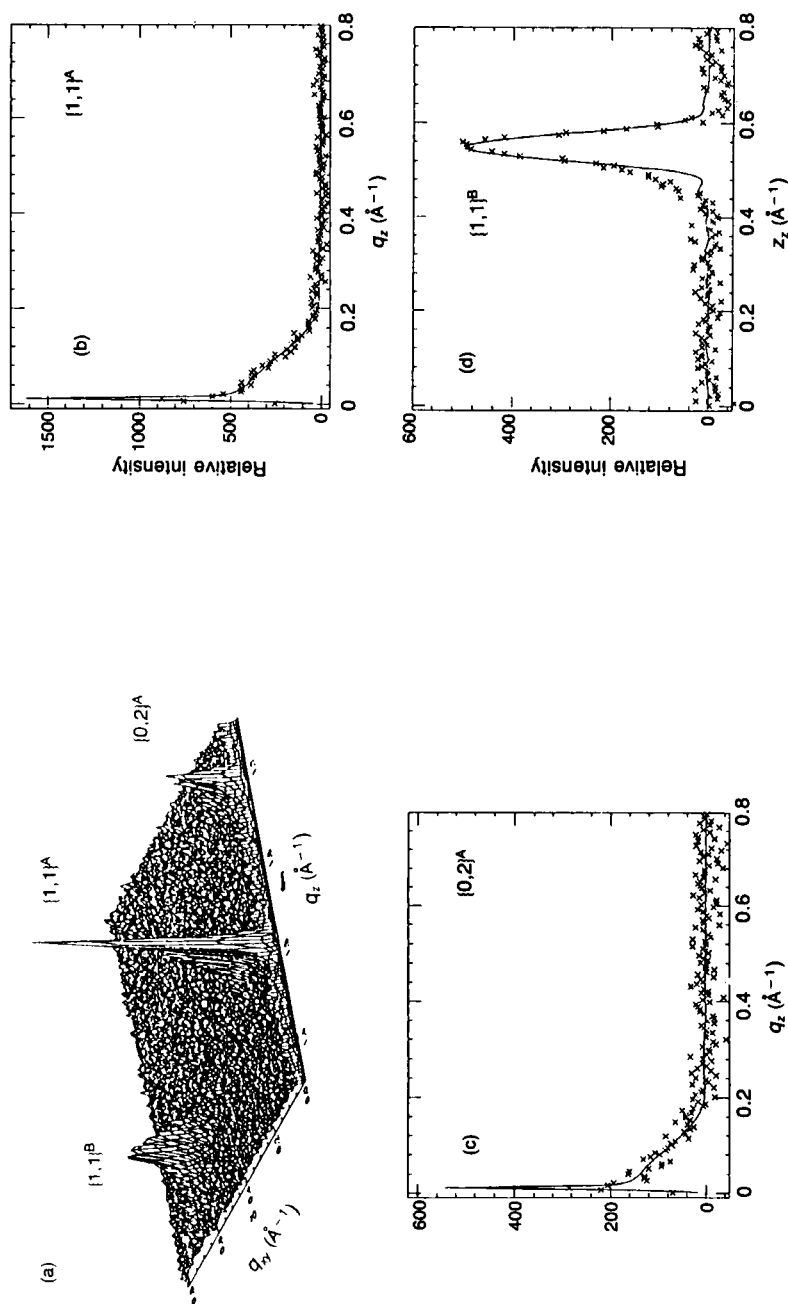


Figure 45 (a) Two-dimensional intensity plot (along q_x and q_z) of the GID pattern obtained from the $\text{HO}(\text{CH}_2)_{30}\text{OH}$ amphiphile on water at 5°C and 100% coverage. Two phases, monolayer A and multilayer B were spontaneously formed on the water surface. (b, c) Observed and calculated Bragg rod intensity profiles of the $\{1,1\}$ and $\{0,2\}$ reflections of the monolayer phase A. (d) Observed and calculated Bragg rod intensity profiles of the $\{1,1\}$ reflection of the multilayer phase B. The relative occupancy of the multilayer *vis-à-vis* the monolayer on the water surface was 15%

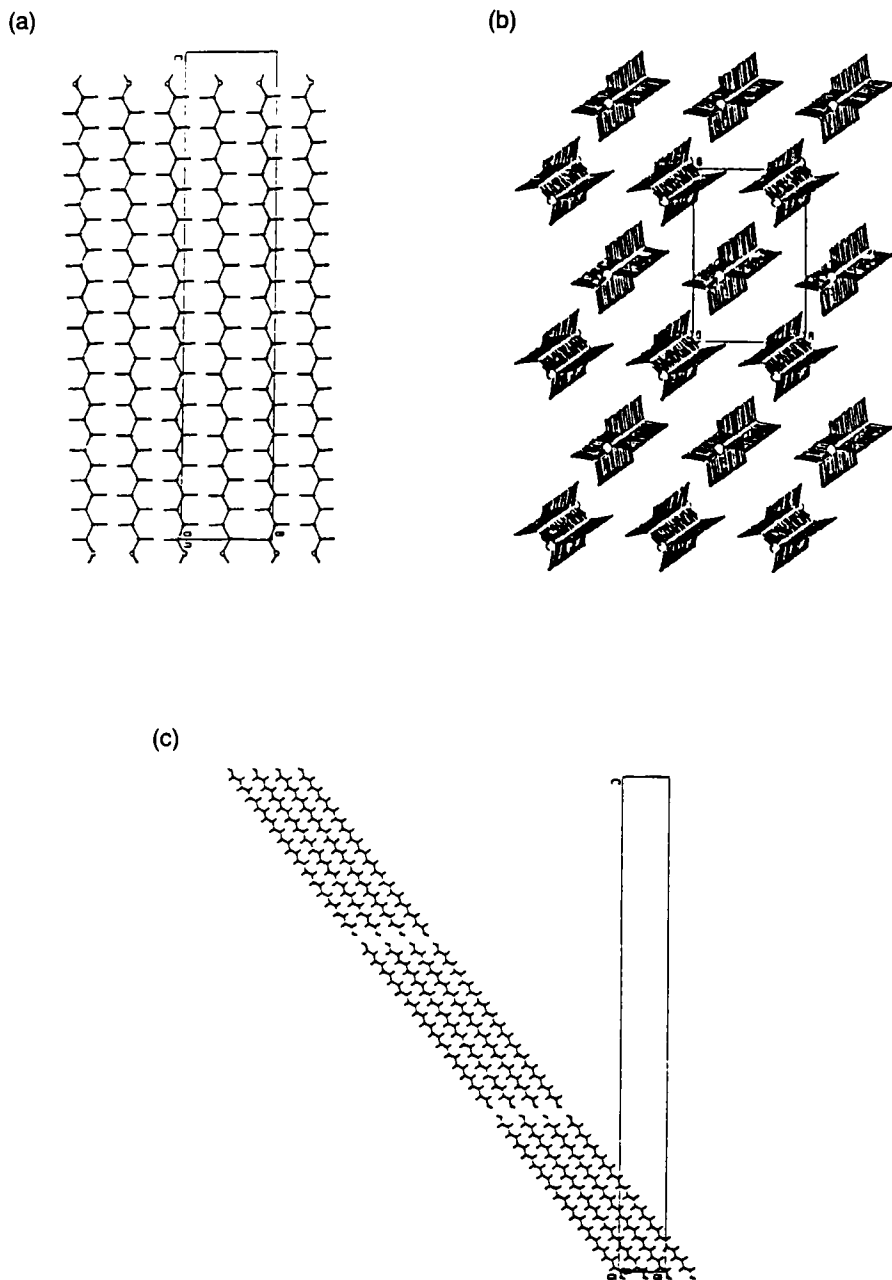


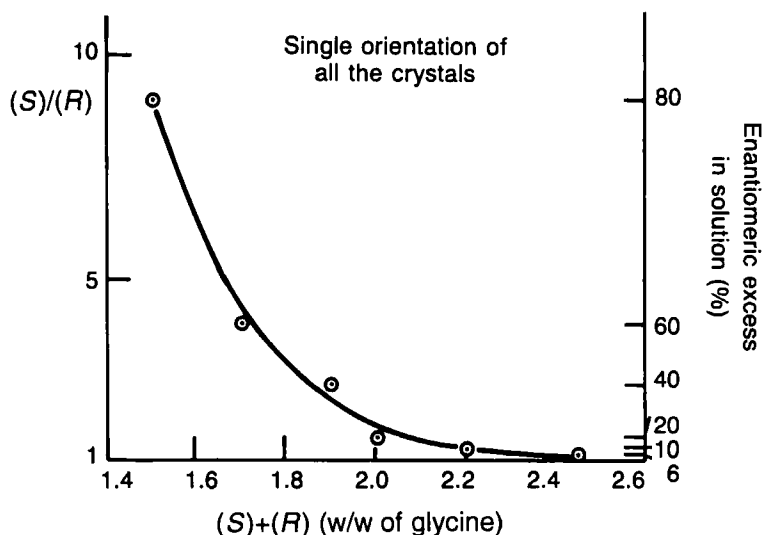
Figure 46 (a, b) Packing arrangements of the monolayer phase of $\text{HO}(\text{CH}_2)_{30}\text{OH}$ viewed along the a -axis and c -axis, respectively. (c) Packing arrangement, viewed along the a -axis, of the multilayer phase of $\text{HO}(\text{CH}_2)_{30}\text{OH}$, assumed to be a trilayer to give the best fit to the data

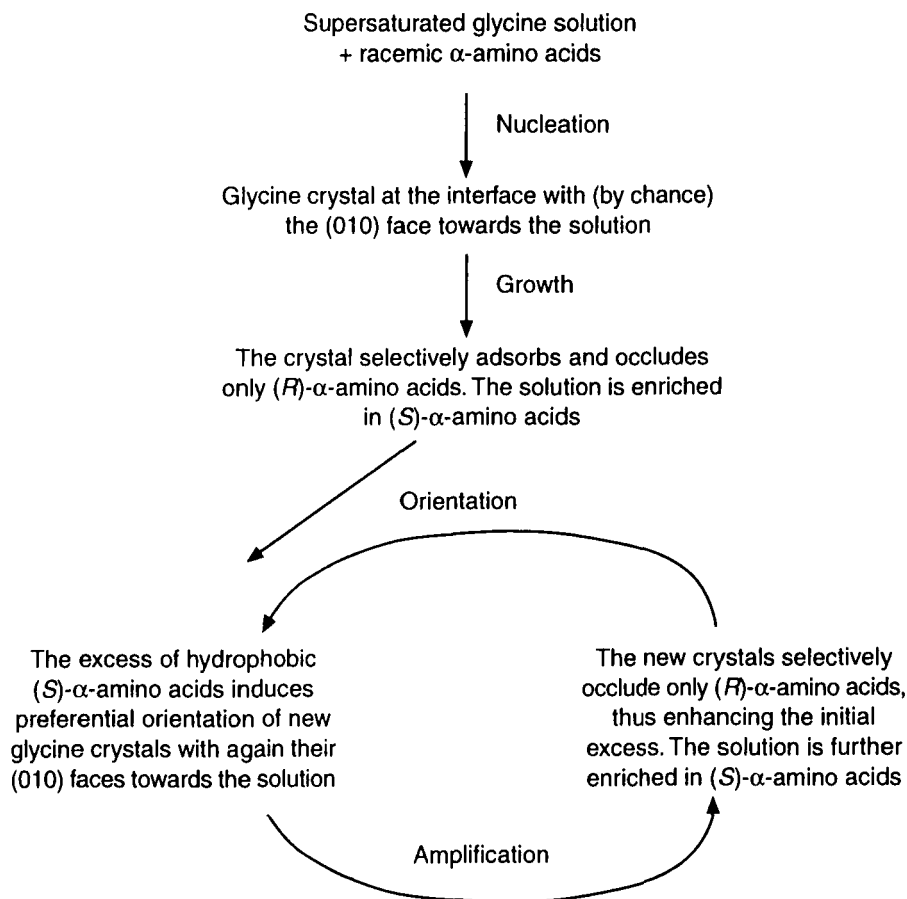
Table 5 Orientation by kinetic effect of floating glycine crystals in the presence of a resolved hydrophilic α -amino acid and 1% (R,S)-leucine

(S)-Additive content (% w/w Gly)	Degree of orientation ^a of floating crystals (%)	Face exposed to air	Enantiomeric excess ^b of occluded leucine (%)
0	0	(010) ₁ , (0 $\bar{1}$ 0)	0
1% alanine	70–75	(0 $\bar{1}$ 0)	60
2% alanine	80	(0 $\bar{1}$ 0)	80
3% alanine	100	(0 $\bar{1}$ 0)	90
1% serine	~0	(010) ₁ , (0 $\bar{1}$ 0)	~0
2% serine	70	(0 $\bar{1}$ 0)	70
3% serine	90	(0 $\bar{1}$ 0)	90
4% serine	100	(0 $\bar{1}$ 0)	>95

^a Checked by X-ray.^b HPLC enantiomeric analysis of occluded leucine provides an independent check of orientation.

(1–4% w/w) (Table 5). The racemic hydrophobic additive induced the formation of crystals floating on the solution surface. When small amounts of the hydrophilic additives were introduced, two types of floating {010} glycine crystal plates were formed, one thick with the (0 $\bar{1}$ 0) face oriented towards the air and so '(0 $\bar{1}$ 0) oriented', and the other thin and '(010) oriented'. The thick plates were greater in number and increased with the amount of the hydrophilic additive

**Figure 47** Correlation between the initial leucine enantiomeric ratio and the total concentration needed for complete (0 $\bar{1}$ 0) orientation of the floating crystals of α -glycine



Scheme 10

in solution. When 4% w/w hydrophilic (*S*)- α -amino acid was present in solution, only the thick (0 $\bar{1}$ 0)-oriented plates were observed. At this concentration, the kinetic retardation of growth was so strong that the inhibition of (010)-oriented crystals was complete (Table 5).

Crystallization of glycine in the presence of partially enriched mixtures of (*R,S*)-leucine in various concentrations showed that complete (0 $\bar{1}$ 0) orientation of α -glycine could be attained by adjusting the initial enantiomeric ratio with the total additive concentration, as shown in Figure 47.

In a regime of high enantiomeric excess of one enantiomer in solution, a low total concentration is required for inducing a single orientation of the floating glycine crystals. When leucine was present at an enantiomeric excess as low as 6% [i.e. (*R*)/(*S*) ratio of 53:47], in solution its total concentration had to be raised above 2.4% w/w of glycine, in order to obtain complete (010) orientation of the glycine

crystals floating on the solution surface. The possibility of inducing such complete orientation of the glycine crystals provides evidence for the feasibility of an overall model for the spontaneous generation and amplification of optical activity under prebiotic conditions as depicted in Scheme 10.

Let us assume that an aqueous solution of a prebiotic mixture of glycine with a variety of α -amino acids is slowly evaporating. Spontaneous fluctuations of some parameter of the system, such as temperature or concentration, can promote the appearance of the first single crystal or a small number of glycine crystals with the same {010} orientation in contact with the solution. The enantioselective occlusion of the hydrophobic α -amino acid additives into these crystals will generate a small enantiomeric excess of molecules of opposite handedness in solution. The above results suggest that this excess might be sufficient to trigger amplification such that all subsequent glycine crystals at the surface will be oriented in the same way. The newly formed crystals contain only one enantiomer, the solution contains both enantiomers, but the enantiomeric excess in solution will increase with continued crystallization. This process could, in principle, eventually lead to a pronounced separation of chiral territories.

6. CONCLUSION AND OUTLOOK

Here we have presented some general ideas and experiments associated with the self-organization of molecules forming two-dimensional and three-dimensional embryonic crystallites *en route* to crystallization.

The growth of these embryonic clusters can be perturbed by the presence of foreign molecules in the environment, since they recognize and so interact stereospecifically with the crystallite surfaces. The recognition by the various sites on the surfaces of the crystallites is akin to that of the kind we are familiar with in biology, such as in enzyme–substrate or antibody–antigen systems.

Once a foreign molecule is strongly bound to such a crystallite surface, it will perturb its growth and probably induce disintegration. Crystallites which do not stereoselectively interact with these additives can grow to completion. This mechanism involving selective growth or disintegration through the presence of minor quantities of additives is most appealing and demonstrates the selectivity required for a cooperative process such as the amplification of small fluctuations from a racemic state.

Another central property linked with self-aggregates and crystals and which is relevant to the present discussion is associated with the nature of cooperative phenomena. For example, as was demonstrated in the diene crystalline systems by the presence of a eutectic and through the autocatalytic resolution of α -amino acid enantiomers inside α -glycine crystals at interfaces, once a low enantiomeric excess is achieved, the system can be driven efficiently towards homochirality.

A severe drawback of the many experiments related to the origin of chirality that involve minerals and molecules which might have been relevant to the origin of life is that they have been performed, by and large, by 'mix and try'. Little consideration has been given to the rules of surface recognition. For example, the performance of chiral crystals of quartz for asymmetric catalysis or enantioselective adsorption of α -amino acids has been disappointing. Clearly one should try to assess the weak interactions at the various liquid–solid interfaces at the planning stage of such experiments. Furthermore, the present studies have also shown that the search for minerals which may interact with chiral molecules should not be limited to chiral crystals but should also encompass nonchiral minerals which express chiral surfaces, such as gypsum, kaolite, talc, etc.

A deeper understanding of solid–liquid interactions in future studies may be expected, since modern experimental and computational techniques will permit one to probe the structure and monitor the dynamics of small clusters almost on the molecular level.

Finally, as hinted in the introduction, the question of the origin of chirality may not be resolved, and no experiment will provide definitive evidence for unraveling this problem. Nevertheless, the search for possible routes should be most rewarding, since it inspires the design of new experiments that would not have been done otherwise, possibly resulting in unexpected discoveries.

7. ACKNOWLEDGMENTS

We thank our colleagues Lia Addadi, Jens Als-Nielsen and Kristian Kjaer with whom we have collaborated on various topics described herein. We are grateful to Edna Shavit, Jarek Majewski, Susan Weinbach and Ivan Kuzmenko for their recent contributions. We acknowledge financial support from the Israel Academy of Sciences and Humanities and the Minerva Foundation, Munich.

8. REFERENCES

1. A. W. Bonner, in *Topics in Stereochemistry* (eds E. Eliel and S. H. Wilen), Vol. 18, John Wiley, New York, 1988.
2. T. L. V. Ulbricht, *Q. Rev. Chem. Soc.*, **13**, 48 (1959).
3. T. L. V. Ulbricht, *Orig. Life*, **6**, 303 (1975).
4. Y. Yamagata, *J. Theor. Biol.*, **11**, 495 (1966).
5. Y. Yamagata, in *Proc. Int. Symp. on Generation and Amplification of Asymmetry in Chemical Systems*, Jülich, Germany, Thiemann, 1974, p. 233.
6. S. F. Mason and G. E. Trauter, *Proc. R. Soc. London, Ser. A.*, **397**, 45 (1985).
7. S. F. Mason, *Nouv. J. Chim.*, **10**, 239 (1986).
8. F. C. Frank, *Biochem. Biophys. Acta*, **11**, 459 (1953).
9. F. F. Seelig, *J. Theor. Biol.*, **31**, 355 (1971).

10. P. Decker, *J. Mol. Evol.*, **2**, 137 (1973).
11. P. Decker, *Nature*, **241**, 72 (1973).
12. P. Decker, *J. Mol. Evol.*, **4**, 49 (1974).
13. P. Decker, *Origins of Optical Activity in Nature*, Elsevier, New York, 1977, p. 109.
14. M. Calvin, *Chemical Evolution*, Oxford University Press, Oxford, 1969, p. 149.
15. E. Havinga, *Chem. Weekbl.*, **38**, 642 (1944).
16. E. Havinga, *Biochem. Biophys. Acta*, **13**, 171 (1954).
17. K. Penzien and G. M. J. Schmidt, *Angew. Chem., Int. Ed. Engl.*, **8**, 608 (1969).
18. B. S. Green and L. Heller, *Science*, **185**, 525 (1974).
19. A. C. D. Newman and H. M. Powell, *J. Chem. Soc.*, 3747 (1952).
20. R. E. Pincock and R. R. Wilson, *J. Am. Chem. Soc.*, **97**, 1474 (1975).
21. G. M. J. Schmidt, *Pure Appl. Chem.*, **51**, 647 (1971).
22. M. Hasegawa, *Chem. Rev.*, **83**, 507 (1983).
23. L. Addadi and M. Lahav, *J. Am. Chem. Soc.*, **100**, 2838 (1978).
24. L. Addadi and M. Lahav, *J. Am. Chem. Soc.*, **101**, 2152 (1979).
25. L. Addadi and M. Lahav, *Pure Appl. Chem.*, **51**, 1269 (1979).
26. L. Addadi, J. van Mil and M. Lahav, *J. Am. Chem. Soc.*, **104**, 3223 (1982).
27. J. R. Scheffer and P. R. Pokkuluri, in *Photochemistry in Organized and Constrained Media* (ed. V. Ramamurthy), VCH, New York, 1991, p. 185.
28. F. Toda, *Top. Curr. Chem.*, **149**, 211 (1988) and references cited therein.
29. Y. Ohashi, in *Reactivity in Molecular Crystals* (ed. Y. Ohashi), VCH, Weinheim, 1993, p. 263.
30. J. van Mil, E. Gati, L. Addadi and M. Lahav, *J. Am. Chem. Soc.*, **103**, 1248 (1981).
31. J. van Mil, L. Addadi, E. Gati and M. Lahav, *J. Am. Chem. Soc.*, **104**, 3429 (1982).
32. I. Weissbuch, L. Addadi, M. Lahav and L. Leiserowitz, *Science*, **253**, 637 (1991).
33. M. Lahav and L. Leiserowitz, *J. Phys. D, Appl. Phys.*, **26**, B22 (1993).
34. L. Addadi, Z. Berkovitch-Yellin, I. Weissbuch, J. van Mil, L. J. W. Shimon, M. Lahav and L. Leiserowitz, *Angew. Chem. Int. Ed. Engl.*, **24**, 466 (1985).
35. Z. Berkovitch-Yellin, L. Addadi, M. Idelson, M. Lahav and L. Leiserowitz, *Angew. Chem., Int. Ed. Engl.*, **21**, 631 (1982).
36. Z. Berkovitch-Yellin, J. van Mil, L. Addadi, M. Idelson, M. Lahav and L. Leiserowitz, *J. Am. Chem. Soc.*, **107**, 3111 (1985).
37. I. Weissbuch, R. Popovitz-Biro, M. Lahav and L. Leiserowitz, *Acta Crystallogr.*, in press (1994).
38. L. Addadi, Z. Berkovitch-Yellin, N. Domb, E. Gati, M. Lahav and L. Leiserowitz, *Nature*, **296**, 21 (1982).
39. M. Vaida, L. J. W. Shimon, Y. Weisinger-Lewin, F. Frolow, M. Lahav, L. Leiserowitz and R. McMullan, *Science*, **241**, 1475 (1988).
40. Y. Weisinger-Lewin, F. Frolow, R. K. McMullan, T. F. Koetzle, M. Lahav and L. Leiserowitz, *J. Am. chem. Soc.*, **111**, 1035 (1989).
41. M. Vaida, I. Weissbuch, M. Lahav and L. Leiserowitz, *Isr. J. Chem.*, **32**, 15 (1992).
42. J. M. McBride, *Angew. Chem., Int. Ed. Engl.*, **28**, 377 (1989).
43. P. Gopalan, M. L. Peterson, G. Crundwell and B. Kahr, *J. Am. Chem. Soc.*, **115**, 3366 (1993).
44. M. Vaida, L. J. W. Shimon, J. van Mil, K. Ernst-Cabrera, L. Addadi, L. Leiserowitz and M. Lahav, *J. Am. Chem. Soc.*, **111**, 1029 (1989).
45. L. J. W. Shimon, M. Vaida, F. Frolow, M. Lahav, L. Leiserowitz, Y. Weisinger-Lewin and R. K. McMullan, *Faraday Discuss. Chem. Soc.*, **95**, 307 (1993).
46. M. D. Cohen, R. Cohen, M. Lahav and P. L. Nie, *J. Chem. Soc.*, 1095 (1972).
47. I. Weissbuch, L. Addadi, Z. Berkovitch-Yellin, E. Gati, S. Weinstein, M. Lahav and L. Leiserowitz, *J. Am. Chem. Soc.*, **105**, 6613 (1983).
48. I. Weissbuch, L. Addadi, Z. Berkovitch-Yellin, E. Gati, M. Lahav and L. Leiserowitz, *Nature*, **310**, 161 (1984).

49. J.-P. Legros and Å Kvik, *Acta Crystallogr. B*, **36**, 3052 (1980).
50. I. Weissbuch, L. J. W. Shimon, L. Addadi, Z. Berkovitch-Yellin, S. Weinstein, M. Lahav and L. Leiserowitz, *Isr. J. Chem.*, **25**, 353 (1985).
51. I. Weissbuch, Z. Berkovitch-Yellin, L. Leiserowitz and M. Lahav, *Isr. J. Chem.*, **25**, 362 (1985).
52. I. Weissbuch, L. Leiserowitz and M. Lahav, *J. Am. Chem. Soc.*, **113**, 8941 (1991).
53. I. Weissbuch, R. Popovitz-Biro, J. L. Wang, G. Berkovic, L. Leiserowitz and M. Lahav, *Pure Appl. Chem.*, **64**, 1263 (1992).
54. D. Jacquemain, S. Grayer Wolf, F. Leveiller, M. Deutsch, K. Kjaer, J. Als-Nielsen, M. Lahav and L. Leiserowitz, *Angew. Chem., Int. Ed. Engl.*, **31**, 130 (1992).
55. J. Als-Nielsen, *Phys. Rev.*, in press (1994).
56. I. Weissbuch, K. Kjaer, J. Als-Nielsen, M. Lahav and L. Leiserowitz, submitted (1994).
57. I. Weissbuch, G. Berkovic, L. Leiserowitz and M. Lahav, *J. Am. Chem. Soc.*, **112**, 5874 (1990).
58. I. Weissbuch, G. Berkovic, K. Kjaer, J. Als-Nielsen, M. Lahav and L. Leiserowitz, submitted (1994).
59. M. Gavish, R. Popovitz-Biro, M. Lahav and L. Leiserowitz, *Science*, **250**, 973 (1990).
60. R. Popovitz-Biro, J. L. Wang, J. Majewski, E. Shavit, L. Leiserowitz and M. Lahav, *J. Am. Chem. Soc.*, **116**, 1179 (1994).
61. D. Jacquemain, F. Leveiller, S. Weinbach, M. Lahav, L. Leiserowitz, K. Kjaer and J. Als-Nielsen, *J. Am. Chem. Soc.*, **113**, 7684 (1991).
62. J. Majewski, L. Margulis, D. Jacquemain, F. Leveiller, C. Bohm, T. Arad, Y. Talmon, M. Lahav and L. Leiserowitz, *Science*, **261**, 891 (1993).
63. J. Majewski, R. Popovitz-Biro, K. Kjaer, J. Als-Nielsen, M. Lahav and L. Leiserowitz, *J. Phys. Chem.*, **98**, 4087 (1994).
64. J. Majewski, R. Popovitz-Biro, K. Kjaer, J. Als-Nielsen, M. Lahav and L. Leiserowitz, submitted (1994).
65. J. L. Wang, F. Leveiller, D. Jacquemain, K. Kjaer, J. Als-Nielsen, M. Lahav and L. Leiserowitz, *J. Am. Chem. Soc.*, **116**, 1192 (1994).
66. I. Weissbuch, F. Frolow, L. Addadi, M. Lahav and L. Leiserowitz, *J. Am. Chem. Soc.*, **112**, 7718 (1990).
67. E. M. Landau, M. Levanon, L. Leiserowitz, M. Lahav and J. Sagiv, *Nature*, **318**, 353 (1985).
68. E. M. Landau, R. Popovitz-Biro, M. Levanon, L. Leiserowitz and M. Lahav, *Mol. Cryst. Liq. Cryst.*, **134**, 323 (1986).
69. E. M. Landau, S. Grayer Wolf, M. Levanon, L. Leiserowitz, M. Lahav and J. Sagiv, *J. Am. Chem. Soc.*, **111**, 1436 (1989).
70. S. Grayer Wolf, L. Leiserowitz, M. Lahav, M. Deutsch, K. Kjaer and J. Als-Nielsen, *Nature*, **328**, 63 (1987).
71. S. Grayer Wolf, M. Deutsch, E. M. Landau, M. Lahav, L. Leiserowitz, K. Kjaer and J. Als-Nielsen, *Science*, **242**, 1286 (1988).
72. D. Jacquemain, S. Grayer Wolf, F. Leveiller, M. Lahav, L. Leiserowitz, M. Deutsch, K. Kjaer and J. Als-Nielsen, *J. Am. Chem. Soc.*, **112**, 7724 (1990).
73. J. W. Mullin, *Crystallization*, 3rd edn, Butterworth-Heinemann, Oxford, 1993.
74. L. J. W. Shimon, F. C. Wireko, J. Wolf, I. Weissbuch, L. Addadi, Z. Berkovitch-Yellin, M. Lahav and L. Leiserowitz, *Mol. Cryst. Liq. Cryst.*, **137**, 67 (1986).
75. L. J. W. Shimon, M. Vaida, L. Addadi, M. Lahav and L. Leiserowitz, *J. Am. Chem. Soc.*, **112**, 6215 (1990).
76. L. J. W. Shimon, M. Vaida, L. Addadi, M. Lahav and L. Leiserowitz, *Organic Crystal Chemistry*, IUCR Crystallographic Symposia 4, Oxford University Press, Oxford, 1991, p 74.
77. F. C. Wireko, L. J. W. Shimon, F. Frolow, Z. Berkovitch-Yellin, M. Lahav and L. Leiserowitz, *J. Phys. Chem.*, **91**, 472 (1987).

