Ciba Foundation Symposium 158

HOST-GUEST MOLECULAR INTERACTIONS: FROM CHEMISTRY TO BIOLOGY

Ciba Foundation
Ciba Foundation
Ciba Foundation
Ciba Foundation
Ciba Foundation
Ciba Foundation
Ciba Foundation
Ciba Foundation
Ciba Foundation
Ciba Foundation
Ciba Foundation
Ciba Foundation

HERIOT-WATT
UNIVERSITY
LIBRARY
EDINBURGH

HERIOT-WATT UNIVERSITY
3 8007 00108 4889

WITHDRAWN

6/6/91

HERIOT-WATT UNIVERSITY LIBRARY

You should bring this book back by the last date on this label. You may be asked to return it sooner, if someone else asks for it to be recalled.

15. JAN. 1993
-3. JUN. 1994
29. NOV. 1996
24. JAN. 1997
-7. MAR 1997
-2 MAY 1997
-5. JUN 1998
23. JAN 1998
-5. JUN. 1998

16 OCT 1998
13 OCT 2000
23 APR 2004
-1 JUN 2007
30 MAY 2008

HOST-GUEST MOLECULAR INTERACTIONS: FROM CHEMISTRY TO BIOLOGY

The Ciba Foundation is an international scientific and educational charity. It was established in 1947 by the Swiss chemical and pharmaceutical company of CIBA Limited—now CIBA-GEIGY Limited. The Foundation operates independently in London under English trust law.

The Ciba Foundation exists to promote international cooperation in biological, medical and chemical research. It organizes about eight international multidisciplinary symposia each year on topics that seem ready for discussion by a small group of research workers. The papers and discussions are published in the Ciba Foundation symposium series. The Foundation also holds many shorter meetings (not published), organized by the Foundation itself or by outside scientific organizations. The staff always welcome suggestions for future meetings.

The Foundation's house at 41 Portland Place, London W1N 4BN, provides facilities for meetings of all kinds. Its Media Resource Service supplies information to journalists on all scientific and technological topics. The library, open five days a week to any graduate in science or medicine, also provides information on scientific meetings throughout the world and answers general enquiries on biomedical and chemical subjects. Scientists from any part of the world may stay in the house during working visits to London.

Ciba Foundation Symposium 158

HOST-GUEST MOLECULAR INTERACTIONS: FROM CHEMISTRY TO BIOLOGY

A Wiley-Interscience Publication

1991

JOHN WILEY & SONS

Chichester · New York · Brisbane · Toronto · Singapore

HERIOT-WATT
UNIVERSITY
LIBRARY
EDINBURGH

14·0256/02.

©Ciba Foundation 1991

Published in 1991 by John Wiley & Sons Ltd.
Baffins Lane, Chichester
West Sussex PO19 1UD, England

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, G.P.O. Box 859, Brisbane,
Queensland 4001, 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

Suggested series entry for library catalogues:
Ciba Foundation Symposia

Ciba Foundation Symposium 158
ix + 278 pages, 80 figures, 21 tables, 15 structures

Library of Congress Cataloging-in-Publication Data
Host-guest molecular interactions: from chemistry to biology.
p. cm.—(Ciba Foundation symposium; 158)
Proceedings of the Symposium on Host-Guest Molecular Interactions: from Chemistry to Biology held at the Ciba Foundation, London, Jul. 3–5, 1990.
Includes bibliographical references and index.
ISBN 0-471-92958-1
1. Ligand binding (Biochemistry)—Congresses. 2. Drugs—Design-Congresses. 3. Protein engineering—Congresses. 4. Protein binding—Congresses. I. Symposium on Host-Guest Molecular Interactions: from Chemistry to Biology (1990: Ciba Foundation) II. Series.
QP517.L54H67 1990
574.19'245—dc20 91-330
CIP

British Library Cataloguing in Publication Data
Host-guest molecular interactions: from chemistry to biology.—(CIBA Foundation Symposia)
I. Sutherland, I. O. II. Series
541.2

ISBN 0 471 92958 1

Phototypeset by Dobbie Typesetting Limited, Tavistock, Devon.
Printed and bound in Great Britain by Biddles Ltd., Guildford.

Contents

---

*Dr Montenay-Garestier was unable to attend the symposium. Her paper was presented by Dr Michael Waring.

Participants

W. J. Albery University College, Oxford OX1 4BH, UK

C. I. Brändén Swedish University of Agricultural Sciences, Department of Molecular Biology, Uppsala Biomedical Center, PO Box 590, S-751 24 Uppsala, Sweden

E. Breslow Department of Biochemistry, The New York Hospital–Cornell Medical Center, 1300 York Avenue, New York, NY 10021, USA

R. Breslow Department of Chemistry, Columbia University, New York, NY 10027, USA

R. I. Carey Institut de Chimie Organique, Université de Lausanne, Rue de la Barre 2, CH-1005 Lausanne, Switzerland

M. H. Caruthers Department of Chemistry and Biochemistry, University of Colorado, Boulder, CO 80309-0215, USA

A. Collet Stéréochimie et Interactions Moléculaires, Ecole Normale Supérieure de Lyon, 46 Allée d'Italie, F-69364 Lyon cedex 7, France

M. J. Crumpton Imperial Cancer Research Fund Laboratories, PO Box 123, Lincoln's Inn Fields, London WC2A 3PX, UK

J. D. Dunitz Laboratorium für Organische Chemie, ETH-Zentrum, Universitätstrasse 16, CH-8092 Zürich, Switzerland

A. B. Edmundson Harrington Cancer Center, 1500 Wallace Boulevard, Amarillo, TX 79106, USA

G. W. Gokel Department of Chemistry, University of Miami, Coral Gables, FL 33124, USA

C. D. Gutsche Department of Chemistry, Texas Christian University, Fort Worth, TX 76129, USA

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

R. M. Kellogg Department of Organic Chemistry, University of Groningen, Nijenborgh 16, Groningen 9747 AG, The Netherlands

O. Kennard University Chemical Laboratories, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, UK

R. U. Lemieux Department of Chemistry, Faculty of Science, University of Alberta, E5-52 Chemistry Building East, Edmonton, Alberta, Canada, T6G 2G2

W. Müller Institute of Organic Chemistry, University of Mainz, J J Becher Weg 18-22, D-6500 Mainz 1, Federal Republic of Germany

D. Parker Department of Chemistry, Science Laboratories, University of Durham, South Road, Durham DH1 3LE, UK

J. Rebek Department of Chemistry, Massachusetts Institute of Technology, Building 18-207, 77 Massachusetts Avenue, Cambridge, MA 02139, USA

D. N. Reinhoudt Laboratory of Organic Chemistry, Twente University, PO Box 217, NL-7500 AE Enschede, The Netherlands

G. C. K. Roberts Department of Biochemistry, University of Leicester, Adrian Building, University Road, Leicester LE1 7RH, UK

C. Sacht (*Ciba Foundation Bursar*) Department of Chemistry, University of Cape Town, Private Bag, Rondebosch 7700, Cape Province, South Africa

J. F. Stoddart Department of Chemistry, University of Sheffield, Sheffield S3 7HF and The School of Chemistry, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK

I. O. Sutherland (*Chairman*) Department of Chemistry, University of Liverpool, PO Box 147, Liverpool L69 3BX, UK

J. G. Vinter Molecular Recognition Unit, University Chemical Laboratories, University of Cambridge, Lensfield Road, Cambridge CB2 1EW and Department of Chemistry, University College London, 20 Gordon Street, London WC1H 0AJ, UK

J. P. Waltho Department of Molecular Biology and Biotechnology, University of Sheffield, Sheffield S10 2TN, UK

M. J. Waring Department of Pharmacology, University of Cambridge, Tennis Court Road, Cambridge CB2 1QJ, UK

Introduction

Ian Sutherland

Department of Chemistry, University of Liverpool, PO Box 147, Liverpool L69 3BX, UK

This meeting is ambitious in its breadth; it ranges from biology to chemistry, but I suspect the actual scope of the meeting will be defined by the papers and the interests of the participants. Otherwise, it would be impossibly broad. The origins of the meeting date back to my postgraduate days in the 1950s, when, unknown to me, many of the foundations of molecular biology were being laid down in nearby laboratories. At that time, the crystal structures of proteins were being examined and the double helix of DNA was discovered. Molecular biology really took off from there and has grown explosively ever since. A number of molecular biologists are here to represent the biological end of the spectrum. I believe that the basic principles are chemical and that there is complete overlap of biology and chemistry, but I have to recognize that different people emphasize different aspects of our topic. In the 1950s, again unknown to me, work on clathrates was proceeding in Oxford. H.M. Powell began the work on clathrates and host–guest chemistry in the solid state, and this is now an active and expanding area. At this meeting we are not going to talk much about clathrates, but we will talk a great deal about complexes in solution. Also in the 1950s, Fritz Cramer was working on cyclodextrins. Cyclodextrin chemistry was the first example of host–guest chemistry using relatively small host molecules, but this work did not attract as much attention at the time as it deserved.

Organic chemists became interested in the field of host–guest chemistry after the discovery of crown ethers in 1967. The impact crown ethers had was quite considerable, but the ripple created then has since grown enormously. Crown ethers have sparked off a great deal of synthetic effort and imagination in the chemical community; new compounds and families of compounds have appeared, many of which will feature in this meeting.

It certainly seemed appropriate to bring these areas together in a single meeting. Intermolecular interactions and molecular assemblies will play an extremely important role in chemistry in the future. For biologists, of course, they have always been the most important thing—all biology is based on molecular assemblies—but they are now coming into the orbit of the synthetic chemist. We shall hear about methods for making organized assemblies of synthetic molecules, which may have specific purposes. New materials and molecular electronics are certainly going to require synthetic molecular assemblies

and possibly also assemblies based on biomolecules. Measurement science is related to this area and molecular sensors are particularly relevant. These involve transport, response, selective detection and a process that is termed transduction. Transduction has its counterpart in biology, but biology has a wider range of transduction processes than I believe are currently used in molecular sensors, so maybe we shall hear some interesting comments from our molecular biology colleagues about transduction and response. At this stage, sensors use both synthetic host molecules and biomolecules; responses can be optical or electrical, but electrical responses will probably feature more prominently in discussion than optical responses. Work on sensors may also provide ideas for those of us who are becoming interested in molecular electronics as a science of the future.

The underlying theme of this meeting is molecular recognition involving small molecules and large molecules. We shall range from ionophores and ions, perhaps the simplest system of recognition but nevertheless quite difficult to achieve, through host–guest chemistry involving synthetic hosts, to the vancomycin group of antibiotics, host molecules that Nature has made, which are comparable in size to those made by synthetic chemists but which function in a more elegant way. Proteins and DNA, it goes without saying, play a major role in molecular recognition. A number of chemists still regard these recognition processes as essentially 'lock and key', a term that has been with us for a long time. The lock and the key are of course much more complex than a real lock and key, not only because they are molecular, which might define their dimensions and fit, but also because they have dynamic properties—they do not necessarily exist in single conformations.

Virtually everything we shall look at takes place in solvents, and solvation must play a very important role in molecular recognition. I hope that we shall be able to shed some light on solvation effects during the meeting. These are difficult effects to discuss in detail because they involve charges. All of us who are interested in putting molecules together are aware of charge distribution and charge–charge interactions, as well as solvation. That takes us into the real world, as opposed to the simplified model of molecules interacting in the gas phase.

We can't avoid thermodynamics if we are to understand intermolecular interactions. Ultimately, virtually all science is based upon numbers, and the numbers that characterize the binding between molecules have their basis in enthalpy and entropy, areas often avoided by organic chemists. I hope that there are participants here who can enlighten us about the details of intermolecular interactions, particularly regarding solvation effects and charge distribution, which are key factors when molecules come together.

We shall hear about recognition by small molecules. Although someone may describe a more elegant performance by a small molecule, I think that the vancomycin antibiotics are extremely well designed host molecules. The credit in this case has to go to evolution and not to the chemist, but chemists have

also been smart in their molecular design. I draw your attention to synthetic molecular clefts and ditopic systems; these are simple strategies which generate good complexes, guest recognition and catalysis of reaction—things that chemists hope to achieve with their synthetic host molecules. A lot of progress has been made in this area even if we haven't quite reached the level of sophistication of vancomycin.

Molecular recognition by large molecules will also feature. I don't think that many of us are able to design large molecules comparable with DNA or proteins. We tend to think of DNA as a double helix, but of course it is also an interesting host molecule, although Nature may not have had that in mind when evolving its structure. DNA is an excellent host molecule because it possesses a lot of 'sticky' functional groups on its surface.

I believe that no chemist can yet design proteins that will act as hosts. Possibly we shall eventually be able to design protein sequences that are different from natural sequences, and we shall hear some approaches to this. We know a lot about the structure of proteins in the solid state and, through NMR studies, about structures of proteins in solution. We even know something about the dynamics of proteins and we can model protein dynamics, given large enough computers and good enough graphics. Despite this, although a lot of progress has been made, we are still unable to predict from a sequence exactly how a protein will fold. Folding is a difficult topic because it involves kinetics as well as equilibria. How do proteins fold and how can they select the conformation that they eventually find from the millions of potential conformations that exist? There is no possibility of equilibrium—the process has to be controlled by kinetics. This is a problem that people like myself who work with simple molecules never have to face, yet it is obviously a serious one for protein chemists. I doubt whether we can solve this problem here, but we may well hear some views on the subject, and on how close we are to protein design. Studying synthetic peptides is one approach to understanding proteins. I believe that we shall find it possible to predict peptide–peptide interactions and approach protein folding using this information.

Last, but certainly not least, we come to molecular models. All scientists use models of varying complexity. In chemistry, we are forsaking the classical space-filling molecular models that we used to hold in front of us at meetings, because they are cumbersome and are not numerate—they are good analogue computers but they don't provide numbers. We are now adopting the more sophisticated molecular modelling provided by computer graphics. We will not cover quantum chemistry, although it provides us with the parameters for molecular mechanics and dynamics, but we shall certainly hear about models that are based on mechanics and dynamics.

I have already referred indirectly to the problem of conformational space. We are perplexed by the amount of conformational space that is available to even the small flexible molecules that we make. Some of us, perhaps wisely,

concentrate on molecules in which conformational space is restricted by structural rigidity. I have referred already to charge distribution and solvation. These may feature in our models, but in some cases our models are related more to the gas phase. Molecular models will certainly appear from time to time during the meeting, and we shall hear an account of molecular modelling at the end. The question I put to you is, how successful are we in making predictions using our models? Such predictions, I believe, are less common than the rationalization of the experimental facts that we learn in the laboratory.

I have covered the content of our meeting, but I am sure you all have your own ideas about what we should discuss. The scope is, to me, frighteningly broad. As a chemist, I view biology as impossibly complex, but there may be some molecular biologists here who are not entirely happy with synthetic organic chemistry! We shall see how we get on with one another.

Molecular self-assembly processes

J. Fraser Stoddart

*Department of Chemistry, The University, Sheffield S3 7HF, UK**

Abstract. A logical and stepwise approach to the establishment of the concept of self-assembly in the synthesis of wholly unnatural products is proposed. The approach is based on the use of irreversibly interlocked molecular systems in the shape of catenanes and rotaxanes as the vehicles through which to transfer from host–guest chemistry the knowledge and experience gained on relatively small molecules to much larger molecules including polymers in which the molecular components are reversibly intertwined. The proposal presents a manifesto for making the transformation from supramolecular to polymolecular chemistry. A number of recent template-directed syntheses of catenanes and rotaxanes are presented as examples of structure-directed synthesis, to illustrate that there are inherently simple ways of making apparently complex unnatural products from appropriate substrates without the need for reagent control or catalysis.