78. I. Weissbuch, L. Leiserowitz and M. Lahav, *Adv. Mater.*, in press.
79. E. Staab, L. Addadi, L. Leiserowitz and M. Lahav, *Adv. Mater.*, **2**, 40 (1990).
80. L. Addadi, S. Weinstein, E. Gati, I. Weissbuch and M. Lahav, *J. Am. Chem. Soc.*, **104**, 4610 (1982).
81. D. Zbaida, I. Weissbuch, E. Shavit-Gati, L. Addadi, L. Leiserowitz and M. Lahav, *React. Polym.*, **6**, 241 (1987).
82. I. Weissbuch, D. Zbaida, L. Addadi, M. Lahav and L. Leiserowitz, *J. Am. Chem. Soc.*, **109**, 1869 (1987).
83. I. Weissbuch, I. Kuzmenko, M. Vaida, S. Zait, M. Lahav and L. Leiserowitz, *Chem. Mater.*, in press (1994).
84. S. P. Weinbach, K. Kjaer, J. Als-Nielsen, M. Lahav and L. Leiserowitz, *J. Phys. Chem.*, **97**, 5200 (1993).
85. S. Weinbach, K. Kjaer, G. Grubel, J. F. Legrand, J. Als-Nielsen, M. Lahav and L. Leiserowitz, *Science*, **264**, 1566 (1994).
86. R. Popovitz-Biro, J. Majewski, L. Margulis, S. Cohen, L. Leiserowitz and M. Lahav, *J. Phys. Chem.*, **98**, 4970 (1994).
87. R. Popovitz-Biro, J. Majewski, L. Margulis, S. Cohen, L. Leiserowitz and M. Lahav, *Adv. Mater.*, in press.
88. J. Majewski, R. Edgar, R. Popovitz-Biro, K. Kjaer, J. Als-Nielsen, V. Bouwman, M. Lahav and L. Leiserowitz, in preparation.
89. I. Weissbuch, L. Addadi, L. Leiserowitz and M. Lahav, *J. Am. Chem. Soc.*, **110**, 561 (1988).

Chapter 7

A Model of the Origin of Life and Perspectives in Supramolecular Engineering

H. KUHN

Ringoldwilstrasse, CH3656 Tschingel ob Gunten, Switzerland

AND J. WASER

6120 Terryhill Drive, La Jolla, CA 92037, USA

1. POSING THE PROBLEM

Modern chemistry finds itself at the outset of a fascinating transition period, in which interest is directed more and more toward the construction of organized systems of cooperating molecules, supramolecular devices in which each molecule has its distinct location and specific role. This new objective has great potential in the areas of microelectronics, integrated optics, sensors and actuators, memories with high storage capacity, information-processing devices, catalysts for complex chemical reactions and energy conversion devices. The most intricate devices of this kind are living systems, and the question of how they develop is a most intriguing problem. Any insight into their evolution provides an important source of ideas in this new and rapidly developing field of molecular engineering.

Simple prototypes of artificial systems, in which different kinds of molecules cooperate to form functional entities, have already been obtained by designing and synthesizing molecules that interlock like the parts of a jigsaw puzzle (lock-and-key principle [1–7]). The interlocking can further be supported by appropriate changes of environmental conditions. For example, in contemplating the construction of organized systems one may look for molecules that are capable

of fitting together and subject them to a program of external manipulations, such as confining them to a water surface and pushing them together. They change position by diffusing and interlocking in the intended way whenever opportunities arise. The monolayer is transferred to a solid support, followed by other layers of different kinds in a planned sequence (programmed-environmental-change principle [8]). Successive iterations of this program lead to growing insight and further and further refinement. Our basic assumption is that for problems of this nature, i.e. the construction of organized systems or the origin of life, the roles of complementary interlocking and of detailed and continued influences by the surroundings, are of crucial importance.

In biological systems supramolecular devices are the result of Darwinian evolution. With this in mind, our intent is to examine the origin of life, beginning with prebiotic conditions on early earth, in order to identify the underlying engineering principles that may be of value in searching for new categories of supramolecular devices.

In our inquiry into the origin of life the lock-and-key and programmed-environmental-change principles will therefore be important guiding principles. Their implementation must be specified, beginning with a simple model. We must analyze their implications and then modify and further specify the model in a process that will require many iterations. In this fashion we will be able to elucidate, step by step, the basic mechanisms in the evolution of lifelike systems.

Questions we are seeking to answer are as follows. How does life arise from prebiotic conditions? What is needed to start the crucial processes? How does complexity arise from simplicity? How can entities that form copies of themselves, subject to processes of variation and selection, approach and finally achieve life? What properties of the basic units of life, and what features of the environment, are crucial for promoting the development of entities that exhibit the characteristics of life? Lifelike systems, systems that have the capability of reproducing and of a continued increase in complexity, display a distinctive fundamental pattern of organization. The emergence of such systems requires exceptional conditions and circumstances. It is these circumstances and conditions that we endeavor to trace.

Our model considerations are based on the present-day understanding of prebiotic chemistry and on the chemical properties of biomolecules [9–11]. We will examine, step by step, the fundamental engineering aspects of the beginnings of life. A characteristic and probably inescapable trait of our model is that again and again it leads to dead ends that are then overcome by breakthroughs in completely new directions. Such breakthroughs will result in an overall system that becomes increasingly complex and sophisticated, while they require new properties of the basic units, new kinds of units or new properties of the environment. It should be apparent that it takes very particular locations for life to develop, for crucial advances to occur, locations for which a number of required conditions must be fulfilled.

In modeling the origin of life [12–14], as well as in searching for perspectives in supramolecular engineering, a consideration of fundamental engineering aspects is more fruitful in leading to new insights than an elaboration of the details of their execution. These details are important in their own right, and we will come to them later in Section 4. In Section 2 we discuss the ground rules, so to speak, the fundamental, overall settings upon which our model is based. Section 3 is concerned with an analysis of what appears to be an almost inescapable sequence of events that is bound to lead to lifelike systems. Section 5, finally, concerns the prospects for supramolecular engineering based on the ideas developed in the earlier sections.

2. GENERAL PREMISES OF THE MODEL

A fundamental prerequisite is that by a special event in a special location, a few suitable monomers have coalesced into a strand that can serve as a template for strand replication. Other requirements include specific temporal and spatial environmental conditions. In more detail, we consider a solution of energy-rich monomers and assume the following.

2.1 Formation of a Template Strand

A solution exists that contains monomers capable of covalently linking together into short strands. During an unusual event, such as drying up of the solution and redissolving, many different strands are formed, at least one of which is capable of acting as a template in the formation of further strands. Details about how such a very special event might have occurred—whether in solution or on a surface, perhaps even preceded by an earlier evolution in the mineral kingdom, on silicate or another mineral surface [11, 15–17], perhaps accompanied by the formation of pyrite from FeS and H₂S as a source of electrons and energy, eventually evolving into a primer of a reductive cycle using CO₂ [16]—will not be considered here.

2.2 Template-Directed Reproduction

This unusual strand diffuses into a region with a solution of energy-rich monomers of the same variety, their activation being perhaps caused by a special group R' which they are linked to.* The unusual strand then serves as a template

*An example concerns the condensation of monomers by oxygen bridges. In the 'special event' this bridge formation may be the result of heating that causes the loss of H₂O, e.g. HO–R–OH + HO–R–OH → HO–R–O–R–OH + H₂O. At room temperature the monomers must first be activated by linkage to an activated group R', i.e. HO–R–OH + HOR' → HO–R–OR' + H₂O. While this removal of H₂O requires energy, subsequent condensation reactions such as HO–R–OR' + HO–R–OR' → HO–R–O–R–OR' + HOR' (and so on for trimers etc.) no longer require an extra source of energy. The activation process can, of course, occur outside the critical region.

in the synthesis of additional strands; the monomers of the template strand weakly attract others in the solution, e.g. by hydrogen bonding, and these monomers interlink as they line up on the template. Each new strand can in turn serve as a template for further strands. (In the special event in Section 2.1 there was abundant external energy present and hence no real need for the monomers to be rich in energy.)

The region under consideration must satisfy special conditions to be given shortly. Other strands formed in the unusual event may enter the region as well. They are of no consequence if they cannot function as templates; if they can there will be competition for a while, until the most effective template strand has taken over. The later entry into this region of a new strand capable of template function has no effect once template-directed strand reproduction is well under way.

2.3 Reproduction Driven by Intricate Cycles of Temperature Changes

A regime of periodic temperature changes permits, at the higher temperatures, the breaking apart of the two strands, permitting each to serve in template-aided synthesis of further strands at the lower temperatures (Figure 1).

2.4 Compartmentation

The solution described, of energy-rich monomers containing a template strand, inundates a region that provides compartmentation, e.g. as furnished by porous rock. While monomers can diffuse into or out of pores with ease, outward diffusion of short strands synthesized in pores with small openings is hampered, and retention of these short strands allows them to multiply and increase their concentration (Figure 2).

2.5 Microdiversity of the Environment to Serve as an Evolutionary Gradient

There must be diversity in the sizes of pore openings. Pores with large openings do not confine short strands but can retain longer ones and therefore can be colonized by such strands. Pores with yet larger openings can be colonized by yet longer strands or by other structures large enough to be constrained from being lost by diffusion. In this way neighborhoods of regions with different-sized pores provide an evolutionary gradient for strand lengthening and, later, in a neighborhood of yet larger pores, for the synthesis of other large structures (Figure 2).

Note that the intermixed presence of a variety of pores does not serve as an evolutionary gradient because, in larger pores surrounded by more narrow ones, the short strands are still prevented from being lost by diffusion. Short strands

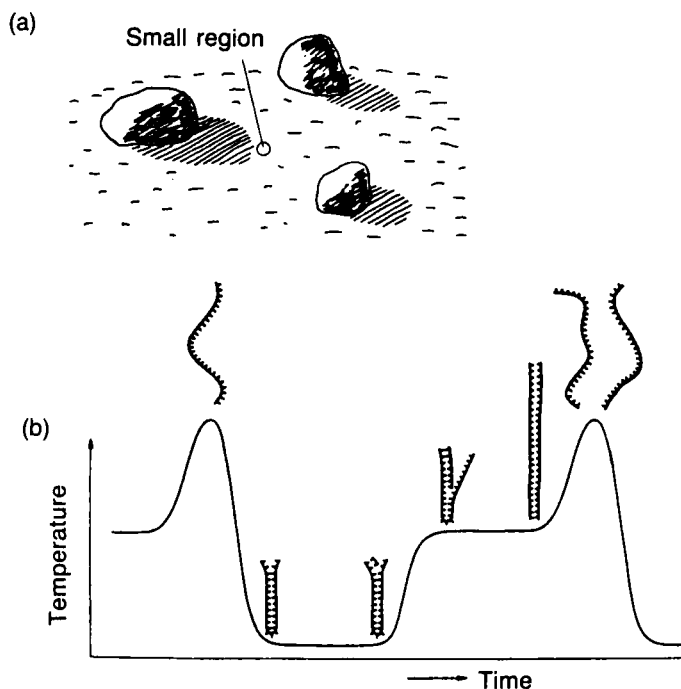


Figure 1 (a) Realization of a regime of periodic temperature changes by shadow-casting rocks. (b) Schematic temperature variation required for strand replication. Reproduced with permission from *Biophysics* (eds W. Hoppe, W. Lohmann, H. Markl and H. Ziegler), Springer, Berlin, 1983, p. 832

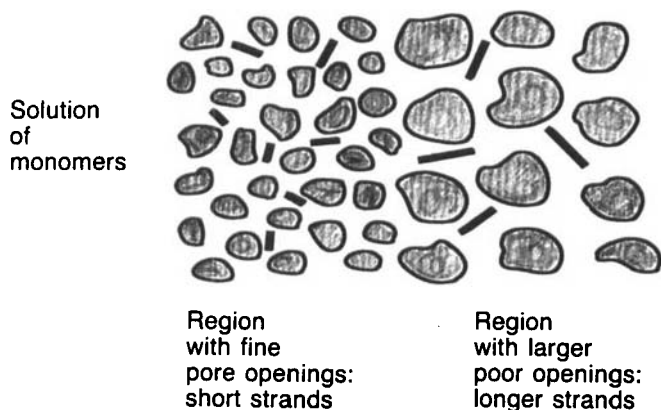


Figure 2 Regions of pores with openings of different size. The region with the larger openings provides an evolutionary gradient for the formation of longer strands (monomers not shown)

can thus freely multiply in the region of larger pores as well. Longer strands, occurring with much lower probability than shorter ones, might rarely be formed in the larger pores, but even so they cannot compete against the shorter strands for the available supply of monomers and will be lost again.

The kind of spatial evolutionary gradient required is the presence of a region that cannot be populated by any of the reproducing devices existing at the time. If then there occurs one of the rare 'beneficial' random errors somewhere, perhaps in one of the copying stages, enabling the resulting daughter copy to colonize the region, it will do so. The meaning of 'beneficial' here is that the daughter strand must be capable of utilizing the opportunities available in the hitherto uncolonized region (such as of not becoming lost by diffusion in the present situation) to survive and multiply.

2.6 Rare Copying Errors are of the Essence

Although most copying errors are lethal (or at least detrimental), rarely and by chance they will promote the colonization of regions of diverging properties. The frequency of copying errors must be sufficiently small (but not zero!) so as to avoid the loss of strands with beneficial properties. Once an improved strand arises, its progeny will displace other strands with inferior properties, a process called natural selection. Terms such as 'beneficial' and 'improved' here refer to the capacity of making faithful copies at a rapid rate, and are not meant to imply that evolution requires any special 'vital force'. Neither are similar anthropomorphic terms we use later. Indeed, random copying errors and natural selection are all we require.

A slowly changing environment provides an evolutionary gradient similar to environmental diversity, provided that changes are sufficiently slow so that the collection of viable strands is not destroyed altogether.

2.7 Multiplication, Occasional Variation by Copying Mismatch, Reconstruction and Selection, all Driven by Intricate Cycles of Environmental Changes

The environmental changes required to drive the propagation of simple strands are modest (see Section 2.3), but later on, when evolution has progressed to more-complicated devices consisting of many strands and involving monomers of different varieties, much more intricate programs are required. During each of the elaborate cycles the complex devices fall into their component strands that are then replicated by template-directed interlinkage of monomers. (Only strands and monomers exist at this point.) Later on the devices are reconstructed from the strands and exposed to environmental conditions that allow only a fraction of them to survive. Certain strands afflicted by a copying error may be rejected in this reconstruction phase, unless certain criteria are met. (For example, a few of the newly made strands may be incapable of being templates

on account of copying errors; these strands simply have no progeny and thus disappear. Other strands may have lost the capability of being made part of a device and diffuse away.) It is most important to remember that some of the strands with copying errors can nevertheless serve as templates and be used as components of devices. Sometimes, although rarely, these new devices may be associated with more beneficial properties than those connected to the devices existing at the time and will replace them. Without such copying errors there would be no evolution.

Sections 2.1 to 2.7 describe the foundations for the origin of self-reproducing entities that then evolve through selection toward increasingly complex structures displaying the features characteristic of present-day living organisms.

3. FUNDAMENTAL ENGINEERING ASPECTS

A search for fundamental engineering aspects should consider appropriate basic elements and suitable environmental conditions that favor a convergence to lifelike systems. Here then is a sequence of model stages that eventually leads to the characteristics of life. Many of them end in stagnation, an evolutionary impasse in which there is no further significant change. Each of these dead ends is overcome eventually by advancement in a novel direction.

3.1 Probability Aspects of Strand Length

The formation of suitable primary strands that can serve as templates is considered to be a rare event, while template-assisted copying occurs with relative ease. It is important that formation of primary strands be sufficiently rare so it does not compete with strand copying. Let α be the probability for the correct linking of two monomers. The probability for the formation of a correct strand containing n monomers (i.e. a strand of n monomers capable of serving as a template) is then $\alpha \exp [-(n-1)]$. If α equals, say, 0.01, then that probability becomes very small as n increases; it is 10^{-10} for $n=6$, 10^{-20} for $n=11$, etc. For this reason only short strands can be formed in this spontaneous fashion. The exact values are of no concern here; the point is that long strands are very unlikely.

Once a strand is formed it can multiply by template-assisted copying until all pores with narrow openings have become saturated with such short strands; a steady state is reached between monomers and short strands that are formed and lost by decomposition and diffusion.

Pores with larger openings are also full of monomers, but short strands are infrequent in them because they diffuse away too fast for efficient copying. For longer strands this situation would not exist, but the very low probability of their direct synthesis from monomers presents an obstacle. However, there is a way out in that strands with $2n$ monomers can be formed by the fusion of two strands of n monomers. Such an event might take place during occasional drying of the solution and redissolving, or by some other unusual occurrence. Strands of yet greater length, capable of colonizing pores with larger openings, can

be formed by further fusion once the concentration of shorter strands in a smaller neighboring pore is large enough to overcome the small probability of such fusions.

Eventually a limit is reached, as copying errors become intolerably burdensome, because the longer the strand the larger the likelihood of 'lethal' copying errors occurring, and strands are generated that cannot act as templates for copying. The probability of such a lethal error rapidly increases with the length of the strand. For example, if there is a 5% chance of a copying error per monomer incorporated during the template-assisted genesis of a strand, with $n = 15$ only half the final strands are error-free [$(0.95)^{14} = 0.5$], and this number drops to one in five for $n = 32$ [$(0.95)^{31} = 0.2$]. If there is a 1% chance of a copying error, these numbers become $n = 70$ for one error-free strand in two and $n = 160$ for one in five [$(0.99)^{69} = 0.5$ and $(0.99)^{159} = 0.2$].

3.2 With one kind of monomer a dead-end is reached at a certain strand length, because longer strands are afflicted with too many copying errors

If just one kind of monomer is available the daughter strands are identical with the original strand, even if the monomers are directional. In this latter case and if the top of one monomer is linked to the bottom of the next, the resulting strand is also directional. Template-directed copying can now occur in two ways (Figures 3a and 3b); the direction of the daughter strand can be the same as or opposite to that of the mother strand (parallel or antiparallel). Mother and daughter strands are identical in either case.

Strands consisting of just one kind of monomer offer little chance of variation; only errors in the way in which elements are linked can occur. Linkage errors that violate the lock-and-key principle result in strands that can no longer serve as templates and are thus 'lethal'; they are replaced by those that can be reproduced. Only a small variety of 'viable' structures can develop with just one kind of monomer, and an evolutionary dead end is reached.

3.3 In a new model we assume two kinds of monomer that are complementary to each other. Strands with distinctive folding conformations evolve that allow intramolecular pairing. Maximum pairing is achieved with a 'hairpin' conformation; it requires monomers that can interlock in an antiparallel fashion

More possibilities are afforded by the availability of two kinds of monomers. We assume them to be complementary, and that complementary monomers can form weak lateral bonds to each other, in this way assisting in the template-driven strand replication (lock-and-key principle). With two kinds of monomers a daughter strand is different from the original strand, but its replica is again a true copy of the original strand, and so forth. These two kinds of strands are arbitrarily called the (+)-strand and the (−)-strand.

While most linkage errors still lead to nonreplicable strands and thus are lethal, another error, the incorporation of the wrong kind of monomer into the daughter

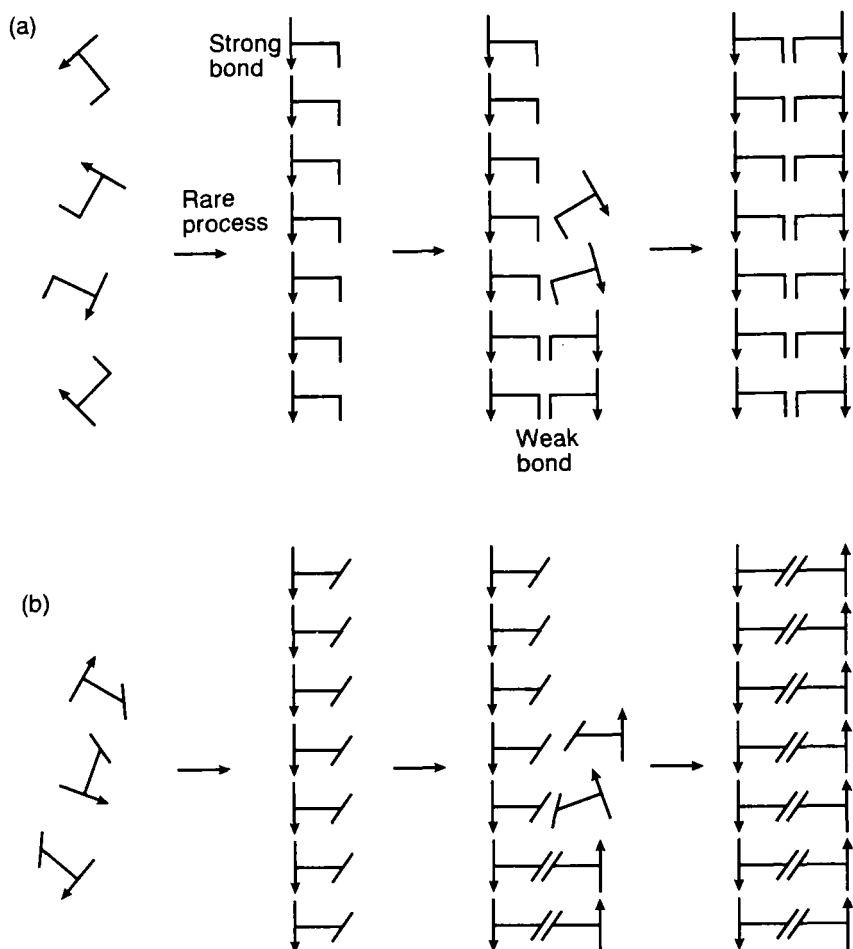


Figure 3 Replication with parallel and antiparallel attraction of monomers; one kind of monomer

strand, is more interesting. That is, an element may on rare occasions be built into a daughter strand that is identical rather than complementary to its companion element in the template strand. Such a mismatch need not prevent replication of the resulting strand, and there would be no evolution without rare errors.

We further assume the monomers and the strands they form to be directional. Two possibilities exist, depending on the lateral complementary bonding between monomers that may require parallel or antiparallel orientation (Figures 4a and 4b). In the first case the direction of a daughter strand will be parallel to that of the mother strand, because of the particular structural properties of the monomers, and in the second case it will be antiparallel. (We disallow a third case,

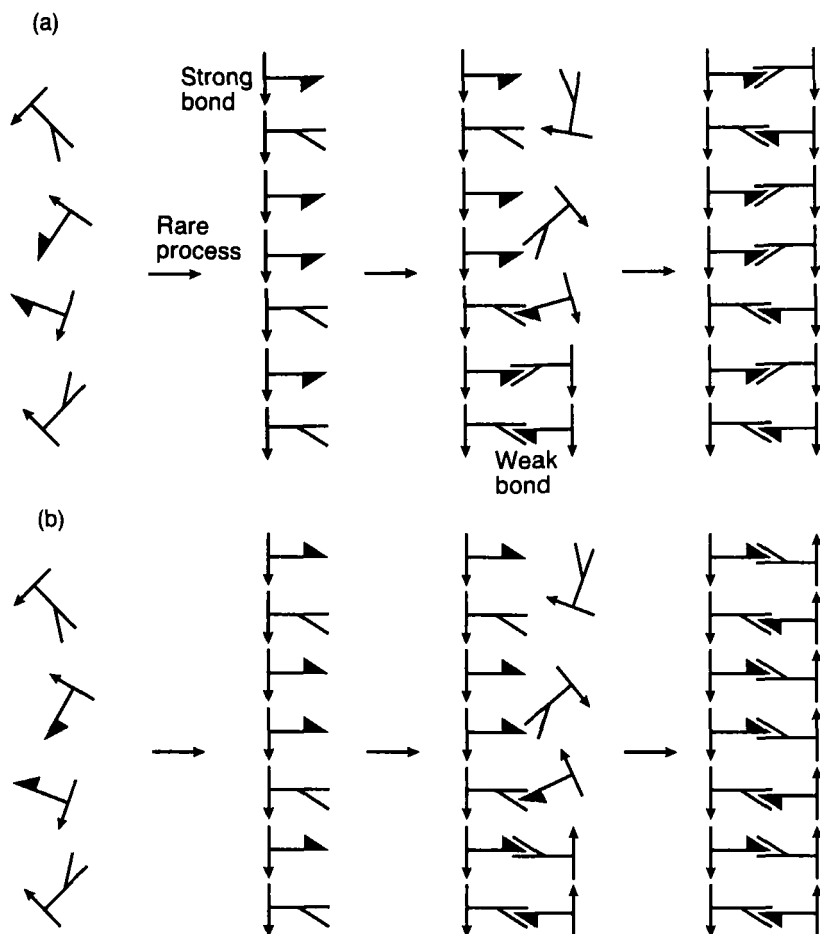


Figure 4 Replication with parallel and antiparallel attraction of monomers; two kinds of monomer

lateral bonding that is independent of orientation, because this possibility has essentially the same structural aspects as having nondirectional monomers.)

After a sufficient number of cycles of reproduction, there will be a pool of strands with many different monomer sequences. Some of these are able to fold back onto themselves in different ways, a fact that will have interesting consequences. This folding occurs when adjacent elements in different strand portions, and in proximity because of bends in the strand, form weak complementary bonds with each other (Figure 5a). Such strands are favored because the insides of the regions of complementary pairing are protected against hydrolytic or other attack from the outside.

The conformation that provides maximum pairing of the monomers of a strand is that of a hairpin, characterized by complete pairing along the 'legs' (lock-and-key principle; Figure 5b); only the monomers in the hairpin 'loop' are not

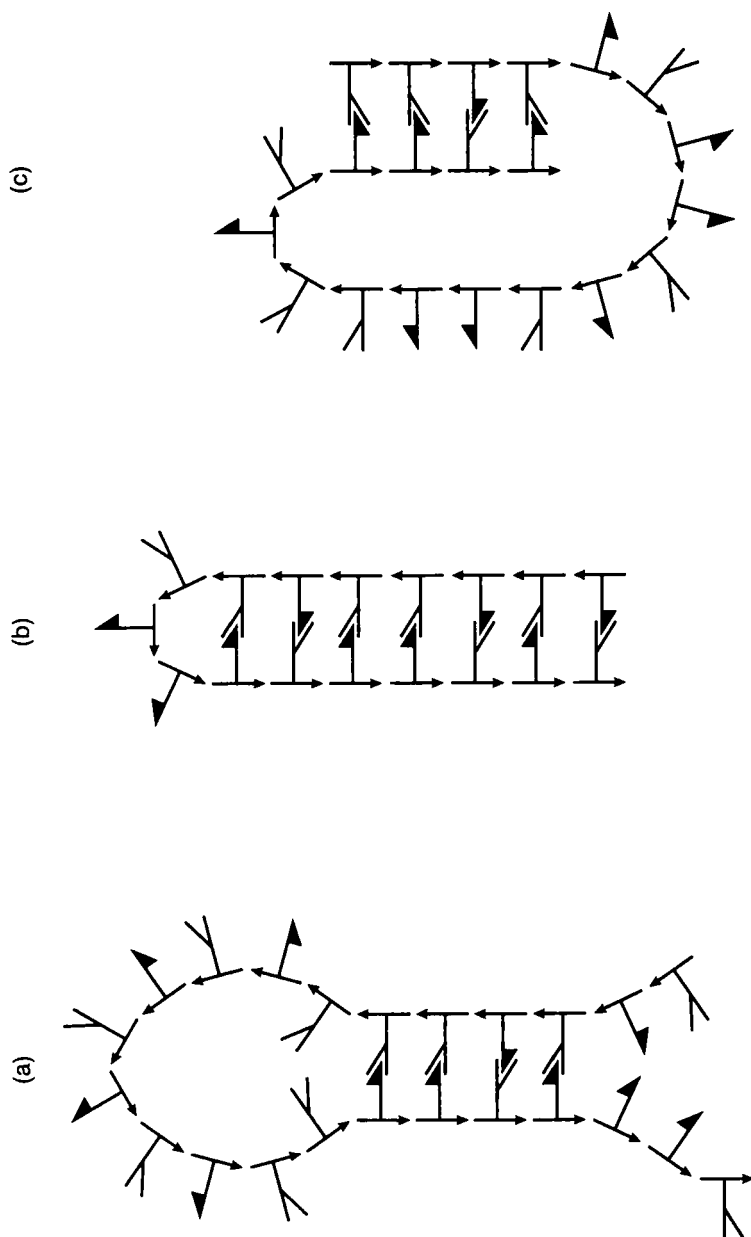


Figure 5 Hairpin formation: (a) incomplete hairpin conformation, with antiparallel attraction of monomers; (b) complete hairpin conformation (note that this conformation requires an antiparallel attraction of monomers); and (c) parallel attraction of monomers does not allow a hairpin conformation and therefore does not afford maximum protection against outside chemical attack

paired. Such a conformation affords maximum protection against external attack and is thus highly favored. Note, however, that a hairpin conformation requires that the strand sections involved be antiparallel; that is, there must be antiparallel monomer pairing. In the parallel case pairing can occur only between strand sections of equal direction. The result would leave large strand portions open and unprotected (Figure 5c). We therefore postulate that an antiparallel orientation is required for lateral monomer bonding because it permits the hairpin conformation. It follows that strand replication must also be antiparallel.

Strands with a hairpin conformation may have arisen gradually, step by step, by means of replication errors that result in new strand conformations. Any new conformation that provides improved protection against chemical attacks from the outside will swamp the earlier, less beneficial conformation in evolutionary competition, a process that converges toward hairpins. It is also possible that strands with a hairpin conformation might first arise more directly, for example as follows. Picture the accidental fusion of two double strands, each consisting of a (+)-strand and a (–)-strand, perhaps just after the replication of single strands and before the separation of mother and daughter strands (Figure 6). Assume that this fusion attaches the end of the (–)-strand of one double strand to the beginning of the (+)-strand of the other, and vice versa. Separation of these two new single strands of double length then yields monomer sequences that are exactly right for the hairpin conformation. That is, we have obtained two hairpin strands. [Fusion of a (+)-strand and a (–)-strand would also do, but temperatures favoring such a rare fusion would favor double strands.] Replicas of hairpins formed in either way are, of course, hairpins as well. It is only the genesis of the first hairpin that requires special events. Hairpins are also favored for another reason that will shortly emerge.