1991 Host–guest molecular interactions: from chemistry to biology. Wiley, Chichester (Ciba Foundation Symposium 158) p 5–22

The art of chemical synthesis is about to undergo one of the most profound changes of a conceptual nature that it has witnessed during its long and impressive development as a scientific discipline. The reason for this renaissance is that more and more chemists are beginning to appreciate that the conventional synthetic methodology of making compounds atom by atom and group by group, employing reagents or catalysts to make or break covalent bonds and so manipulate functional groups and transform molecular structures is not going to be enough. In all its many synthetic pathways and transformations, Nature does not have to rely on the use of protecting groups in the manner and practice of synthetic organic chemistry. Indeed, most biological systems contain the information that is necessary for their own chemical formation and physical organization. One of the keys to the efficient operation of biological systems is their ability to self-replicate and self-assemble. In essence, the chemist needs to learn, in principle at least from Nature, how to assemble compounds molecule by molecule from molecular components. Physicists with an eye to the future and looking particularly to the age of the molecular computer have been imploring

**Present address*: The School of Chemistry, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK

chemists to address the problems of self-assembly and self-replication for some time now. In his search for a modular chemical approach to molecular electronic devices Carter (1984) commented that 'this challenge will probably only be mastered when we have learned the principles of self-organization and self-synthesis from the biological world and have applied them more broadly to both organic and inorganic chemistry'. There is little doubt that a new science—call it molecular cybernetics if you wish—awaits development. It will be involved intimately with a fundamental investigation of the chemical sciences, which will enable the future development of mechanoelectrical and photoelectrical communication systems and devices (Lehn 1988, Balzani & Scandola 1990).

I am convinced that it is from an appreciation of host–guest molecular interactions (Cram 1988) that a modular chemical approach to the construction of new materials will emerge in the near future. The approach will be based on the idea of assembling carefully designed molecular components in a template-directed manner. The molecular components will have to be held together not by classical covalent bonds, but rather by non-convalent bonds. However, the problem with the best-understood chemical systems that display host–guest interactions is that they are small compared with the biological systems that exhibit molecular recognition. Some kind of bridge or scaffolding between the biological and chemical worlds has to be built. It occurred to us that the kind of self-assembly process employed in biological systems might be easier to establish in chemical systems if the relatively small molecules contained an element of irreversible mechanical entanglement in addition to the non-covalent bonding interactions that would also have to be present between the molecular components of the man-made assemblies. It seemed prudent to begin by exploring the mechanical principles of interlocking between molecular components and then go on to relax this irreversible mode of molecular association to the more reversible ones of threading and intertwining. By following this logical thought process we were led inexorably to consider the possibility that molecular compounds known as catenanes and rotaxanes (Schill 1971, Dietrich-Buchecker & Sauvage 1987) could serve as prototypes for the construction of large, ordered and structured molecular assemblies. When a guest is complexed by a host (Fig. 1a), the 1:1 complex is usually in equilibrium in solution with the molecular components. On the other hand, formation of a rotaxane (a ring compound with a chain through it that is terminated by large groups) by either a threading (Fig. 1b) or a clipping (Fig. 1c) process will be irreversible, as will the formation of a catenane (a linked-ring compound) by a clipping process (Fig. 1d). I shall now show how weak, non-covalent bonding interactions between molecular components not only preside over the synthesis of molecular assemblies, but also persist in these molecular assemblies that they help to create. This is an obvious, but nevertheless important, feature of these new chemical systems. The order they acquire during their formation can obviously become the basis for the subsequent storage and transfer of information at a molecular level.

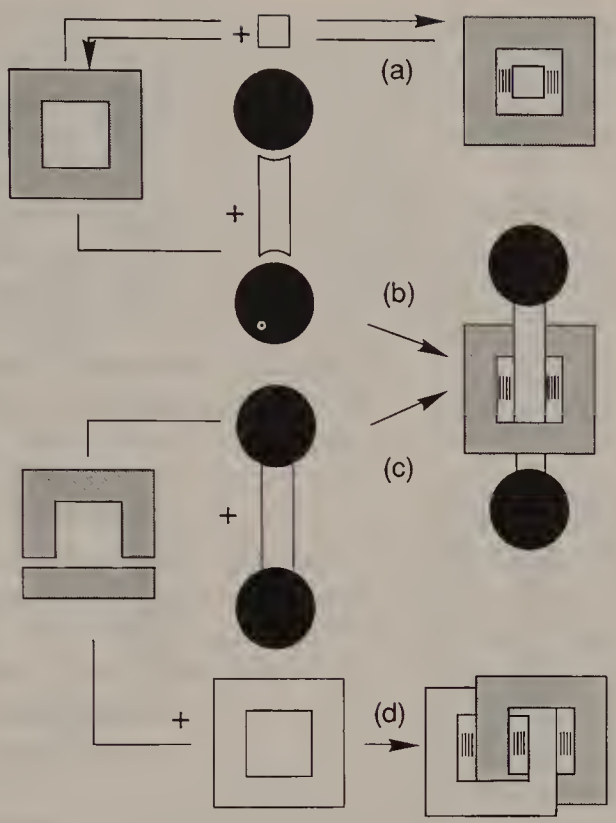


FIG. 1. Diagrammatic representation of (a) formation of a reversible 1:1 complex between a host (shaded) and a guest, maintained by non-covalent interactions that are illustrated as vertical striations; (b) production of a rotaxane, a compound formed by a threading process in which an extended guest is covalently capped with large groups (filled circles) after it has threaded itself through the ring-shaped host, preventing reversal; (c) production of a rotaxane by a clipping process in which the complete guest molecule complexes with a partially formed ring, which is then closed; (d) production of a catenane composed of two linked macrocyclic rings from a complete ring and a clipped ring (shaded) which is completed after complexation.

Background

We had established during the 1980s that the redox-active bipyridinium dication (Summers 1980) paraquat $[PQT]^{2+}$ forms a 1:1 complex with the cyclophane-like macrocyclic polyether, bisparaphenylene-34-crown-10 (BPP34C10) (Fig. 2; Stoddart 1988). The stability constant ($K_a = 730\ M^{-1}$) in acetone for this deep orange-coloured complex, which is formed at room temperature when the counter ions are soft hexafluorophosphates (PF_6^-), corresponds to a complexation free energy of 16.3 $kJmol^{-1}$. The complex $[BPP34C10.PQT]^{2+}$ is stabilized by (i) electrostatic interactions, including [C—H···O] hydrogen bonding between the guest dication and the crown ether oxygen atoms in the host and (ii) dispersive forces, including charge transfer interactions between the

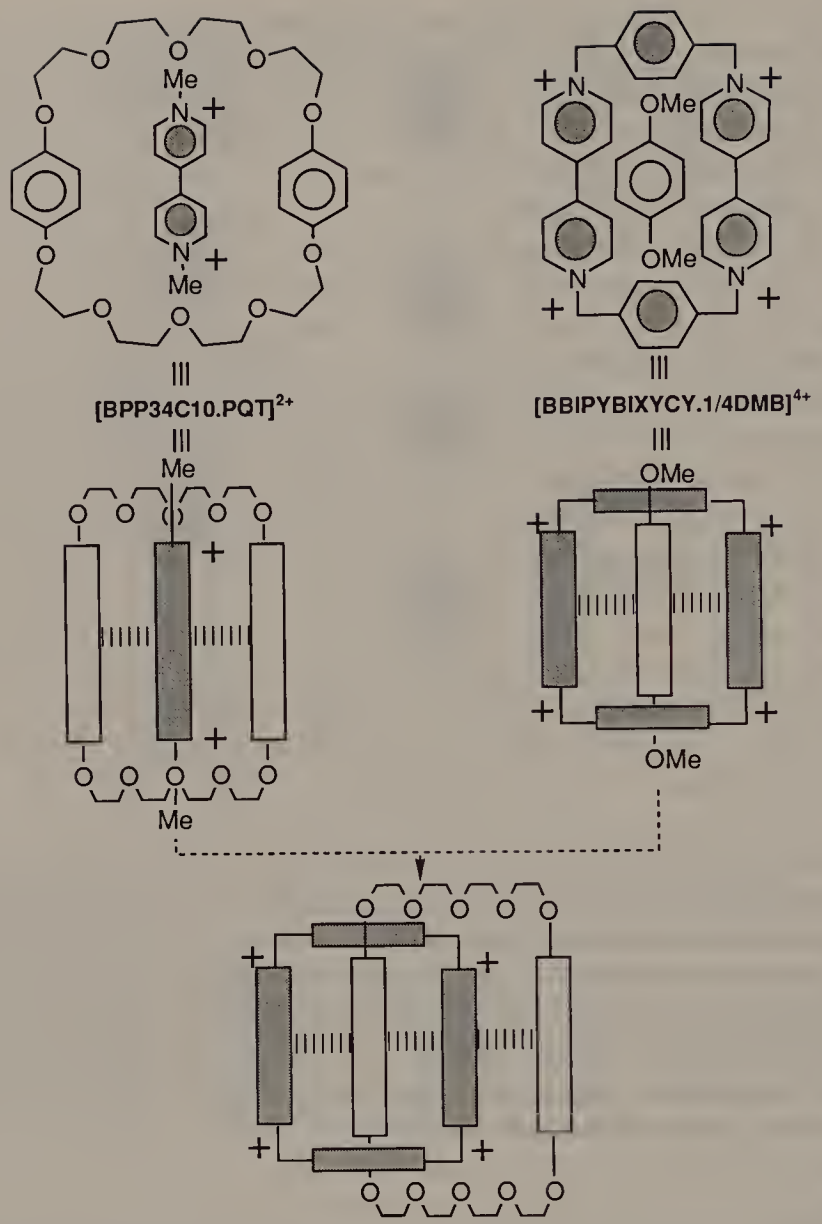


FIG. 2. The host–guest chemistry behind the design of a [2]catenane. The formation of complexes between PQT and BPP34C10 and between BBIPYBIXYCY and 1/4DMB suggested a strategy for the production of a [2]catenane (*bottom*), {[2]-[BPP34C10]-[BBIPYBIXYCY]-catenane}$^{4+}$ (see Fig. 3 for the synthesis of this catenane). BPP34C10, bisparaphenylene-34-crown-10; PQT, paraquat; BBIPYBIXYCY, a bisbipyridinium-bixylylcyclophane; 1/4DMB, 1,4-dimethoxybenzene.

π-electron-accepting bipyridinium moiety of the $[PQT]^{2+}$ dication and the π-electron-donating hydroquinol rings of BPP34C10. The charge transfer interactions account for the deep orange colour of the crystalline 1:1 complex, which has been fully characterized by X-ray diffraction studies in the solid state.

The centrosymmetrical structure has a rotaxane-like character with the $[PQT]^{2+}$ dication inserted through the middle of the BPP34C10 ring; such structures can be termed pseudorotaxanes. The distance between the mean planes of the parallely aligned π-donating hydroquinol residues and the π-accepting bipyridinium ring system is 3.7 Å.

The next question we asked ourselves was (Fig. 2) whether we could reverse the roles of the paraquat residues and the hydroquinol rings by incorporating the former in a host tetracation such as $[BBIPYBIXYCY]^{4+}$ (bisbipyridinium-bixylylcyclophane) and including the latter in the form of a simple neutral guest like 1,4-dimethoxybenzene (1/4DMB). The tetracationic cyclophane $[BBIPYBIXYCY]^{4+}$ was synthesized (Odell et al 1988) and was found to complex weakly with one molar equivalent of 1/4 DMB (Ashton et al 1988). The K_a value of 17 M^{-1} in acetonitrile at room temperature corresponds to a free energy of complexation of 6.9 kJ mol^{-1}. On the basis once again of an X-ray crystal structure, it may be concluded that the $[BBIPYBIXYCY.1/4DMB]^{4+}$ tetracationic complex is probably stabilized by weak edge-to-face interactions (Jorgensen & Severance 1990) between the 1/4DMB and the orthogonally aligned paraphenylene units in the $[BBIPYBIXYCY]^{4+}$ tetracation (with the hydrogen of one aromatic ring pointing towards the centre of the neighbouring aromatic ring), as well as by the dispersive interactions associated with π-stacking. Similar edge-to-face T-geometry has been observed (see, for example, Gould et al 1985) in crystals of benzene, naphthalene and anthracene as well as in proteins that are rich in aromatic amino acids, in which about 70% of the side chains of phenylalanine and tyrosine residues exhibit stabilizing edge-to-face interactions (Burley & Petsko 1988). Clearly, the electrostatic component is not so significant in stabilizing the $[BBIPYBIXYCY.1/4DMB]^{4+}$ tetracationic complex because of the lack of polyether chains (see below). The solid state structure of the 1:1 complex formed between $[BBIPYBIXYCY][PF_6]_4$ and 1/4DMB shows that once again the guest is inserted pseudorotaxane-like through the centre of the tetracationic macrocycle.

The observations that BPP34C10 forms a complex with the $[PQT]^{2+}$ dication (Fig. 2) and that the tetracationic cyclophane $[BBIPYBIXYCY]^{4+}$ forms a complex with 1/4DMB suggested the possibility of combining these different features in the design of a [2]catenane (Ashton et al 1989).

Template-directed synthesis

When equimolar amounts of the bis(pyridinium) salt $[BBIPYXY][PF_6]_2$ (BBIPYXY, bispyridiniumxylyl) and bis(bromomethyl)benzene (BBB) were stirred in acetonitrile at room temperature in the presence of three molar equivalents of BPP34C10, the [2]catenane shown in Figs. 2 and 3 was isolated in the remarkably high yield of 70% as a crystalline tetrakishexafluorophosphate after purification by column chromatography and counter ion exchange with

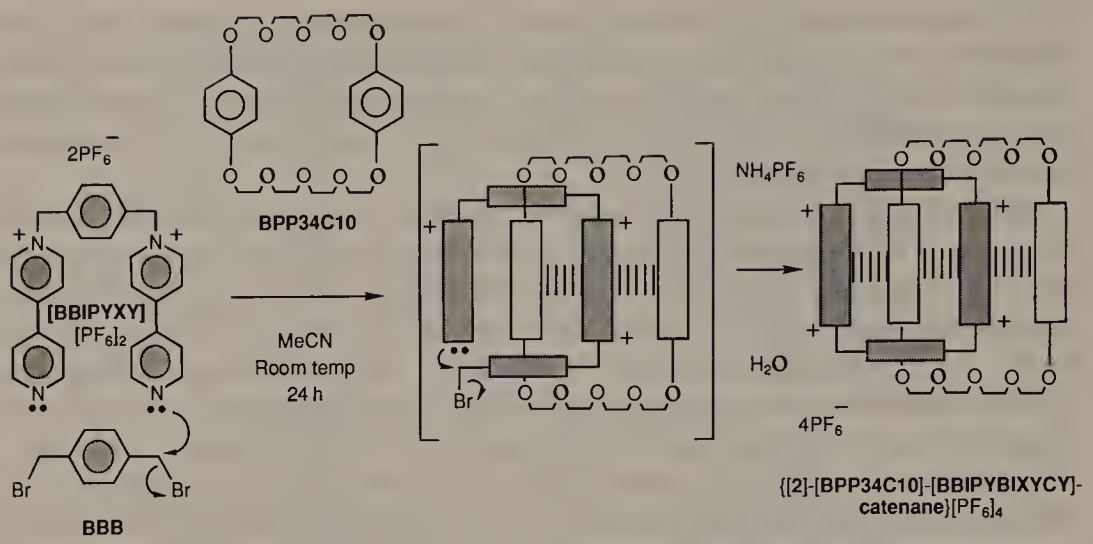


FIG. 3. A [2]catenane made to order. Using the features of the complexes shown in Fig. 2, a [2]catenane was constructed from BPP34C10, $[BBIPYXY]^{2+}$ and BBB, bis(bromomethyl)benzene.

ammonium hexafluorophosphate. In the absence of BPP34C10 the maximum yield of $[BBIPYBIXYCY][PF_6]_4$ that we have ever managed to isolate is 12%, and that was only after heating the acetonitrile solution containing the $[BBIPYXY][PF_6]_2$ and BBB for several days under reflux. Thus, it is much easier to construct the $\{[2]\text{-}[BPP34C10]\text{-}[BBIPYBIXYCY]\text{-catenane}\}^{4+}$ tetracation than it is to make one of its components, namely the tetracationic cyclophane $[BBIPYBIXYCY]^{4+}$, by itself. The implications of this observation are considerable and augur well for the future of self-assembly processes employing this kind of template-directed synthesis. The [2]catenane acts as a template for its own formation. The first nucleophilic substitution between the $[BBIPYXY]^{2+}$ dication and BBB produces a trication (Fig. 3), which, because of its similarity to the $[PQT]^{2+}$ dication, will thread its way through the centre of the BPP34C10 macrocycle. The second nucleophilic substitution is then favoured entropically and possibly even enthalpically. Whatever the precise details of the mechanism, the reaction is amazingly efficient. What is more, dynamic ^1H NMR spectroscopy indicates that the non-covalent bonding order, which presided over the formation of the [2]catenane in the first place, 'lives on' when the compound is dissolved in polar solvents such as acetonitrile and acetone. In a relatively slow site-exchange process involving 'inside' and 'alongside' hydroquinol rings, the BPP34C10 macrocycle feeds its way through the tetracationic cyclophane $[BBIPYBIXYCY]^{4+}$ and in so doing experiences a free energy of activation ($\Delta G^{\ddagger}$) of 65 kJ mol^{-1}. Simultaneously, in a faster exchange process involving 'inside' and 'alongside' bipyridinium rings, the neutral macrocycle is pirouetting around the tetracationic cyclophane with a

$\Delta G^{\ddagger}$ barrier of 51 kJ mol^{-1}. The solid state structure of the {[2]-[BPP34C10]-[BBIPYBIXYCY]-catenane}$^{4+}$ tetracation reveals both π-stacking and edge-to-face interactions involving the appropriate aromatic rings, that is, the structure is that of a highly ordered [2]catenane. This order is maintained in the crystal beyond the molecule in the form of an alternating stack of the π-electron-rich and π-electron-deficient aromatic rings repeating continuously in one of the crystallographic directions. This supramolecular order, coupled with a little imagination, suggests possible molecular designs for the construction of polyrotaxanes and double helicates.