3.4 The earlier dead-end, caused by the detrimental rise of copying errors with increasing strand length, is overcome by evolution moving in a new direction. Hairpins assemble by weak lateral attraction into aggregates that resemble picket fences, presuming that the basic monomers have properties that allow for such lateral attraction between hairpins (not to be confused with lateral bonding between complementary monomers)

While copying errors are needed to provide a pool of variations from which selection can take place, errors that prevent hairpin folding are disadvantageous. The colonization of pores with larger openings by strands created in this fashion is thus limited, perhaps at $n = 50$ or $n = 100$. However, larger structures that permit the colonization of such pores can arise in the following way. We make the assumption that two hairpins can be weakly bonded together laterally (Figure 7a), perhaps through van der Waals forces or electrostatically, mediated by mutual interaction with intercalated ions (lock-and-key principle).

Such aggregation into picket-fence-like aggregates would offer additional protection against hydrolysis or other attack on the sides of the hairpins. Another reward is that the size of these structures permits the colonization of pores with

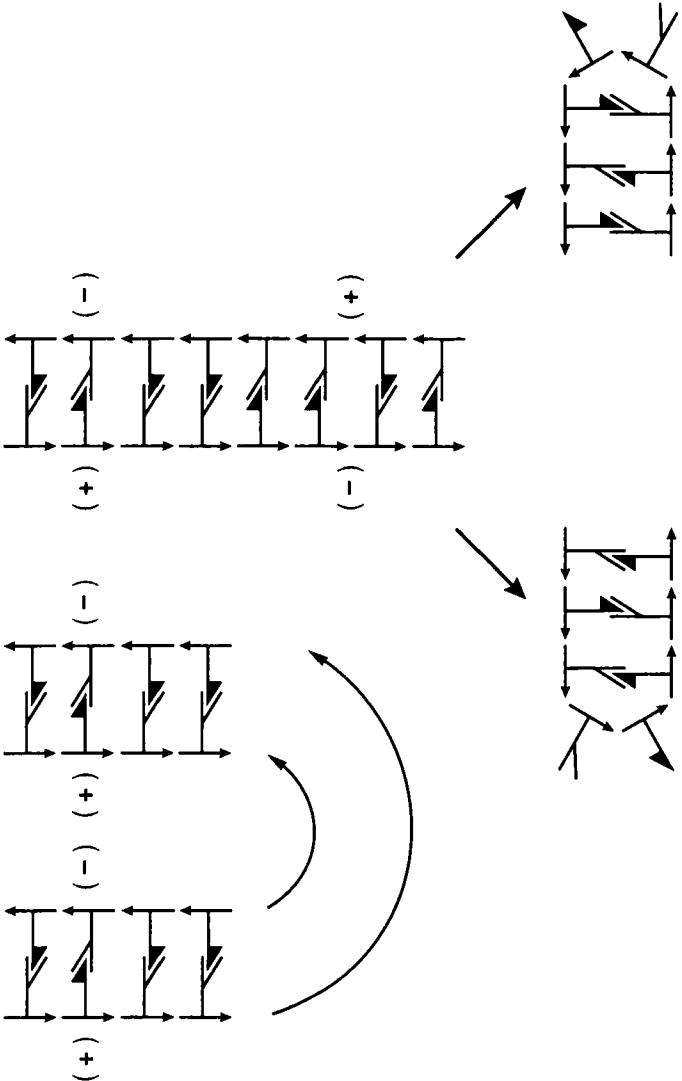
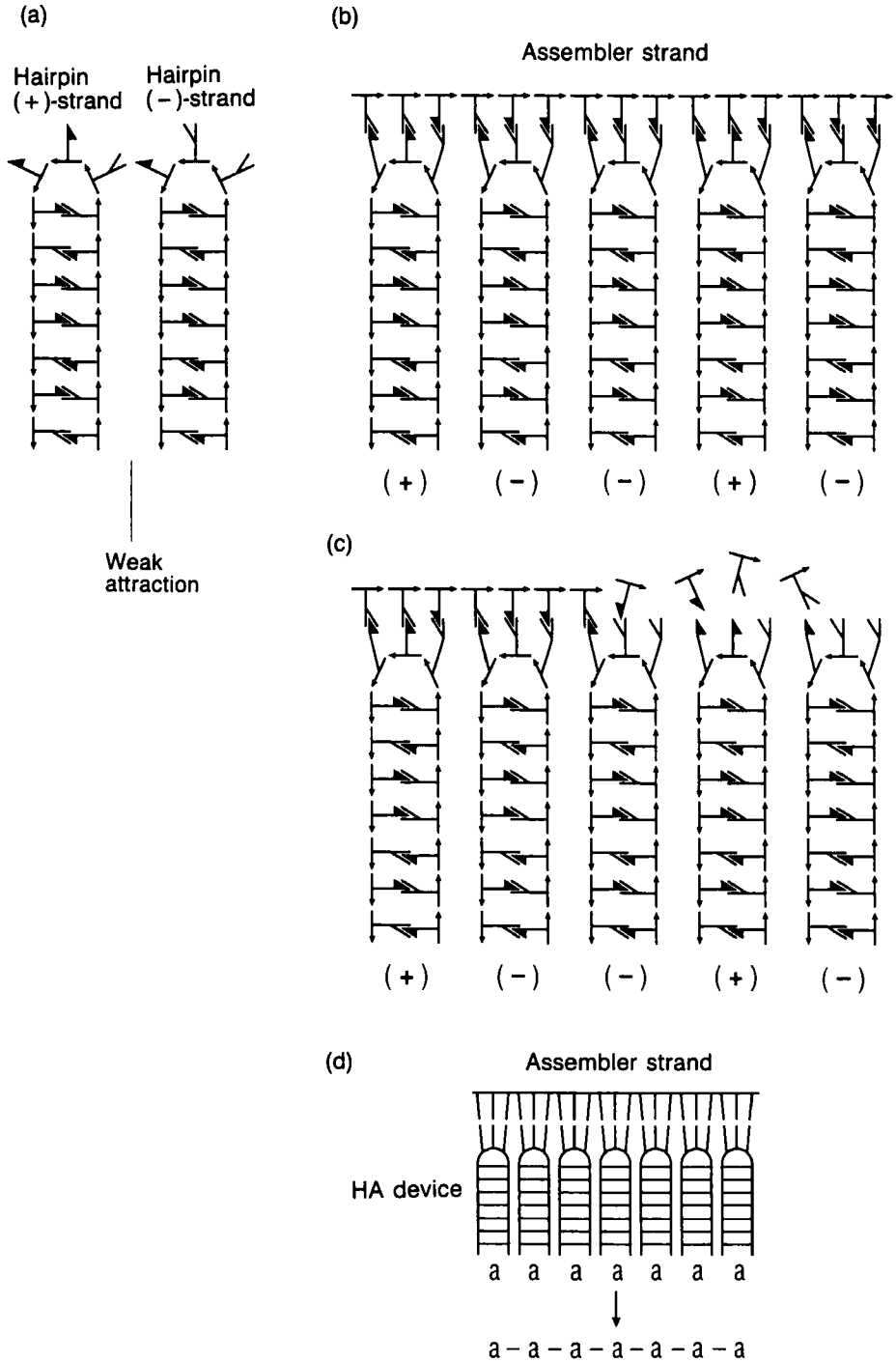


Figure 6 Automatic generation of strands capable of assuming a hairpin conformation by the appropriate fusion of two double strands



larger openings than was possible with just strands. However, of crucial importance is another feature of such aggregate formation, that of providing an error filter. Any new hairpin that is added to an aggregate must fit it with great exactitude to be acceptable. Strands incapable of forming hairpins or forming unacceptable hairpins are not built in and diffuse away. Errors are kept to a constant level in this way. Complete error suppression is, of course, not acceptable because it does not provide variation, i.e. a pool of 'mutants' with slightly different sequences of monomers from which selection can take place to permit adaptation to changes in the environment and colonization of regions with different properties (programmed-environmental-change principle).

3.5 A special 'assembler-strand' guides aggregation of hairpins into a 'hairpins-assembler-strand device'. This provides a strong selective advantage by speeding up the formation of picket-fence-like aggregates of interlocking hairpins

An important improvement would be the existence of a strand of monomers to which the monomers in the loops of the hairpins could attach themselves. Such a strand, an 'assembler strand' (A-strand), would greatly speed up the formation of interlocking picket-fence-like aggregates from hairpins and would thereby provide a strong selective advantage. We term the resulting combination of assembler strand and attached hairpins picket fence the 'hairpins-assembler strand device' (HA device; Figure 7b). (We assume this loose attachment to be by triplets of monomers in the hairpin loops that are tied to complementary monomer triplets in the A-strand for reasons that will be explained in Section 3.10.) We imagine that such an A-strand might work as a guide, leading a new hairpin to an HA device in the process of being built, the hairpin being only loosely attached to the A-strand and therefore still mobile and capable of changing position along the A-strand. In this way the A-strand would convert three-dimensional into one-dimensional diffusion. It may also be that each new hairpin is first attracted laterally to the hairpin incorporated previously into the picket fence aggregate. It can then slide along and up to the A-strand because of thermal motion, and becomes incorporated when it is able to form a suitable attachment to the A-strand. In either situation the presence of the A-strand would be crucial in the firm fixation of a new hairpin into the aggregate, complementarity of several monomers in the hairpin loop with the monomers along the A-strand being required for solid incorporation of the new hairpin. Once in existence, such strand-aided assemblage of hairpins would be greatly favored by selection.

Figure 7 (a) Lateral weak attraction of hairpins leads to the formation of a picket-fence-like aggregate. Shown are a (+)-hairpin and a (-)-hairpin, but a weak attraction obtains between identical hairpins as well. (b) An assembler strand greatly assists in the formation of a picket-fence-like aggregate. (c) Possible genesis of an assembler strand. (d) Catalysis of the synthesis of an a-oligomer

Note that open (i.e. unfolded) strands always exist because of errors in the replication of hairpins. They are usually lost by diffusion but if, by chance, one of them has the appropriate sequence for serving as an A-strand, even for the binding of only two or three hairpin loops, it becomes part of the system; that is, many of its replicas would still diffuse away, but others would be kept by being part of HA devices. It may also be that the loop portion of an already existing aggregate could have served as the template for the formation of an appropriate A-strand (Figure 7c; see also [18]).

3.6 A new variety of monomers promotes independence from the hitherto present compartmentation of the environment exemplified by porous rock. Interlinkage of these monomers is catalyzed by the HA-device and the oligomers produced assist compartmentation by agglutination, thereby narrowing pore openings

Eventually an impasse is reached. The size of the hairpin aggregates has an upper limit and pores with openings that do not sufficiently impede the loss of these aggregates by diffusion cannot be colonized by them. A breakthrough would occur if the aggregate could somehow improve the confinement provided by the environment and even furnish its own confinement, permeable just to monomers. While oligomers that assist in the confinement by agglutination in pore passages might possibly be synthesized from the same variety of monomers that make up the present strands and aggregates, the logistics problems involved in the separation of two competitive tasks, the reproduction of HA devices and the formation of such agglomerates, would be very great. An entirely different variety of monomer would suit the new task much better. Potential confusions in the logistics of the entire scheme would be sidestepped in this way. Additionally, a new variety of monomer would vastly increase the scope of structures that could be synthesized.

We assume that this is what happens, that another type of monomer has been available all the time or becomes accessible at this point. Hitherto unpopulated regions containing suitable monomers can be colonized and a considerable evolutionary advance takes place.

(a) Oligomers of the second variety impede diffusion from pores by agglutinating inside pore channels

In our model this task is accomplished in the following way. We will denote the new monomers by a_1 , a_2 , etc., and will call them monomers of the second variety to distinguish them from the first variety of monomers that make up the HA device. They are to be endowed with the capability of becoming linked to the open ends at the bottoms of the (+)-hairpins and (–)-hairpins, either before or after assemblage of the hairpins into the HA device. In the HA device they are then side by side, and their proximity to each other greatly facilitates their becoming linked by covalent bonds (Figure 7d).

Assuming that these very simple oligomers are capable of agglutination, they would assist the restraining function of pores by narrowing pore openings, making a new group of larger pores accessible to being populated. The exact

monomer sequence in these oligomers would not be important at this stage as long as there is agglutination.

(b) Oligomers of the second variety assist in strand replication by intercalation

At a later time other beneficial oligomers would be formed from the same new variety of monomers. Their function might be to assist in the reproduction of strands and hairpin aggregates, perhaps by some kind of intercalation during the copying process, making this process more accurate and faster. Of course, all such 'favorable' oligomers would arise only rarely, by chance, and for the time being their evolutionary advantages would be of temporary value only because the very copying errors that led to their creation would eventually also destroy them. What will be needed is some way in which advantageous sequences in such oligomers can be 'remembered'.

The question may be asked: why not begin evolutionary modeling with coagulates or coacervates that have been demonstrated to form cell-like shells, as originally proposed by Oparin [19]? The answer is that in such schemes there would not be any evolutionary inducement to develop machinery for fabricating coagulating oligomers. While the presence of such oligomers *ab initio* would be of obvious advantage by itself, the existence of machinery capable of producing them has far greater portents because such machinery can produce many other oligomers that can serve countless other purposes in a step-by-step evolution. Step by step, the machinery itself would also evolve.

The chances of such machinery being generated at once, by chance, are astronomically small. This is different in an evolution by small steps, each of attainable probability and each providing an evolutionary reward, even though their realization takes an immense number of iterations of periodic changes of some kind, as for example represented by cyclic cooling and warming, to drive the many cycles of synthesis and decay. Also required is an environment of subtly different regions, the colonization of which is made possible by the rewards of unused monomers, or 'food' being present. However, such spatial and temporal environments exist, for example, in the form of porous rock subject to many large and small cyclic temporal changes that obtain on earth in a great variety in different locations and at different times (programmed-environmental-change principle).

3.7 The sequence of a_1 and a_2 in the oligomers becomes related to the sequence of (+) and (–) hairpins so that a rigorous relationship exists between the sequence of monomers in the hairpin loops and the a_1 and a_2 -monomers

The next major step is that the HA device becomes in effect an interface, a translation apparatus, between the monomers in the hairpin loops (and thereby the complementary monomers in the A-strand) and the monomers of the second variety that make up the oligomers. The way this might come about is as follows.

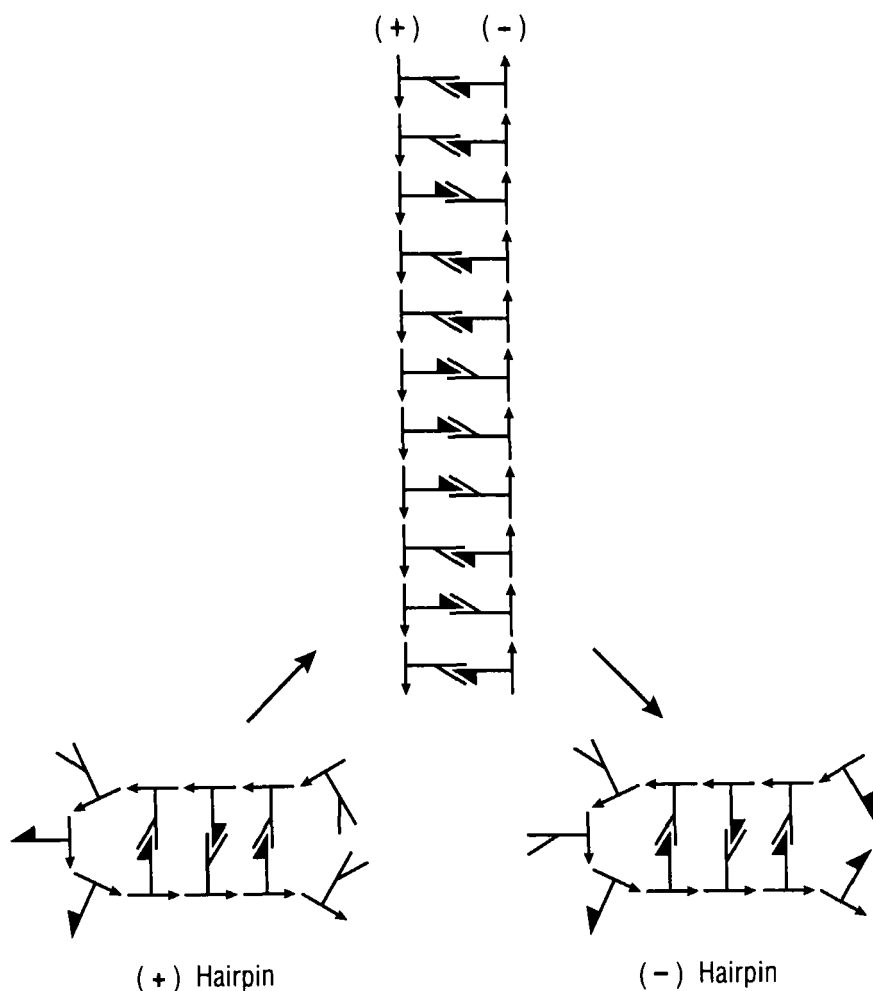


Figure 8 Template-assisted strand replication and assumption of a hairpin conformation by the two separated strands. Note that the two monomer sequences are identical, with two exceptions. The monomers in the center of the loop are complementary, and the monomers at the two ends of a given hairpin are the same and complementary to the monomers at the end of the other hairpin. The reason for the difference in the two ends is that only such a difference allows an attachment to the different hairpins of different α -monomers, a_1 and a_2 , as further discussed in the text and in the legend to Figure 9

An important aspect of (+)-hairpins and (-)-hairpins is that their legs are identical: each leg of a (+)-hairpin has the same monomer sequence as the corresponding leg of its complementary (-)-hairpin (Figure 8). The monomers in the loops of (+)-hairpins and (-)-hairpins are complementary, by their genesis. This hairpin property is most remarkable, and it must be remembered

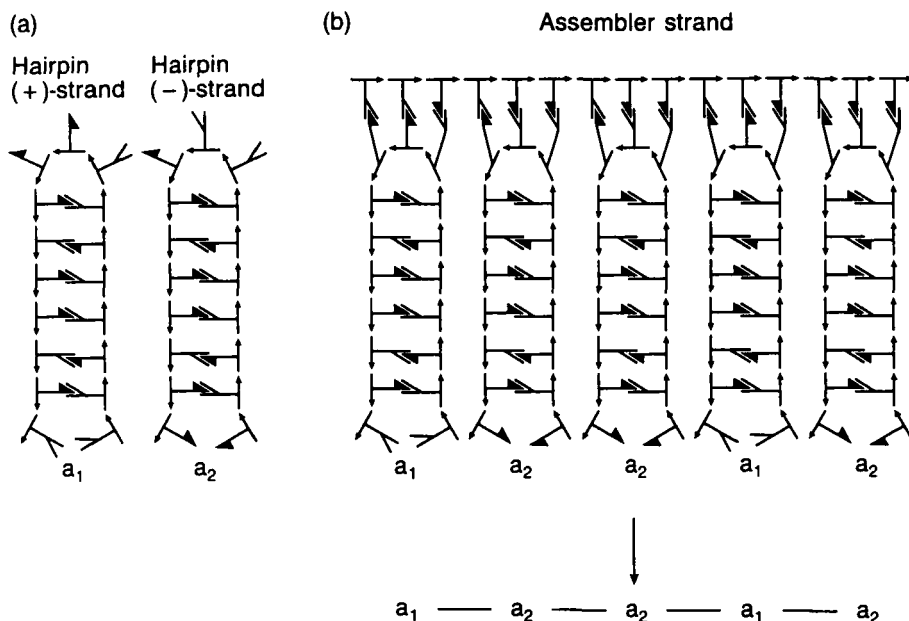


Figure 9 (a) (+)-Hairpin and (-)-hairpin with different a-monomers attached to the different leg-ends of the two kinds of hairpin. The difference in hairpin ends is necessary for an attachment of different a-monomers. The evolutionary reward for such a difference is the scope in different a-oligomers allowed in this way. (b) Catalysis of the synthesis of a specific a-oligomer

that all this has its root entirely in the fact that hairpins represent the most stable strand conformation and that for them strand replication must be in the anti-parallel direction. The two kinds of hairpins may also differ at the ends of the legs (Figure 8) because copying errors in this region are not very deleterious, at least not until this region takes on a very important feature, i.e. that of carrying the new variety of monomers.

We postulate that there are two prevalent kinds of the new monomers, a_1 and a_2 . Then, starting out with just one kind of (+)-hairpin and (-)-hairpin, evolution will put a premium on the development of a relationship between, say, the monomer in the middle of the loop of a hairpin and the monomer of the second variety attached to its bottom end (cf. also Section 3.10). For example, (+)-strands might carry a_1 -monomers at their ends and (-)-strands a_2 (Figure 9a). Moreover, the sequence of monomers of the first variety on the A-strand is complementary to that in the loop of the hairpin linked to them, and in this way there exists a one-to-one relationship between the sequence of monomers along the A-strand and the sequence of monomers of the second variety in the oligomers (Figure 9b). A-strands carry the pattern, the prescription, for the a-oligomers. While in this way a translation apparatus has come into existence as a by-product of an error filter, this device is of no benefit at this stage because any special sequence of monomers in the A-strand that leads to an oligomer with an especially beneficial sequence of a_1 and a_2 is not preserved, because of copying errors.

- 3.8 An HA-device (the HAE₁-device) evolves that furnishes an oligomer functioning as a replication catalyst E₁ that reduces copying errors by virtue of a special sequence of monomers a₁ and a₂. Its involvement in template-guided copying stabilizes the monomer sequence in the A-strand to a degree that permits the retention of favorable special sequences. The translation apparatus becomes conserved. Machinery evolves that produces other catalysts, E₂, E₃, etc., each having a distinct sequence of the a-monomers**

A major step beyond the production of agglutinating oligomers is the evolution of a special sequence of a-monomers which acts as a replication catalyst E₁ that increases the precision of reproduction, perhaps by intercalation that, for example, may improve contacts during replication. Of particular importance is precision in the replication of A-strands, the blueprints for the monomer sequence in a-oligomers. Eventually, per chance, a degree of precision is reached that permits the retention of the replication catalyst E₁. What is required is the retention of a special sequence during the time needed for selection, i.e. the time required for the special HA device, the HAE₁ device, to acquire a sufficient concentration so that its further replication is assured.

A new meritorious oligomer E₂ may arise next, again through copying errors. E₂ may, for example, provide assistance in the precise attachment of a-monomers to the bottoms of the hairpins, perhaps again by intercalation. These two catalysts, E₁ and E₂, may arise in the same region or, more likely, in two areas that are very close to each other so that the two oligomers can easily intermix by diffusion (Figure 10). Also suppose that E₂ and the HAE₂ device responsible for it are retained sufficiently long to be selected. Such a selection is greatly helped by a cooperation between E₁ and E₂ that will exist from the outset. E₂ will assist in the linkage of a-monomers to the leg-ends of hairpins, speeding up the production of oligomers, while E₁ will be responsible for the faithful reproduction and eventual retention of the beneficial E₂-sequence. The efficiency of this 'cooperative' of HAE₁ and HAE₂ devices speeds up the reproduction of its components and permits it to invade and take over the regions occupied by HAE₁ devices alone.

Another possibility is that the catalysts E₁ and E₂ are one and the same E₁ at the outset, a catalyst that both improves replication and assists in the attachment of a-oligomers to the ends of hairpin legs, both perhaps by intercalation. At a later stage there is a split of tasks; an HAE₂ device evolves that yields a variant of E₁, a variant that further improves a-monomer attachment. The improved efficiency has the consequence that the cooperatives of HAE₁ and HAE₂ devices win out over just HAE₁ devices.

- 3.9 A cooperative of translation machines evolves. Each of them produces an a-oligomer that acts as a catalyst E₁, E₂, E₃, etc. These catalysts are mutually supportive because HAE₁-devices, HAE₂-devices, etc. exist in close-by regions, so that the a-oligomers produced are of mutual benefit**

Even though there is competition for monomer 'food' between HAE₁, HAE₂ and ordinary HA devices furnishing the needed agglomeration, they happily

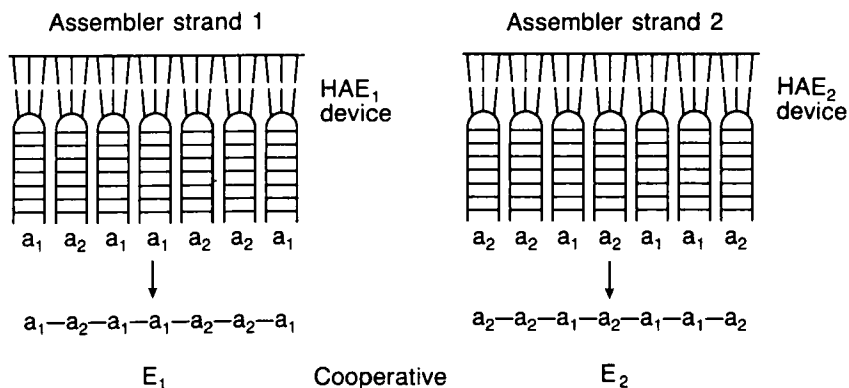


Figure 10 HAE_i devices which produce two different catalysts, E₁ and E₂, that are mutually supportive, being capable of diffusing into each other's neighborhoods

coexist as long as there are more advantages than disadvantages to their simultaneous presence: in short, there is cooperation. There are also new benefits associated with the confinement that was a necessity, from the beginning, to prevent excessive loss by diffusion of multiplying strands (and later of HA devices). The advantage provided by an HA device producing an unusual, beneficial new oligomer is, of course, also conferred to the entire neighborhood. In time the number of evolving HAE_i devices will increase; every time a new E_i conveys an improvement in efficiency the resulting cooperative replaces its older 'progenitor'. In this way a cooperative of increasing complexity develops. Further evolution favors an increase of both the number of sorts of monomers of the second variety a_k and the diversity of the oligomers formed by them.

Note that our translation apparatus is made possible by the selective advantages afforded by the entry of the new a -oligomers, whose initial usefulness was in assisting compartmentation and in diminishing copying errors (Sections 3.6 and 3.7). Its emergence is of great evolutionary benefit. It sparks the beginning of a 'genetic' code, of a memory for specialized useful a -oligomers that have arisen fortuitously. The HA device has become capable of 'reading' the monomer sequence along its A-strand, a sequence that codifies the sequence of a -monomers in a -oligomers. Machinery for the continuous production of replication and other catalysts E_i is now in place. These portentous events permit a process of integration that is a fundamental requirement for the emergence and, later, the diversification of life. The advent of this translation machinery represents a process of integration that is as fundamental as was the aggregate formation described in Section 3.4. At that point the major step was the appearance of a collective of units that collaborate in the rejection of unsuitable strands; here it is the advent of a functioning cooperative consisting of collaborating HAE_i devices that produce catalysts E_i of mutual benefit. However, this process eventually ends in most serious impasses to be discussed soon.

We may again address the question of what the environment must look like so that such a cooperative system must evolve by necessity and indeed go on evolving. While single strands can exist in regions of fine compartmentation only, hairpin aggregates are able to colonize regions with fewer constraints, with less-constraining compartmentation. In the present situation of cooperating HAE_i devices even larger compartments become 'habitable' and the presence of a stable translation apparatus allows the evolution of ever more complex oligomers, permitting the invasion of further hitherto 'uninhabitable' regions. The environmental gradients needed were, at first, areas of different sizes of compartmentation. In later stages any as yet unpopulated region that is 'within reach' provides an evolutionary gradient, where we mean by 'within reach' that a minor change in the existing system, such as a new catalyst E_j, permits the colonization of the region. For example, a most important (much later) step will be the advent of oligomer catalysts that promote the production of suitable monomer 'food'. Regions with an insufficient supply of such monomers can then be invaded. Thus there is an enormous evolutionary gradient towards systems with metabolism.

3.10 In an elaboration of the HA-device we justify the earlier assumption of a hairpin loop attachment to the A-strand by three complementary monomer pairs by considering the requirements for the evolution of a code

In Sections 3.5 and 3.7 we described how the monomer sequence along the A-strand portion of an HA device is translated into a sequence of monomers in an a-oligomer through the mediation of the hairpins attached to the A-strand. We assumed that the firm integration of each hairpin requires the linkage of three adjacent monomers in the hairpin loop to three complementary monomers along the A-strand, in addition to a weak lateral attraction between neighboring hairpins. The monomers of an A-strand are thus arrayed into adjoining triplets, and to each is attached a triplet in the loop of a hairpin. This 'triplet' assumption is based on arguments that address the following question. What is the optimal number of monomers in the A-strand (to which can be attached an equal number of monomers in hairpin loops) so that changes in the sequence of these monomers can code for a sufficient variety of a-oligomers to assure evolutionary survival? We now consider some important postulates.

(1) *There is a premium on a code that is dependable and elaborate enough to accommodate all needs for survival in a given environment, yet no more complicated than is necessary.* Systems that evolve such a code would have a huge evolutionary advantage over systems that do not, because it would provide for great elasticity in responses to yet to be pioneered or to changing environmental conditions. It would permit as much diversity and multitude in a-monomer components as might be needed for such adaptations. On the other hand, a code that is more elaborate than needed for survival would be burdensome in evolutionary terms; any step in such a direction would not lead to selection. Another simple postulate, without which not even the simplest code could exist, is the following.