Molecular modelling indicates that by replacing the phenylene rings in the tetracationic cyclophane [BBIPYBIXY]$^{4+}$ with biphenylene units it should be possible to thread two BPP34C10 rings simultaneously through the centre of this 'expanded' tetracationic cyclophane. Indeed, it is possible to isolate {[3]-[BPP34C10]-[BBIPYBIBTCY]-[BBP34C10]-catenane}$[PF_6]_4$ (BBIPYBIBTCY, bispyridiniumbisbitolylcyclophane) in 20% yield after reacting equimolar amounts of the bis(bipyridinium) salt $[BBIPYBT][PF_6]_2$ with 4,4′-bisbromomethylbiphenyl (BBBP) in the presence of a three molar excess of BPP34C10 and subsequent chromatography and counter ion exchange (Fig. 4; P. R. Ashton,

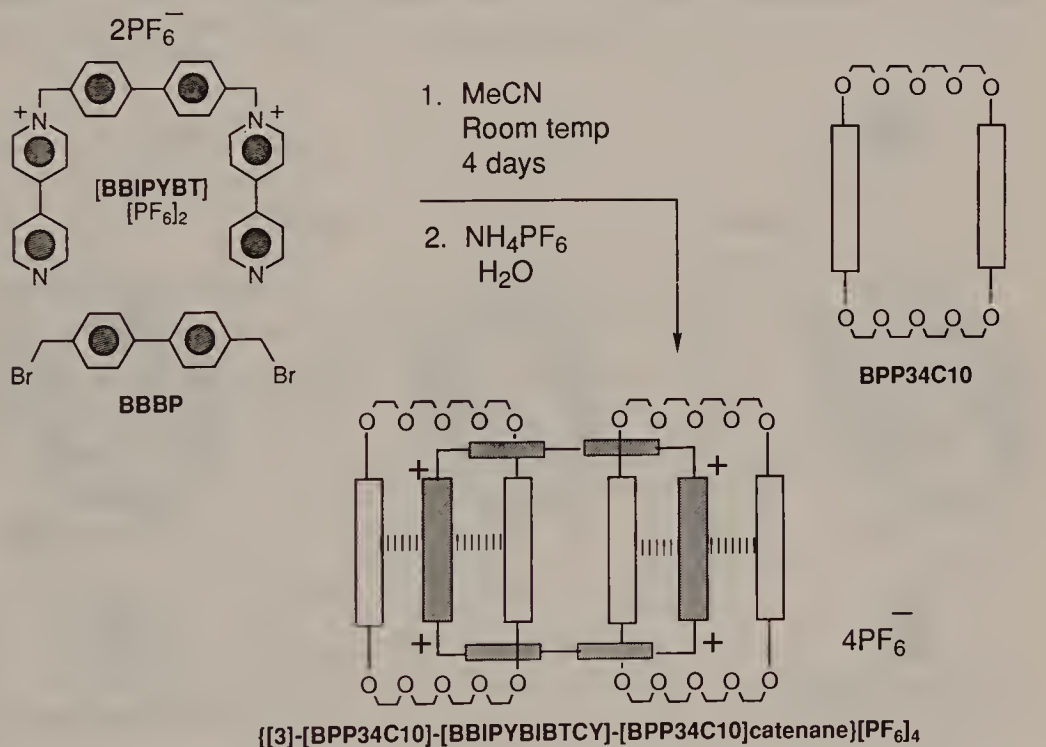


FIG. 4. A [3]catenane made to order. Expansion of the tetracationic cyclophane by incorporating a biphenylene unit permits threading through of two BPP34C10 rings to form a [3]catenane. [BBIPYBT]$^{4+}$, bispyridiniumbitolyl; BBBP, bisbromomethylbiphenyl.

C. L. Brown, E. J. T. Chrystal, T. T. Goodnow, A. E. Kaifer, K. P. Parry, A. M. Z. Slawin, N. Spencer, J. F. Stoddart & D. J. Williams, unpublished work 1989). The fact that this [3]catenane can be prepared with such relative ease is all the more remarkable when it is recognized that all attempts to obtain [BBIPYBIBTCY][PF_6]$_4$ by itself have so far been unsuccessful. Here is an intriguing example of a molecular assembly being made despite the fact that its central molecular component has yet to be isolated and characterized; molecular self-assembly is a highly desirable event. The solid state structure indicates that this [3]catenane, like the [2]catenane, is highly ordered. However, an interesting additional feature of this assembly is the loss of the alternating pattern of π-donors (D) and π-acceptors (A) observed in the [2]catenane. The π-stacking is associated with a DADDAD sequence in the [3]catenane and, what is more, the sequence extends beyond the molecule throughout the crystal, producing a continuously stacked array. Although once again the molecular order persists in solution (as indicated by the temperature dependence of the ^{1}H NMR spectrum) a process in which both BPP34C10 macrocycles are revolving simultaneously around the tetracationic cyclophane [BBIPYBIBTCY]$^{4+}$, so exchanging the sites of the hydroquinol rings 'inside' and 'alongside' the tetracation, is occurring with the relatively low free

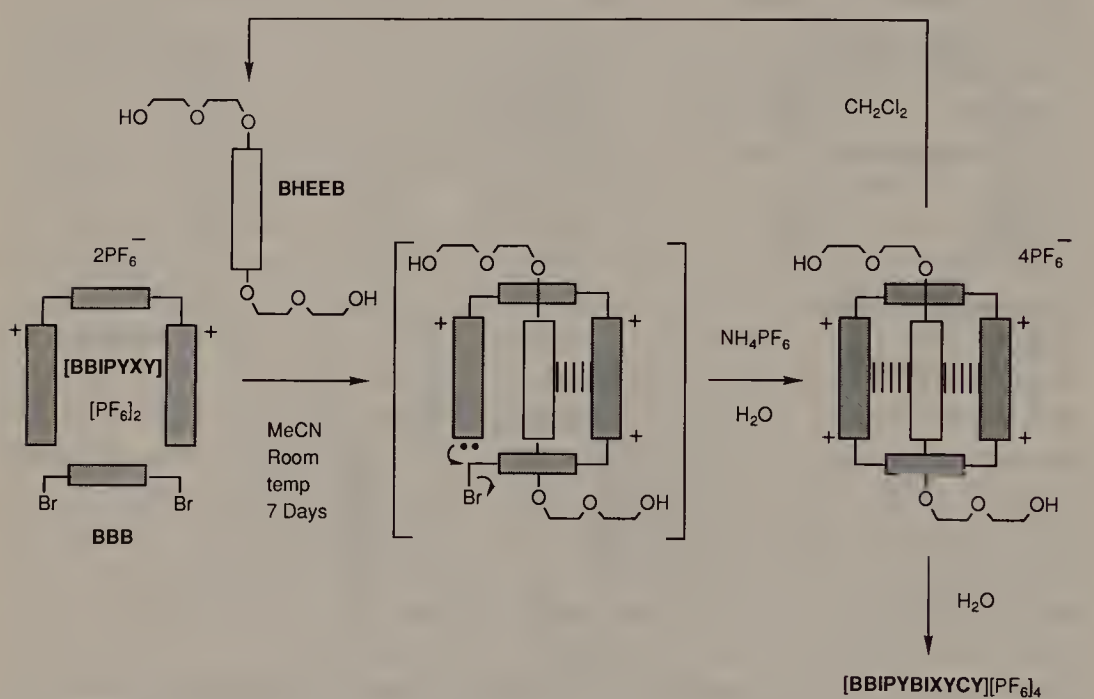


FIG. 5. The template-directed synthesis of a molecular component. Synthesis of the tetracationic cyclophane [BBIPYBIXYCY]$^{4+}$ is facilitated by the use of BHEEB, bishydroxyethyleneoxyethyleneoxybenzene, as a template. The template molecule can be separated from the product and recycled.

energy of activation of 45 kJ mol^{-1}. Placing two hydroquinol rings inside the [BBIPYBIBTCY]$^{4+}$ in the [3]catenane would seem to be less favourable than placing one hydroquinol ring inside the tetracationic cyclophane [BBIPYBIXYCY]$^{4+}$ in the [2]catenane.

A possible approach to the synthesis of the tetracationic cyclophanes that are difficult to prepare on their own is to employ a neutral compound as a template and then remove it from the charged macrocycle by unthreading when the reaction is complete (P. L. Anelli, P. R. Ashton, N. Spencer, A. M. Z. Slawin, J. F. Stoddart & D. J. Williams, unpublished work 1989). Although 1/4DMB is ineffective as a template in the synthesis of [BBIPYBIXYCY][PF_6]$_4$, compounds in which the hydroquinol rings carry polyether chains do act as neutral molecule templates. For example, BHEEB (bishydroxyethyleneoxyethyleneoxybenzene) increases the yield of the tetracationic cyclophane from 12% to 35% (Fig. 5). A pure sample of [BBIPYBIXYCY][PF_6]$_4$ can be separated easily from the template (BHEEB), which can be recycled by continuous liquid–liquid extraction of an aqueous solution of the highly coloured 1:1 complex with dichloromethane. As the extraction proceeds to completion, the colour is dispelled from the aqueous solution. When acetonitrile is replaced by dimethylformamide in the reaction scheme outlined in Fig. 5 and a catalytic amount of sodium iodide is added, the yield of the tetracationic cyclophane is increased to 45%. Finally, on carrying out the same reaction at ultra high

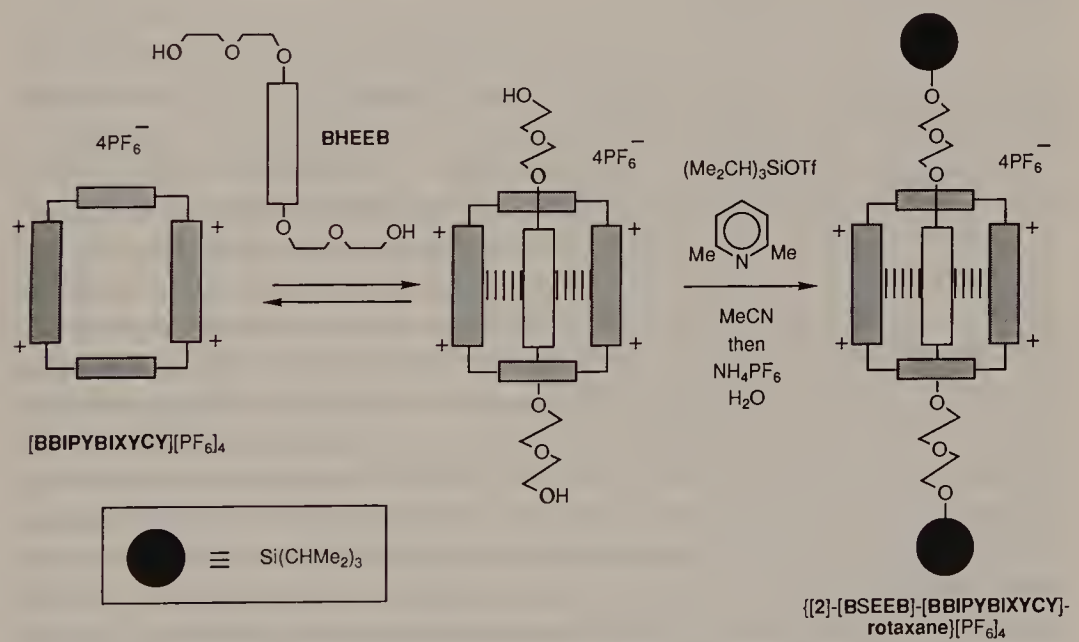


FIG. 6. A [2]rotaxane made to order by the threading procedure. {[2]-[BSEEB]-[BBIPYBIXYCY]-rotaxane}$^{4+}$ was prepared by treatment of BHEEB, bishydroxyethyleneoxyethyleneoxybenzene, with triisopropylsilyl triflate in acetonitrile containing BBIPYBIXYCY and lutidine.

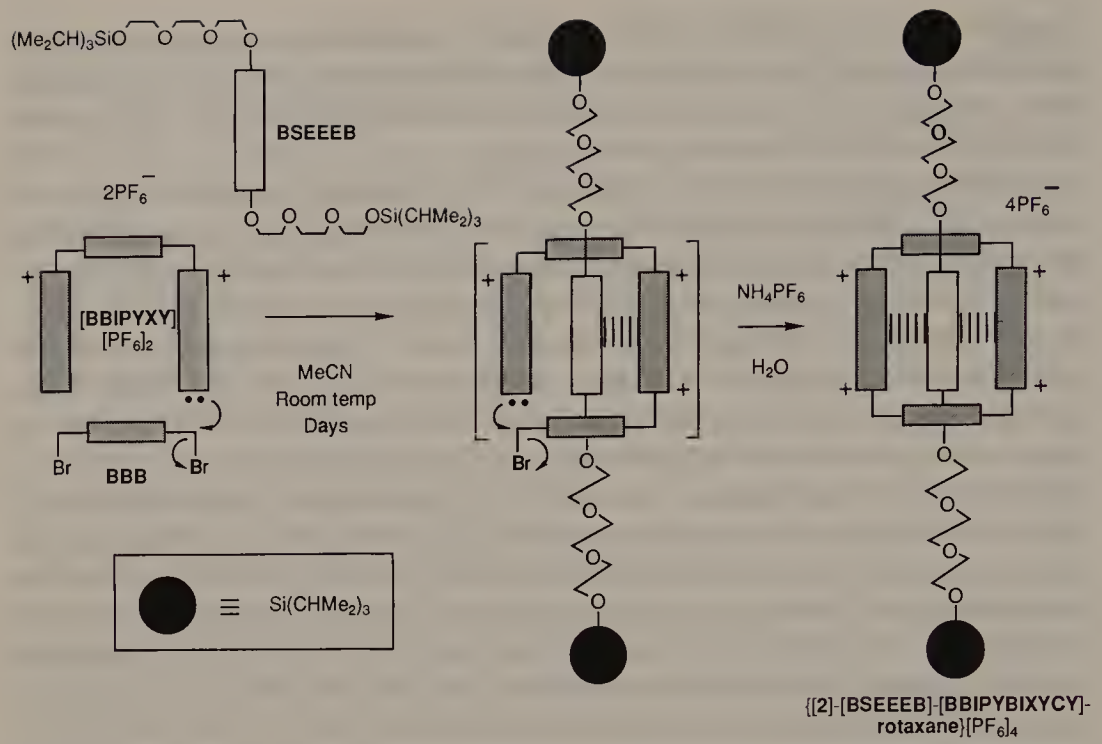


FIG. 7. A [2]rotaxane made to order by the clipping procedure. BSEEEB, BBIPYXY and BBB, bis(bromomethyl)benzene were used to prepare {[2]-[BSEEEB]-[BBIPYBIXYCY]-rotaxane}$^{4+}$.

pressures (10 kbar) a yield of 62% of [BBIPYBIXYCY][PF_6]$_4$ is obtained. When BHEEB was treated with triisopropylsilyl triflate [$(Me_2CH)_3SiOTf$] in acetonitrile containing a molar equivalent of [BBIPYBIXYCY][PF_6]$_4$ and a suitably hindered base (lutidine), {[2]-[BSEEB]-[BBIPYBIXYCY]-rotaxane}[PF_6]$_4$ was isolated following counter ion exchange (Fig. 6). This is an example of the preparation of a [2]rotaxane by the threading procedure (Fig. 1b). Although it is not possible to prepare the [2]rotaxane incorporating BSEEB (bissilyloxyethyleneoxyethyleneoxybenzene) by the clipping procedure (Fig. 1c), {[2]-[BSEEEB]-[BBIPYBIXYCY]-rotaxane}[PF_6]$_4$ was obtained in 14% yield by reacting [BBIPYXY][PF_6]$_2$ and BBB in acetonitrile in the presence of three molar equivalents of BSEEEB followed by counter ion exchange in the usual manner (Fig. 7). The solid state structures of the [2]rotaxanes indicate that in both cases the hydroquinol ring is placed centrosymmetrically inside the tetracationic cyclophane. Following on from the characterization of these [2]rotaxanes, the design and synthesis of a highly ordered molecular shuttle has been achieved and its operation has been demonstrated by dynamic ^{1}H NMR spectroscopy (P. L. Anelli, P. R. Ashton, N. Spencer, A. M. Z. Slawin, J. F. Stoddart & D. J. Williams, unpublished work 1989). The basis for the design is a [2]rotaxane composed of a polyether

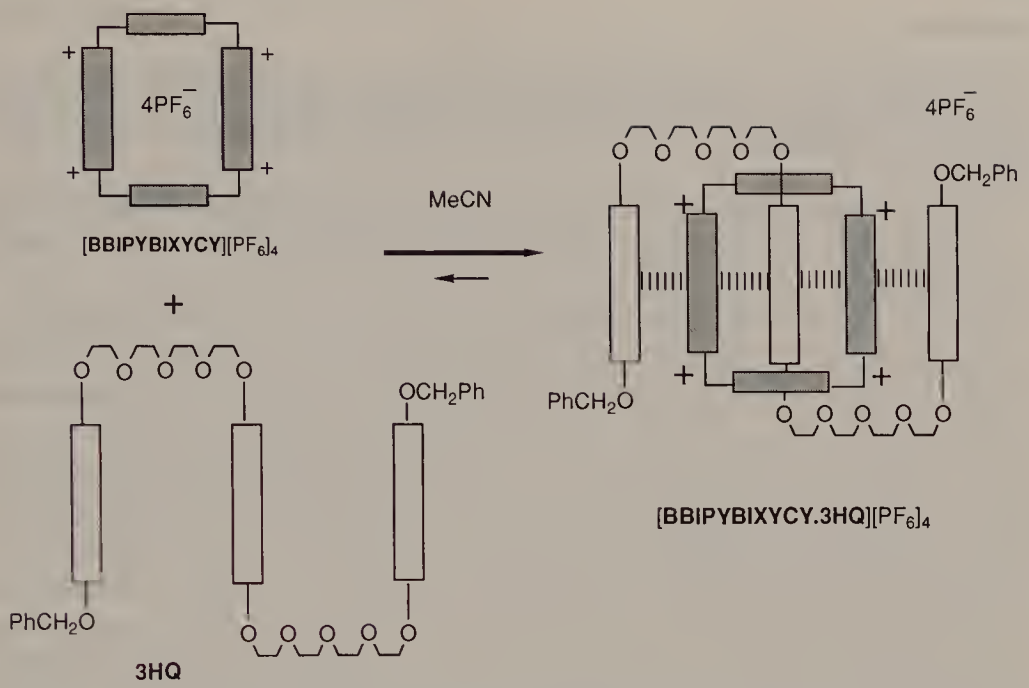


FIG. 8. The self-assembly of a molecular bead, $[BBIPYBIXYCY]^{4+}$, and a molecular thread, 3HQ, an acyclic polyether containing three hydroquinol rings and terminated by benzyl groups.

chain interrupted by two symmetrically placed π-electron-rich hydroquinol rings and terminated by triisopropylsilyl end groups and the tetracationic cyclophane $[BBIPYBIXYCY]^{4+}$. Using ^1H probes in both components of the [2]rotaxane, a barrier of about 44 kJ mol^{-1} to the shuttling process of the tetracationic cyclophane back and forth between the two hydroquinol rings has been calculated. This molecular shuttle may be looked upon as a prototype for a new range of molecular mechanical and electronic devices.

It is also highly encouraging that pseudorotaxanes will self-assemble spontaneously. An example is given in Fig. 8, which involves the tetracationic cyclophane $[BBIPYBIXYCY]^{4+}$ and an acyclic polyether (3HQ) containing three hydroquinol rings and terminated by benzyl groups. The 1:1 complex $[BBIPYBIXYCY.3HQ][PF_6]_4$ has been fully characterized (P. L. Anelli, P. R. Ashton, N. Spencer, A. M. Z. Slawin, J. F. Stoddart & D. J. Williams, unpublished work 1989). The X-ray crystal structure suggests that polyrotaxanes and double helicates should self-assemble spontaneously from the appropriate molecular components. If this next stage in the research programme is successful, then we will have demonstrated that highly ordered catenanes and rotaxanes are the ideal vehicles at the small molecule level around which to experiment with a view to designing larger systems that will self-assemble in not only a spontaneous but also in a reversible mode. In this way, information can be transferred at the molecular level as well as stored.

Reflections

All the examples of template-directed synthesis discussed in this paper illustrate the principle of structure-directed synthesis (Kohnke et al 1989) which states that:

> There are inherently simple ways of making apparently complex unnatural products from appropriate substrates without the need for reagent control or catalysis.

It implies that the required stereoelectronic information is preprogrammed into the substrates of a chemical reaction; this allows the efficient and precise spontaneous self-assembly of these 'intelligent' substrates, thus yielding either large molecules with completely defined superstructures or, as in the case of the catenanes and rotaxanes made to order, an organized molecular assembly composed of unique molecular components. The concept of structure-directed synthesis challenges the dogma of present day synthetic chemistry, which is often viewed as a search for more sophisticated reagents and catalysts to carry out well-known chemical reactions. It calls for a change in our thinking about the art of synthetic chemistry which I suggest now has to undergo a major conceptual revolution with fundamental implications for the practice of chemical synthesis as we have perceived it in the classical and traditional sense. Following the examples provided by Nature, we must begin to learn how to build large molecules and molecular assemblies molecule by molecule in addition to the familiar atom-by-atom and group-by-group approaches involving the time-consuming and often not so efficient functional group manipulations and transformations that rely upon the delicate and intricate use of protecting groups.

Acknowledgements

I am deeply indebted to Drs Neil Spencer and Pier Lucio Anelli for leading the research programme I have described in this paper with such drive and success, to Mr Peter R. Ashton for his imaginative and intuitive use of fast atom bombardment mass spectrometry in the characterization of high molecular weight complexes and assemblies, and to Dr David J. Williams and Miss Alexandra M. Z. Slawin (Imperial College, London) for their apparently effortless solutions to many extremely demanding X-ray crystal structure problems. We were all admirably supported by many talented contributions from postgraduate research students, namely Christopher L. Brown, Douglas Philp, Mark V. Reddington, and Cristina Vicent. I thank the Science and Engineering and the Food and Agricultural Research Councils for their generous financial support. The intellectual input of senior scientists at both ICI Agrochemicals (Drs Keith P. Parry and Ewan J. T. Chrystal) and Shell Research (Dr Barbara Odell) contributed greatly to everyone's understanding.

References

Ashton PR, Odell B, Reddington MV, Slawin AMZ, Stoddart JF, Williams DJ 1988 Isostructural alternately-charged receptor stacks. The inclusion complexes of hydroquinone and catechol dimethyl ethers with cyclo(paraquat-*p*-phenylene). Angew Chem Int Ed Engl 27:1550–1553

Ashton PR, Goodnow TT, Kaifer AE et al 1989 A [2]catenane made to order. Angew Chem Int Ed Engl 28:1396–1399

Balzani V, Scandola F 1990 Supramolecular photochemistry. Ellis Horwood, Chichester

Burley SK, Petsko GA 1988 Weakly polar interactions in proteins. Adv Protein Chem 39:125–189

Carter FL 1984 The molecular device computer: point of departure for large scale cellular automata. Physica 10D:175–194

Cram DJ 1988 The design of molecular host, guests, and their complexes (Nobel Lecture). Angew Chem Int Ed Engl 27:1009–1020

Dietrich-Buchecker CO, Sauvage JP 1987 Interlocking of molecular threads: from the statistical approach to the templated synthesis of catenands. Chem Rev 87:795–810

Gould RO, Gray AM, Taylor P, Walkinshaw MD 1985 Crystal environments and geometries of leucine, isoleucine, valine and phenylalanine provide estimates of minimum nonbonded contact and preferred van der Waals interaction distances. J Am Chem Soc 107:5921–5927

Jorgensen WL, Severance DL 1990 Aromatic–aromatic interactions: free energy profiles for the benzene dimer in water, chloroform, and liquid benzene. J Am Chem Soc 112:4768–4774

Kohnke FH, Mathias JP, Stoddart JF 1989 Structure-directed synthesis of new organic materials. Angew Chem Int Ed Engl Adv Mater 101:1129–1136

Lehn JM 1988 Supramolecular chemistry. Scope and perspectives. Molecules, supermolecules, and molecular devices (Nobel Lecture). Angew Chem Int Ed Engl 27:89–112

Odell B, Reddington MV, Slawin AMZ, Spencer N, Stoddart JF, Williams DJ 1988 Cyclobis(paraquat-*p*-phenylene): a tetracationic multipurpose receptor. Angew Chem Int Ed Engl 27:1547–1550

Schill G 1971 Catenanes, rotaxanes, and knots. Academic Press, New York

Stoddart JF 1988 Conception and birth of new receptor chemistry from dibenzo-18-crown-6. Pure Appl Chem 60:467–472

Summers LA 1980 The bipyridinium herbicides. Academic Press, New York

DISCUSSION

Ron Breslow: Could I raise the question of function? You talked about the need to develop these new classes of synthetic molecular assemblies because of their interesting potential functions. What are you going to do that cannot be done already?

Stoddart: One example concerns catalysis. By and large, Nature has devised large molecular assemblies, in the form of the enzymes, to preside over catalytic events. It could be by putting some effort into the self-assembly of high molecular weight synthetic receptors that we shall be able to improve on the catalytic performances and reaction selectivities of synthetically derived enzyme mimics. You, of course, are one of the world's leading authorities in the area

of enzyme modelling. Although you and others have been using, for example, chemically modified cyclodextrins as the basis for constructing enzyme mimics, I hope you would concede that there might be some virtue in being able to build up very much larger molecular weight synthetic receptors. We need to uncover the synthetic methodology that will allow us to construct large and ordered synthetic polymers that have the ability to complex both specifically and selectively with substrates larger than have so far been subjected to enzyme modelling. We should, for example, spend more time with artificial enzymes, addressing the challenge of forming covalent bonds between more than one substrate molecule to create even larger molecules, rather than spend so much time trying to devise ways of breaking covalent bonds in single substrate molecules. I think there is some virtue in us going bigger in the design of enzyme mimics. In this context, it is interesting that generally speaking, 'big is beautiful' in biological systems.

Ron Breslow: People have been making polymers for a long time, but I suppose they haven't had the structural control that you have.

Stoddart: That is exactly my argument. Polymer chemistry during the last 50–60 years has been concerned with the constitutions and structures of synthetic polymers at a primary level. Generally, it has been coincidental if there have been some stabilizing interactions between the chains of synthetic polymers. It has been noted occasionally that chains 'talk' to each other in synthetic polymers. For example, many of the more recent aromatic polymers investigated display interactions between their chains (Tanner et al 1989). It is important that in synthetic polymer chemistry we now start trying to build in non-covalent bonding interactions within and between the polymer chains, so that the molecules have much more of the character of DNA, proteins, polysaccharides and glycoproteins. Synthetic polymers of the next generation will need to have function as well as form, like these biopolymers.

Dunitz: You presented many fascinating details about your marvellous molecules. In particular, you referred to edge-to-face interactions between aromatic rings, and said that the hydrogens of one aromatic ring point towards the hole in the centre of another. You mentioned that edge-to-face interactions are a characteristic feature of the crystal structures of the hydrocarbons benzene, naphthalene and anthracene, as well as of the packing of aromatic rings in proteins. I cannot speak for the aromatic rings in proteins, and, anyway I am not sure that the resolution would be sufficient to establish such fine details of the packing there, but the detailed crystal structures of benzene, naphthalene and anthracene are known and have been studied intensively. The main feature of these structures is the herring-bone pattern made by the edge-to-face packing of the molecules. In the benzene structure it is true that one type of hydrogen sits nearly above the centre of a neighbouring ring and makes six essentially equal carbon–hydrogen distances. But these are *not* the shortest non-bonded C...H distances; there are others up to 0.2 Å shorter. In naphthalene and

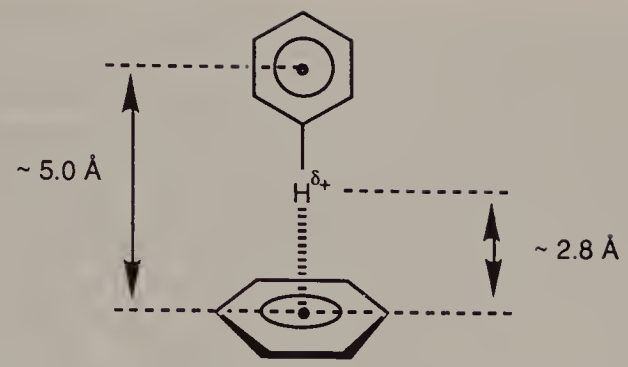


FIG. 1. (*Stoddart*) Edge-to-face interaction of neighbouring aromatic rings, such as that observed in $[BBIPYBIXYCY.1/4DMB]^{4+}$.

anthracene, the edge-to-face pattern persists, but the aromatic rings no longer have hydrogen atoms vertically above or below their centres. The packing energies of these crystals arise from many interactions and all nearest neighbours contribute in comparable amounts. For example, for benzene, one recent estimate is that four neighbours contribute 10.4% each, four others 7.1% each and four others 6.2% each (Gavezzotti & Desiraju 1988). I know there is a widespread belief that the hydrogens of one aromatic ring tend to point towards the hole in the centre of another. According to my calculations, this may be the case in the benzene dimer, but it is not a very good guide to the packing of aromatic rings in extended structures, where many other types of interaction have to be considered.

Stoddart: The small-molecule recognition work (Stoddart 1988, Ashton et al 1988, 1989) which we have been doing on the 1:1 complexes, rotaxanes and catenanes has allowed our X-ray crystallographer, Dr David Williams at Imperial College, London, to locate precisely on aromatic rings hydrogen atoms which form a kind of hydrogen bond to the π-face of a neighbouring aromatic ring. To a first and good approximation, the interacting electropositive hydrogen atom ($H^{\delta+}$) is placed above and in the middle of the projected hexagon that describes the six carbon atoms of the benzenoid ring (Fig. 1). The $H^{\delta+}$ is usually located at around 2.8 Å from the centroid of the benzenoid ring with which it is interacting and the corresponding centroid–centroid distance between the two more-or-less orthogonally related benzenoid rings is about 5.0 Å.

As regards the extent to which this interaction is stabilizing, we have not, so far, been able to partition out this edge-to-face interaction from other stabilizing interactions, such as those associated with π-stacking, in a way that satisfies us. I don't feel that the edge-to-face interaction is a highly stabilizing one. I would guess that it's not worth more than a couple of hundred calories per mole, but this guess is based more on my knowledge and experience of related systems than on any experimental data for the systems discussed in my paper. However, we and other people here have uncovered a large number of these

aromatic edge-to-face interactions in small-molecule recognition systems over the last few years.

Dunitz: I'm not saying that it doesn't happen in your molecules; I'm only trying to establish how important it is for benzene, naphthalene and anthracene.

Stoddart: All I was referring to was the herring-bone structure observed in the packing of benzene, naphthalene and anthracene molecules in their crystalline states. As far as I am aware, this packing is associated with edge-to-face interactions between benzenoid rings. I still feel it is not unlikely that the edge-to-face interactions we see are related to those observed in the crystalline states of benzene, naphthalene and anthracene. However, in so far as the crystal structures of naphthalene and anthracene are not the best models for our molecular and supramolecular systems, I would not wish to press the comparison to extremes, because, after all, the benzene dimer is more similar to our systems where the aromatic components are phenyl or 'isolated' naphthalene rings.