(2) *For the evolution of a code the (+)-hairpin strand and (-)-hairpin strand must be different in the loop.* Let us then first consider attachment by singlets or doublets of complementary monomer pairs. Singlets hardly offer any scope for a code. With doublets the two loop monomers may be different or they may be identical. In the first case we have two possibilities, $\langle R_1 R_2 \rangle$ and $\langle R_2 R_1 \rangle$, where R_1 and R_2 denote monomers of the first, original variety and the angular symbols ($\langle \rangle$) indicate strand direction. Consider the first of them. The doublet portion of the replica hairpin is $\rangle R_2 R_1 \rangle$ which, noting opposite directionality, is identical with $\langle R_1 R_2 \rangle$, and therefore the (+)-strand and (-)-strand are identical in the loop. The same is, of course, true for the second possibility and hence both are ruled out by requirement (2). If both monomers in the loop doublet are the same, the doublets would be $\langle R_1 R_1 \rangle$ or $\langle R_2 R_2 \rangle$. We rule out this case because of the paucity of coding possibilities it affords. It soon would fall by the wayside in an evolutionary competition against the code opportunities made possible, for example, by triplets.

(3) *More than three monomers are ruled out for reasons of parsimony.* Triplets offer $2^3 = 8$ possibilities with two R-monomers, R_1 and R_2 , and $4^3 = 64$ possibilities with four R-monomers, permitting to code for some 60 kinds of a_k -monomers, thus allowing ciphers for an enormous number of a-oligomers. Moreover, an attachment involving triplets greatly favors the interlocking of components, permits rapid assembly and endows the resulting picket fence structure with rigidity. An attachment involving more than triplets is not optimal; the higher the number of complementary monomer pairs providing the attachment, the more complex the accompanying logistics, the more wobbly and rickety the resulting picket-fence-like aggregate and the less rapid its assembly. Unless really required for coding needs, an attachment by more than three monomer pairs would represent an evolutionary burden rather than an advantage.

It is of great importance to realize that an attachment by triplets, although optimal, is possible only if the monomers involved in a realization of this model satisfy certain energetic and geometrical requirements. For example, the linkage between complementary monomers and the postulated lateral forces between complementary monomers and the postulated lateral forces between hairpins must, in concert, effectively counteract any destabilizing temperature motion, yet must be flexible enough to permit disengagement. The strands formed by the monomers must be flexible enough to permit folding into hairpins without requiring too many monomers in the loop.

While this leaves triplets we impose another demand. There has to be a way, certainly in the early stages of evolution, of telling where a triplet along the A-strand begins and ends, i.e. there has to be a 'reading frame'. Specifically, we demand (letting N stand for R_1 or R_2 and, in later stages, also R_3 etc.) the following.

(4) *In the early stages of evolution there must exist a reading frame consisting of fixed first and third monomers in the triplets along the A-strand, i.e. these triplets must be of the form $\rangle R_1 N R_2 \rangle$.* Consider then, for example, the sequence $\langle R_2 R_1 R_1 \rangle$ in the hairpin loop; template-driven replication yields a hairpin with the loop triplet $\rangle R_1 R_2 R_2 \rangle$, which is the same as $\langle R_2 R_2 R_1 \rangle$, so that (+)-hairpins and (-)-hairpins differ in the centers of their loops, i.e. they satisfy requirements (1) to (4). (Interchanging R_1 and R_2 yields the remaining possibility that would work

equally well.) It is remarkable that these unique sequences follow by necessity from such simple arguments, the advantages of the hairpin conformation, the blessings of a code and a reading frame, and the insistence on parsimony.

The reading frame requirement also ensures firm and precise incorporation of the hairpins into the HA device. In later stages of evolution, when catalysts or even more complex devices that improve the attachment of hairpins to A-strands have come into existence, both stability and reading frame desiderata can be relaxed. An occupancy of the two end positions in a triplet by other monomers then gradually becomes possible. This permits the evolution of a code that at first can utilize two of these three positions for use in coding, and eventually all three of them. The sophisticated codes made possible in this way will allow many more than two a-monomers to be used in the synthesis of evermore sophisticated catalysts E_i .

We arbitrarily assign the (+) label to hairpins with the triplet $\langle R_2 R_2 R_1 \rangle$ in their loop, and the a_1 label to the a-monomer kind that is attached to their bottoms. The triplet $\langle R_2 R_1 R_1 \rangle$ is then associated with (–)-hairpins, and they carry the a-monomer kind a_2 (Figure 11). The first-selected such attachment might have arisen by chance or through chemical or other circumstances. These then are the beginnings of a code. An R_1 in the middle of an A-strand triplet, being complementary to an R_2 in the center of the hairpin loop, serves to code for a_1 . The ‘codon’ for a_1 is therefore $\rangle R_1 R_1 R_2 \rangle$, and $\langle R_2 R_2 R_1 \rangle$ (or $\rangle R_1 R_2 R_2 \rangle$) is its ‘anticodon’. Similarly, $\rangle R_1 R_2 R_2 \rangle$ is the codon for a_2 and $\langle R_2 R_1 R_1 \rangle$ (or $\rangle R_1 R_1 R_2 \rangle$) the anticodon, with directions indicated by the angular symbols (Figure 11).

The foregoing discussion is only about the logic of hairpin loop attachment to the assembler strand by triplets, not about how evolutionary gradients have led to a triplet-based code. The details of such an evolution are always variations on the same theme—small accidental steps, each of which provides an advantage that is selected for. The specific advantages of importance here concern the relationship between the monomer sequences on the assembler strand and a-oligomers, accidental at first but codified later because of the benefits of a memory. In short, the advent of a translation apparatus provides benefits from the very beginning, benefits that are slowly increased by the evolution of a more and more detailed code that in the end makes use of all positions in the loop triplet.

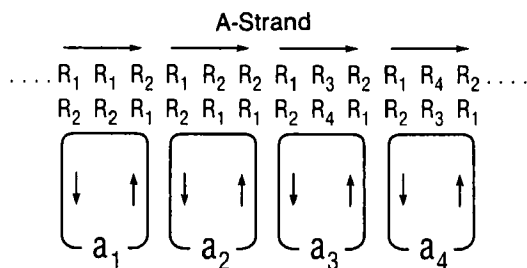


Figure 11 One-to-one relationship between R-triplets in the A-strand, R-triplets in the hairpin loops (top) and a-monomers attached to the hairpins (bottom)

3.11 A dead-end is reached at a certain level of complexity because nonsensical translation products accumulate. While a useful catalyst is represented by the message carried by a given A-strand, assign a (+) to it, the message carried by its (–) A-strand replica is nonsensical

The complementary replicas of the A-strands involved in E_i production also furnish oligomers, but the chance that these oligomers have any merit aside from contributing to agglutination is essentially zero. As long as only a few HAE_i devices participate in the functional cooperative that exists in a given region, the presence of these (–)-strands presents no problem; they assist in agglutination, which is assumed to be a general property of oligomers of arbitrary monomer sequences. (We have arbitrarily placed a (+) label with the A-strands furnishing useful catalysts E_i and a (–) label with their replicas.) However, as the number of HAE_i devices grows, so will the number of 'plain' HA devices associated with such 'nonsensical' (–)-strands. Eventually a point is reached in which, for example, an increase in agglutination is detrimental; there are too many plain HA devices that produce too many oligomers that have no special purpose and the catalysts E_i are swamped by them.

This is a major disaster, a very serious matter that needs to be resolved if there is to be any further evolution. How in the world can we conceive of a way (in our model) to avoid the accumulation of such waste material that dilutes the effectiveness of any catalyst E_i and the synthesis of which consumes valuable a-monomers? Any further progress is hampered and even terminated unless this new dead end is overcome.

A way out would be an arrangement in which the (–)-strands no longer function as A-strands; as a consequence, HA devices based on (–)-assembler strands would no longer be made and translation of (–)-assembler strands into oligomers would be eliminated. At an earlier stage the introduction of a new variety of monomers diminished the dependence on compartmentation (see Section 3.6). We now follow the same path and assume the appearance of another new variety of monomers that permits the separation of the two tasks, reproduction and translation of A-strands.

3.12 The dead-end is overcome by separating the two tasks, reproduction and translation of strands. This is achieved by the appearance of a third monomer variety, D, that is a modified version of the original monomers of the first variety, already designated by R. A modified E_1 -catalyst, E_1' , drives the replication into D-strands of assembler-strands and hairpins, made of R-monomers, while a similar catalyst, E_1'' , governs the reverse replication, of D into R-strands. We postulate that D-strands are unable to serve as assembler strands, their only purpose being to function as templates for the replication of R-strands. Since (–) R-assembler strands are no longer needed as templates and their presence is deleterious, they disappear; (–) HA-devices and nonsensical a-oligomers are eliminated in this way

We assume that a new variety of monomers makes its appearance in the region considered, monomers that differ only slightly from those of the first variety.

Perhaps they have arrived by diffusion from the outside, or perhaps there were small chemical changes in the monomers of the first variety, engendered by some new catalyst E_i . For reasons to become clear in Section 4, the biological implementation of the model under discussion, we have called monomers of the first variety R_1 , R_2 , etc. and now call those of the new or third variety D_1 , D_2 , etc. (We shall relate them later to RNA and DNA.)

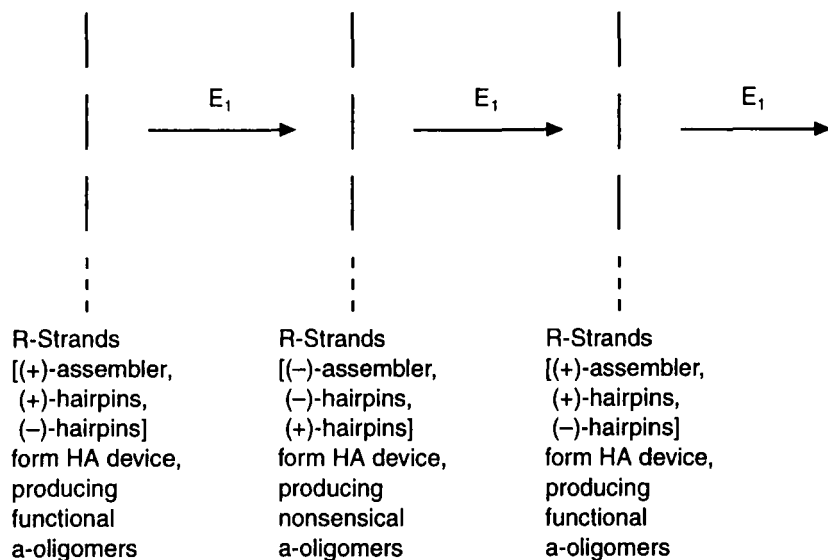
We have supposed R_1 and R_2 to be complementary to each other and now assume similar complementarity between D_1 and D_2 , D_1 and R_2 , and D_2 and R_1 , implying weak, antiparallel, lateral bonding between all of these pairs of monomers, such as by van der Waals forces or hydrogen bonds. We also postulate that replication shifts so that in the end R-strands serve as templates for D-strands rather than for new R-strands, a process driven by a catalyst E_1' that is a variant of the original replication catalyst E_1 . The reverse replication, in which D-strands serve as templates for R-strands, is assumed to be governed by another variant E_1'' of E_1 . Finally, we assume that the small difference between R-monomers and D-monomers has the consequence that D-strands are no longer capable of functioning as assembler strands and of thus participating in the synthesis of a-oligomers. The production of nonsensical a-oligomers is avoided in this way (Figures 12a and 12b).

But how can such an enormous changeover take place? Slightly modified versions of the catalyst E_1 always appear, the result of rare but ever-present errors in the monomer sequences of A-strands. The overwhelming majority of them are not beneficial, but some of them improve the adaptation of the system to the environment and then win out over the older variants. This is what we assume for the new catalysts E_1' and E_1'' —that they improve the efficiency of replication, and that they therefore replace the older catalyst E_1 .

(a) The ratio of (–)-assembler strands to (+)-assembler strands decreases and the production of nonsensical a-oligomers is diminished

At the beginning, useful HAE_i devices and potentially harmful HA devices are still produced in equal numbers, respectively being assembled under the control of (+)-assembler strands and (–)-assembler strands that are incorporated into them. Now, in the different pores of a region there are, of course, fluctuations in the ratio of (+)-assembler strands to (–)-assembler strands. Consider, then, a pore in which, by chance, there are substantially more (+)-assembler strands than (–)-assembler strands and that the ratio of the corresponding D-strands is skewed in the same way. The cooperative in this pore is at an advantage over those in neighboring pores—it is less gummed up by messy a-oligomers furnished by its (–)-assembler strands and multiplies faster than rival cooperatives in the neighborhood. If it remains intact for a sufficient number of cycles it will be selected and will populate all pores in the region. This is the first step, and many similar steps in the same direction follow, diminishing the ratio of (–)-assembler

(a) Situation before changeover



(b) Situation after complete changeover

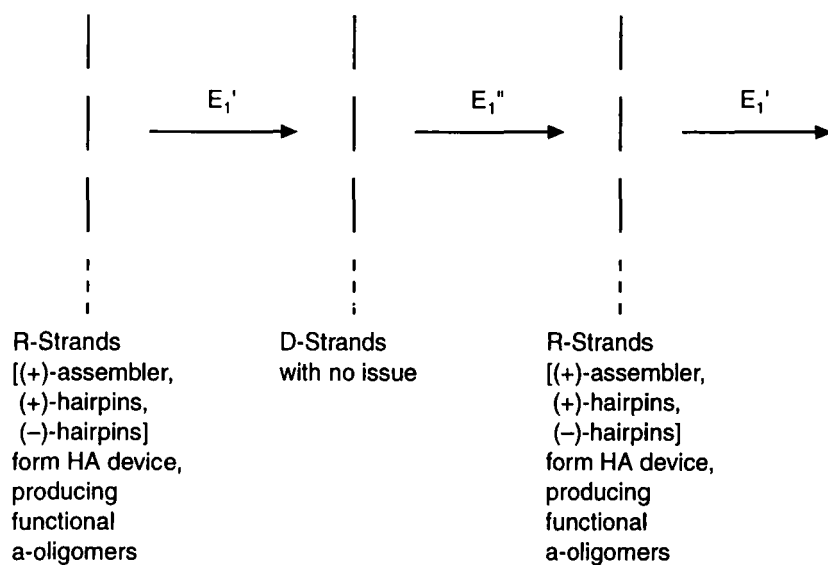


Figure 12 (a) Replication and assemblage of HA devices, both under the control of R-strands. (b) Control of replication assumed by D-strands, eliminating the need for (-)-R-strands and thus avoiding pollution by nonsensical a-oligomers

strands to (+)-assembler strands, each step thus reducing the concentration of nonsensical oligomers, thereby improving the new cooperative system so that it will become selected.

Eventually the production of (–)-assembler strands ceases altogether and the creation of nonsensical a-oligomers is eliminated in this way. The *translation* of blueprints into a-oligomers has become separated completely from the *reproduction* of blueprints; the message for translation is carried by (+)-assembler strands and that for reproduction by their D-templates, themselves replicated on top of (+)-assembler strands. Replication of (+)-hairpins and (–)-hairpins is also accomplished on D-templates and vice versa.

(b) New monomers a_3 and a_4 are needed as the sophistication of catalysts E_i increases, but this also requires new R-monomers and D-monomers

As the system develops, the benefits conferred upon it by the emergence of new catalysts, such as an increase in the precision of replication, lead to an ever-growing complexity and sophistication. Consider, for example, the two relatively simple catalysts E_1' and E_1'' . It may well be that oligomers consisting just of a_1 -monomers and a_2 -monomers are of sufficient complexity to fulfill the tasks required of these catalysts, but it is also possible that more kinds of a-monomers, say a_3 and a_4 , are needed. Certainly an advantage for the inclusion of more kinds of a-monomers will arise at some point. From the foregoing it is clear that the coding for such monomers also requires a pair of new, complementary R-monomers and D-monomers. Thus, maintaining the $\rangle R_1 N R_2 \rangle$ reading frame, the a-monomers a_3 and a_4 could be coded for by the codons $\rangle R_1 R_3 R_2 \rangle$ and $\rangle R_1 R_4 R_2 \rangle$, the respective anticodons being $\langle R_2 R_4 R_1 \langle$ and $\langle R_2 R_3 R_1 \langle$ (or $\rangle R_1 R_4 R_2 \rangle$ and $\rangle R_1 R_3 R_2 \rangle$) (Figure 11). The reproduction of R-strands containing R_3 and R_4 requires, of course, the existence of corresponding D_3 -monomers and D_4 -monomers and of appropriate D-strands containing them.

(c) How do new kinds of monomers appear on the scene when needed?

It may be that suitable R-monomers R_3 and R_4 had been present for some time or that their arrival sparks a new evolution along an existing evolutionary gradient. Similar possibilities exist for the a-monomers a_3 and a_4 that either were already attached to R_3 or R_4 monomers or could easily be joined later to hairpins consisting of old and new R-monomers. The D-monomers corresponding to R_3 and R_4 might have been derived by variation of their R-counterparts with the help of some catalyst E_i , perhaps the same catalyst that earlier facilitated the modification of some R_1 -monomers and R_2 -monomers to become D_1 -monomers and D_2 -monomers. Occasional replication errors make R_3 and R_4 part of R-strands every so often, but they will

become incorporated into the system (and the code) only at a point in time in which such incorporation becomes beneficial and is therefore selected for. This is the crucial time. As stated, the new monomers did not need to arrive or become available at that point, they could have been around long before, but would become part of the system by selection only once they were conferring an evolutionary benefit upon it.

While by the separation of reproduction and translation of strands a very serious situation has been resolved, another one is waiting just around the corner, i.e. the occasional loss of strands that contain the blueprints of a partner in a cooperative, a loss that may be lethal to the system concerned. Its resolution requires the emergence of yet another descendant of catalyst E_1 and will be attended to next.

3.13 The blueprints of the components of a cooperative, contained in D-strands, are united into one 'superstrand' that contains the blueprint of the entire cooperative in one piece. The loss by dispersion of blueprint portions is thereby prevented. A third descendant E_1'' of E_1 catalyzes the replication of D'-superstrands on D-superstrand templates and vice versa

A serious difficulty arises very soon as cooperatives grow in complexity. With an ever-increasing number of cooperating components it becomes increasingly probable that a D-strand, carrying the message for the reproduction of an assembler strand, is lost by dissipation. It may, for example, diffuse away or disintegrate into its constituents. Loss of this D-strand would mean that the complementary assembler strand could no longer be reproduced and that the corresponding HAE_i device would no longer be built. The absence of any partner in a cooperative would, of course, be most detrimental to its functioning.

One way to get around this problem would be to tie D-strands carrying the blueprints of a cooperative together into longer D-superstrands that carry replicas of several R-strands [(+)-assembler strands as well as (+)-hairpins and (-)-hairpins], perhaps connected by some kind of bridge monomers, symbolized by X. This is the idea we propose to pursue in our model (Figure 13a).

(a) Emergence of a D-superstrand that serves as a template for many R-strands

We attribute to the catalyst E_1' , or to one of its descendants, an additional property. (E_1' catalyzes the interlinking of D-monomers on top of an R-strand template.) We assume that occasionally, under the guidance of E_1' , a longer D-superstrand arises that represents more than one R-strand. That is, every so often a replicate of a certain R-strand is made to which then are appended, under the control of E_1' , some 'meaningless' or 'neutral' monomers, denoted by X. To this appendix then is added the replica of a second R-strand and so on. Thus every so often a D-strand arises consisting of several R-replicas linked by XXXX bridges (Figure 13a).

Of course, such D-superstrands must then be capable of serving as templates for the R-strands they represent, assisted by what is perhaps a variant of E_1'' . It is not unreasonable to assume in our model that a descendant of E_1'' develops the capability of recognizing the newly produced XXXX sections on

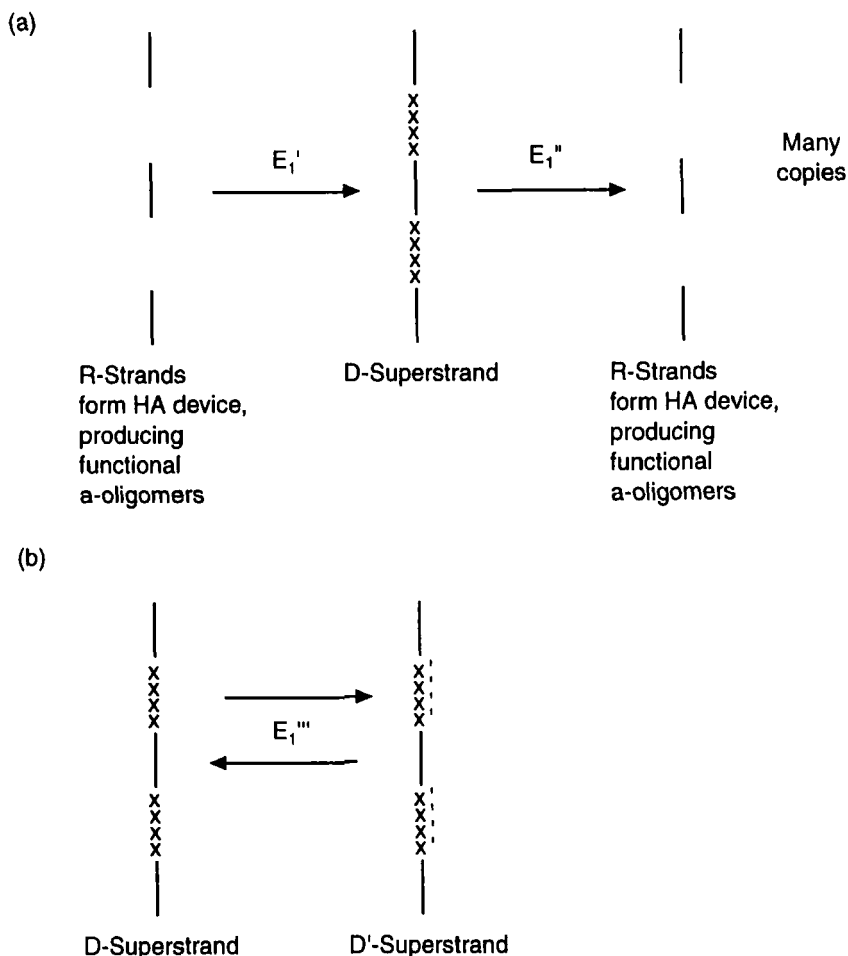


Figure 13 (a) The catalyst E_1' facilitates the formation of a D-superstrand from R-strands (a rare event) with interleaved pieces of X-monomers. E_1'' in turn guides the replication of individual R-strands, a D-superstrand serving as the template. For simplicity the replication of only three R-strands is shown. (b) Replication of a D'-superstrand on a D-superstrand template and vice versa, driven by the catalyst E_1'''

the new D-superstrand as signals for starting and stopping the production of R-strands and using the sections between the XXXX bridges as templates*. The evolutionary benefits are that transcription of the D-superstrand into its component R-replicas by E_1'' is more efficient than the piecemeal transcription of the earlier D-strands and minimizes loss of the R-blueprint pages.

*The fact that in the RNA world even short monomer progressions may be recognized by complementary sequences is used, for example, in isolating strands with specific monomer successions from a pool of random sequences [20].

(b) Replication of D-superstrands

The difficulty that arises at this point is the lack of a mechanism by which the new D-superstrand is replicated directly, because its genesis is a unique event that is expected to occur only rarely. What is needed is another catalyst E_1''' , a descendant of E_1 (or of E_1' or E_1''), that facilitates replication in terms of D-monomers alone; that is, it promotes the making of D-replicas of our D-superstrands, call them D'-superstrands, and of D-replicas of these D'-superstrands. Moreover, in order to guarantee replication of entire D-superstrands and D'-superstrands, it is necessary that such (antiparallel) replication begin at the very end of a parent strand. E_1''' must therefore contain a recognition site for strand-ends, a demand the fulfillment of which is not unrealistic at this evolutionary stage. (The problem of incomplete strand replication probably exists in earlier stages as well, but becomes crucial for the long D-superstrands.)

Moreover, it is necessary to restrict replication from the D-world to the R-world in such a way that only D-superstrand sections are replicated into R-strands, but not D'-superstrand sections. Since the complementary replicas of the XXXX bridges are X'X'X'X' bridges, we therefore postulate that these latter cannot act as recognition sites for E_1'' . As a consequence, E_1'' cannot transcribe D'-superstrands into R-strands. No avenue to nonsensical a-oligomers will therefore open up. The emerging new mode of replication is more efficient than the older one; it replaces it and its scope slowly increases. The catalysts E_1'' and E_1''' evolve to accommodate yet longer D-superstrands that emerge and eventually represent the blueprints of a cooperative of about five assembler strands, say, plus a larger number of (+)-hairpin and (-)-hairpin strands, presumably containing several replicas of all these R-strands.

At the outset each of the evolving replication modes is only a small operation, proceeding in a niche, so to speak, of the then dominant replication protocol, but its improved efficiency is such that eventually it takes over. As D-superstrands become longer and the cooperatives larger, the benefits of D-superstrands and their attendant catalysts E_1'' and E_1''' are such that all of replication soon rests on this mode alone.

(c) Replication, transcription and translation separated

Undoubtedly all the steps described in Section 3.12 and in the present section do not occur one after the other, in the orderly sequence of our account. The superstrand scheme just described has its start at about the time that D-monomers and modified catalysts enter the picture, slowly taking over the task of replication, as discussed in these sections. At the same time a more intricate code begins to develop from the primitive beginnings detailed in Section 3.10. The sequence is always the same: random errors in this or that direction provide the variation needed for selection, based on improvements in survival conditions—a slow, tedious, but inescapable mechanism.

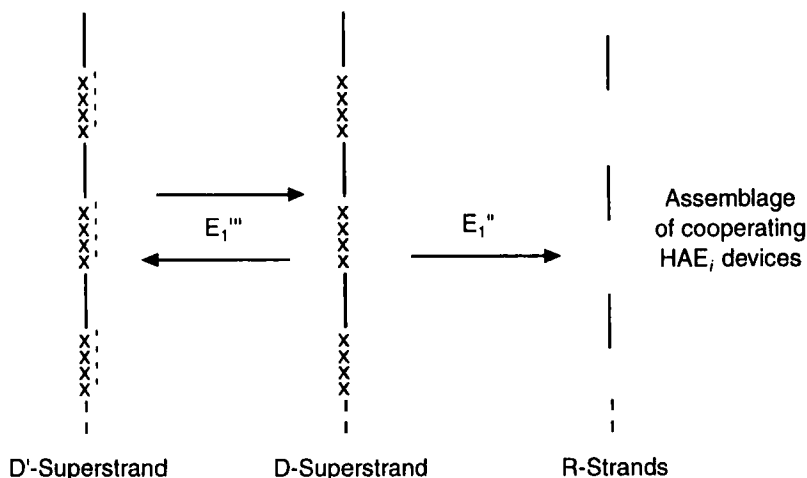


Figure 14 Relationship between catalysts E_1''' and E_1'' , D-superstrands and D'-superstrands, R-strands and cooperatives of HAE_i devices

This then is the 'final' picture. The evolving descendant of E_1'' 'uses' the XXXX sections on the long D-superstrand template as recognition segments and 'utilizes' the sections between these bridges as templates for the R-strands they represent (Figure 13a). The collection of assembler strands, replicated from a D-superstrand, operates to aggregate appropriate (+)-hairpins and (–)-hairpins (also replicated from a D-superstrand) into the HAE_i devices that constitute the cooperative under consideration. The catalyst E_1''' drives the synthesis of D-superstrands on top of D'-superstrands and vice versa (Figure 13b), E_1'' the creation of (+)-assembler strands and (+)-hairpins and (–)-hairpins on sections of D-superstrands, while E_1' disappears because it is no longer needed (Figure 14). The D-superstrands thus carry the blueprint pages of a cooperative, while the translation of partial 'photocopies' of this master blueprint, of the assembler strands and (+)-hairpins and (–)-hairpins, into a-oligomers is still done by HAE_i devices.

At the outset each of the evolving replication modes is only a small operation, proceeding in the shadow of the then dominant replication protocol, but its improved efficiency is such that eventually it takes over. As D-superstrands become longer and the cooperatives larger, the benefits of D-superstrands and their attendant catalysts E_1'' and E_1''' are such that all of replication soon rests on this mode alone.

We imagine all of these steps as occurring at first in a small, insignificant way, as minor and rare sidesteps of the main replication routine. At each step, once preliminary 'blunderings' have been overcome and a reasonable efficiency has evolved, the advantages of the new replication protocol become large enough for it to take over. In this way reproduction and transcription eventually become completely separated. A further necessity is that E_1''' evolves into a copying

device for D-superstrands, a device that is different and separate from the HAE_i devices that translate assembler strands into a-oligomers.

(d) Supramolecular engineering ideas result in systems that resemble biomachinery

Note that the device that emerges resembles the translation machinery of molecular biology. This is remarkable, even though its conception and the assumptions specifying the lock-and-key and the programmed-environmental-change principles have been influenced to some extent by the givens of molecular biology and by planetary aspects. Many of the underlying ideas were not, but rather were the result of the necessity of overcoming seemingly inescapable dead ends. Perhaps the most noteworthy of these postulated events was the need for an error filter, the picket-fence-like aggregate (in which the lock-and-key principle played a major role), which then almost automatically evolved into a translation device. Other important ideas were the necessity of coping with the confusion brought about by the rank proliferation of nonsensical a-oligomers and the disastrous loss of pieces of blueprint during multiplication. Examples of traits that seem to be an inescapable consequence of these general ideas are the antiparallel direction of strand replication, the triplet attachment of hairpins to the A-strand in HA devices and the abandonment of (–)-assembler strands through the introduction of D-monomers and D-superstrands.

The ideas presented thus provide a rationale of how the translation machinery of molecular biology might have evolved, each proposed step being compatible with the laws of physics and chemistry and small enough to be feasible. Our next task is to challenge this model with the known facts of molecular biology, and especially to consider whether the machinery described can provide insight into how the present-day genetic code might have arisen. This is done in Section 4.5.