Rebek: With these non-covalently linked materials you have a unique opportunity to provide vehicles for functional groups for catalysis. You could put one function on one end of the rotaxane or on one of the components of the catenane, and a second function on the other and, through the dynamics of the system, a catalytic site would be created at certain times. This would be more effective than having the functions on separate molecules.

The characteristic feature of naturally occurring systems that form assemblies is that they are self-complementary in an intermolecular sense. To build assemblies with your molecules I think you will need to elaborate them to include some divergent functionality. Part of the charm of being an organic chemist is that if you can find things that don't have a naturally occurring counterpart, that's so much the better.

Stoddart: For the reasons you mentioned, we are obviously thinking about how we can introduce functionality both in a convergent and in a divergent sense into catenanes, rotaxanes and, indeed, double helicates. To be more precise, we are considering portions of their structures where we can easily change their constitutions. The polyether loops in the hydroquinol components are, of course, very attractive in this respect because replacement of the central $-CH_2OCH_2-$ unit by a 2,6-di-substituted pyridyl unit provides a convenient way of preserving the necessary five heteroatom repeat unit and also offers, through the 4-position, an aromatic platform on which to site divergent functional groups. We are trying to use the non-covalent bonding present in the middle of the molecules to assemble them from their matching molecular components. We hope that these internal stabilizing interactions will ultimately allow us to have considerable control over the relative arrangements of any divergent groups that are located on the outside of the catenanes, the rotaxanes or the double helicates.

Gokel: I was quite struck by the fact that in one of your complexes there appeared to be two electron-rich substrates sandwiched between two electron-poor

ones, with a repeating unit of two rich, one poor. What do you know about the energetics of this situation? It is an interesting question whether there are stabilizing interactions involved or if these interactions are primarily a void-filling mechanism.

Stoddart: The stacking of the two π-electron-rich hydroquinol rings in the centre of the [3]catenane is intriguing. One experimental finding which shows how easily two π-electron-rich rings stack is that the shortest distance, 3.35 Å, is an intermolecular one between hydroquinol rings in adjacent [3]catenane molecules in the crystal. This result surprised me initially. Perhaps in assemblies such as these catenanes, in which the positive charge is localized in one ring system—the tetracationic cyclophane—with the neutral ring systems composed of both ethereal oxygen atoms and π-electron-rich hydroquinol rings solvating the positive charge, the stacking of the π-systems substantially changes the electronic character of the π-electron-rich components so that they are not as unfavourably disposed towards each other as you might expect—that is, the character of the hydroquinol rings is altered by their association with the dicationic bipyridinium rings. To be honest, we need some help from theoretical chemists to explain how we are able to bring two hydroquinol rings face to face with each other at a distance as close as 3.35 Å.

Hamilton: In your work with double helices, I wonder what problems you might encounter in terms of forming overlapping polymeric structures, rather than discrete 1:1 complexes. That brings in a more general question about the choice of π-stacking interactions as the primary intermolecular binding interaction, as opposed, for example, to hydrogen bonds. Do you see any potential for introducing specificity into your designs by using more electron-rich systems interacting with more electron-deficient systems, to provide a direction for the formation of the helix?

Stoddart: I take your point. You are asking if we will get very large molecular assemblies as a result of overlapping of helical chains within the double helicates that are being formed. I think we may find self-recognition between strands of complementary sizes, as Lehn & Rigault (1988) have found with copper-coordinated synthetic double helices, but we have not yet done the necessary experiments so I can only anticipate that result. Your point concerned whether, if there is a random distribution of overlapping chains of different lengths, we would be able to control this phenomenon by altering the π-electronic character of the donor and acceptor systems. I agree that we should be able to control matching of chains at this level. I would just like to add that we are investigating systems with π-stacking interactions rather than hydrogen bonding for a number of reasons. There are many people, yourself included, who are doing elegant assembly work using hydrogen-bonding interactions. We felt that the π-stacking interaction was one that needed to be developed and exploited from the vantage point of self-assembly.

References

Ashton PR, Odell B, Reddington MV, Slawin AMZ, Stoddart JF, Williams DJ 1988 Isostructural alternately-charged receptor stacks. The inclusion complexes of hydroquinone and catechol dimethyl ethers with cyclo(paraquat-*p*-phenylene). Angew Chem Int Ed Engl 27:1550–1553

Ashton PR, Goodnow TT, Kaifer AE et al 1989 A [2]catenane made to order. Angew Chem Int Ed Engl 28:1396–1399

Gavezzotti A, Desiraju G 1988 A systematic analysis of packing energies and other packing parameters for fused-ring aromatic hydrocarbons. Acta Crystallogr Sect B Struct Sci 44:427–434

Lehn JM, Rigault A 1988 Helicates: tetranuclear and pentanuclear double helix complexes of Cu-I and poly(bipyridine) strands. Angew Chem Int Ed Engl 27:1095–1097

Stoddart JF 1988 Conception and birth of new receptor chemistry from dibenzo-18-crown-6. Pure Appl Chem 60:467–472

Tanner D, Fitzgerald JA, Philips BR 1989 The Kevlar study: an advanced materials case-study. Angew Chem Int Ed Engl Adv Mater 28:649–654

Self-assembly in supramolecular systems

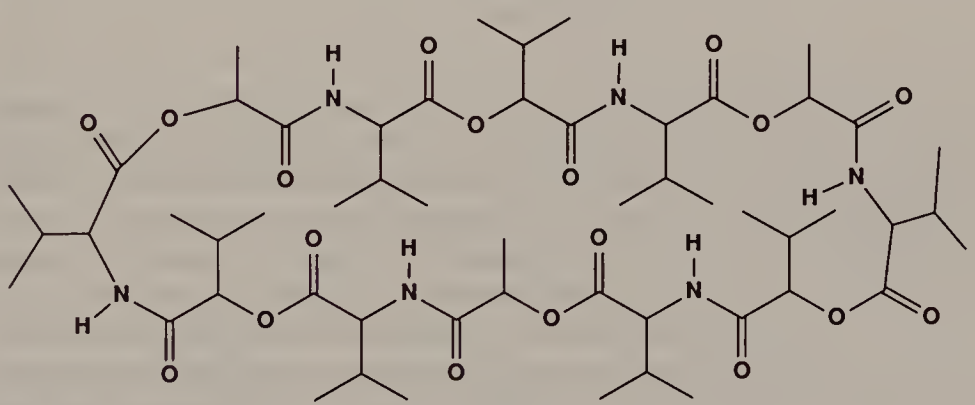
George W. Gokel

Department of Chemistry, University of Miami, Coral Gables, Florida 33124, USA

Abstract. The complex structural and functional properties of many natural molecules have spawned innumerable attempts to understand and mimic biological activity. This has often involved preparing extremely complex structures of carefully designed geometries. In natural systems the primary structure of a protein (the amino acid sequence) establishes all of the structural relationships within the molecule, although many of these are not apparent until the molecule folds, coils, or otherwise adopts the appropriate conformation. Nature has selected suitable amino acid sequences for various applications during eons of evolution. In this paper, we report our efforts to achieve similar results by providing all of the required structural elements on a flexible framework. This concept is illustrated in three ways: the design and preparation of a redox-switched vesicle and a small-molecule molecular receptor (both based on the ferrocene system) and of a functional cation channel.

1991 Host–guest molecular interactions: from chemistry to biology. Wiley, Chichester (Ciba Foundation Symposium 158) p 23–38

Valinomycin (Brockmann & Schmidt-Kastner 1955) is a naturally occurring cyclododecadepsipeptide that functions as a 'transport antibiotic'. Functionally, valinomycin makes membranes permeable to K^+, interfering with the process of oxidative phosphorylation in mitochondria. Valinomycin is thus a 'cation carrier'. It complexes K^+ and effectively conducts it through a lipid bilayer



1 (Valinomycin)

membrane. The molecule (**1**) has several quite interesting properties that permit it to function in this fashion.

The compound has nine isopropyl groups and three methyl groups. This makes the exterior of the molecule hydrophobic, while the dozen carbonyl groups make the interior quite polar. Although the compound, when considered in two dimensions, has a cavity size far larger than the 2.6 Å diameter of K^+, it is very selective for that cation over either Na^+ or Ca^{2+}, the other common ions found in serum. The compound contains six ester and six amide carbonyl groups. If only six of these were to provide an octahedron of donors about K^+, one might expect them to be the amide carbonyls, since these are more polar than the ester carbonyl groups. Instead, it is the ester donors that bind K^+. Finally, D-amino acids are unknown in proteins but half of the residues in valinomycin have this stereochemistry. Indeed, the compound's chirality can be represented as $(D,D,L,L)_3$.

The apparently too large size of valinomycin is not a problem: the compound folds into a 'tennis ball seam' arrangement that envelops K^+ three-dimensionally. When the cation is absent, this conformation is held in place by six hydrogen bonds between the amide NH groups and the opposite amide carbonyl groups. In addition to holding the pre-binding conformation, this ties up the more polar amide carbonyl groups which would undoubtedly favour either Ca^{2+} or Na^+ over K^+. The folding of valinomycin provides an encompassing environment for the cation but it also fulfils another important requirement for cation transport. Cation binding is usually considered in terms of the equilibrium constant but not in terms of the complexation and release rates that govern K_{eq} or K_S. The paradox of transport is that complexation

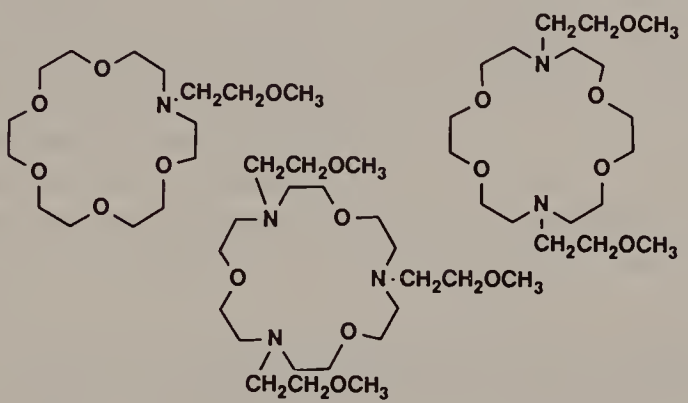
$$\text{Ligand} + \text{Cation} \underset{}{\overset{K_S}{\rightleftharpoons}} \text{Complex}$$

$$K_S = \frac{k_1 \text{ or } k_{\text{complex}}}{k_{-1} \text{ or } k_{\text{release}}}$$

must be strong and k_1 should be fast at the source phase, but complexation should be weak and k_{-1} should be large at the receiving phase. Valinomycin strikes the appropriate compromise by having intermediate values of rate and complexation constant. This is achieved, in part, by flexibility in the ring system. This, in turn, is accomplished by using the change in chirality (D,D,L,L) to make the requisite molecular bends without compromising flexibility. These remarkable features of valinomycin suggested to us that we could set the structural stage for complexation and transport by incorporating the appropriate elements and permitting Nature to accomplish some of the organization.

Results and discussion

We began the effort described here by designing compounds that we called 'lariat ethers' because of the appearance of their molecular models and because they 'rope' (with the macro-ring) and 'tie' a cation (with the side-arm) (Gokel et al 1980). Lariat ethers having two (Gatto & Gokel 1984, Gatto et al 1986) or three arms (Miller et al 1989) are called bibracchial or tribracchial lariat ethers. Most of the effort undertaken so far involves compounds attached at nitrogen which we call nitrogen-pivot compounds (Schultz et al 1985). The three basic nitrogen-pivot structures are shown (2).



2

When any of these compounds complexes a cation, the ring interacts with it in the usual fashion and then the side-arm swings over to afford a third dimension of solvation (Gandour et al 1986). The two-armed, 18-membered ring systems bind sodium cation in a pseudo-crypt arrangement with the cation roughly centred between the ring and the two side-arm donor groups (Arnold et al 1987). The potassium cation, for steric reasons, is bound by the macro-ring and side-arm donor groups on opposite sides.

The example of valinomycin has permitted us to develop some useful, divalent, calcium-selective bibracchial lariat ethers (White et al 1989). These are all derivatives of 4,13-diaza-18-crown-6 (see 2) which have $-CH_2-CO-NH-CHR-CO-OR'$ side-arms. Four of the derivatives have side-arms as follows: -Gly-Gly-OR', Gly-Ala-OR', Gly-Val-OR' and -Gly-Leu-OR'. A -Gly-Lac-OR' derivative was also prepared to assess the importance of the amide carbonyl in this series. The cation binding for the -Gly-Ala-OCH_3 compound was appriximately 10^7 for Ca^{2+} in water compared to less than 10^3 for Na^+ in the same solvent. Thus the Ca^{2+}/Na^+ selectivity is more than 10,000-fold—a hitherto unachieved binding selectivity for a neutral ionophore.

Although most of the extensive effort so far described in the synthetic ionophore field has been directed to carrier molecules, some limited study of

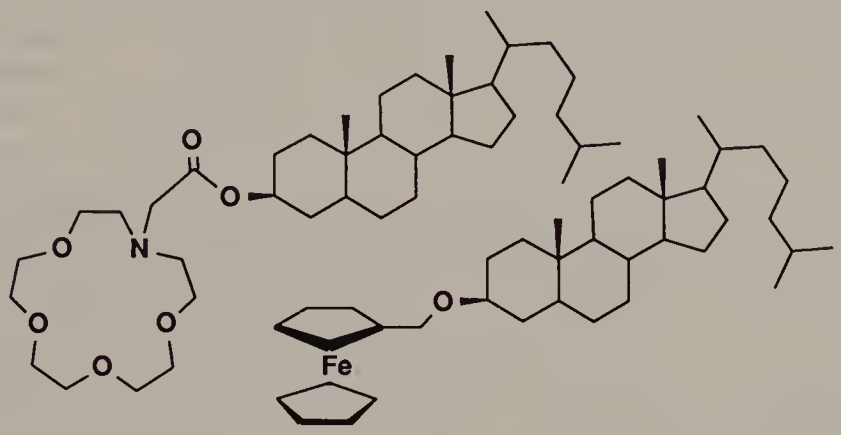
cation-conducting channels has also been reported (Behr et al 1982, Tabushi et al 1982, Kragton et al 1985, Führhop et al 1985, Jullien & Lehn 1988, Carmichael et al 1989). It might be more appropriate to call such systems 'pore-formers' rather than synthetic channels. Urry et al (1986) have summarized their observations concerning channels based on the gramicidin A transmembrane channel as follows: 'Thus, the requirements of such a transmembrane channel for lipid bilayers would be (a) a lipophilic exterior; (b) a low energy conformation that is sufficiently flexible to allow local ion-induced relaxations of conformation; (c) a length sufficient to span the lipid portion of the bilayer and reflect the membrane thickness-dependent conductances; and (d) the capacity to interact with ions without the perfect sequestering observed in the carrier mechanism, that is, the structure should be one in which the interacted ion can readily, in a directional manner, exchange coordinations and thereby allow flow along the transmembrane channel.' The application of the flexible framework strategy to this problem led us to consider placing three macrocylic rings on a flexible framework in such a way that one of them could serve as an entrance, one as an exit, and the third as a cation relay. Numerous questions concerning this approach had to be answered. A few of the prominent ones are listed below.

Can a crown ether ring serve as a polar head group in a lipid bilayer?
If so, will the macro-ring be too polar to reside in the centre of the bilayer?
What must be the spacing of the three rings to fit a membrane?
How far can a cation jump? That is, would a single, central ring be sufficient or would other 'relay points' be required?
What would be the role of water in such a system?

It seemed to us that the first question requiring to be answered was whether or not a crown ether could serve as the head group for synthetic bilayer (vesicle) formation. This question was answered by the preparation (Gokel et al 1987) and aggregation (Echegoyen et al 1988) of aza-15-crown-5 attached to a steroid via a glycine residue. These compounds were found to aggregate into unilamellar vesicles of about 300 Å diameter. They also exhibited 1–2% volume entrapments and were generally similar to phosphatidylcholine vesicles. One significant difference was their essential rigidity (Fasoli et al 1989). Since all the lipophilic residues are steroids, this result was not surprising. We also noted that the formation of aggregates could be altered by the complexing of cations—that is, by changing the charge state of the head group.

Having succeeded in forming such systems, we wondered whether another means of changing the charge state would be to oxidize a neutral, non-polar substrate. The ferrocene system seemed an ideal candidate. The system shown in **3** could be oxidized either electrochemically or by using chemical reagents and could then be induced to aggregate. When oxidized, the ferrocinium cation system aggregated in multilamellar vesicles having membrane thicknesses of

about 45 Å, as judged from electron micrographs. Once formed, the aggregates could be broken by either electrochemical or chemical methods and the unchanged starting monomer could be recovered. The steroid crown referred to above is shown at the left in **3** and the ferrocenyl steroid is illustrated at the right.



3

Having satisfied ourselves that a crown ether could serve effectively as the head group in a lipid bilayer we resolved to undertake the syntheses of several related structures. The question of span length next required consideration. Fortunately, X-ray crystal structures of the gramicidin A system were recently reported by Wallace & Ravikumar (1989) and Langs (1989) and their data are summarized in Table 1. The structure of gramicidin A is O=CH-L-Val-Gly-L-Ala-D-Leu-L-Ala-D-Val-L-Val-D-Val-L-Trp-D-Leu-L-Trp-D-Leu-L-Trp-D-Leu-L-Trp-ethanolamine.

We decided to use the phosphatidylcholine monomer as a model, because egg lecithin vesicles are widely studied and well understood. We were also

TABLE 1 Comparison of gramicidin A channel and pore structures

Feature	*Channel*[a]	*Pore*[b]
Gross structure	Dimer	Dimer
Structure length	31 Å	26 Å
Hole size	3.85–5.47 Å	4.9 Å
Residues/turn	7.2	6.4
Cation(s)	None	Cs^+
Solvent	None	38 solvent sites
Structure resolution	0.86 Å	2.0 Å
Coil interior	Hydrophilic	Hydrophilic
Coil exterior	Hydrophobic	Hydrophobic
Hydrogen bonds	28	28

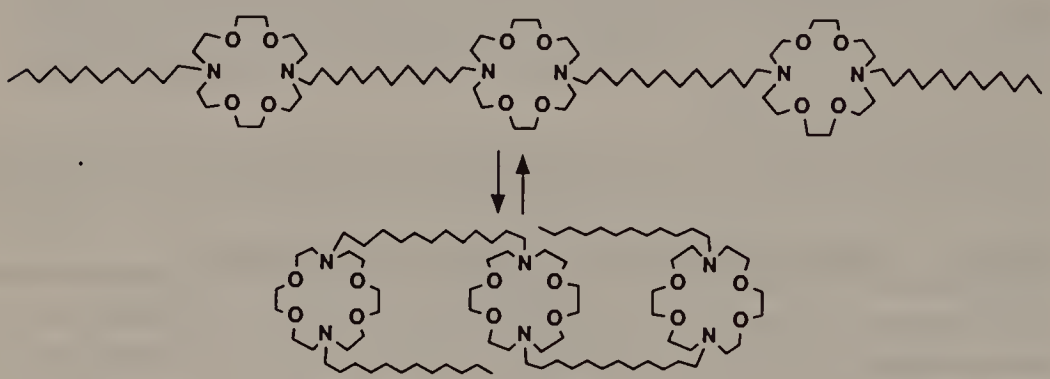
[a]Wallace & Ravikumar 1989.
[b]Langs 1989.

TABLE 2 ^{23}Na cation transport rates in phosphatidyl choline vesicles[a]

Compound	*Rate (s^{-1})*	*Relative rate*
C_{12}-<18N>-C_{12}	0.31	1
<18N>-C_{12}-<N18N>-C_{12}-<N18>	<0.2	<1
C_{12}-<N18N>-C_{12}-<N18N>-C_{12}-<N18N>-C_{12}	13.5	44
Gramicidin	1400	4500

[a]Some data from Nakano et al 1990.

fortunate that Riddell & Hayer (1985) had recently reported an NMR technique for assessing transport rates in gramicidin and had applied it to the egg lecithin system. We therefore simply averaged the length values shown in Table 1 ([31 Å + 26 Å]/2 = 28.5 Å). From Corey-Pauling-Koltun (CPK) molecular models we determined that a compound such as that shown in **4** should have a transmembrane span of nearly this value. It was anticipated that the compound would fold with the two exterior rings at the lipid–aqueous phase boundary. The fate of the inner macro-ring was less clear: it might form the third ring of a cylinder or it might elongate and function merely as a relay and not as any part of an imagined tunnel. If the latter happened, the net polarity of the macrocycle would be reduced, because it would be spread over several ångströms rather than concentrated in the centre of the bilayer.



4

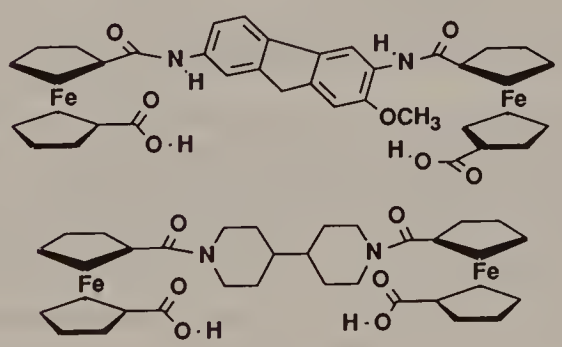
The synthesis of this compound was far from simple but relatively straightforward. Once the compound was in hand, studies were undertaken to determine whether or not it increased membrane permeability. Two means are commonly used to assess cation flux: steady-state methods and unidirectional flow techniques. The latter methods often present problems because liposome breakage may be interpreted as cation transport or a transmembrane potential may be observed, apart from various other problems. In our case, we felt somewhat more comfortable with the NMR technique that Riddell & Hayer (1985) had developed. It was used by Buster et al (1988) to assess Na^+ cation

flux in phosphatidylcholine vesicles. In the steady-state method that we chose, a rate constant and kinetic order are determined by using ^{23}Na NMR methods. Dysprosium tripolyphosphate was used as a shift reagent to distinguish $^{23}Na_{internal}$ from $^{23}Na_{external}$ ($[^{23}Na]_{int}/[^{23}Na_{ext}] = 1{:}4.5$). The various substrates were incorporated by microlitre injection followed by incubation at 50 °C for 0.5–1 h. Incubation was critical to the success of these studies. The kinetic order for each system could be obtained by varying the ionophore concentration.

The cation transport ability of four compounds was assessed by the dynamic NMR method. Among them were gramicidin and the compound illustrated in **4**. The latter can be represented schematically as C_{12}-<N18N>-C_{12}-<N18N>-C_{12}-<N18N>-C_{12}. In addition to these, a cation carrier that corresponds to the centre of the molecule (C_{12}-<N18N>-C_{12}) and a triple ring structure lacking the external side-arms (<18N>-C_{12}-<N18N>-C_{12}-N18>), were examined. The results are summarized in Table 2.

The kinetic orders were also of considerable interest. For gramicidin, known to be dimeric, the kinetic order was two. For C_{12}-<18N>-C_{12} it was also two, presumably because of a flip-flop transport mechanism. The channel-former, C_{12}-<N18N>-C_{12}-<N18N>-C_{12}-<N18N>-C_{12}, exhibited first-order kinetics, suggesting that it was spanning the membrane as anticipated.

Finally, our general interest in transport has caused us to consider the question of the selective binding of small molecules. Rebek and co-workers (Rebek 1991, this volume) have developed a family of receptors based on the Kemp triacid. Two carboxyl residues face each other across a spacer of certain length and electronic character. Our own approach to this problem involves the ferrocene molecule in which two aromatic rings are separated by an iron atom that serves as a molecular ball-bearing. Small molecule complexation by the two receptors illustrated in **5** has been demonstrated. We anticipated that both compounds



5

shown in **5** would bind small molecules but that the top compound might take advantage of π-stacking, an interaction not available to the bottom structure. Binding studies were conducted on diamines in $CDCl_3$ and perdeuteriotetrahydrofuran (THF-d_8) and the range of K_S was about 400–4000 M^{-1}.

Many questions about cation and molecule complexation and transport remain to be answered. Our approach of using feeble forces in flexible frameworks has so far permitted us to prepare highly calcium-selective cation binders, various novel vesicle systems, a molecule that enhances transmembrane cation flux, and small molecule complexing agents based upon the ferrocene system. The emphasis of our future work will be in the latter three areas.

Acknowledgements

I warmly thank the many co-workers in my group whose names appear on the literature cited. In particular, this work would not have been possible without the expert and extensive intellectual and experimental collaboration of Professor Luis Echegoyen, who, with his co-workers, is largely responsible for the spectroscopic, electrochemical and aggregation studies reported here. Professor Jerry Atwood and his co-workers provided the crystal structures which have proved indispensible. Finally, I thank the NIH for their support of our programme through grants GM-36262, GM-33940 (joint with Professor Echegoyen), AI-27179 and AI-30188.

References

Arnold KA, Echegoyen L, Fronczek FR et al 1987 Evidence for cryptand-like behavior in bibracchial lariat ether (BiBLE) complexes obtained from X-ray crystallography and solution thermodynamic studies. J Am Chem Soc 109:3716–3721

Behr JP, Lehn JM, Dock AC, Moras D 1982 Crystal structure of a polyfunctional macrocyclic K^+ complex provides a solid-state model of a K^+ channel. Nature (Lond) 295:526–527

Brockmann H, Schmidt-Kastner G 1955 Valinomycin XXVII Matterl über Antibiotica aus Actinomyceten. Chem Ber 88:57–61

Buster DC, Hinton JF, Millett FS, Shungu DC 1988 ^{23}Na-Nuclear magnetic resonance investigation of gramicidin-induced ion transport through membranes under equilibrium conditions. Biophys J 53:145–152

Carmichael VE, Dutton PJ, Fyles TM James TD, Swan JA, Zojaji M 1989 Biomimetic ion transport: a functional model of a unimolecular ion channel. J Am Chem Soc 111:767–769

Echegoyen LE, Hernandez JC, Kaifer A, Gokel GW, Echegoyen L 1988 Aggregation of steroidal lariat ethers: the first example of nonionic liposomes formed from neutral crown ether compounds J Chem Soc Chem Commun 836–837

Fasoli H, Echegoyen LE, Hernandez JC, Gokel GW, Echegoyen L 1989 Direct evidence from electron paramagnetic resonance studies for virtual immobility in niosomes derived from steroidal lariat ethers. J Chem Soc Chem Commun 578–580

Führhop J-H, Liman U, David HH 1985 Sealing and opening porous monolayer vesicle membranes. Angew Chem Int Ed Engl 24:339–340

Gandour RD, Fronczek FR, Gatto VJ et al 1986 Solid state structural chemistry of lariat ether and BiBLE cation complexes: importance of topology rather than cavity size in binding. J Am Chem Soc 108:4078–4088

Gatto VJ, Gokel GW 1984 Syntheses of calcium-selective, substituted diazacrown ethers: a novel, one-step formation of bibracchial lariat ethers (BiBLEs). J Am Chem Soc 106:8240–8244

Gatto VJ, Arnold KA, Viscariello AM, Miller SR, Morgan CR, Gokel GW 1986 Syntheses and binding properties of bibracchial lariat ethers (BiBLEs): survey of synthetic methods and cation selectivities. J Org Chem 51:5373–5384

Gokel GW, Dishong DM, Diamond CJ 1980 Lariat ethers. Synthesis and cation binding properties of macrocyclic polyethers possessing axially disposed secondary donor groups. J Chem Soc Chem Commun 1053–1055

Gokel GW, Hernandez JC, Viscariello AM et al 1987 Steroidal lariat ethers: a new class of macrocycles and the crystal structure of N-(cholesteryloxycarbonyl)aza-15-crown-5. J Org Chem 52:2963–2968

Jullien L, Lehn J-M 1988 The 'chundle' approach to molecular channels. Synthesis of a macrocycle-based molecular bundle. Tetrahedron Lett 29:3803–3806

Kragton UF, Roks MFM, Nolte RJM 1985 Facilitated ion transport in a biomembrane model. Artificial ion channels as pores in dihexadecyl phosphate vesicles. J Chem Soc Chem Commun 1275–1276

Langs DA 1989 Three-dimensional structure at 0.86 Å of the uncomplexed form of the transmembrane ion channel peptide gramicidin A. Science (Wash DC) 241:188–191

Miller SR, Cleary TP, Trafton JE, Smeraglia C, Fronczek FR, Gokel GW 1989 Tribracchial lariat ethers, 'TriBLEs' based on 4,10,16-triaza-18-crown-6: an apparent limit to sidearm contributions in lariat ether molecules. J Chem Soc Chem Commun 806–808

Nakano A, Xie Q, Mallen JV, Echegoyen L, Gokel GW 1990 Synthesis of a membrane-insertable, sodium cation-conducting channel: kinetic analysis by dynamic ^{23}Na NMR. J Am Chem Soc 112:1287–1289

Rebek J Jr 1991 Clefts as receptor and enzyme analogues. In: Host–guest molecular interactions: from chemistry to biology. Wiley, Chichester (Ciba Found Symp 158) p 98–114

Riddell FG, Hayer MK 1985 The monensin-mediated transport of sodium ions through phospholipid bilayers studied by sodium-23 NMR spectroscopy. Biochim Biophys Acta 817:313–317

Schultz RA, White BD, Dishong DM, Arnold KA, Gokel GW 1985 Nitrogen-pivot lariat ethers: syntheses, equilibrium binding, and dynamic properties. J Am Chem Soc 107:6659–6668

Tabushi I, Kuroda Y, Yokoto K 1982 A,B,D,F-Tetrasubstituted β-cyclodextrin as artificial channel compound. Tetrahedron Lett 23:4601–4604

Urry DW, Trapane TL, Venkatachalam CM 1986 Potassium-39 NMR of K^+ interaction with the gramicidin channel and NMR-derived conductance ratios for Na^+, K^+ and Rb^+. J Membrane Biol 89:107–111

Wallace BA, Ravikumar K 1989 The gramicidin pore: crystal and structure of a cesium complex. Science (Wash DC) 241:182–187

White BD, Mallen J, Arnold KA et al 1989 Peptide side-arm derivatives of lariat ethers and bibracchial lariat ethers: syntheses, cation binding properties, and solid state structural data. J Org Chem 54:937–947

DISCUSSION

Reinhoudt: When talking about the cation channel, you didn't mention the role of the anion. This is something that worries me when people talk about channels, because as soon as the cations begin to penetrate the organic environment, charge separation develops. Of course, this can be partly

compensated for by strong solvation by water. I would like to hear your comments on this. Have you varied the anions in your organic phase?

Gokel: We haven't really had these molecules long enough or in sufficient quantity to do much besides demonstrate efficacy. That's why I qualified my statements by saying that they were conjecture. I am confident that at least some water is involved in the cation transport process. I suspect that the solvated cations are travelling through the channel as partially hydrated species. I think that the presence of some water compensates for the absence of Lewis basic relay points along the channels interior. I can't say at the moment whether the cation is actually being transported through with its anion.

Reinhoudt: Which anion did you use?

Gokel: Chloride.

Reinhoudt: That's a hard anion.

Gokel: Yes.

Ron Breslow: The alternative is to set up a pH gradient across the membrane. You can pump ions with a pH gradient, or pumping of ions can produce a pH gradient. Is the pH inside different from that outside? If the anion doesn't travel you have to neutralize the charge separation somehow, and one way to achieve that is for the protons to come back.

Reinhoudt: For that, you need a proton-ionizable group somewhere in the system.

Ron Breslow: You could put hydroxide inside to act as a counter ion, in which case the pH should be higher inside than outside.

Gokel: We have not measured the pH, so we are ignorant about the details of the mechanism.

Albery: I totally agree, of course, that electroneutrality must be preserved. What sort of quantities of sodium ions have you measured being transported using your NMR method? That might begin to answer the question about pH.

Gokel: This is a difficult question to answer because the rate depends on the concentration of 'channel-former' present in the vesicles. We compared cation flux rates at a concentration of 40 μM. The range of observed rates covered three orders of magnitude. The value observed was $13.5\ s^{-1}$, in comparison with a value of $1.4 \times 10^3\ s^{-1}$ for gramicidin and one of $0.31\ s^{-1}$ for the carrier molecule that corresponds to the central ring with two side chains.

Albery: Are there other cations besides protons that could go back in the opposite direction to the sodium?

Gokel: There are three cations present in the system. The dysprosium cation is the shift reagent and the sodium ion is directly observed by NMR. In order to prepare the shift reagent, potassium tripolyphosphate is added as well.