3.14 Further Steps: Synthesis of ‘Food’ Monomers, Cell Envelopes, etc.

A most important evolutionary step is the emergence of oligomers that can act as catalysts in the production of suitable monomer ‘food’, e.g. by modifying monomers that are already present in abundance or by using them as building blocks for larger monomers, perhaps by building on an early mode of supply of R-monomers, a-monomers and D-monomers, substituting, step by step, each reaction in a prebiotic production cycle by an a-oligomer-catalyzed reaction that is more beneficial. Regions with an insufficient supply of actual ‘food’ monomers but sufficient monomers that are suitable as raw material now become accessible. A significant evolutionary gradient points towards systems with metabolism.

Of the greatest importance is the liberation from specific surroundings, a liberation that must not provide complete isolation, however. It requires the

appearance of oligomers that go beyond assisting confinement by agglutination, oligomers that promote the formation of actual envelopes which, on the one hand, contain entire cooperatives and, on the other hand, still permit an interchange of monomers or other raw materials with the outside world, perhaps through some channels that allow control over 'cell' influx and output. Such a 'cell membrane' keeps all of the collaborating parts of a cooperative together and they become integrated entities. Its advent signals the existence of organisms. It liberates the functional units in its interior, the organism they represent, from the necessity of an environment that provides compartmentation. The colonization of vast new regions is now possible. Metabolism can become increasingly sophisticated and the genetic machinery can evolve into a complex entity.

As the contents and the sizes of such primitive cells grow, they include ever more catalysts E_i and larger D-superstrands. A critical point will be reached at which such cells break into two, each of which needs, for survival, to contain at least one D-superstrand and a full complement of HAE_{*i*} devices etc. Life can now explode into full bloom, it is liberated from any need for external confinement, it can diverge, thousands or millions of new environments can be populated.

At this point it is useful to recall some of the major steps that this evolution is based on. The beginning is a relatively short strand of monomers that can serve as a template for replication. The strands are lengthened by strand doublings, but this must come to an end because replication errors increase with strand length. The impasse is overcome when strands, in a hairpin conformation, organize themselves into picket-fence-like aggregates that serve as error filters. These aggregates evolve into devices that catalyze the synthesis of a-oligomers from a new variety of monomers, oligomers that extend external compartmentation and, in rare instances, provide other important survival assists. The transformation of the aggregates into translation devices establishes a three-to-one correspondence between R-monomers and a-monomers. This metamorphosis is crucially assisted by one or several catalysts E_1 that improve replication to the extent that R-assembler strand 'blueprints' survive and can establish themselves. The problems caused by the undesirable accumulation of nonsensical a-oligomers and the occasional loss of blueprint pages are overcome by the advent of D-monomers and suitable catalysts E_1 that permit the separation of replication and translation.

It is important to recognize that the advent of integrated functional units, enclosed in an envelope, had to be waited for before the explosive emergence of 'life' and its divergence in many different directions became possible. Prior to this sweeping change any evolutionary advance required the presence of appropriate monomers and the existence of specific spatial and temporal conditions, not necessarily in the same location for each breakthrough. Regions that did not provide these necessities were condemned to stagnation, but the integration just described had the consequence of a drastic reduction in all these conditions.

It must also be emphasized that all this is not the result of any grand design. It is purely a matter of chance variations and a selection that is provided by an environment of suitable diversity of compartmentation, the whole being driven by intricate cycles of temperature variations. Among ideas that are pregnant with implications, that of an error filter is surely a most unexpected application of the lock-and-key principle.

Finally, the processes considered here must clearly be distinguished from processes that lead to complex structures by self-organization of molecules. It is essential to note that the processes leading to lifelike systems are induced by the spatial and temporal structure of the environment.

4. REALIZATION IN THE ACTUAL WORLD: THE ORIGIN OF LIFE

Section 4 is about the real world, a comparison of the model arrived at in Section 3 with the facts of molecular biology as well as with the results of some experiments and computer simulations.

4.1 Identification of Components of the Model with Entities of Molecular Biology

The analogies that emerge from a comparison of the model hitherto discussed with what might be the precursors of characteristic elements of molecular biology are not altogether surprising, but nevertheless still remarkable. A comparison between the two leads to the following correspondences. The precursor of the DNA double helix is a D-superstrand intertwined with a D'-superstrand. Moreover, we equate the R-monomers R_1 through R_4 with the nucleotides cytidine (C), guanosine (G), uridine (U) and adenosine (A), respectively, the a-monomers with amino acids and the a-oligomers with (simple) proteins. The directionality indicated by the chemical 5'-3' designation we identify with the earlier angular symbol ($\circ$).

The assembler strand (or A-strand), an integral part of HA devices, plays the role of an early precursor of messenger RNA (mRNA), and the hairpins function as primitive transfer RNA (tRNA). Recall that these hairpins constitute the main body of the HA devices, in which triplets of nucleotides in the hairpin loops are attached to triplets of complementary nucleotides in the assembler strand. The ribosome, the translation machinery of the modern cell, may be looked upon as a highly developed derivative of the HA device, the primitive machinery that catalyzes the production of a-oligomers and mediates the correspondence between the codon sequence along A-strands and the order of a_k -monomers along a-oligomers. The paramount role of ribosomal RNA (rRNA) in present-day ribosome points to an ancestral ribosome precursor [14] that did not involve a-oligomers (i.e. proteins). In other words, there must have

existed primitive early machinery of the kind described by HA devices that was slowly modified by increasingly sophisticated rRNAs and proteins.

The catalysts E_i are primitive enzymes. Specifically, E_1''' and E_1'' would be early precursors of DNA-dependent DNA polymerase and DNA-dependent RNA polymerase, respectively, the enzymes that drive the replication of DNA and its transcription into RNA. The early catalyst that promotes the joining of amino acids to hairpin legs (we have called it E_2 in Section 3.7) would be the ancestor of synthetases, the enzymes that assist this attachment today. DNA consists of sections of which some are used as templates in transcription while others are not. We hypothesize that the XXXX portions of D-superstrands are precursors of the omitted portions. Present-day proteins consist of linear arrays of selections among 20 kinds of amino acids. Their sequence is related to the sequence of nucleotide triplets in the messenger RNA responsible for their synthesis, according to an elaborate code. The question arises of how this code might have evolved, and it is an important test of the model to see whether it permits a plausible reconstruction of this history. We address this problem in Section 4.5.

4.2 Spatial and Temporal Environmental Requirements

An important early spatial requirement is compartmentation of a graduated nature, with details as described in Section 2.4, in order to provide an evolutionary gradient. This condition appears to be easily satisfied considering the rocky nature of our planet, e.g. by many clays and meteoritic material that was fairly abundant in early millennia.

The complexity of the cyclic environmental change program needed is, however, considerable. The intermediate steps involved, e.g. in the reproduction of HA devices prior to the advent of D-superstrands, are roughly as follows. First, the temperature must be raised to cause the disintegration of the HA devices into their A-strand and hairpin components, the latter of which, in addition, will have to unfurl before they can be used as templates to assist in the synthesis of complementary strands, a step that requires lower temperatures. (Periodic pH changes, affecting the equilibrium between single and double strands, might also play a role.) Following strand multiplication, additional delicately balanced cooling and heating sequences are required to get the strands to fold into hairpins again (with the important exception of assembler strands) and to reconstitute devices from the hairpins by assembler-strand-assisted aggregation. Cyclic changes in the concentrations of ions entering the pores by diffusion may play a role similar to that of temperature changes. (The stability of double strands of nucleotides is known to be very sensitive to the concentration of cations such as Na^+ and K^+ in the 0–100 mM range and Mg^{2+} and Ca^{2+} in the 0–1 mM range.) It should, however, be realized that with increasing complexity of the HAE_i devices, no corresponding increase in the complexity of the temperature and other programs need be required. This is

because emerging enzymes improve the replication process and increasingly assume control of replication.

4.3 Availability of Suitable Monomers and Attempts at the Realization of Replicating Strands

(a) Amino acids and nucleotides

The question arises of whether nucleotides and amino acids existed in sufficient concentrations under prebiotic conditions in or near the location where the critical events took place, i.e. the genesis of the first strand that could serve as a template and its template-driven replication. The prebiotic conditions required in the context of our paradigm consist in the existence of a very particular, minute location in which all required conditions are met. While the syntheses of the amino acids Gly, Ala and Asp and the nucleic acid base adenine need simple conditions only and no special programs of environmental change, the syntheses of other amino acids and other nucleic acid bases do [9–11]. As is detailed in the footnotes to Table 1 in Section 4.5 some very particular conditions are required that differ from case to case and often involve sequential steps in anhydrous and aqueous media or even on surfaces, as well as specific conditions for driving accumulation processes. A notable possibility is the production of relevant compounds in a redox cycle in which CO_2 is reduced by H_2S on an FeS surface with the formation of pyrite (FeS_2) and the synthesis of successively larger organic molecules. This is believed to have been a very primitive precursor to the citric acid cycle [16].

These considerations must be included in an exploration of the possible available starting materials and in a search for reasonable syntheses. Thus for many years a main difficulty in understanding how nucleotides might have existed in primeval times was the experimental difficulty of obtaining ribose. However, Eschenmoser *et al.* [21] recently obtained ribose 2,4-diphosphate in 17% yield in an alkaline solution of glycolaldehyde phosphate and formaldehyde. The prevailing optimism that nucleotides could have been available under prebiotic conditions appears therefore to be justified.

(b) Template-directed polymerization

The next step, the condensation into strands of nucleotides under prebiotic conditions using polyphosphates, organic condensing agents or heat, has also been studied [9, 22], as well as polynucleotide formation by template-directed synthesis of activated nucleotides. It has been shown that an efficient incorporation of complementary bases is achieved in the presence of Zn^{2+} [23]. This is of interest because many DNA and RNA polymerases are known to be zinc metalloenzymes. Moreover, in the absence of metal ions, activated G-nucleotides that polymerize on a poly(C) template can give rise to over 90%

of 3',5'-oligo-G-nucleotides, with chain lengths of up to 40 bases [24]. Template-directed polymerization using a random poly(GC) template has also been achieved [25], and an autocatalytic system in which a hexanucleotide is produced by the linking of trinucleotides has been reported [26]. Orgel [27] and Maizels and Weiner [27] have considered the possibility that RNAs endowed with aminoacyl groups at the 3'-OH terminal might be better substrates for replication than bare RNAs. In aminoacylated RNA this would prevent the initiation of replication at internal positions of the template. Such a scenario would entail an attachment of amino acids to RNA strands prior to their serving as templates. This possibility would have facilitated the emergence of HAE_i devices and, assuming linkage of Gly to (+)-strands and Ala to (–)-strands prior to hairpin formation, it would also have hastened the emergence of a primitive translation device.

(c) Biological suitability parallels ease of formation

A surprising recent development is that quite complex biomolecules can be obtained under prebiotically reasonable conditions. A particularly impressive example concerns uroporphyrin III, an important intermediary in the biosynthesis of vitamin B₁₂. Conditions can be arranged so that this coenzyme is a highly specific reaction product in a prebiotic synthesis. The ring contraction required in B₁₂ synthesis takes place spontaneously in metal complexes of model compounds [28]. Nucleic acid bases, ribose and the amino acids Gly, Ala and Asp, as mentioned above, also show this remarkable coincidence of biological functionality and ease of formation.

An important approach in this regard is what Eschenmoser has termed 'selfinstruction', the creation of complex structures from small molecules by a small number of changes in external conditions, each step being guided by lock-and-key situations that provide precise positioning of the molecules created in the preceding step. Of the many possible reactions essentially only one is therefore realized in each step. In this approach the synthesis of structurally quite complex molecules turns out to be remarkably simple. Thus, as described in Section 4.4 the lock-and-key situation that so greatly facilitates template-driven strand replication provides and enforces the correct chirality. All these possibilities should encourage further attempts at the synthesis of nucleotides and the realization of nonenzymatic replication, keeping in mind the broad variety of possible prebiotic conditions.

(d) Contrast with primeval soup paradigm

In view of all this it is quite reasonable to suppose that the environmental conditions and the sequence of programs of environmental changes triggering the emergence of life could have been realized at some point in time in a very special, tiny region of this huge planet. This is in sharp contrast to the primeval soup paradigm, which requires the needed ingredients to exist in solution, at a reasonable concentration, in whatever large region that is taken up by this solution. It must also be emphasized that this basic assumption underlying our

paradigm, i.e. the existence of a minute suitable location, makes modeling of the origin of life quite insensitive to chemical theories concerning the overall conditions on prebiotic earth. The particular tiny location needed can exist under a broad variety of such global conditions.

(e) Overall prebiotic conditions of minor importance

While the required programs and conditions are indeed complex and intricate, it is specific local situations that are important and these are myriad. Our knowledge of the early history of the planet leaves little doubt that the basic necessities were available, such as liquid water and solids capable of providing microcompartmentation. Thus the oldest-known rocks are sedimentary, attesting to the presence of liquid water and continents. There also existed solar, volcanic and other energy sources [9, 11, 15, 16], and cyclic changes were available as well, such as the night-and-day cycle and cycles associated with shadow-casting rocks. Our planet is vast and the millions of years required were also available, until by chance all the required conditions were met.

4.4 Chirality of Nucleotides and its Evolutionary Benefit

Nucleotides are chiral because ribose, one of their components, is chiral, and nucleotides with D-ribose have the opposite chirality from those containing the L-isomer. It turns out that strands consisting of nucleotides of the same chirality, either D or L, are more appropriate for precise template-assisted replication than are strands containing monomers of mixed chirality. With monomers of just one chirality, the template and growing daughter strand can intertwine into a double helix, a situation that provides special stabilization for each new monomer that arrives at the location at which it will be attached to the monomer that preceded it (lock-and-key principle, selfinstruction). Moreover, monomers of incorrect chirality are rejected in this buildup of new daughter strands: they do not fit into the niche at the growth position created by the template strand and the already-existing portion of the daughter strand. No such propitious relationship exists if the template contains monomers of mixed chirality. The replication precision and efficiency of strands containing just D-ribose or just L-ribose has the consequence that after a short time only these two types of strand can survive.

Another most significant aspect of strand chirality is that hairpins consisting of nucleotides with D-ribose have a helical twist that is opposite to that of hairpins made up of nucleotides with L-ribose. An important implication of this fact is that picket-fence-like hairpin aggregates must consist of hairpin strands of the same chirality; hairpins of opposite twist just do not interlock properly (lock-and-key principle).

It is most likely that D-ribose and L-ribose were present in equal amounts on prebiotic earth, and this would therefore be true for nucleotides as well.

In this view the fact that present-day RNA contains the D-isomer rather than the L-isomer of ribose is the result of a 'frozen accident': the HAE_i device that won out over all others contained monomers with D-ribose. Strands containing just L-ribose might still arise later on or in other places, but they would have no effect on evolution and would again disappear.

Table 1 Sequence of amino acids in the order in which they were coded for. Columns 1 and 2 are ordered by availability (footnotes *a* and *b*) and additional criteria (footnotes *c*–*f*). Note that the situations described under footnotes *b* and *c* below fit the present paradigm very well, while they are unsuited to the primeval soup paradigm. A solidus between two amino acids indicates that in the early code no distinction was made between them. In column 3 are shown the amino acids added in the late stages, in decreasing average order of occurrence in present-day proteins [40], implying a decrease in general importance. It is noteworthy that other arguments independently suggest the same sequence in which these amino acids were coded for. Prebiotically available Phe and Tyr (footnote *b*) came first; His, available under special conditions (footnote *g*), came next; the S-containing Cys and Met (footnote *h*) followed; and Trp, with no known abiotic synthesis (footnote *i*), came last (see also footnote *j*)

Gly ^{<i>a,f</i>}	Lys ^{<i>b</i>}	Phe ^{<i>b,k</i>}
Ala ^{<i>a,f</i>}	Arg ^{<i>b</i>}	Tyr ^{<i>b,k</i>}
Asp/Glu ^{<i>a,f,e</i>}	Gln/Asn ^{<i>c,e</i>}	His ^{<i>g,k</i>}
Val ^{<i>a,f</i>}	Pro ^{<i>a,d</i>}	Cys ^{<i>h,k</i>}
Leu/Ile ^{<i>a,e</i>}		Met ^{<i>h,j,k</i>}
Ser ^{<i>a</i>}		Trp ^{<i>i,j,k</i>}
Thr ^{<i>a</i>}		

^a Thermal synthesis involving CH₄, NH₃ and water on quartz sand at 950 °C [41]. The concentrations of the resulting amino acids decrease from the top down in the table. Similar data from various attempts to simulate prebiotic conditions and from analyses of the Murchison meteorite are compiled elsewhere [42].

^b Thermal synthesis under nonaqueous conditions [42]. The concentrations of the resulting amino acids decrease from the top down.

^c Easily obtained from Glu and Asp but unstable against UV radiation or heat [42], thus absent in a prebiotic soup under steady-state conditions but present in a special location in a distinctive environment in which amino acids continuously emerge. Possibly formed from Glu and Asp by amidation after binding to tRNA [38, 42].

^d The abundance of Pro would place it further up in the table by five amino acids. However, its steric properties strongly affect the structure of an oligopeptide containing it and we assume that this fact restricts its being utilized.

^e Very similar amino acids; same codon at first, separate codons later.

^f Metabolic pathways for generating these amino acids were established early, providing independence from abiotic sources—a fact that supports their early introduction into the code [36].

^g Synthesis difficult and requiring special conditions [43], thus assumed to occur at a late stage.

^h Sulfur-containing compounds are believed not to be incorporated into early proteins [39].

ⁱ No reasonable abiotic synthesis is known; the occurrence of Trp is assumed to depend on the presence of a sufficiently sophisticated metabolism.

^j One of two amino acids that are coded for by a single codon, suggesting a late change of these codons from Ile and stop to Met and Trp. A late use of these codons for Met and Trp is supported by studies indicating that the tRNAs for Met and Trp were derived from other tRNAs by mutation [44]. Differences between mitochondrial codes and the universal code indicate a relative lack of code conservation, e.g. supporting a late differentiation of Met from Ile and of Trp from stop. This variability in code is related to the simplicity of mitochondria putting them under less selective stress than their hosts. For example, in all nonplant mitochondria except those of coelenterates and lower organisms on the one hand and echinoderms on the other hand, AUA is used for Met instead of for Ile [38]. Coelenterates branched off the main line of the phylogenetic tree early, echinoderms later. The shift of AUA from Ile to Met took place after the coelenterates had branched off and was conserved in all higher organisms, except that in echinoderms there was a reversal from Met to Ile.

^k Amino acids that occur in present-day proteins with an averaged frequency of 4% or less, in decreasing order. The present-day occurrences in columns 1 and 2 are above 4% [40].

4.5 Rationale for the Modern Genetic Code

The next question we wish to address at length is how the present-day code might have arisen. (For quite different approaches to the code problem see [29–39].)

We assume that there existed a crucial location that offered the all-important compartmentation by porous rock and the needed cyclical changes driving the replication of strands and aggregates. This location was inundated by amino acids and by energy-rich nucleotides G, C, A and U that were continuously synthesized in its surroundings from compounds exposed to specifically changing anhydrous as well as aqueous conditions (some details are presented in Section 4.3 and in the footnotes to Table 1. In the earliest phase G and C were most important because they base pair with three rather than just two hydrogen bonds, and the most prevalent amino acids were Ala and Gly.

As outlined in Sections 3.10 and 4.1, we assume that in the original stage of coding evolution there were two codons available, 5'-GCC-3' and 5'-GGC-3'. In later stages, as the evolving ribosome began to stabilize the attachment of hairpins to the assembler strand, position 1 was placed into coding service as well, and finally all three positions were so used.

We conjecture that the sequence of incorporation of the amino acids was governed, in the main, by their availability, tempered to some degree by structural aspects (Table 1). We moreover assume that another important factor was the polarity of the amino acids [45]. We hypothesize that the present-day relationship of a purine base in position 2 to a polar amino acid, and of a pyrimidine base to a nonpolar or weakly polar amino acid, has existed since the earliest stages (the 'polarity assumption'). Thus in the earliest stage of evolution we assume that G in position 2 associates with Gly and C with Ala. In later stages we associate A with the more polar and G with the less polar of a pair of amino acids, and for nonpolar and weakly polar acids we assign the nonpolar one to U and the weakly polar one to C. Finally, whenever there are alternatives in assignments, we are guided by considerations involving the present-day code.

As a brief overview, the developments in later evolutionary stages are these. While in stage A (Figure 15) position 2 has to be occupied by G or C, in later stages A or U also becomes acceptable, leading to the codons GNC in stage B, where N denotes G, C, A or U. Next, the need for restrictions in position 3 is relaxed and this position can be occupied by G, C, A or U (stage C, GNN). Later, the nucleotide in position 1 need no longer be G but still must be one of the two purine bases A or G, either of them denoted by Pu (stage D, PuNN). Finally, all base sequences are allowed (stages E, F, and G, NNN). The following sections give more details on stages B–G.

(a) Stage B, codons GNC

Depending on the particulars of the environment, the use of A and U as code letters, in addition to G and C, can have selective advantages, because the transition temperature from hairpin to open conformation (needed in strand

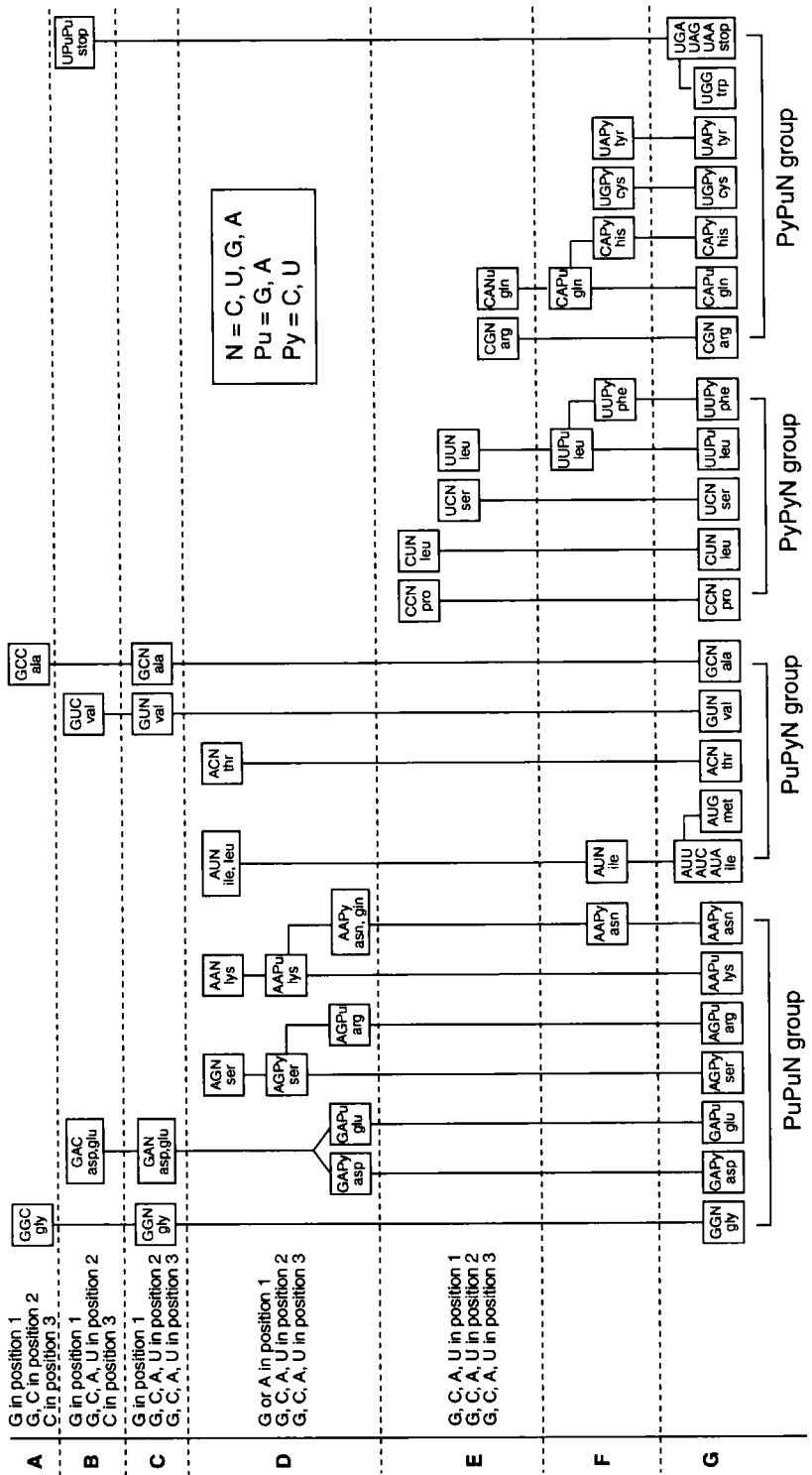


Figure 15 Stages in the evolution of the modern genetic code

replication) strongly depends on the GC/AU ratio of the hairpin, and a certain ratio is optimal for given environmental conditions. Nevertheless, the importance of a GNC reading frame* at this early stage demands that only in position 2 can G and C be supplemented by A and U.

The amino acids that follow Gly and Ala in availability are Asp, Glu and Val (Table 1). According to our polarity assumption, the codon GUC will be associated with the nonpolar Val and GAC with the polar Asp or Glu, the last two amino acids being interchangeable at this stage.

As more and more amino acids are incorporated there is a need for stop codons. Such codons should differ as much as possible from GNC. Instead of the purine G in position 1 there should be a pyrimidine (C or U, the weakly binding U being presumed best); in position 3 should be a purine (G or A). Position 2 should hold a purine, because it is known that hairpins with Py in the loop bind less well to the complementary Pu in the assembler strand than vice versa [46]. This yields the stop codons UPuPu.

(b) Stage C, codons GNN

The sequence GNC, used to maintain the reading frame, gradually loses importance because of the increased efficiency of the translation device. However, only position 3 is thereby opened for all bases; positions 1 and 2 retain their importance for the proper incorporation of new hairpins.

(c) Stage D, codons PuNN

The maintenance of the triplet reading frame no longer requires that position 1 be taken up by G, and the purine A can serve the same purpose. Hence the codons AUN, AGN, ACN and AAN become available for the amino acids Leu or Ile, Ser, Thr and Lys, next in sequence in Table 1. Using our polarity assumption we associate the codons AGN and AAN with Ser and Lys (moderately and strongly polar) and AUN and ACN with Leu or Ile and Thr (nonpolar and weakly polar).

The next amino acids are Arg and Gln or Asn. We assume that later in stage D complementarity in position 3 gradually becomes important again to increase the number of available codons. However, at first only a Pu-Py complementarity is needed; that is, for a Pu in the codons there needs to be a Py in the anticodons and vice versa. The code letter for Ser, the N in AGN, settles on Py. Thus, Pu is no longer used for Ser and AGPu becomes available for Arg. Similarly, AAPy is no longer used for Lys and becomes available for Asn or Gln. Moreover, a distinction is made between Asp and Glu, GAPy coding for the former and GAPu for the latter.

*Our GNC frame resembles Crick's PuNPY comma-free reading frame [31]. His model, however, does not involve the hairpins-assembler strand aggregate, but is based on a flipping mechanism.

(d) Stage E, codons NNN

As the final stage G is approached, the Pu requirement for position 1 (for maintaining the reading frame) is relaxed and position 1 gradually can be occupied as well by Py (U or C). However, UPUu remains reserved for stop codons. This means that CCN, CUN, UCN, UUN, CGN, CAN, UGPy and UAPy become available. The new possibilities are utilized in our model for two purposes: to gain access to amino acids not yet incorporated, and to provide additional codons for important amino acids, perhaps as protection against excessive mutations and/or to provide a large pool of tRNAs for these amino acids.

Codons CCN and CUN, differing from the stop codons UPUu at positions 1 and 2, are now available for nonpolar amino acids. CCN codes for Pro and CUN for Leu; Pro is prebiotically abundant but has not been pressed into use up to now because its use in oligopeptides presents steric difficulties (it cannot participate in α -helical conformations); Leu is the amino acid with the highest frequency in present-day proteins [40] and now is supplied with additional codons. Following our polarity assumption, codons UCN and UUN, differing from UPUu in position 2, are used for weakly polar or nonpolar amino acids. We assign UCN to Ser and UUN to Leu. This provides additional codons for both; Ser is also quite prevalent in present-day proteins [40]. Codons CGN and CAN, differing from UPUu in position 1, become available for polar amino acids; CGN (in addition to other codons) is used for Arg and CAN for Gln (previously not distinguished from Asn).

(e) Stage F

Once the new codons CAN come into use for Gln, AAPy can code exclusively for Asn. Similarly, the use of CUN and UUN for Leu permits the restriction of AUN to Ile. The availability of codons thus liberated permits amino acids whose prebiotic availability is less certain (Table 1) to be added to the program. We assume that in positions 3 in the codons UUN for Leu and CAN for Gln Py falls into disuse. At this point the codons UUPy and CAPy become available for new polar and nonpolar amino acids, respectively. We assign the first to nonpolar Phe and the second to polar His. For the next amino acids in Table 1, Cys and Tyr, the codons UPUy that differ from UPUu in position 3 remain. Of these, UGPy is assigned to less polar Cys and UAPy to more polar Tyr [45]. All 64 codons have now been assigned.

(f) Stage G

In this last stage AUN and UPUu differentiate. Of the four AUN codons for Ile, AUG becomes used for Met, and UGG, one of the four stop codons UPUu, becomes associated with Trp. This brings us to the present-day code.

The approach we have used here reconstructs the modern code through a parsimonious sequence of incremental stages. Each stage is fully plausible in that

it follows the known laws of chemistry and physics and rests upon the conditions of the stage that preceded it. The insight thus afforded strongly suggests that the ancestral self-replicating system resembled that described by the HA device.