Ron Breslow: You initially had an equal concentration of sodium ions on each side and you were looking simply at line broadening. It's only if you produce a sodium ion gradient that there is this problem of electroneutrality.

Gokel: That's correct; we are looking at a steady-state process in which we observe sodium cation (^{23}Na) exchange (inside the vesicle versus outside). The sodium peak heights differ because the absolute amount of sodium inside and outside differs even though the concentration is the same.

We simply don't know whether the cation is accompanied by an anion. The distance across the membrane is probably more than 30 Å. An anion may well accompany the cation, but we have no direct evidence either way.

Reinhoudt: Your carriers have a fairly strongly basic nitrogen atom. In water, at pH 7, they will be protonated.

Gokel: They will be at least partly protonated.

Reinhoudt: Do you know what the pK_a of your system is?

Gokel: The pK_a values of the tertiary nitrogen atoms are probably about 11. As you know, tertiary amines typically exhibit pK_a values in the range of 10–11. I doubt that ours is exceptional, but we have not measured it.

Reinhoudt: If it is protonated the concentration of receptor molecules will be very low.

Gokel: We should try to assess the extent of protonation. That also has a bearing on the issue of ligand solubility, because as the extent of ligand nitrogen protonation increases, solubility in water should also increase.

Reinhoudt: We have done experiments in liquid-immobilized membranes. Below about pH 5 or pH 6 we lose the carriers completely. Therefore, when they are not very hydrophobic, the partition depends on the structure of the carrier.

Rebek: This is really a daunting problem, to build a molecule or an array that can actually look like and function as a channel. What strikes me about the structure of gramicidin is that the interior of the sleeve is lined with π-bonds; there aren't any convergent lone pairs, so this is a lubricated sort of channel that doesn't provide ions with good places to park. If you create a series with parking places, you will face a dynamic problem that comes from slow release of the ions. I think an effective channel will be one which is propped open and doesn't have a high affinity for the ions along the interior.

Gokel: It has been said that the gramicidin channel functions by a sort of peristaltic action in which the cation is 'grabbed' by several amide donor groups. It is difficult to imagine this situation, because the amide donor group is a potent donor group in aqueous solution. If several amide groups bound a cation at a particular site, the next cation coming through the channel would require the kinetic energy of a billiard ball to push the first one through.

It would be deceptive if I had said that we designed the cation channel-former I showed because we thought it would mimic the mechanism of gramicidin. My original idea was to try to build a system with carefully set up relay sites; these would be ethers, which are poorer donors than amides. The present system was designed to be the 'base line' compound with which more successful designs could be compared. When it was successful, it occurred to me that perhaps the

central ring is elongated, rather than being in the crown conformation. Such an elongation would reduce the local polarity within the bilayer mid-plane. The central ring extension might give some additional residence points for the cation. We are now making an analogous structure which has a single poly(oxyethylene) strand in the centre rather than a macro-ring. Our concept has evolved from something that you 'look through', to something in which there is a hole propped open, a hole on either side of the membrane and a groove inside through which the cation can slide.

Müller: Have you thought of using the self-assembly properties of liquid-crystalline compounds to get the channels? This would enhance the stacking tendency of the channel-forming molecules.

Gokel: Yes, we have. We originally devised some of the steroidal lariat ethers for just that reason: to use the steroids to organize the holes in the attached macro-rings. Our idea was to organize the stacked steroids. A thermally induced change in helicity would change the relative overlap of the macro-rings. Thus, instead of having a binary (on-off) switch, you would have a rheostat. As the change in helicity altered the orientation of the holes with respect to each other, the size of any channel in the structure would gradually diminish, as would the size of any object that could pass through. This may well be a valid concept; unfortunately, we did not complete work on that project.

Müller: I was thinking of aza-crowns, for example. You can substitute aza-crowns to yield liquid-crystalline compounds which could possibly form liquid-crystalline phases in the lipid bilayer.

Gokel: Praefcke and others have made some systems that have radial symmetry, but they're not based on crown ethers (Kohne & Praefcke 1984).

Müller: Mertesdorf & Ringsdorf (1989) have made some.

Stoddart: The synthetic channel molecules that you described lacked preorganization. I recognize that you are going to introduce steroids to increase their rigidity, but I was concerned about the way you illustrated the channels, given that there is so much dynamic flexibility in the hydrocarbon chains. How concerned are you about the lack of preorganization in these structures?

Gokel: I did illustrate in structural form rather speculative conformations that I probably should have characterized as 'cartoons'.

When this work began, we considered the problem in functional terms. We wondered if we could prepare a compound that had what we believed to be all the functional requirements to act as a channel, but arrayed on a flexible framework. We knew that there would be infinite structural possibilities in a flexible system. Even so, we thought that if all the necessary components were present, Nature might use hydrophobic forces to organize the system so that it would be functional. In other words, we were letting the molecule be subject to a sort of Darwinianism. This notion is contrary to the idea of rigid preorganization that has, I believe, been the dominant thinking in this area for

many years. I think it is fair to represent the idea as we conceived it, especially because we have evidence of functionality and no evidence as yet that contradicts the basic idea. Of course, concept is not evidence. I certainly agree with you that there are innumerable possible conformations and it may be that none of those that you or I might draw will prove to be the functional one.

However, we do know some things. For example, we know that a number of crown ethers and cryptands can form micelles or vesicles. It is therefore reasonable to assume that the macro-rings on both surfaces of the membrane are roughly in the D_{3d} arrangement. Spanning of the membrane requires at least 30 Å of superstructure because the egg lecithin membrane is that thick. This distance and the fact that the distal rings must be on the membrane surface mean that there must be a more-or-less extended arrangement of hydrocarbon chains. As I have said, we have no evidence about the central or proximal ring's conformation. I suspect it is extended to give a greater length to the overall structure, but at the moment this is only a suspicion. I hope we shall be able to obtain firm structural evidence that will guide future syntheses.

Stoddart: Crystallinity is one of Nature's ways of achieving organization, but the bis-ferrocene doesn't form a molecular cleft. It forms an extended structure in which the binding sites are as far away from each other as possible.

Gokel: I think the answer to that is that we have shown a crystal structure of the ligand (**5**) (p 29) rather than the complex. Since there is nothing inside the cleft or cavity, the molecule adopts a conformation to minimize energy in the solid state. We have not been lucky enough to obtain good crystals of the complex. If and when we do, I am confident from solution data that the two ferrocene rings will be orientated in the binding conformation shown in **5**.

Stoddart: Your system might be organized in a supramolecular fashion; that's another model you could invoke to account for a channel structure. You are providing an explanation at the molecular level. However, you could look to a supramolecular level for an explanation of how the components of the 'jigsaw' fit together in the membrane.

Gokel: Your work has shown us that we must have a broader prospective, but don't forget that an important inspiration for much of the channel work up until now was Lehn's recognition that the crystal structure of tetracarboxy-18-crown-6 formed a channel-like structure (Lehn 1979); that finding focused thinking on a covalent approach to channels. The kind of organization involved in such a structure is beautiful, but don't forget that Nature has made membranes that are sloppy. Can you imagine a stained-glass window that's fluid mosaic? The membrane is a dynamic system but I feel we tend to think more rigidly.

Parker: Perhaps you need to begin with a disorganized system, not with a very organized system. There are some strong analogies between ionic conduction through a channel of the type that you are trying to get, and ionic conduction in the solid state. For example, with lithium salts in polyethylene oxide, it was

initially thought that lithium ions were hopping along a helical solvating structure and that that was the rate-limiting step. The ionic conduction in the solid state actually occurs in the amorphous phase, in which there are small energy barriers between different binding sites and polymer segmental motion is involved. One primary concern is, as mentioned earlier, the interaction between the cation and its counter ion. Perchlorate and triflate are the preferred anions to inhibit counter ion association, particularly in the case of lithium. Although the idea of starting with a beautifully ordered system is very attractive, I think it could be just as productive, or perhaps even more productive, to begin with a beautifully disorganized system!

Gokel: I don't mean to say that a well-conceived, well-organized system isn't a thing of beauty—it is indeed. I have been struck, for example, by the calixarene complexes that Jerry Atwood has worked with (Atwood et al 1989). These molecules tend to form what I think could be characterized as lipid bilayers; the calixarenes are oriented in such a way that there is a polar phase made up of water and sodium ions and the calixarenes form bilayers. Perhaps coincidentally, the thickness of the polar layer is the same as that in natural clays. I think these results show that there are certain natural driving forces that can be used to advantage.

Collet: If you disturb a bilayer, you modify its properties. For example, you change certain transition temperatures between the rigid and less rigid conformations of the side chains. Perhaps you are inducing leaks in the bilayer; if so, you will immediately observe an increase in transport. Did you check, by calorimetry for example, that the transition temperature is unaffected, that it is the same before and after addition of the crown ether? Your crown ethers are probably tensioactive molecules—they may change the surface tension of your system, and that effect alone might be responsible for the change in transport kinetics you observe.

Ron Breslow: It's not even necessary that you change the bulk properties of the bilayer. We have done some experiments with molecules that incorporated into bilayers which were reasonably well organized; the probe induced local disorder, yet the bulk properties of the bilayer were not obviously affected. There's a risk of disordering around the probe, but not seeing any significant change in bulk transition temperature or bulk properties.

Gokel: These are important issues. I admit that we have not done any calorimetry. Working with Professor Echegoyen, we have duplicated the results reported for a number of related systems. We have carefully reproduced the work on gramicidin (Buster et al 1988, Riddell & Hayer 1985) and I don't think transport in the case of gramicidin is attributed to leakage. We feel reasonably confident that leakage is not involved, but the danger is always present.

Kellogg: It is nice to combine natural components with unnatural components. You attached two short peptide chains to an aza-crown ether. Another possible

strategy involving use of a mixture of natural and synthetic components is to combine, for example, the repeating α-hydroxy acid–amino acid–α-hydroxy acid–amino acid units found in valinomycin with unnatural chains (polyethylene glycol, for example), repeat this one, two or three times, finally closing the ring to make a macrocycle. I would add that there are, of course, many cyclic systems found in Nature that are built up from repeating units. The functions of many of these are not known. You could use alternating unnatural and natural components, cyclize, and work from that to obtain a partially synthetic crown ether with the desired complexing properties. What is your view of such a synthetic strategy?

Gokel: The total synthesis of valinomycin reported by Merrifield and his co-workers (Gisin et al 1969) involved many steps and yielded less than 200 mg product.

Kellogg: I am talking about using a replicating system to get away from the misery of total synthesis; you could, at least in principle, use very simple components like amino acid segments tied to a synthetic chain, and combine these units in a repetitive manner simply via a classical templated crown ether synthesis.

Gokel: If you mean a crown ether-based system that used some polyether strands and some peptide strands, I see no reason why such a system could not be prepared. Indeed, this might be a clever strategy, but there are some potential problems. An example of such a difficulty arose in some work we completed recently. We were asked by a colleague to measure cation binding strengths for certain crown ethers containing peptide-like side arms. The problem with these particular structures was that although amide carbonyl groups were available in the side arms, the amide linkage was used to attach them to the macrocycle. As a result, the cation binding ability of the ring was diminished. In the system you suggest, the balance of rigidity and flexibilty would have to be considered very carefully.

We have thought about hexapeptides in my group. Last year, we conducted an on-line search of Chemical Abstracts that revealed about 90 reports from 1970 to 1989 of cyclohexapeptide syntheses, isolation or properties. Some of the compounds reported were found to be cation binders. Our interest in such structures is that amino acids are Nature's basic building blocks. We think it may be possible to use the amino acid unit as a building block to achieve properties we have been searching for in other systems.

References

Atwood JL, Coleman AW, Zhang H, Bott SG 1989 Organic clays—synthesis and structure of Na_5(calix[4]arene sulfonate).$12H_2O$, K_5(calix[4]arene sulfonate).$8H_2O$, Rb_5(calix[4]arene sulfonate].$5H_2O$, and Cs_5(calix[4]arene sulfonate).$4H_2O$. J Inclusion Phenom 7:203–211

Buster DC, Hinton JF, Millett FS, Shungu DC 1988 ^{23}Na-Nuclear magnetic resonance investigation of gramicidin-induced ion transport through membranes under equilibrium conditions. Biophys J 53:145–152

Gisin BF, Merrifield RB, Tosteson DC 1969 Solid-phase synthesis of the cyclododecadepsipeptide valinomycin. J Am Chem Soc 91:2691–2695

Kohne B, Praefcke K 1984 Liquid-crystalline compounds. 15. Hexa-*o*-alkanoyl-scyllo-inositols, the 1st alicyclic, saturated, discotic liquid crystals. Angew Chem Int Ed Engl 23:82–83

Lehn JM 1979 Macrocyclic receptor molecules—aspects of chemical reactivity—investigations into molecular catalysis and transport processes. Pure Appl Chem 51:979–999

Mertesdorf C, Ringsdorf H 1989 Self-organization of substituted azacrowns based on their discoid and amphiphilic nature. Liq Cryst 5:1757–1772

Riddell FG, Hayer MK 1985 The monensin-mediated transport of sodium ions through phospholipid bilayers studied by sodium-23 NMR spectroscopy. Biochim Biophys Acta 817:313–317

General discussion I

Flexibility or rigidity?

Stoddart: David Parker raised a very good point when he asked us to think about flexibility (p 36). We should also think about rigidity imposed at a supramolecular level in our systems. We can see considerable flexibility at the molecular level in the design of the molecular components but, if we are going to address functional systems such as membranes or molecular electronics, we need to introduce some order between the molecules. Rigidity at a supramolecular level involving 'correctly' positioned intermolecular interactions is called for. Wolfgang Müller brought up that point when he asked us to think about the way that liquid crystals might be used to orient molecules into channels (p 34). We should think more about order at the supramolecular level.

Dunitz: At the most elementary level, the competition between rigidity and flexibility is really one between stability and kinetics. For complexation to happen reasonably quickly the molecules must be flexible enough to give a low activation energy. But a very stable complex has to be fairly rigid, because, in general, the stronger the binding (the deeper the potential energy well), the higher is the corresponding force constant. You can see how this works by considering the desolvation of an alkali cation that has to occur in the course of complexation with an ionophore. The hydration energy of sodium, for example, is around 100 kcal mol^{-1}. If the ionophore were rigid, the activation energy for complexation in aqueous solution would be of this order of magnitude and the reaction would not occur in a reasonable time. Flexibility is needed for displacement of the water molecules, essentially one at a time, by the oxygens of the ionophore.

Gokel: Presumably, at least some of the water molecules must be replaced so that the ligand can interact with the cation.

Dunitz: For the cation to go from the hydrated form to become encapsulated in the ionophore, the oxygen atoms of the water molecules have to be replaced by the oxygen atoms of the ligand. If this is to happen at a reasonable rate, flexibility is necessary. For efficient ionophores, one is not interested in getting ultra-stable complexes; very stable complexes cannot act as carriers because the off-rate will be too low. For each problem, one may have a different goal, depending on how much thermodynamic stability and how much kinetic flexibility is needed.

Gokel: Although we have not published much in this area, we have actually done considerable work on the thermodynamics of crown complexes. We are able to measure the equilibrium binding constant between crowns and cations

at various temperatures. We can use the van't Hoff relationship to obtain both the enthalpy and entropy of binding. Sometimes there is a quite remarkable difference in the strength of the enthalpic interaction between two complexes that have exactly the same stability constant. Of course, this is related to changes in solvent order. In any event, I agree that the magnitude of the enthalpy and the balance of enthalpy and entropy can be critical for the overall success of binding and transport.

Dunitz: The stability constant is a ratio of two rates; the need for a good on-rate puts limitations on the off-rate.

Gokel: Or on the magnitude of the stability constant.

Reinhoudt: In principle that is true, but only for monomolecular reactions. There are many cases, siderophores for example, in which the stability constant is extremely high, yet the rate of exchange of different ions is fast because the exchange is not a monomolecular but a bimolecular process in which there is an S_N2-type substitution. Bimolecular exchange has also been shown in complexes of simple macrocycles.

Gokel: The huge stability constants ($>10^{50}$) quoted for the siderophores are calculated from bond energies, not ratios of rates; they are not stability constants in the kinetic sense.

Reinhoudt: They are stability constants because they define the overall equilibrium constant irrespective of whether there are six reactions involved or one. If there are two or three deprotonation steps there can be a very high overall stability constant for that particular complex, but at the same time the rate per individual step can be very high.