4.6 Some Findings in Favor of the Proposed Model

(a) Early amino acids associated with more codons

In a study of the evolution of the genetic code, based on the criterion that the stage of incorporation of an amino acid is related to the number of codons used for it and its availability, Baumann and Oro [39] arrived at a sequence that corresponds well with ours. Their stages 1, 2 and 3 correspond to our stages A–D, E–F and G, respectively, and in their stage 1 the average purine content of the codons is higher than the pyrimidine content, but the pyrimidine content increases in stage 2. In our model, this corresponds in stages A–D to the restriction of position 1 to purines and in stage E to the opening up of position 1 to pyrimidines.

(b) Stem recognition

Some synthetases are capable of specific aminoacylation of rudimentary tRNAs that comprise the acceptor stem only [47]. This has led to the proposal that the recognition system of ancestral tRNA synthetases primarily involved the acceptor stem, and that anticodon recognition was added later [48]. These findings favor the idea that synthetases emerged from our catalyst E₂ (Sections 3.7 and 4.1).

(c) RNA came first

Additional support for our model is provided by the ‘RNA prior to DNA hypothesis’, in favor of which a number of facts are offered in a review paper by Oro *et al.* [49]. Some of these facts are as follows. Transcriptases (also called DNA-dependent RNA polymerases) can use (under slightly modified laboratory conditions) RNA instead of DNA as templates in RNA replication [50], supporting the view that they are derivatives of an RNA replicase. This interpretation of transcriptase, as originating prior to the appearance of DNA, has also been indicated elsewhere [51]. A common origin of DNA and RNA polymerases, as well as of reverse transcriptase (the enzyme that drives the replication of DNA on an RNA template), is further suggested by the high evolutionary conservation of the region of nucleotide sequences of the genes coding for these enzymes [52]. The claim that DNA replicase evolved from RNA replicase is further supported by the discovery of a vestige of the original reverse transcriptase in a eubacterial subunit (eubacteria are among the oldest-known bacteria). These results specifically support our contentions that there

existed RNA-based protein synthesis prior to the appearance of DNA and that the catalysts E_1' , E_1'' and E_1''' were derived from E_1 .

(d) *Reconstruction of a progenote*

Another approach is that of Benner and Ellington [53] who tried to reconstruct a 'progenote', their term for the last common ancestor of modern life forms from which a divergent evolution started. Its attributes were deduced from the metabolic pathways of modern organisms and other modern biochemical facts, which permitted a reasonably detailed reconstruction of its genetics and metabolism. The progenote had an RNA-based protein production device and also had DNA. This again is supportive of our model. The progenote corresponds in our model to the situation when DNA had evolved to the point of serving as an informational macromolecule, a situation resulting in the explosive divergence of life.

Benner and Ellington consider the advent of the progenote as the third episode in the evolution of early life (as they call the different stages). The first episode was based entirely on RNA. The second episode began with the advent of what the authors term a 'breakthrough organism', in which encoded messenger RNA directed the synthesis of proteins that were subject to selection for their catalytic activity. While this aspect of their breakthrough organism also supports our model, there are crucial differences. In contrast to our model and to the above-mentioned results of Lazcano *et al.* [51], their breakthrough organism already used DNA to store genetic information, but there were almost no protein-based enzymes at this stage. A transcription mechanism using catalysts exclusively consisting of RNA was in place before there was extensive use of protein-based catalysts.

The following argument is given for this early use of RNA before a translation-based synthesis of proteins occurred. The ribonucleotide reductases (the enzymes that convert ribonucleotides into deoxynucleotides, i.e. RNA monomers into DNA monomers) are not homologous in different kingdoms—they differ in sequence and structure and evidently do not have a common ancestor. From this, Benner and Ellington draw the conclusion that the original reductase was a ribozyme (an enzyme consisting only of RNA), which was later supplanted by different protein-based enzymes in the different kingdoms. However, there is no reason why the ancient, original reductase could not have been protein based, and why a later protein-for-protein replacement could not have occurred instead.

(e) *Common tRNA ancestor*

The model assumption that the tRNAs of all amino acids have a common ancestor is corroborated by comparative studies [54]. This conjectured ancestor consists mainly of C and G, supporting the hypothesis that these nucleotides

were used first. (A very similar nucleotide sequence is also found in eubacterial tRNA [55]). Suppose that we accept this hypothesis and, accordingly, replace all Us by the structurally similar Cs (both pyrimidines) and all As by Gs (both purines), a total change of less than one in three. If then a hairpin conformation is assumed for this sequence, 85% of the bases in the hairpin legs are paired. This favors the view that the ancestral adapters were in a hairpin conformation. Other recent investigations [56] have shown that during codon–anticodon binding an association of tRNA molecules takes place, suggestive of a remnant of the ancient picket-fence-like arrangement.

(f) Reading frame

A sequence analysis of the DNAs of viruses, prokaryotes and eukaryotes [57] buttresses the view that the assembler strand was the primordial genetic information carrier with a GNC reading frame, where G and C are fixed and N, either G or C, provided the first code. The analysis revealed periodic correlations, pointing to an earlier reading frame of PuNPy (Pu = purine, G or A; Py = pyrimidine, C or U; here N = Pu or Py). The deviations from such a PuNPy reading frame in different organisms were found to be correlated with the time elapsed since their phylogenetic diversion. It has been calculated that the last use of a PuNPy reading frame was some 5×10^9 years ago, a number that is in good agreement with the age of fossil traces of early life. It can therefore easily be imagined that at first G and C were used exclusively to designate the reading frame, and that later G shared its first place in the triplet with another purine, A, because of the steric similarity of these two; similarly, the third place began to be filled by either C or U, both pyrimidines, leading to the early PuNPy reading frame (also see Section 4.5). All these findings give credence to the concept of a vestigial translation device of the type described by our HA device.

4.7 Some Experimental Results in Support of the Proposed Model

(a) Triplet attachment

The model developed in Section 3 from an engineering viewpoint called for R-monomers with properties that are rather severely limited both in terms of geometry and energetics. The question then is: are these requirements fulfilled by the nucleotides G, C, A and U? Hairpins fashioned from RNA are well known and have been characterized experimentally, and it has been possible to demonstrate the binding of synthetic hairpins to an assembler strand made of complementary bases.

One set of experiments involved synthesized assembler strands, hairpin strands and small strand pieces such as trimers and tetramers [46]. The assembler strands were immobilized by fixation to cellulose, and hairpin binding occurred at 0 °C (Figure 16). The strength of binding was determined by warming and

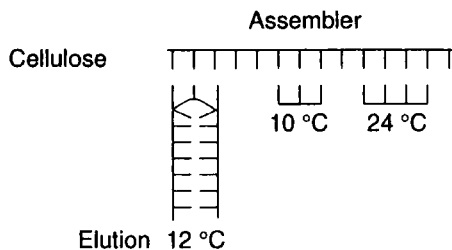


Figure 16 Experimental elution of hairpins, trimers and tetramers attached to an assembler strand fixed on a cellulose column

measuring the temperature of elution with the following results. Hairpins with five deoxyadenosine residues in the loop that were complementary to the monomers along the assembler strand came off the column at 12 °C. The elution temperature of the corresponding trinucleotide was almost the same (10 °C), while the tetranucleotide was eluted at 24 °C, indicating hairpin binding by triplets. (Dinucleotides were not attached at all to the assembler strands on the column, nor was there retention of actual tRNA with an anticodon complementary to the assembler strand codons—not very surprising considering the huge size difference between tRNA and the synthetic hairpins.) Corresponding results were obtained with five deoxyguanoside residues in the hairpin loops and complementary residues along the assembler strand.

Moreover, the temperature of elution was increased upon adding Mg^{2+} to the hairpin solution, pointing to lateral binding between hairpins linked to the assembler strand. Nondenaturing gel electrophoresis of an Mg^{2+} -containing solution of assembler strands and hairpins indicated gapless linkage of hairpins to assembler strands by triplets of complementary basepairs, as required by the model.

An examination of the hairpins by NMR [58] revealed that the three nucleotides in the center of the loop were oriented outward, in a position where they could readily bind to the assembler strand (Figure 16). In contrast, the adjoining nucleotides on either side, just above the helically intertwined hairpin legs, were pointed away from the solvent, precluding any base pairing with the assembler strand.

(b) Amino acid attachment

Another question concerns the chemical bonding of amino acids to the 2'-OH group at the 3'-end of RNA, a prerequisite for the proposed peptide bond formation. Such bond formation has indeed been observed [59]. Peptide bond formation has been observed upon bringing amino acid esters of insoluble alkanols, in a monolayer on a water surface, into close proximity [60].

These experiments support our contention that the initial translation device consisted of the equivalent of an mRNA assembler strand and (+)-hairpins and (-)-hairpins of RNA. In our model the amino acids Gly and Ala, linked to the 2'-OH groups at the 3'-hairpin ends, are thought to have been activated earlier, e.g. by binding Gly to GMP and Ala to CMP (CMP are the low-energy C-nucleotide and G-nucleotide). A possible explanation for the assumption of such specific binding of Ala to CMP and Gly to GMP lies in the fact that Gly and GMP are eluted at the same speed from silica by an aqueous medium [61], while Ala and CMP are eluted at another, significantly different speed. This offers the possibility of *in situ* formation of high-energy GMP-Gly and CMP-Ala on a silica substrate, assisted by some condensing agent that links GMP with Gly and CMP with Ala.

4.8 Computer Simulations and Model Considerations

Computer simulations [62] and perusals of macroscopic models [13, 14] demonstrate that the translation device proposed in Section 3 can easily be constructed from nucleotides and amino acids. In the computer simulation the forces between atoms are assumed to be determined by a set of empirical parameters [63]. The solvent is considered in a global fashion as a dielectric, and the solvated Mg^{2+} ions, assumed to be interposed between the phosphate groups in the backbones of neighboring hairpins, are simulated by charged spheres. The geometry of the system is then established by minimizing the energy.

The essential results of the computer simulation are as follows [62]. Hairpins with legs assumed to be intertwined in the usual, so-called right-handed A-RNA fashion, linked to the assembler strand by triplet base pairs and laterally linked by Mg^{2+} , constitute a stable system. This is of some interest, because in earlier considerations [13, 14] of macroscopic models we had used the 'slim left-handed Z-RNA' conformation, and the question then arose of how a change from a left-handed to a right-handed conformation, as is present in modern tRNA, could have happened at a late stage. This question is therefore no longer at issue.

The specific intercalation of GMP-Gly and CMP-Ala at the bottoms of (+)-hairpins and (-)-hairpins, and the binding of these amino acids to the 2'-OH groups at the 3'-hairpin ends can easily be simulated by macroscopic models [14] and on a computer, and it can be shown that the peptide condensation between the amino acids tied to the bottoms of neighboring hairpins can readily be achieved, both sterically and energetically, through a conformational change [62].

4.9 Chirality

The helical intertwining of the hairpin legs requires that all monomers the hairpin consists of are of the same chirality, i.e. that they all contain just

D-ribose*. Hence the CMP involved is chiral. Moreover, alanine is also chiral, and according to force field calculations CMP-L-alanine is, by $4.2 \text{ kcal mol}^{-1}$ more stable than CMP-D-alanine [62]. This may well explain why it is L-amino acids that are used in the biosynthesis of proteins†.

4.10 Conclusion

The results of experiments, computer simulations and work with macroscopic models, as well as other supportive findings, are most encouraging. Nevertheless, more work, especially with synthesized hairpins and assembler strands, is urgently needed.

5. PERSPECTIVES IN SUPRAMOLECULAR ENGINEERING

It is a great challenge to design and synthesize instructed molecules that carry a message, the know-how that induces their automatic organization into a complex, multicomponent device, thereby pointing the conventional methods of traditional synthetic chemistry in entirely new directions. Current approaches in the construction of such supramolecular devices are based on the synthesis of molecules with regions that are complementary to each other, molecules that 'recognize' each other and form aggregates that have been specifically planned to behave as functional units. This interaction can take place in solution or at interfaces, by spontaneous interaction, interlocking and specific bonding, and it can further be developed and elaborated by exposing the molecules to a programmed sequence of environmental changes.

5.1 Description of the State of the Art

(a) Strategies based on the lock-and-key principle

In recent years a great variety of supramolecular structures have been created, beginning with the synthesis of receptor molecules that can recognize cations by forming cryptate complexes—cage-like compounds that can accommodate these cations in their interior [1]. Other cryptates capable of accepting increasingly sophisticated molecules have been obtained. Concave receptors with two receptor sites have been designed and realized, structures that have permitted

*It could just as well be just L-ribose; which of the two was a matter of chance at the time the first successful hairpin aggregated was selected.

†One may ask why attention is focused here on just CMP; the reason is that while GMP is chiral as well, the amino acid glycine that may be linked to it is not chiral, so that there is only the one GMP-glycine.

the interlinkage of guest molecules, such as in the synthesis of a pyrophosphate by facilitating the bonding of two acetylphosphate molecules. The complexing of several functional components with a cryptate has allowed the construction of supramolecular units that combine several facilities, such as systems capable of photoinduced energy and electron transfer [1, 5, 64, 65].

Cleft structures have been made that permit the specific binding of, for example, cAMP by a cooperation of hydrogen bonding, Coulombic interactions and stacking interactions [66]. A self-replicating complex has been obtained based on the idea that an adenine moiety, bound to a cleverly designed adenine acceptor moiety, can recognize a molecule constituting the other moiety. The complex can thus be used as a template in the autocatalytic synthesis of additional identical molecules from a mixture of the two components. This constitutes a very exciting case of an autocatalytic process [4]. Metal complexes of double helices of different lengths, cleanly separated, have been obtained from a mixture of ingredients by the planned interaction between appropriate metal ions and the other components [67]. Thus Cu^+ and Ag^+ ions, which exhibit tetrahedral coordination, yield double helices, while octahedrally coordinated metal ions cause intertwining into triple helices [68].

It has also been possible to synthesize catenanes, compounds consisting of two or more molecular rings threaded into each other like two or more links of a macroscopic chain. Thus two-link chains can automatically form in a mixture containing ring-shaped molecules in addition to long strand pieces that are often shaped like a horseshoe and tend to thread the rings (like the eye of a needle, because of chemical interactions of a donor-acceptor type or hydrophobicity), and short strand pieces that serve for ring closure. Under suitable conditions the threaded strands react at both ends with the ends of the short strands, the result being catenanes [69]. In a recently synthesized catenane the ring was cyclodextrine, the threading moiety was a polyether chain with intercepted diphenyl and terminal amino groups, and the ring-closing piece was *p*-phenyldiacetyl chloride [70,71]. In an even more unusual catenane three cyclic molecules were tied together, each of them being threaded onto one of the three equal arms of a bicyclic molecule. This catenane, of three-fold symmetry, was obtained by making use of a template effect that involved copper ions [72].

Structures called rotaxanes have been made by threading strands containing donor entities through rings containing acceptor entities and tying bulky groups to the strand-ends. Again these structures form automatically, under the appropriate conditions, in a mixture containing the three ingredients: the rings, the strands and the plugs that prevent the strands from becoming unthreaded. Rotaxanes with remarkable properties have been made, such as electronic switches of molecular size, by employing benzene rings that intercept a polyether chain [3].

Appropriately constructed molecules are capable of self-organization into aggregates of particular architecture [6, 7], and an immense variety of such aggregate structures is known, of which we shall mention only a few examples.

Molecules may stack to form columns with holes that act as tubes for the passage of ions [1, 6]. Stacks of alternating donor and acceptor moieties have been formed [72]. An elegant, automatic synthesis is that of a cylindrical multi-component complex, capped on its end [73]. Another item is the synthesis of a tennis-ball-like structure with a four-fold inversion axis (turn 90° and simultaneously invert) from two identical moieties that automatically interlock when they find themselves in an opposed position and turned 90° relative to each other [74]. Vesicles, micelles and liquid crystals with specifically designed functions have also been obtained by self-organization through the mixing of cleverly designed component molecules [6, 7].

(b) Strategies based on the lock-and-key and programmed-environmental-change principles

While the automatic self-organization of molecules into supramolecular systems through the mixing of components under suitable conditions requires great synthetic ingenuity, it is also possible to generate complex assemblies with preassigned functions through a cleverly planned sequence of external influences that act upon relatively simple molecular components. A most promising future lies in the combination of both intelligently planned syntheses of interlocking components and ingenious programs of environmental changes.

A useful way to assist self-assembly involves spreading molecules with a hydrophilic head and a hydrophobic tail on an aqueous surface and pushing the molecules together at an appropriate speed to produce a monolayer. Layers of increasing complexity can be obtained by co-spreading two or more kinds of molecules [8]. The same methodology can also involve amphiphiles (molecules or ions attracted to an air–water interface) and nonamphiphiles. For instance, the amphiphilic dimyristoylphosphatidic acid and the nonamphiphilic cyclic bis(bipyridinium) tetracation, in the appropriate ratio, form a well-organized, densely packed monolayer [75]. The cyclic tetracation goes below the densely packed amphiphile, the plane of their rings being parallel to the air–water interface. This arrangement can serve as a receptor for appropriate electron-rich guest molecules.

Such differently equipped monolayers can be manipulated in many ways to form increasingly complex arrangements. Moieties that have been used include different kinds of chromophores, electron-donating and electron-accepting groups in precise geometries, and functional units that permit the study of energy transfer, exciton motion, electron tunneling and electrical conduction through molecular wires [8].

Complex monolayer assemblies deposited on glass slides can be cleaved between specifically chosen, distinct monolayers, for example, by depositing a poly(vinyl alcohol) film on top of the assembly and then lifting it, the film serving as a transfer agent for the selected portion of the assembly thus removed. The lifted-off structure can then be put into contact with a selected target surface.

It can also be turned over and transferred to an aqueous surface that effects the dissolution of the poly(vinyl alcohol) film, thus providing access for further manipulations at the underside of the removed layer [8].

It is also possible to induce highly specific, planned reactions within the superstructure of a well-designed monolayer assembly, such as (*in situ*) syntheses, polymerizations, in-plane or interplane ligations, etc. [76]. This methodology is an important tool in the armory used in the construction of machinery that has specific operational functions.

Monolayers with functional groups that stick into an aqueous subphase can act as receptors for the recognition of specific groups or molecules dissolved in the subphase. For instance, a monolayer consisting of a barbituric acid lipid can specifically bind 2,4,6-triaminopyrimidine by forming six hydrogen bonds to it; this process then triggers the cleavage of a carbon-carbon double bond that links the barbituric acid to the lipid [77]. Another case of a specific interaction concerns the binding of the protein streptavidin to a monolayer of biotin lipid, forming a well-ordered two-dimensional crystal of the protein [6]. The crystal has open binding sites for biotin that face the aqueous subphase, allowing the binding of an additional biotin-linked protein to this layer, etc. In this way, the formation of a layered structure that consists of programmed functional component protein monolayers is facilitated.

The direct usage of biological constructs is an important aspect. Efforts in this direction involve the use of assemblies of functional proteins and enzymes [78]. A recent example pertains to the two-dimensional assembly of membrane proteins, such as bacteriorhodopsin, and shows great promise in areas as diverse as molecular electronics and optical switching, molecular sieves, and the creation of nanometric lithographic patterns. By packing in dry films it has been possible to keep such assemblies operative to as high as 140 °C [79].

Yet another route in the construction of assemblies of monolayers consists in the chemisorption by a solid of appropriate molecules in a solution. The end-groups of such chemisorbed monolayers can then be changed chemically into reactive groups, allowing the chemisorption of a second monolayer on top of the first one, etc. [80]. It is also possible to fabricate layered structures that are intercalated, e.g. with copper ions inserted through lateral transports along hydrophilic planes [81].

5.2 Ideas for Future Development

(a) Prospects for the current paradigm

The current paradigm for constructing supramolecular machines, i.e. through the programmed interaction between specifically designed molecules (lock-and-key and programmed-environmental-change principles), appears to be very promising in the development of new avenues for the synthesis of more and more complex molecules that recognize each other and organize into assemblies

of increasing ingenuity. Assistance to interlocking by programmed sequences of external manipulations will be of mounting importance with the growing sophistication of planned machineries.

Molecular assemblies The manipulation of packages of organized assemblies by accurate cleavage and merger on a molecular scale and the use of biomolecules and combinations of biomolecules and otherwise synthesized molecules as interlocking functional components should play important roles in the development of increasingly ingenious devices. The advantages of biomolecules, as used, for example, in the development of sophisticated sensors, are ease and rapidity of supply, while laboratory-synthesized molecules afford a broader variety and scope.

Two-dimensional templates The use of microlithographic methods and surface treatments, such as changing hydrophilic portions of a surface to hydrophobic, should be most fruitful in the development of information-processing devices. They permit the imprinting of submicrometer patterns onto solid supports that can then serve as templates in the formation of monolayers with a programmed in-plane organization. Template patterns that allow the positioning of individual molecules present a great challenge. An exciting development toward this goal is the generation of scanning patterns with molecular-scale resolution by an electron beam [82]. It can be produced, for example, on the tip of a tunneling electron microscope and focused by an electron lens [83].

Three-dimensional structures from two-dimensional templates An important direction of future supramolecular engineering strategies is the use of two-dimensional patterns in the buildup of complex three-dimensional structures. The development of molecular sensing devices, information-processing devices and molecular wiring is a great challenge, as is the use of signals provided by moving particles (electrons, protons, photons, polarons) in the control of supramolecular switching and information-processing systems. What is desired is a replacement of the currently available two-dimensional devices for the mass storage of data and for signal processing by three-dimensional devices created by self-organization on two-dimensional templates. A basic problem relates to the assemblage of a complex three-dimensional structure on a template. While the production of submicrometer templates is feasible, there is still the question of how to achieve accurate registering between neighboring layers on a molecular scale. It might be accomplished through a sequence of steps: by bonding proteins or other macromolecules to a suitable substrate, by designing these macromolecules so that they have recognition sites for other macromolecules, by letting each layer react with the appropriate next sort of macromolecule, each layer arranging itself in a preassigned pattern determined by the preceding layer, all in a carefully planned manner until the desired array is achieved.

Printing on a molecular scale Building a complex array of molecular accuracy on a mesoscopic template is one facet of the interplay between the macroscopic and molecular worlds. Another facet is the reading and writing of information on a molecular scale. Towards this end the addressing of a single dye molecule with light has been achieved in spectral hole burning [84], and it appears to be feasible to so address a dye molecule in a distinct position within an array [85]. Recent developments in near-field light microscopy [8, 86] are of great importance in this regard. The addressing of distinct molecules by electrons has also been demonstrated [87].

Alternative to wires Another potential difficulty is precise wiring on a molecular level. Here an approach inspired by biology may be useful. Much of the switching in biosystems is achieved by means of receptor sites on enzymes to which certain specific molecules can bind, receipt of which turns on and off the production of molecules acting as substrates for receptor sites on other enzymes. Communication thus takes place by means of these specific molecules that carry the 'turn on' or 'turn off' message to the receptors, simulating, so to speak, a wireless connection between signal-releasing enzymes and receptors on other enzymes. This process can be repeated, signaling in sequence from one enzyme to the next. This simplifies the geometrical conditions that would have to be met by a complex network of wires, including problems of insulation. A system for information processing might well be designed on such principles, supplemented perhaps by molecular wires for the transfer of some of the signals.

Near-future objectives The storage and processing of information is certainly one of the most challenging long-term aims of supramolecular engineering. However, some immediate goals are no less intriguing, such as the development of microsensors, materials for integrated optics and optoelectronics, materials based on cooperative effects involving, for example, ferromagnetism or ferroelectricity, highly specific surface catalysts, and systems that model problems in chemistry and physics or that model bioprocesses in an effort to elucidate design principles, as in a simulation of the reaction center of photosynthesis by an exploration of the optimal arrangement and energetics of chromophores and the ideal number of construction elements in the design of an electron pump [88].

(b) Search for new paradigms

A key feature of the mechanism that leads to the emergence of lifelike systems is not mimicked in present-day efforts in molecular engineering. It is the growth in complexity and sophistication by the cycle of multiplication, variation and selection, and it is important to ponder such possible molecular engineering aspects in the origin of life that might inspire advances in the construction of supramolecular devices. Basic to any advance in this direction is the availability

of copying mechanisms for blueprints that contain the information for the building of the machinery, copying mechanisms that are error-prone to just the right degree to supply sufficient variation to permit useful selection. Such blueprints might be one-dimensional or two-dimensional arrays, and the development of such one-dimensional [26] or two-dimensional [8, 89] artificial replicating systems would be important in promoting these perspectives. Another and probably much faster approach that also may provide rapid convergence toward useful devices is the use of biosystems to synthesize the components of artificial machinery. This avenue is, however, restricted essentially to the use of protein components.

The direct application of the Darwinian evolution principle is also very fruitful. For example, *in vitro* selection and evolution techniques have recently been used to generate new classes of ribozymes capable of enhancing reaction rates by factors like 7×10^6 [20]. Darwinian evolution can equally be applied to more specific molecular engineering tasks. It is clear that there is a great future in supramolecular devices and that critical thinking along the lines pointed out here will lead to important advances. The solution of engineering problems in unforeseen directions and the synthesis of complex systems that had not even been thought of before will become possible.

5.3 The Origin of Life: A Coda

The earlier considerations of the model steps that have led to systems with a genetic apparatus provide a great deal of insight into the course of this first phase in the evolutionary process, an inevitable sequence of events as long as the special conditions in a special location are given, conditions that also contribute significantly to the details of the direction the evolution takes. The model is transparent because of the simplicity of its assumptions, and it therefore permits an assessment of how the evolutionary path is governed by accidental changes in a monomer sequence and by existing specific environmental conditions. Only the advent of systems with a self-sustaining set of mutually beneficial translation devices provides liberation from restrictive environmental situations. It is noteworthy that the basic traits of this translation machinery, which inevitably develops wherever the precise chemical, spatial and temporal prerequisites exist, strongly resemble those of the genetic apparatus of biosystems. All other potential evolutionary pathways, pathways that do not end up in a genetic apparatus with these traits, inevitably turn out to be doomed to end up in an impasse, or a cul-de-sac.

Worthy of emphasis is the success that is associated with the approach used here, the methodology of a search for small steps, each of feasible probability, and a search for evolutionary gradients pointing to adaptive improvements that reward these steps. The approach leads to a final stage from which a cataclysmic diversification of lifelike systems can occur, even though little can be said about the details of the steps followed by terrestrial biology.

The detailed considerations also emphasize the fact that the enigma of the origin of life cannot be resolved by general reflections on the self-organization of matter or by computer simulations of the development of order out of chaos, both of which are of great interest but fail because they do not provide any detailed paths and no questions regarding details are posed or answered. It is of the greatest importance to ask again and again, after even the smallest step, about what comes next. What evolutionary circumstances provide a gradient towards further evolution and how are these gradients utilized? Such a procedure, a modeling of detailed steps, each of them feasible on chemical, physical and probabilistic grounds and rewarded by evolutionary improvements, is fundamentally different from an approach that searches for self-organizing processes that lead to dissipative structures in homogeneous phases, such as in a primeval ocean. It is very striking that the tightly confined path described earlier inescapably appears to lead to a very specific genetic apparatus and code, and to the subsequent explosive emergence of different life forms.

It also should be appreciated that the possibilities of evolution are greatly limited because of its restriction to a narrow path of minute steps, each small advance being selected only if it bestows an evolutionary benefit upon the system. In contrast, molecular engineering is not hampered by this restriction; its limits are human imagination and creativity alone, and the possibilities are endless. The sky is the limit.

6. REFERENCES

1. J.-M. Lehn, *Angew. Chem.*, **100**, 91 (1988); *Angew. Chem., Int. Ed. Engl.*, **27**, 89 (1990); *Angew. Chem.*, **102**, 1347 (1990); *Angew. Chem., Int. Ed. Engl.*, **29**, 1304 (1990).
2. C. Dietrich-Buchecker, B. Frommberger, I. Lürer, J. P. Sauvage and F. Vögtle, *Angew. Chem.*, **105**, 1526 (1993); *Angew. Chem., Int. Ed. Engl.*, **32**, 1434 (1993).
3. R. Ballardini, V. Balzani, M. T. Gandolfi, L. Prodi, M. Venturi, D. Philp, H. G. Ricketts and J. F. Stoddart, *Angew. Chem.*, **105**, 1362 (1993); *Angew. Chem., Int. Ed. Engl.*, **32**, 1301 (1993).
4. J. Rebek, *Angew. Chem.*, **102**, 261 (1990); *Angew. Chem., Int. Ed. Engl.*, **29**, 245 (1990).
5. C. Seel and F. Vögtle, *Angew. Chem.*, **104**, 542 (1992); *Angew. Chem., Int. Ed. Engl.*, **31**, 528 (1992).
6. M. Ahlers, W. Müller, A. Reichert, H. Ringsdorf and I. Venzmer, *Angew. Chem.*, **102**, 1310 (1990); *Angew. Chem., Int. Ed. Engl.*, **29**, 1269 (1990); H. Ringsdorf, B. Schlarp and J. Venzmer, *Angew. Chem.*, **100**, 117 (1988); *Angew. Chem., Int. Ed. Engl.*, **27**, 113 (1988).
7. T. Kunitake, *Angew. Chem.*, **104**, 692 (1992); *Angew. Chem., Int. Ed. Engl.*, **31**, 709 (1992).
8. H. Kuhn and D. Möbius, in *Physical Methods of Chemistry*, 2nd edn, Vol. 9B, Part 6 (ed. B. W. Rossiter and R. C. Baetzold), Wiley, New York, 1993, p. 375.
9. J. Oro, S. L. Miller and A. Lazcano, *Annu. Rev. Earth Planet. Sci.*, **18**, 317 (1990).
10. A. Eschenmoser and E. Loewenthal, *Chem. Soc. Rev.*, **21**, 1 (1992); A. Eschenmoser, *Nova Acta Leopoldina*, **67**, 201 (1992).

11. C. de Duve, *Blueprint for a Cell: The Nature and Origin of Life*, Patterson, Burlington, North Carolina, 1991; H. Föllman, *Chemie und Biochemie der Evolution*. Quelle and Meyer, Heidelberg, 1981.
12. H. Kuhn, *Angew. Chem.*, **84**, 837 (1972); *Angew. Chem., Ed. Engl.*, **11**, 798 (1972); *Naturwissenschaften*, **63**, 68 (1976).
13. H. Kuhn and J. Waser, *Angew. Chem.*, **93**, 495 (1981); *Angew. Chem., Int. Ed. Engl.*, **20**, 500 (1981); *Experientia*, **39**, 834 (1983); H. Kuhn, *Mol. Eng.*, **1**, 377 (1992).
14. H. Kuhn and J. Waser, in *Biophysics* (ed W. Hoppe, W. Lohmann, H. Markl and H. Ziegler), Springer, Berlin, 1983, p. 830.
15. A. G. Cairns Smith, *Genetic Takeover and the Mineral Origin of Life*, Cambridge University Press, Cambridge, 1982; *J. Theor. Biol.*, **10**, 53 (1966). N. Lahav, *Heterogeneous Chemistry Reviews*, **1**, 159 (1994).
16. G. Wächtershäuser, *Proc. Natl. Acad. Sci. USA*, **85**, 1134 (1988); G. Wächtershäuser, in *Early Life on Earth, Nobel Symposium 84* (ed. S. Bengtson), Columbia University Press, New York, 1994.
17. G. F. Joyce, *Nature*, **338**, 217 (1989).
18. N. Lahav, *J. Theor. Biol.*, **151**, 531 (1991).
19. A. I. Oparin, *Ration. Annu.*, **148** (1929).
20. D. P. Bartel and J. W. Szostak, *Science*, **251**, 1411 (1993).
21. D. Müller, S. Pitsch, A. Kittaka, E. Wagner, C. E. Wintner and A. Eschenmoser, *Helv. Chim. Acta*, **73**, 1410 (1990).
22. J. Oro and E. Stephen-Sherwood, in *Cosmochemical Evolution and the Origins of Life* (eds J. Oro, S. L. Miller, C. Ponnamperna and R. S. Young), Reidel, Dordrecht, 1974, p. 523; J. Oro and A. Lazcano-Araujo, in *Cyanide in Biology* (eds B. Vernesland, E. E. Conn, C. J. Knowles, J. Westley and F. Wissing), Academic Press, London, 1981, p. 517.
23. P. K. Bridson and L. E. Orgel, *J. Mol. Biol.*, **144**, 567 (1980); P. K. Bridson, H. Fakhrai, R. Lohrmann, L. E. Orgel and M. van Roode, in *Origin of Life* (ed. Y. Wolman), Reidel, Dordrecht, 1981, p. 613.
24. G. F. Joyce, T. Inoue and L. E. Orgel, *J. Mol. Biol.*, **176**, 279 (1984).
25. G. F. Joyce and L. E. Orgel, *J. Mol. Biol.*, **188**, 433 (1986).
26. G. v. Kiedrowski, B. Wlotzka, J. Helbing, M. Matz and S. P. Jordan, *Angew. Chem.*, **103**, 456 (1991); A. Terfort and G. v. Kiedrowski, *Angew. Chem.*, **104**, 626 (1992).
27. L. E. Orgel, *J. Mol. Evol.*, **29**, 465 (1989); N. Maizels and A. M. Weiner, *Proc. Natl. Acad. Sci. USA*, **84**, 7383 (1987).
28. A. Eschenmoser, *Angew. Chem.*, **100**, 5 (1988); *Angew. Chem., Int. Ed. Engl.*, **27**, 5 (1988).
29. C. Woese, D. H. Dugre, W. C. Saxinger and S. A. Dugre, *Proc. Natl. Acad. Sci. USA*, **55**, 966 (1966); C. Woese, *Nature*, **226**, 817 (1970).
30. F. H. C. Crick, *J. Mol. Biol.*, **38**, 387 (1968).
31. F. H. C. Crick, S. Brenner, A. Klug and G. Pieczenik, *Orig. Life*, **7**, 389 (1976).
32. M. Eigen and R. Winkler, *Naturwissenschaften*, **68**, 217 (1981).
33. M. Eigen, B. F. Lindemann, M. Tietze, R. Winkler, A. Dress and A. Haeseler, *Science*, **244**, 673 (1989).
34. J. C. Lacey and D. W. Mullins, *Orig. Life*, **13**, 3 (1983).
35. T. H. Jukes, *Life Sci. Space Res.*, **21**, 177 (1984).
36. F. I. R. Taylor and D. Coates, *BioSystems*, **22**, 177 (1989).
37. T. J. Gibson and A. T. Lamond, *J. Mol. Evol.*, **30**, 7 (1990).
38. S. Osawa, T. H. Jukes, K. Watanabe and A. Muto, *Microbiol. Rev.*, **56**, 229 (1992).
39. U. Baumann and J. Oro, *BioSystems*, **29**, 133 (1993).
40. K. Wada, S. Aota, R. Tschuiya, F. Ishbashi, T. Gojobori and T. Ikemura, *Nucl. Acids Res.*, **18**, 2367 (1990).
41. K. Harada and S. W. Fox, *Nature*, **201**, 335 (1964).
42. J. T. F. Wong and P. M. Bronskill, *J. Mol. Evol.*, **31**, 167 (1979).
43. C. Shen, S. I. Miller and J. Oro, *J. Mol. Evol.*, **31**, 167 (1990).

44. J. R. Lacey and M. P. Staves, *Orig. Life*, **20**, 303 (1990).
45. R. Wolfenden, L. Andersson, P. M. Cullis and C. C. B. Southgate, *Biochemistry*, **20**, 849 (1981).
46. U. Baumann, U. Lehmann, K. Schwelbnuss, J. H. van Boom and K. Kuhn, *Eur. J. Biochem.*, **170**, 267 (1987); U. Baumann, R. Frank and H. Blöckler, *Biochem. Biophys. Res. Commun.*, **157**, 986 (1988).
47. D. D. Buechter and P. Schimmel, *Biochemistry*, **32**, 5267 (1993).
48. M. E. Saks, J. R. Sampson and J. N. Abelson, *Science*, **263**, 191 (1994).
49. A. Lazcano, R. Guerrero, L. Margulis and J. Oro, *J. Mol. Evol.*, **27**, 283 (1988).
50. V. Llaco, E. Silva, A. Lazcano, L. M. Rangel, P. Gariglio and J. Oro, in *Prebiological Selforganization of Matter* (eds C. Ponnampuruma and F. R. Eirich), Deepak, New York, 1990, p. 247.
51. A. Lazcano, J. Fastag, P. Gariglio, C. Ramirez and J. Oro, *J. Mol. Evol.*, **27**, 365 (1988).
52. A. Lazcano, V. Llaco, G. E. Fox and J. Oro, in *Abstr. 6th ISSOL Meet.*, 1989, p. 194.
53. S. A. Benner and A. D. Ellington, *Nature*, **328**, 295 (1987); *Proc. Natl. Acad. Sci. USA*, **86**, 7054 (1989).
54. M. Eigen and R. Winkler, *Naturwissenschaften*, **68**, 217, 282 (1981).
55. M. W. Kilpatrick and R. T. Walker, *Nucl. Acids Res.*, **9**, 4387 (1981).
56. D. Pörschke and D. Labuda, *Biochemistry*, **21**, 53 (1982).
57. J. C. W. Shephard, *Proc. Natl. Acad. Sci. USA*, **78**, 1596 (1981); *J. Mol. Evol.*, **17**, 94 (1981).
58. C. A. G. Hassnot and J. H. van Boom, *J. Biomol. Struct. Dyn.*, **1**, 115 (1983).
59. A. T. Profy and D. A. Usher, *J. Mol. Evol.*, **20**, 147 (1984); *J. Am. Chem. Soc.*, **106**, 5030 (1984).
60. K. Fukuda, Y. Shibasaki and H. Nakahara, *Thin Solid Films*, **99**, 87 (1983).
61. U. Lehmann and H. Kuhn, *Adv. Space Res.*, **4**, 153 (1984); U. Lehmann, *BioSystems*, **17**, 139 (1985).
62. E. v. Kitzing, Dissertation, University of Göttingen, 1985; *Molekülsimulation mit Hilfe von Kraftfeldrechnungen am Beispiel der Aggregation von Nukleinsäuren verschiedener Konformation zu einem Komplex mit Uebersetzungsfunktion*, Edition Herodot, Raderverlag, Aachen, 1986.
63. P. Weiner, A. Kollman, *et al.*, *J. Am. Chem. Soc.*, **106**, 765 (1984).
64. J.-M. Lehn, *Angew. Chem.*, **103**, 1365 (1991).
65. F. Vögtle, W. M. Müller, V. Müller, M. Bauer and K. Rissanen, *Angew. Chem.*, **105**, 1356 (1993); *Angew. Chem., Int. Ed. Engl.*, **32**, 1295 (1993).
66. G. Deslongchamps, A. Galan, J. de Mendoza and J. Rebek, *Angew. Chem.*, **104**, 58 (1992); *Angew. Chem., Int. Ed. Engl.*, **31**, 61 (1992).
67. B. Dietrich, P. Viout and J.-M. Lehn, *Macrocyclic Chemistry. Aspects of Organic and Inorganic Supramolecular Chemistry*, VCH Verlag, Weinheim (1992).
68. A. Pfeil and J.-M. Lehn, *J. Chem. Soc., Chem. Commun.*, 838 (1992); J.-M. Lehn and A. Rigault, *Angew. Chem.*, **100**, 1121 (1988); *Angew. Chem., Int. Ed. Engl.*, **27**, 1095 (1988); R. Krämer and J.-M. Lehn, unpublished results.
69. J. F. Stoddart, *Angew. Chem.*, **104**, 860 (1992); *Angew. Chem., Int. Ed. Engl.*, **31**, 846 (1992).
70. D. Armspach, P. R. Ashton, C. P. Moore, N. Spencer, J. F. Stoddart, T. J. Wear and D. J. Williams, *Angew. Chem.*, **105**, 944 (1993); *Angew. Chem., Int. Ed. Engl.*, **32**, 854 (1993).
71. G. Wenz, *Angew. Chem.*, **106**, 851 (1994); *Angew. Chem., Int. Ed. Engl.*, **33**, 803 (1994).
72. D. B. Amabilino, P. R. Ashton, M. S. Tolley, J. F. Stoddart and D. J. Williams, *Angew. Chem.*, **105**, 1358 (1993); *Angew. Chem., Int. Ed. Engl.*, **32**, 1295 (1993).
73. R. Krämer and J.-M. Lehn, unpublished results; P. Baxter, J.-M. Lehn, A. de Cian and J. Fischer, *Angew. Chem.*, **105**, 92 (1993).

74. R. Wyler, J. de Mendoza and J. Rebek, *Angew. Chem.*, **105**, 1820 (1993); *Angew. Chem., Int. Ed. Engl.*, **32**, 1699 (1993).
75. R. C. Ahuja, P. L. Caruso, D. Möbius, G. Wildburg, H. Ringsdorf, D. Philp, J. A. Preece and J. F. Stoddart, *Langmuir*, **9**, 1534 (1993).
76. A. Ruau-del-Teixier, *J. Chim. Phys.*, **85**, 1067 (1988).
77. R. Ahuja, P. L. Caruso, D. Möbius, W. Paulus, H. Ringsdorf and G. Wildburg, *Angew. Chem.*, **105**, 1082 (1993); *Angew. Chem., Int. Ed. Engl.*, **32**, 1033 (1993).
78. M. Aizawa, K. Owaku, M. Matsuzawa, H. Shinohara, Y. Ikariyama, *Thin Solid Films*, **180**, 227 (1989). M. Aizawa, *IEEE Engineering in Medicine and Biology*, **13**, 94 (1994).
79. Y. Shen *et al.*, *Nature*, **366**, 48 (1993).
80. J. Sagiv, *J. Am. Chem. Soc.*, **102**, 92 (1980); C. D. Bain and G. M. Whitesides, *Angew. Chem. Adv. Mater.*, **101**, 522 (1989); C. D. Bain, E. B. Troughton, Y. T. Tao, J. Evall, G. M. Whitesides and R. G. Nuzzo, *J. Am. Chem. Soc.*, **111**, 321 (1989).
81. R. Maoz and J. Sagiv, in *Proc. Int. Symp. Functional LB Films*, Tokyo, 1992 (ed. K. Fukuda, M. Sugi), p. 23. R. Maoz, R. Yam, G. Berkovic and J. Sagiv, in: *Organic Thin Films and Surfaces* (ed A. Ulman) Vol. 1, Academic Press, Boston (1993).
82. H. P. Zingsheim, *Ber. Bunsen-Ges. Phys. Chem.*, **80**, 1185 (1976).
83. H. W. Fink, H. Schmid, H. J. Kreuzer and A. Wierzbicki, *Phys. Rev. Lett.*, **67**, 1543 (1991); H. J. Kreuzer, K. Nakamura and A. Wierzbicki, in press.
84. W. E. Moerner and T. Basche, *Angew. Chem.*, **105**, 537 (1993); *Angew. Chem., Int. Ed. Engl.*, **32**, 457 (1993); M. Orrit and J. Bernard, *Phys. Rev. Lett.*, **65**, 2716 (1990).
85. H. Kuhn, *Thin Solid Films*, **178**, 1 (1989).
86. A. Lewis and K. Lieberman, *Nature*, **345**, 214 (1991); E. Betzig and J. K. Trautman, *Science*, **257**, 189 (1992).
87. D. M. Eigler and J. K. Schweizer, *Nature*, **344**, 524 (1990).
88. H. Kuhn, *Phys. Rev. A.*, **34**, 3409 (1986).
89. D. Möbius, *Z. Phys. Chem. Neue Folge*, **154**, 121 (1987).

Chapter 8

Perspectives in Supramolecular Chemistry—From the Lock-and-Key Image to the Information Paradigm

JEAN-MARIE LEHN

Université Louis Pasteur, Strasbourg and Collège de France, Paris, France

1. FROM MOLECULAR TO SUPRAMOLECULAR CHEMISTRY

For more than 150 years, since the synthesis of urea by Friedrich Wöhler in 1828 [1], molecular chemistry has developed a vast array of highly sophisticated and powerful methods for the construction of ever more complex molecular structures by the making or breaking of covalent bonds between atoms in a controlled and precise fashion.

Organic synthesis has grown rapidly and masterfully, leading to a whole series of brilliant achievements in the great syntheses of the last 50 years, where elegance of strategy has been combined with feats of efficiency and selectivity. Molecular chemistry has thus established its power over the covalent bond. The time has come to do the same for noncovalent intermolecular forces. Beyond molecular chemistry based on the covalent bond there lies the field of supramolecular chemistry [2, 3], whose goal it is to gain control over the intermolecular bond. It is concerned with the next step in increasing complexity beyond the molecule towards the supermolecule and large, organized molecular systems held together by noncovalent interactions.

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 (Figure 1). It is '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 [2, 3].

The objects of supramolecular chemistry are supramolecular entities, supermolecules possessing features as well defined as those of molecules themselves. One may say that supermolecules are to molecules and the intermolecular bond what molecules are to atoms and the covalent bond [2, 3].

It was Paul Ehrlich who recognized that molecules do not act if they do not bind, thus introducing the concept of receptor [4]. Fixation requires interaction, affinity between the partners, that may be related to the idea of coordination introduced by Alfred Werner [5], supramolecular chemistry being in this respect a generalization of coordination chemistry [6]. Finally, binding must be selective, a notion that was enunciated by Emil Fischer in 1894 [7] and very expressively presented in his celebrated 'Lock-and-Key' image of steric fit, implying geometrical complementarity, that is the basis of molecular recognition.

With these three concepts—fixation, coordination and recognition—the foundations of supramolecular chemistry are laid.

Molecular interactions form the basis of the highly specific recognition, reaction, transport, regulation and other processes that occur in biology, such as substrate binding to a receptor protein, enzymatic reactions, assembling of multiprotein complexes, immunological antigen-antibody association, intermolecular reading, translation and transcription of the genetic code, regulation of gene expression by DNA-binding proteins, entry of a virus into a cell, signal induction by neurotransmitters, cellular recognition and so on. The design of artificial, abiotic systems capable of displaying processes of the highest efficiency and selectivity requires the correct manipulation of the energetic and stereochemical features of the noncovalent, intermolecular forces (electrostatic interactions, hydrogen bonding, van der Waals forces, etc.) within a defined molecular architecture. In doing so, the chemist finds inspiration in the ingenuity of biological events and encouragement in the demonstration that such high efficiencies, selectivities and rates can indeed be attained. However, chemistry is not limited to systems similar to those found in biology, but is free to create unknown species and to invent novel processes.

Supramolecular chemistry is a highly interdisciplinary field of science covering the chemical, physical and biological features of chemical species of greater complexity than molecules themselves that are held together and organized by means of intermolecular (noncovalent) binding interactions. It has been defined, conceptualized and structured into a coherent system. Its roots extend into organic chemistry and the synthetic procedures for molecular construction; into coordination chemistry and metal ion-ligand complexes; into physical chemistry and the experimental and theoretical studies of interactions; into biochemistry and the biological processes that all start with substrate binding and recognition; and into materials

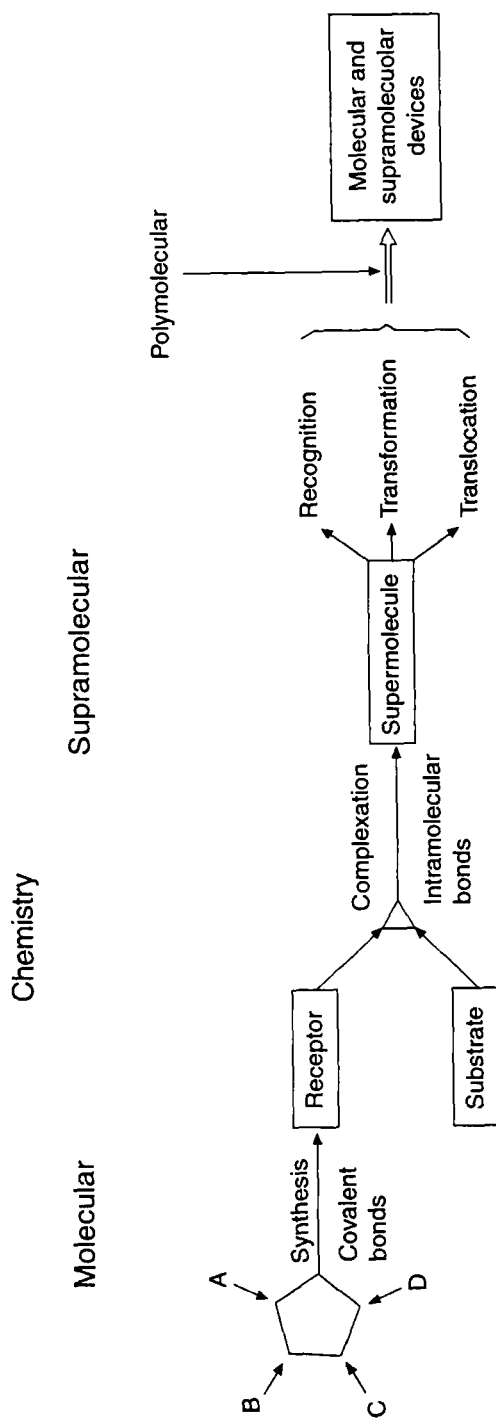


Figure 1 From molecular to supramolecular chemistry

science and the mechanical properties of solids. A major feature is the range of perspectives offered by the cross-fertilization of supramolecular chemical research resulting from its location at the intersection of chemistry, biology and physics. Drawing on the physics of organized, condensed matter and expanding over the biology of large molecular assemblies, supramolecular chemistry becomes a supramolecular science. Such wide horizons are a challenge and a stimulus to the creative imagination of the chemist.

2. RECOGNITION, INFORMATION AND COMPLEMENTARITY

Molecular recognition is defined by the energy and the information involved in the binding and selection of substrate(s) by a given receptor molecule; it may also involve a specific function. It implies a pattern recognition process through a structurally well-defined set of intermolecular interactions. A complex or supermolecule is characterized by its (thermodynamic and kinetic) stability and selectivity, i.e. by the amount of energy and information brought into operation.

Molecular recognition thus involves the molecular storage and supramolecular readout of molecular information. It implies geometrical and interactional complementarity between the associating partners, i.e. optimal information content of a receptor with respect to a given substrate. This amounts to a generalized double-complementarity principle extending over energetic as well as geometrical features the 'lock-and-key' steric fit concept of Emil Fischer [7]. It may be useful to devise a word for complementary partners. One might propose *pleromers* (from the Greek *πλήρωμα* meaning complement and *μέρος* meaning part), i.e. parts that complement each other.

The balance between rigidity and flexibility is of particular importance for the binding and the dynamic properties. Rigid 'lock-and-key'-type receptors are expected to present very efficient recognition, i.e. both high stability and high selectivity. On the other hand, flexible receptors that bind their substrates by an 'induced fit' [8] process may display high selectivity but have lower stability, since part of the binding energy is used up in the change of conformation of the receptor. Processes of exchange, regulation, cooperativity and allostery require a built-in flexibility so that the receptor may adapt and respond to changes. Flexibility is of great importance in biological receptor-substrate interactions, where adaptation is often required for regulation to occur.

It has already become apparent that information is a key notion of supramolecular chemistry, in fact the most fundamental and general one, that constitutes the common thread running through the whole field. In this respect supramolecular chemistry may be considered as a chemical information science or molecular 'informatics' concerned with the molecular storage and supramolecular readout and processing of information via the structural and temporal features of molecules and supermolecules [3, 9]. These characteristics lead to the notion of programmed supramolecular systems (see Section 5).

3. MOLECULAR RECOGNITION, CATALYSIS AND TRANSPORT

Supramolecular chemistry started with the selective binding of alkali metal cations by natural [10] and synthetic macrocyclic and macropolycyclic ligands, namely the crown ethers and cryptands [2, 3, 11, 12]. This led to the emergence of molecular recognition as a new domain of chemical research that expanded over other areas and became supramolecular chemistry [2, 3, 9]. It underwent rapid growth with the development of synthetic receptor molecules of numerous types for the strong and selective binding of cationic, anionic or neutral complementary substrates of organic, inorganic or biological nature by means of various interactions (electrostatic, hydrogen bonding, van der Waals, donor-acceptor).

Many types of receptor molecules have already been explored (crown ethers, cryptands, spherands, cavitands, calixarenes, cyclophanes, cryptophanes, etc.) [2, 3, 10–13]. Still many others may be imagined for the binding of complementary substrates of chemical or biological significance, for instance for the development of substrate-specific sensors or for the recognition of structural features in biomolecules (nucleic acid probes, affinity cleavage reagents, enzyme inhibitors, etc.).

The combination of recognition features with reactive functions generates supramolecular reagents and catalysts that operate in processes involving two main steps: substrate recognition followed by its transformation into products [3, 12b]. Because of their relationship with enzymatic catalysts, they present biomimetic features. By nature they are abiotic reagents that may perform the same overall processes as enzymes without following the same mechanistic pathways. More importantly, they may also effect highly efficient and selective reactions that enzymes do not perform. This represents a very important area for further development which may lead to a range of reactive receptor molecules combining substrate specificity with high reaction efficiency and selectivity. Much work remains to be done that should contribute very significantly to the understanding of chemical reactivity and to its application in industrial processes.

Suitably modified receptors act as carriers for the selective transport of various types of substrates through artificial or biological membranes [3]. Again, many further developments may be envisaged, concerning for instance the construction of selective membrane sensors or the transport of drugs through biological barriers, which may include targeting if suitable target-selective recognition groups are introduced.

Recognition, reactivity and transport represent the three basic functional features of supramolecular species.

A further important line of development concerns the design of molecular and supramolecular devices built on photoactive, electroactive or ionoactive components operating respectively with photons, electrons or ions [3, 9]. Thus, a variety of photonic devices based on photoinduced energy and electron transfer

may be imagined [9, 14]. Molecular wires, ion carriers and channels facilitate the flow of electrons and ions through membranes. Such functional entities represent entries into molecular photonics, electronics and ionics, which deal with the storage, processing and transfer of materials, signals and information at the molecular and supramolecular levels.

A whole field at the interface with physics, microelectronics and microoptics lies here which has barely been explored and remains wide open, presenting intriguing goals such as the development of storage (battery), amplification, switching and rectification devices. The chemistry of signal generation, processing, transfer, conversion and detection, namely semiochemistry, touches upon both physical and biological signaling processes.

4. SELF-ASSEMBLY AND SELF-ORGANIZATION

In line with the key role played by molecular information in the field, the most recent developments in supramolecular chemistry concern the explicit implementation of molecular recognition as a means of controlling the evolution of supramolecular species and devices as they build up from their components and operate through self-processes. Self-assembly and self-organization have mainly been a characteristic feature of biological processes. Supramolecular chemistry now provides ways and means for chemical science to explore this area and apply its power of design and control [3, 9].

Supramolecular chemistry has relied on more or less rigidly organized, synthetically built molecular receptors for effecting molecular recognition, catalysis and transport processes and for setting up molecular devices. Beyond such preorganization lies the design of systems undergoing self-organization; that is, systems that are capable of spontaneously generating a well-defined (functional) supramolecular architecture by self-assembly from their components in a given set of conditions. This is designed assembly into a discrete supramolecular species, unlike the spontaneous formation of molecular layers, films, membranes, etc. Considerations on self-assembly and self-organization, as well as on the content of other concepts and terminology in supramolecular chemistry, may be found elsewhere [15].

The molecular information necessary for the process to take place and the algorithm that it follows must be stored in the components and operate through selective molecular interactions. Thus, these systems may be termed programmed supramolecular systems, which generate organized entities following a defined plan based on molecular recognition events [9].

Self-assembly and self-organization have recently been achieved in several types of organic and inorganic systems [3, 9, 16–20]. By clever use of metal coordination, hydrogen bonding or donor–acceptor interactions for templating or self-assembly, the formation of a variety of novel and intriguing species has been achieved such as inorganic double and triple helices [17, 18b] termed

helicates, catenanes [18], threaded entities (rotaxanes), cage compounds [19] and inorganic ion grids [20].

Inorganic self-assembly involves the spontaneous generation of well-defined supramolecular architectures from organic ligands and metal ions. The latter serve both as cement that holds the ligands together and as centers orienting them in a given direction. In the process, full use is made of the structure and coordination features of both types of components, which in addition convey redox, photochemical or chemical functionality depending on their nature. Metal ions have properties of special interest as linkers for self-assembly. They provide a range of coordination geometries, binding strengths, and formation and dissociation kinetics, as well as a variety of photophysical and electrochemical properties.

Organic self-assembly makes use of interactions other than metal ion coordination, such as electrostatic, hydrogen bonding, van der Waals, stacking or donor-acceptor effects found in proteins, nucleic acids, liquid crystals and molecular complexes. The designed use of these forces for the directed self-assembly of a given structure requires knowledge of their strength and their dependence on distances and angles.

The spontaneous generation of organized structures depends on the design of molecular components capable of self-assembling into supramolecular entities presenting the desired architectural and functional features. The type of species obtained will result from the information stored in the components. Thus, the self-assembly process may be directed by molecular recognition between two or more complementary subunits so as to form a given supramolecular architecture. If these molecular units incorporate specific optical, electrical, magnetic or binding properties their ordering may induce a range of novel features. Depending on the subunits involved the association may lead either to supermolecules or to organized assemblies, such as membranes, molecular layers and films, mesophases, polymeric species or solid-state lattices.

5. PROGRAMMED MOLECULAR SYSTEMS

Recognition-directed self-processes may be considered to represent molecular information-handling procedures. The information necessary for a process to take place and the algorithm (the 'Aufbau' rules) that it follows must be stored in the components and operate via selective molecular interactions. These systems have been termed programmed supramolecular systems [9], a designation that stresses explicit design, steering and predictability.

Molecular programming thus involves the incorporation into molecular components of instructions whose operation through specific recognition algorithms generates the desired supramolecular architecture. The program is molecular and the information is stored in the covalent framework; its operation through noncovalent recognition interactions is supramolecular. The processing

of molecular information via molecular recognition events implies the passage from the molecular to the supramolecular level.

Three levels of information input may be distinguished: (1) molecular recognition for the selective binding of complementary components; (2) orientation in order to allow growth through sequential binding of the components in the correct relative disposition; and (3) termination of the process, requiring a built-in feature, a stop signal, that specifies the end point and signifies that the process has reached completion; this may be a closure relation generating a closed structure. In addition, temporal information may be involved if the progressive buildup of the final superstructure occurs through a defined sequence of molecular instructions and algorithms, a given component or recognition event coming into play at a well-defined stage in the total process.

Beyond programmed systems lie systems responding to external stimuli, keeping a trace of the actions and adjusting their reaction, i.e. systems capable of elementary 'learning' and of being trained.

6. SELF-RECOGNITION: INSTRUCTED CHEMICAL SYSTEMS

In addition to high efficiency, selectivity and cooperativity, another basic feature characterizing programmed supramolecular processes is 'self-recognition'—the recognition of like from unlike, of self from nonself—embodied in the spontaneous selection and preferential assembly of like components in a mixture.

With respect to inorganic self-assembly this would involve preferential binding of like metal ions by like ligands in a mixture of ligands and ions. In a study of helicate self-assembly from a mixture of different ligands and different metal cations, it was found that only the 'correct' helical complexes were formed through self-recognition [21]. This results from the spontaneous selection and preferential binding of like metal ions by like ligand strands in a mixture to assemble selectively into the corresponding helicates.

In a broader perspective, these results point to the emergence of a new outlook involving a change in paradigm, from 'pure compounds' to 'instructed mixtures', from 'unity' (pure substance) to 'multiplicity + information' (mixture of instructed components + program). Rather than pursuing mere chemical purity of a compound or a material, one would seek the design of instructed components which, as mixtures, would lead through self-processes to the spontaneous and selective formation of the desired (functional) superstructures. This may recall the side-by-side buildup of complex species (displaying highly integrated functions) that takes place in the self-assembly of the machinery of the living cell.

One may venture to predict that this instructed mixture paradigm will define a major theme of chemical research in the years to come: the spontaneous but controlled buildup of structurally organized and functionally integrated

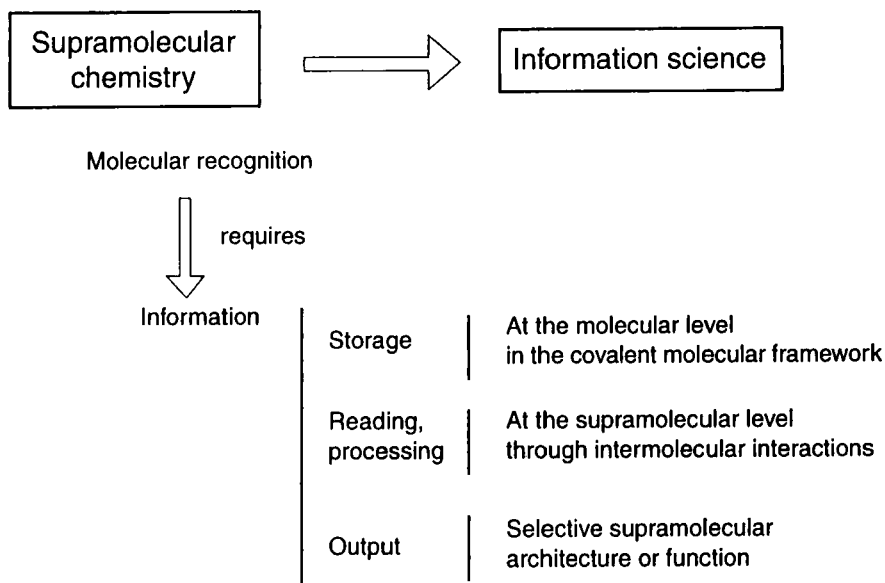


Figure 2 Supramolecular chemistry as an information science

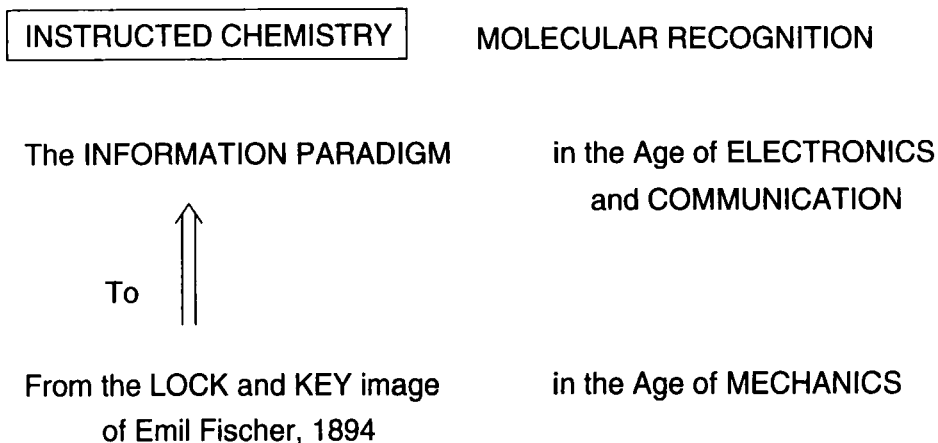


Figure 3 From the lock-and-key image to the information paradigm

supramolecular systems from a preexisting 'soup' of instructed components following well-defined programs and interactional algorithms.

7. FROM THE LOCK-AND-KEY IMAGE TO THE INFORMATION PARADIGM

It has become clear that the keyword of supramolecular chemistry is not structure but information. Thus, supramolecular chemistry is leading the way towards apprehending chemistry as an information science (Figure 2). In the 100 years since 1894, molecular recognition has evolved from the lock-and-key image of the age of mechanics to the information paradigm of the age of electronics and communication (Figure 3). It is quite appropriate that this change is occurring around the centenary of Emil Fischer's celebrated publication. The concepts of instructed chemistry and chemical information processing are here to stay. Their implementation in molecular and supramolecular structures and functions will represent a major line of development in chemistry in the next century and opens wide the perspectives in supramolecular chemistry.

8. REFERENCES

1. F. Wöhler, *Poggendorf's Ann. Phys.*, **12**, 253 (1828).
2. (a) J.-M. Lehn, *Pure Appl. Chem.*, **50**, 871 (1978); (b) *Acc. Chem. Res.*, **11**, 49 (1978).
3. J.-M. Lehn, *Angew. Chem., Int. Ed. Engl.*, **27**, 89 (1988).
4. P. Ehrlich, *Studies on Immunity*, Wiley, New York, 1906.
5. A. Werner, *Z. Anorg. Chem.*, **3**, 267 (1893).
6. J.-M. Lehn, in *Perspectives in Coordination Chemistry* (eds A. F. Williams, C. Floriani and A. E. Merbach), VHCA, Basel and VCH, Weinheim, 1992, p. 447.
7. E. Fischer, *Ber. Dtsch. Chem. Ges.*, **27**, 2985 (1894).
8. D. E. Koshland, *Adv. Enzymol.*, **22**, 45 (1960); *Annu. Rev. Biochem.*, **37**, 672 (1968).
9. J.-M. Lehn, *Angew. Chem., Int. Ed. Engl.*, **29**, 1304 (1990).
10. (a) Yu. A. Ovchinnikov, V. T. Ivanov and A. M. Skrob, *Membrane Active Complexones*, Elsevier, New York, 1974; (b) B. C. Pressman, *Annu. Rev. Biochem.*, **45**, 501 (1976).
11. C. J. Pedersen, *Angew. Chem., Int. Ed. Engl.*, **27**, 1021 (1988); D. J. Cram, *Angew. Chem., Int. Ed. Engl.*, **27**, 1009 (1988); G. Gokel, *Crown Ethers and Cryptands*, Royal Society of Chemistry, Cambridge, 1991.
12. (a) J.-M. Lehn, *Struct. Bond.*, **16**, 1 (1973); (b) J.-M. Lehn, *Science*, **227**, 849 (1985).
13. F. Diederich, *Cyclophanes*, Royal Society of Chemistry, Cambridge, 1991; C. D. Gutsche, *Calixarenes*, Royal Society of Chemistry, Cambridge, 1989; A. Collet, *Tetrahedron*, **43**, 5725 (1987); C. Seel and F. Vögtle, *Angew. Chem., Int. Ed. Engl.*, **31**, 528 (1992); for typical recent developments see *Top. Curr. Chem.*, **165** (1993).
14. V. Balzani and F. Scandola, *Supramolecular Photochemistry*, Ellis Horwood, New York, 1991.
15. J.-M. Lehn, in *Proc. of the Centenary of the Geneva Conf.* (ed. M. V. Kisakürek), VCH, Weinheim, 1993, p. 77.
16. J. S. Lindsey, *New J. Chem.*, **15**, 153 (1991); O. Philp and J. F. Stoddart, *Synlett*, 445 (1991); G. M. Whitesides, J. P. Mathias and C. T. Seto, *Science*, **254**, 1312 (1991).

17. (a) J.-M. Lehn, A. Rigault, J. Siegel, J. Harrowfield, B. Chevrier and D. Moras, *Proc. Natl. Acad. Sci. USA*, **84**, 2565 (1987); (b) J.-M. Lehn and A. Rigault, *Angew. Chem., Int. Ed. Engl.*, **27**, 1095 (1988); (c) R. Krämer, J.-M. Lehn, A. DeCian and J. Fischer, *Angew. Chem., Int. Ed. Engl.*, **32**, 703 (1993); (d) U. Koert, M. M. Harding and J.-M. Lehn, *Nature*, **346**, 339 (1990).
18. (a) J.-P. Sauvage and C. O. Dietrich-Buchecker, *Tetrahedron*, **46**, 503 (1990); (b) E. C. Constable, *Tetrahedron*, **48**, 10 013 (1992).
19. P. Baxter, J.-M. Lehn, A. DeCian and J. Fischer, *Angew. Chem., Int. Ed. Engl.*, **32**, 69 (1993).
20. P. Baxter, J.-M. Lehn, M.-T. Youinou and J. Fischer, unpublished results.
21. R. Krämer, J.-M. Lehn and A. Marquis-Rigault, *Proc. Natl. Acad. Sci. USA*, **90**, 5394 (1993).

Index

- A-type conformation 93
- N*-2-Acetamido-4-nitrophenylpyrrolidine (PAN) 227
- 8-Acetyl-10-methylisoalloxazine 79
- Acid-base catalysts 151, 163
- Acylenzyme 76
- Additives, role in crystallization 179–80
- Adenosines 94
- Affinity distribution
 - measuring experimentally 40
 - models for 36
 - nonspecific binding 42
 - open questions concerning 35
 - parameter fit 40
 - potential experimental measurement 35
 - properties of 34–5
 - universality of 40
 - see also* Receptor affinity distribution (RAD) model
- Affinity distributions 32–6
- Air–solution interface 201–5, 235
- (*R,S*)-alanine
 - crystal growth 221–5
 - crystal morphology 229
 - twinning of 228–31
- N*-Alkyl-4-dihydronicotinamides 79
- Amino acids 6, 14, 65, 85, 283, 287
 - α -amino acids, enantioselective nucleation of α -glycine by 215–19
 - (*R*)-amino acids 199
 - (*R*)- α -amino acid 195–6
 - (*S*)-amino acids 199
 - associated with codons 291
 - attachment 294–5
- Aminoacyl synthetases 9
- Aminolysis 76
- Amphiphilic alcohols, epitaxial nucleation of ice by 209–15
- Amphiphilic molecules
 - self-aggregation at air–solution interfaces 201–19
 - self-organization of 235
- Antibodies
 - as catalysts. *See* Catalytic antibodies
 - as synthetic catalysts 163–6
 - as templates 166–9
 - binding to proteins 162
 - combined with metals 169–70
 - potential to rival enzymes 162
- Antibody multispecificity 32
- Antigen–antibody interactions 25, 29, 155
- Antisense oligonucleotide 15, 16
- Antisense RNA and DNA 15
- Antiviral agents 139
- Aptamers 118–22
- Arachidamide 232, 233
- 3-Arachidamidopropionamide 232, 233
- Aromatic 2-chloroethylamines 123, 125, 128
- (*S*)-asparagine monohydrate 220, 221
- Aspartic proteinase (2APR) 48, 53, 55
- Aspartic proteinase-peptide inhibitor (3APR) 53
- Aspartyl synthetase 9
- Assembler strands 261–2
- Association constant 39
- Asymmetric synthesis in centrosymmetric host crystals 183–5
- Azo dyes 3
- B-lymphocytes 33
- B-type conformation 93
- Bacillus amyloliquefaciens* 64
- Bacillus stearothermophilus* 63
- 'Bait and switch' 167
- Binaphthyl 176
- Binding constant 26, 27, 29
- Binding energy 30–1, 185
- Biocatalysts 73–88
- Biological specificity 26–31
 - meaning of 40–4
- Biology, molecular recognition in 25–71
- Biomolecules, chemical properties of 248

- Bis(bipyridinium) tetracation 298
N-Bromoacetyl 129
 Bromobenzene 2
 α -Bromoketone 79
 2-Bromomethyl-*N*-benzylthiazolium bromide 80
 5-Bromouridine 21
 Brønsted linear free energy relation 60
 Bull testis 21
 Burnet, Sir Macfarlane 31
- Carbohydrate-lectin interactions 18–21
 Carbon-carbon double bonds 176, 183
 Carbonate hydrolysis, antibody catalysis 158
 3-Carboxy-4-nitrobenzenethiol 78
 Catabolite gene activator protein (CAP) 81
 Catalysts 311–12
 antibodies as. *See* Catalytic antibodies
 conversion of macrocycles to 150–4
 macrocycles as 149–72
 Catalytic antibodies 149–72
 development of 154–8
 enzyme characteristics 158–63
 Catenanes 3–4, 297
 Cell envelopes 279–81
 Centrosymmetric crystals 201, 235
 asymmetric synthesis 183–5
 chirality generation in 193–201
 Chemical mutagenesis 75–8
 future directions 85
 Chiral amplification in system glycine-soluble α -amino acids 235–42
 Chiral crystals
 crystallization of nonchiral molecules into 176–9
 structures 177
 Chiral cyclobutane products 178
 Chiral polymeric inhibitors 228
 Chiral solid solutions, transformation of nonchiral crystal hosts into 181–201
 Chirality 173–246, 295–6
 conformational 176
 generation and amplification in solid state 176–201
 generation in centrosymmetric crystals 193–201
 of nucleotides 285–6
 spontaneous generation 174–5
 stable molecular 176
 theories of origin 173
- Chlorobenzene *between* 1–2, 2
 Chymotrypsin 4
 α -Chymotrypsin 77
 Chymotrypsin model 2
E-Cinnamamide-(*E*)-cinnamic acid system 183–5
E-Cinnamamide-(*E*)-2-(β -thienyl)acrylamide system 185–92
 Cinnamic acids 203
cis-trans isomerization 167
 Citrate synthetase 48
 Citric acid cycle 283
 Cleft structures 297
 CMP-D-alanine 296
 Compartmentation 250
 Complementarity 310
 Computer simulations 295
 Copying errors 252, 254
 Copying mismatch 252
 COSMIC technique 66–7
 Coulombic interactions 297
 Crystal growth 175
 of α -glycine 221–5
 of (*R,S*)-alanine 221–5
 solvent effects on 220–1
 Crystal nucleation
 enantioselective inhibition of 235
 molecular recognition during 219–35
 selective inhibition of 227
 tailor-made inhibitors 225
 Crystal polymorphism 225–7
 Crystal surfaces 175
 Crystal symmetry 185
 Crystalline interfaces 173–246
 Crystallites, two-dimensional or three-dimensional 175
 Crystallization, role of additives in 179–80
 Cyclic peptide 154
 Cyclodextrin 1, 3, 5
 α -Cyclodextrin *between* 1–2, 4
 β -Cyclodextrin 2, 4
 Cysteine 75, 76, 77
 Cytosine 93
- Darwinian evolution 248, 302
 $\Delta\Delta G$ values 9
 Deoxyhemoglobin (2HHB) 53
 Deoxyribonucleotides 92
 Derivatization agents 75
 Diazoacetamide 168
 Diels-Alder reaction 163, 166

- 4,4'-Dimethoxy- β -truxinic acid 203
Dimyristoylphosphatidic acid 298
DNA 80-3, 90
 cellular 138
 double-stranded 91, 94, 95, 99, 102, 109, 114, 121
 multistranded 98
 secondary structures 97
 single-stranded 98, 101, 128
 targeting of specific sequences with oligonucleotide derivatives 104-13
 tetrameric structures 97
DNA polymerases 15, 97
DNA replicase 291
DNA sequencing 115-16, 119, 120
Double-mutant cycles 63-5, 66
- E. coli* 81, 130
EDTA 129
EDTA-Fe complex 126, 130
Electrostatic energy 30
Electrostatic interactions 55-6
Elementary free energies, additivity of 38-40
Enantiomers, resolution of 227-8
Enantioselective inhibition of crystal nucleation 235
Enantioselective nucleation of α -glycine by hydrophobic α -amino acids 215-19
Environmental changes 252-3
Enzymatic catalysis, 'lock-and-key' analogy of 73
Enzyme models 2
Enzyme-substrate complexes 29
Enzymes
 general chemical modification 74-5
 modification with cofactors 78-80
Epichlorohydrin 4
Epitaxial nucleation of ice by amphiphilic alcohols 209-15
Ester hydrolysis 2
Evolutionary gradient 250-2
- Fischer, Emil 1, 73, 90
FlavoGAPDH 79
Fluorescent neoglycoprotein 21
Food monomers 279-81
Force field 30-1
Free energy of interaction, modeling 38
Furocoumarins 125
- Genetic code
 rationale for 287-91
 Stage A 287
 Stage B, codons GNC 287-9
 Stage C, codons GNN 289
 Stage D, codons PuNN 289
 Stage E, codons NNN 290
 Stage F 290
 Stage G 290
Genetic engineering 90
(*R,S*)-Glutamic acid 191
Glutamyl synthetase 9, 12
Glutathione peroxidase 77, 78
Glyceraldehyde-3-phosphate dehydrogenase (GAPDH) 79
 α -Glycine 190
 enantioselective nucleation by hydrophobic α -amino acids 215-19
 polymorphic behaviour 225
 γ -Glycine, polymorphic behaviour 225
 α -Glycine- α -amino acids system 193-201
Glycine reductase 77
Glycine-soluble α -amino acids system 235-42
Glycoproteins 18, 19
Grazing incidence X-ray diffraction (GID) 204-5, 207, 212, 213, 219, 232-5
Guanosines 94
- H-DNA 97
HA devices 261-3, 266-8, 270, 271, 273, 281
HAE₁ device 266
HAE₂ device 266
HAE_i devices 267, 268, 271, 272, 275, 278
Hairpins 254-8
 aggregation 258-61
 conformation 258, 259
 formation 257
 loop attachment 270
 religation of 262
 replicas of 258
 translation of sequences 263-6
Hairpins-assembler strand devices. *See* HA devices
Hammett equation 60-3
Hammett plot 60-3
Hapten 155, 158, 162, 163, 168
 binding 155
 design 155
Hemoglobin (2HHB) 15, 18, 48, 52, 55, 79

- Hepatitis delta virus 128
N-Hexyl-4-dihydropyridine 79
 Histidine 77
 HIV protease 85
 $\text{HO}(\text{CH}_2)_n\text{OH}$ 233–5
 Homochirality 173
 Homopurine–homopyrimidine sequences 114
 Hoogsteen base pairs 93, 97
 Hoogsteen-type hydrogen bonding 94
 Human immunodeficiency virus (HIV-1) 139, 140
 Hybridization probes 114–15
 Hydrochloride dihydrate 228
 Hydrogen bonds 272, 297
 Hydrolysis 76
 Hydroperoxides 78
 Hydrophobic dithiols 79
 Hydrophobic interactions 55–6
 Hydroxyalkyl ester monolayers 214
 4-Hydroxybenzoic acid (HBA) 206
 4-Hydroxybenzoic acid (HBA) mono-hydrate, oriented nucleation 206–9

 Ice, epitaxial nucleation by amphiphilic alcohols 209–15
 Imidazole 2, 83, 151
 Immune system, 'emergence' from non-specific binding 44
 Immunoglobulins 33, 41, 42
 Information role in supramolecular chemistry 310, 315
 Intermolecular interactions 63–5
 Intramolecular interactions 63–5
 Iodoanthranilic acid 40
 Iodotyrosine 40
 Iodovanillin 40
 Isoleucine 6, 10, 11
 Isoleucyl synthetase 10, 11

 Lactonization reaction 162
 Lectins 21
 Leucine 240, 241
 Ligand binding, kinetics of 28–9
 Ligand-receptor complex 33
 Liquid surface diffractometer 199
 Lock and key hypothesis 1–23

 Macrocycles, conversion into catalysts 150–4
 Macrocycles as catalysts 149–72

 Magnetic locks 10–15
 Major histocompatibility complexes (MHCs) 26
 Messenger RNA (mRNA) 281
 Methemoglobin (2MHB) 53, 55
 4-Methoxy-(*E*)-cinnamamide 203
 4-Methoxycinnamic acid (MCA)
 crystal nucleation 201–5
 molecular packing characteristics 204
N-Methylarachidamide 233
 Methylchymotrypsin 77
 Methyl 4-nitrobenzenesulfonate 77
 Methyl 4-nitro-*trans*-cinnamate 167
N-Methylporphyrin 170
 Michaelis-Menten catalysis 154
 Michaelis-Menten kinetics 151, 154
 Michaelis-Menten microreversibility 9
 Molecular assemblies 300
 Molecular biology 90
 Molecular devices 113–22
 Molecular information 312, 314
 Molecular programming 313–14
 Molecular recognition 31, 175, 310–12
 during crystal nucleation 219–35
 in biology 25–71
 Molecular surface complementarity, geometric algorithm for identification of 48–53
 Molecular tubes 4, 6
 MOPC315 83
 Multiple-mutant cycle analysis 65–7
 Mutational flow diagram 61

 NAD 79
 NADH 79–80
 Nanocrystals, effect of auxiliaries 231–5
N-palmitoyl-(*R*)-lysine 219
 Nicotinamide 79
 Nicotinamide adenine dinucleotide (NAD^+) 78
 4-Nitrophenylester substrate 158
 NPN43C9 158, 159, 170
 Nuclear Overhauser enhancement spectroscopy (NOESY) 67
 Nucleic acids 89
 affinity modification 123–8
 cellular 138
 chemical modifications of 125
 double-stranded 90
 oligonucleotide derivatives as inhibitors of 131–40
 principles of recognition of 90

- sequence-specific chemical modification 128–31
- single-stranded 90
- specificity of interaction of oligonucleotide derivatives with 107–10
- structure of 91–8
- targeting 98–104
- virus-specific 140
- Nucleotides 283
 - chirality of 285–6
- 4-(Octadecyl)-oxybenzoic acid monolayer 208
- 4-(Octadecyl)-oxycinnamic acid 202, 205
- Olfactory coding 45
- Olfactory receptor repertoire 44–5
- Olfactory threshold polymorphisms 45
- Oligonucleotide addresses 111–13
- Oligonucleotide analogues 98–103, 122
 - potential chemotherapeutic applications 137
- Oligonucleotide-based techniques 113–22
- Oligonucleotide binding 81, 102
- Oligonucleotide derivatives
 - alkylating 135
 - as inhibitors of nucleic acids and proteins 131–40
 - cellular targets 135–40
 - covalent crosslinking 133
 - for targeting nucleic acids 98–104
 - photoreactive 125
 - reactive 130, 131
 - reactive groups 123–8
 - specificity of interaction with nucleic acids 107–10
 - targeting of specific sequences in RNA and DNA 104–13
 - targets for 132
- Oligonucleotide libraries 122
- Oligonucleotide-directed mutagenesis 116–17
- Oligonucleotide-polynucleotide interactions, kinetics and thermodynamics 105–7
- Oligonucleotides 89–147
 - antiviral activity 139
 - as specific tags for labeling other macromolecules and constructions 115
 - attachment of functional groups to 103–4
 - cholesterol-conjugated control 139
 - cooperative interactions 130
 - imperfect complexes of 109
 - in genome research 115
 - in sequencing of nucleic acids 115–16
 - intercalating groups stabilizing complementary complexes of 110
 - polylysine-conjugated 139
 - reactive derivatives of 90
 - recognition potential 90
 - targeting potential 90
 - triple-helix-forming 130–1
- Oligosaccharides 18
- Origin of life 247–306
 - attempts at realization of replicating strands 283–5
 - availability of suitable monomers 283–5
 - fundamental engineering aspects 253–81
 - general premises of model 249–53
 - identification of components of model with entities of molecular biology 281–2
 - modeling 249
 - spatial and temporal environmental requirements 282–3
- Oxaloacetate, decarboxylation of 154
- 6-Oxoheptanal 80
- Pancreatic trypsin inhibitor 48
- Papain 77, 79, 80
- Peptide ligation 77
- Peptide nucleic acids (PNAs) 101, 135
- Phenylalanine 229
- Phenylmethanesulfonyl fluoride (PMSF) 75
- Phosphates 93
- Phosphodiester bonds 100
- Phosphodiester internucleotide connection 100
- Phosphodiester oligonucleotides 139
- Phosphodiester–phosphoramidate oligonucleotide derivatives 100
- Phosphonate 158
- Phosphoramidate linkages 100
- Phosphorothioate oligonucleotides 100, 139
- Polar versus nonpolar crystals 225–7
- Poly[*p*-acrylamido-(*S*)-phenylalanine] 228
- Poly(*N*-acryloyl-12-dodecanoic acid) 227
- Polyamide oligomers 101
- Poly(ethylene glycol) 4
- Polymerase chain reaction (PCR) 10, 117
- Polymorphic systems 225–7

- Pore openings, diversity in sizes 250
- Porphyrins 170
- Position-sensitive detector (PSD) 198
- Prebiotic chemistry 248
- Prebiotic conditions 248
- Printing on a molecular scale 301
- Probabilistic interactions 31–2
- Progenote, reconstruction of 292
- Programmed molecular systems 313–14
- Protein association, conceptual model of 29
- Protein-based catalysts 74
- Protein catalysts, future directions 84
- Protein-engineering experiments 56–9
- Protein–ligand binding 26
- Protein–ligand complexes 29
 - application of algorithm to 53–5
 - association free energies 28
 - crystallographic data 48
 - geometric matching 47
 - predicting formation of 46
- Protein–ligand interactions 25–71
 - geometric fit 46–56
 - molecular surface recognition in 46–8
 - nonadditivity in 60–3
- Protein–ligand recognition 30–1
- Protein–protein complexes, crystallographic data 48
- Protein–protein interactions 25
 - additivity in 59–60
 - total free energy 65
- Proteins
 - affinity modification 123–8, 131
 - oligonucleotide derivatives as inhibitors of 131–40
- Psoralene 125
- Psoralene oligonucleotides 17
- Purine–pyrimidine sequences 93
- 4-Pyridyl-acetic acid, decarboxylation of 168
- Pyrimidines 102
- Pyrophosphate splitting 3
- Rabbit reticulocyte system 15, 18
- Racemic additives 198
- Racemic mixtures 174
- Receptor affinity distribution (RAD) model 36–40
 - experimental fitting and physiological significance of 40
 - inferences from 40–4
 - insight gained from particular systems 44–5
 - signal-to-noise ratio between specific and nonspecific binding 43
 - specificity versus nonspecificity in biological recognition 42–4
- Receptor–ligand binding 29
- Receptor–ligand complex 26
- Receptor–ligand interaction 30–1
 - thermodynamics 29–30
- Receptor repertoires 32–3
 - basic questions related to 40–1
- Replication catalysts 266
- Replication with parallel and antiparallel attraction of monomers 255–6
- α -Rhamnose monohydrate 221, 222
- D-Ribose 285
- L-Ribose 285
- Ribosomal RNA (rRNA) 128, 281
- RNA 80–3, 90, 91
 - double-stranded 91, 93
 - multistranded 98
 - prior to DNA hypothesis 291
 - secondary structures 97
 - self-cleaving structures 127
 - single-stranded 98, 101, 113, 132
 - targeting of specific sequences with oligonucleotide derivatives 104–13
- RNA-based catalysts 74
- RNA-cleaving RNA structures 127
- RNA replicase 291
- RNase P 128
- Rotaxanes 297
- Safety locks 4–9
- Second harmonic generation (SHG) 227
- Selective theory of antibody action 31
- Selenium 77, 78
- Selenoenzymes 78
- Selenosubtilisin 77–8
- SELEX 118
- Self-aggregation 201–19
- Self-assembly 175, 312–13
- Self-organization 235, 297, 298, 312–13
- Self-recognition 314
- Semisynthetic catalysts 83–4
- Semisynthetic enzyme 75, 76
- Semisynthetic selenoenzyme 78
- (*R,S*)-Serine 198
- Serine proteases 75, 76
- Seryl synthetase 9, 12
- Shape space 36
- Single crystals 178, 201, 235
- Site-specific nucleases 80–3

- Solvent effects on crystal growth of solvates 220–1
- Specificity
 concept of 26
 quantitative definition 27
- Spermatozoa 21
- Stacking interactions 297
- Stem recognition 291
- Stereospecificity, quantitative approaches 36
- Strand length, probability aspects 253–4
- Subtilisin 75, 77, 85
- Superstrands 275–9
 emergence of D-superstrand 275–6
 replication of D-superstrands 277
 replication, transcription and translation separated 277–9
 supramolecular engineering 279
- Supramolecular chemistry 307–16
 basic functional features 311
 features of 308
 foundations of 308
 information role in 310, 315
 molecular background 307–10
 objects of 308
- Supramolecular devices 248
- Supramolecular engineering 249, 279, 296–303
 current paradigm 299–300
 future development 299–302
 search for new paradigms 301–2
 state of the art 296–9
 strategies based on lock-and-key and programmed-environmental-change principles 298
 strategies based on lock-and-key principle 296–8
 superstrands 279
- Supramolecular reagents 311
- Surface tension measurements 201
- Target sites, selection of 111–13
- Temperature changes, reproduction driven by intricate cycles of 250
- Template-assisted strand replication 264
- Template-directed copying 254
- Template-directed polymerization 283–4
- Template-directed reproduction 249–50
- Template strand 249
- Tetrachlorothiophene dioxide (TCTD) 163
- Thiazolopapain 80
- Thiol glutathione 78
- Thiolsubtilisin 75, 77
- Threading-in process 2
- Three-dimensional (3D) structures from two-dimensional templates 300
- Three-dimensional (3D) symmetry 181
- Threonine 229
- Thymidine 94
- Tick-borne encephalitis virus 139
- Tobacco ringspot 127
- trans-cis* photoisomerization 203
- Transfer RNA (tRNA) 281, 292–4
 aminoacylation of 6
- Translation machines 266–8
- Transport 311–12
- Triplet attachment 293–4
- Tri-*o*-thymotide 176
- Trypsin 48, 53, 77
- Trypsin-trypsin inhibitor (2PTC) 53, 55
- Twinning of (*R,S*)-alanine 228–31
- Two-dimensional (2D) crystal precipitation 205
- Two-dimensional (2D) crystal structures 214
- Two-dimensional (2D) symmetry 181
- Two-dimensional (2D) templates 300
- Tyrosine adenylate 48
- Tyrosyl tRNA synthetase 48
- Tyrosyl tRNA synthetase-tyrosine adenylate (3TSl) 53
- Valine 6
- van der Waals energy 30
- van der Waals forces 272
- van der Waals interactions 56
- Vancomycin 83
- Viral RNAs 139
- Watson–Crick base pairs 92–5, 114
- Watson–Crick binding 98
- Watson–Crick recognition scheme 90
- X-ray diffraction geometry 198
- Yeast phenylalanine tRNA 6
- Z-conformation 93
- Z-DNA 93, 97