Gokel: My point is that such stability constants should not be compared to those in which a single cation is bound by a single ligand in a single step. If such a process had a stability constant of 10^{50}, and since $K_S = k_1/k_{-1}$, the decomplexation rate would have to be 10^{-66}, even if the complexation rate was near the diffusion-controlled rate of about $10^{10}\,s^{-1}$.

Reinhoudt: Not if complex formation takes place in more than one step.

Gokel: Exactly; the single-step complexation constants, which are typically in the range $10–10^9$, should not be directly compared with the high stability constants reported for siderophores. I *think* I agree with you.

Albery: I don't think I do agree. If there is a sequence of steps, there will be a free energy profile and there has to be a single rate-limiting transition state. You go from a deep valley to a mountain top through a series of high valleys, but the largest free energy difference still has to be to overcome at some step; the rate-limiting step must be from the lowest dip to the highest peak.

Sutherland: It depends on whether we are talking about guest exchange, which can be a bimolecular process, or about the rate of a dissociation process. The two are different; some binding energy is retained in a bimolecular process, so it can be fast.

Reinhoudt: There are some data (de Jong et al 1976) on alkylammonium cations, where simple extraction experiments give wrong information because the rate constant that is measured, by NMR for example, is not a rate constant of the monomolecular dissociation process but includes bimolecular exchange.

Ron Breslow: I'm sure we shall return to this issue of flexibility and rigidity. It's a very interesting problem; if you put too much rigidity into a catalyst, you in fact inhibit the motions that are required for the kinetic process you are trying to accelerate.

Stoddart: One needs to achieve a very fine balance between rigidity and flexibility. We have found that if we take the bridging paraxylylene units out of our charged polymers and make the bridges entirely polyether ones, the helicate structures do not form. To get formation of a supramolecular structure you need just sufficient rigidity yet not too much flexibility.

Ron Breslow: What is wanted is the maximum rigidity that doesn't interfere with the process; that is the key to high rates.

Sutherland: We will no doubt come back to this issue. I think it would be interesting to see what evolution has achieved, because it should have got the balance right!

Reference

de Jong F, Reinhoudt DN, Smit CJ, Huis R 1976 Kinetics and mechanism of complexation between *tert*-butylammonium hexafluorophosphate and 18-crown-6. Tetrahedron Lett 4783–4786

Host-guest interactions in thin membranes: selective ion transport and transduction into electronic signals

D. N. Reinhoudt, W. F. Nijenhuis and P. D. van der Wal

Department of Organic Chemistry, Twente University, P.O. Box 217, 7500 AE Enschede, The Netherlands

Abstract. Synthetic receptor molecules that selectively complex with charged guest molecules can be used to transport salts through liquid membranes and to transduce chemical information into electronic signals. In both cases the receptor molecules are present in thin membranes in contact with aqueous solutions. Extreme lipophilicity of the receptor molecules is therefore required: calix crown ethers and calixspherands meet these requirements. Their synthesis and complexation properties will be discussed. In order to mimic the large rates of transport through biomembranes, thin supported liquid membranes (< 100 μm), in which the receptor molecules are present, were investigated. The selective ion transport has been studied as a function of the experimental parameters and interpreted via computer simulations of the transport processes. The transduction of complexation into electronic signals can be achieved via the 'immobilization' of receptor molecules on the gate surface of an ISFET chip. Parameters that govern the signal transduction in multilayer systems have been studied and simulated.

1991 Host–guest molecular interactions: from chemistry to biology. Wiley, Chichester (Ciba Foundation Symposium 158) p 42–54

The increasing interest in the design and synthesis of molecular receptors and their complexing properties with charged and neutral guests has been stimulated by the possible applications of such synthetic receptors in chemistry, medicine, and material science. In this chapter we shall describe the synthesis of receptor molecules based on calix[4]arenes and their application in two fields: selective transport of salts through liquid membranes, and the direct transduction of chemical interactions into electronic signals.

These two potential applications have been selected from the ongoing research efforts in our group. They will serve to illustrate that the translation of complexation properties at the molecular level into the desired macroscopic properties of a 'chemical system' or a 'molecular material' is not a simple extrapolation, but a scientific challenge in itself. It is not a coincidence that in

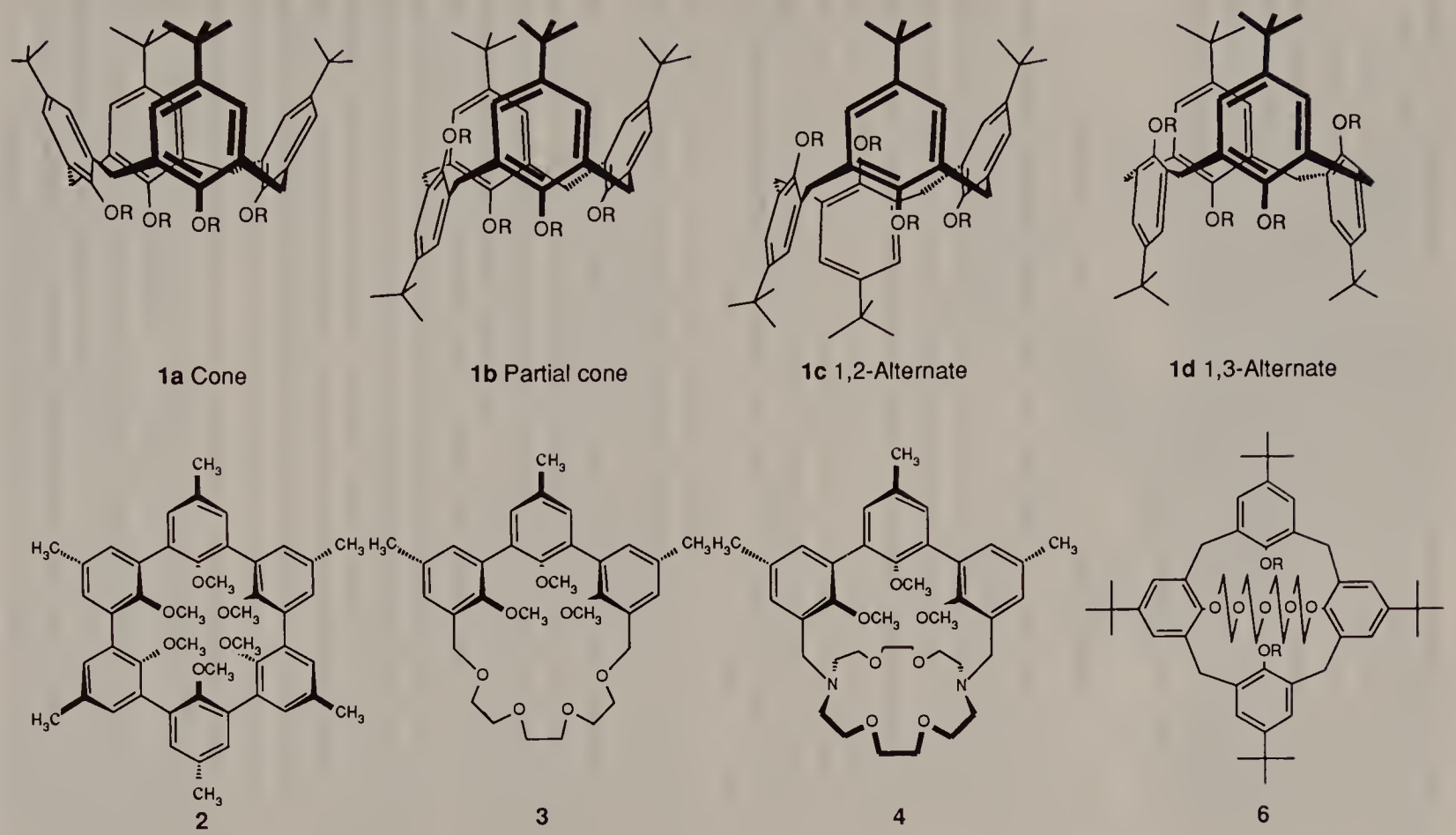


FIG. 1. Structures of macrocyclic receptor molecules; **1a–d**, four different conformations of *p-tert*-butylcalix[4]arene; **2**, spherand; **3**, hemispherand; **4**, cryptahemispherand; **6**, calix crown ether.

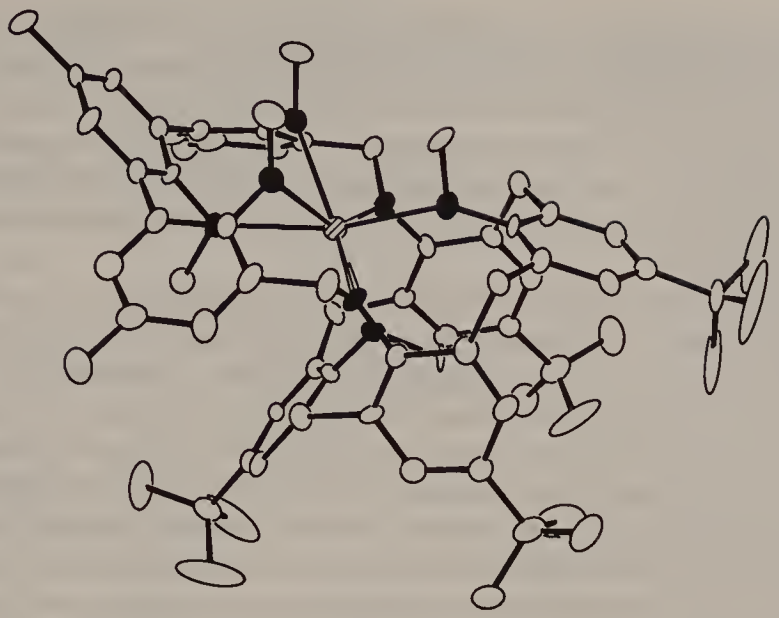
both examples the synthetic receptors are embedded in a thin organic film, because ultimately most applications of molecular receptors will require manipulation at the nanometre level (nanochemistry), or even at the molecular level.

The first part of this chapter will deal with the *synthesis* of receptor molecules based on calix[4]arenes. We decided to investigate this type of building block because the calix[4]arene structure (**1a–d**, Fig. 1) has (i) a potential molecular cleft or cavity, (ii) four functional OH groups and (iii) an intrinsic hydrophobic character that will be needed to avoid destabilization of the ultimate chemical system in contact with aqueous solutions.

Calix[4]arenes as building blocks for synthetic receptor molecules

The calix[4]arenes are easily accessible from the base-catalysed condensation of *p-tert*-butylphenol and formaldehyde, due to the pioneering work of Gutsche and co-workers (Gutsche et al 1989). Subsequent (selective) de-*tert*-butylation and/or reaction of one or more of the phenolic groups leads to a rapidly increasing number of different calix[4]arenes. Gutsche proposed to define the two faces of the calix molecules as the *lower* (phenolic groups) and *upper* rims; we shall use this nomenclature here. The calix[4]arenes with free OH groups are conformationally flexible and the molecules can adopt four different extreme conformations (**1a–d** in Fig. 1). When the phenolic groups are alkylated or acylated with a substituent larger than methyl, the interconversion of the different conformations is no longer possible. Recently we have found that tetra*methyl*calix[4]arenes are mixtures of conformers in which all four forms can be found (L. C. Groenen, unpublished work 1990).

We became interested in calix[4]arenes as building blocks in relation to the synthesis of a receptor for a radioactive isotope of Rb^+, which can be used for organ imaging (Van Herk & De Zeeuw 1978). Since in biological fluids Rb^+ will rapidly exchange with Na^+ and K^+ ions, only a kinetically stable complex of Rb^+ can be used. The spherands (**2**) (Fig. 1) reported by Cram & Lein (1985) form kinetically stable complexes with Na^+ and Li^+ but they reject K^+ and larger alkali cations completely. Hemispherands (**3**) or cryptahemispherands (**4**) (Fig. 1) do not form kinetically stable complexes in aqueous solutions (Dijkstra et al 1988, 1989) because the cations are not sufficiently shielded from the solution. Consequently, we designed a host molecule that combines the shielding ability of the spherands with the cavity size of the cryptahemispherands, by combining a calix[4]arene and a terphenyl bridge to form a calixspherand (**5**) (Fig. 2) that forms kinetically stable complexes with Na^+, K^+, and Rb^+ ions (M^+) (Dijkstra et al 1989). The half-life of decomplexation at room temperature varies between different cations from one year to two hours and the high kinetic stability is attributed to the partial cone conformation (**1b**) of the calix[4]arene moiety in the **5**.M^+ complex, which was observed both in the solid state (see Fig. 2) and by ^1H NMR spectroscopy in solution.



5.M⁺

FIG. 2. The X-ray-derived structure of the complex of an alkali cation ($M^+ = Na$) and a calixspherand (**5**).

Systematic variation of the bridge between O(2) and O(4) oxygen atoms of the 1,3-dimethoxy-calix[4]arene moiety yielded a series of receptor molecules that exhibit extremely high K^+/Na^+ selectivities. These receptors (**6**) (Fig. 1) have been used for the membrane transport experiments that are described in the second part of this chapter and for the selective transduction of complexation reactions into electronic signals, described in the third part. We have found that the selectivity of complexation can be further enhanced by increasing the preorganization (Cram & Lein 1985) of the binding sites in the calix crown ethers (**6**) (Fig. 1). This was achieved by alkylation of **6** (R = H) with more bulky alkyl groups ($R = C_2H_5$, n-C_3H_7, i-C_3H_7, $CH_2C_6H_5$; Ghidini et al 1990). The different conformers—cone, 1,3-alternate and partial cone (Fig. 1)—exhibit different selectivities and thermodynamic stabilities of complexation with different cations.

Receptor-mediated transport of salts through supported liquid membranes

The most remarkable observation by Pedersen (1967) in his studies on crown ethers was the shape-selective complexation of alkali cations—that is, that the stability of the complex is dependent on the size of the cation. Not surprisingly, the use of this property in selective separation of cations via membrane transport

was one of the first potential applications of these compounds to be investigated. Until recently, all experiments with this aim in mind were carried out in bulk liquid membranes (Lamb et al 1980) composed of two aqueous phases separated by a solution of the receptor molecules in an immiscible organic solvent. Although important results on the selectivity and rate of transport were obtained, the frequently mentioned analogy with biological membranes is hardly realistic (Stolwijk et al 1987).

We have therefore studied a system which is much closer to the dimensions of biomembranes, the supported liquid membrane system. The receptor molecules, dissolved in an organic solvent, are immobilized in a thin (micro)porous polymer membrane ($d = 100\,\mu m$). The rate of transport of different salts through such membranes has been studied as a function of a number of parameters of the system (Stolwijk et al 1989a). The initial studies revealed that the partition coefficient of the receptor molecule is extremely important (Stolwijk et al 1989b), together with the association constants of the complexes formed, the diffusion coefficient of the complex, and the partition of the salt between the aqueous and the membrane phases (source phase and receiving phase; Fig. 3). We can now describe the membrane flux for single ion transport with a model based on these independently determined parameters (Fig. 3).

Generally, such membrane systems are not stable because the receptor leaches out of the very thin membrane. In order to improve this stability, molecular properties, other than those of complexation, have to be introduced into the structure of the receptors. The calix crown ether (illustrated in Fig. 1) shows a large selectivity for K^+ in comparison with Na^+ and, when it is substituted with four *p-tert*-butyl groups, its partition coefficient increases to $>10^{10}$, sufficiently large to retain the receptors in the membrane for days. An alternative to the use of *p-tert*-butylcalix crown ethers is the covalent attachment of polysiloxanes to the receptor molecules (Wienk et al 1990).

When we studied the selectivity of the K^+/Na^+ transport through these supported liquid membranes with *o*-nitrophenyl octyl ether as the solvent, we observed a much lower selectivity than was expected on the basis of the complexation data for Na^+ and K^+ in chloroform (W. F. Nijenhuis, unpublished work 1990). A computer simulation of the entire transport process as a function of all parameters, that predicts the selectivity, has been developed.

Our results demonstrate that the extrapolation of molecular complexation properties to macroscopic properties, in this case selective membrane transport, requires a great deal of fundamental research.

Direct transduction of chemical information into electronic signals

In biological systems, selective interactions between organic hosts and (in)organic guests are used to store and transduce chemical information in order to maintain

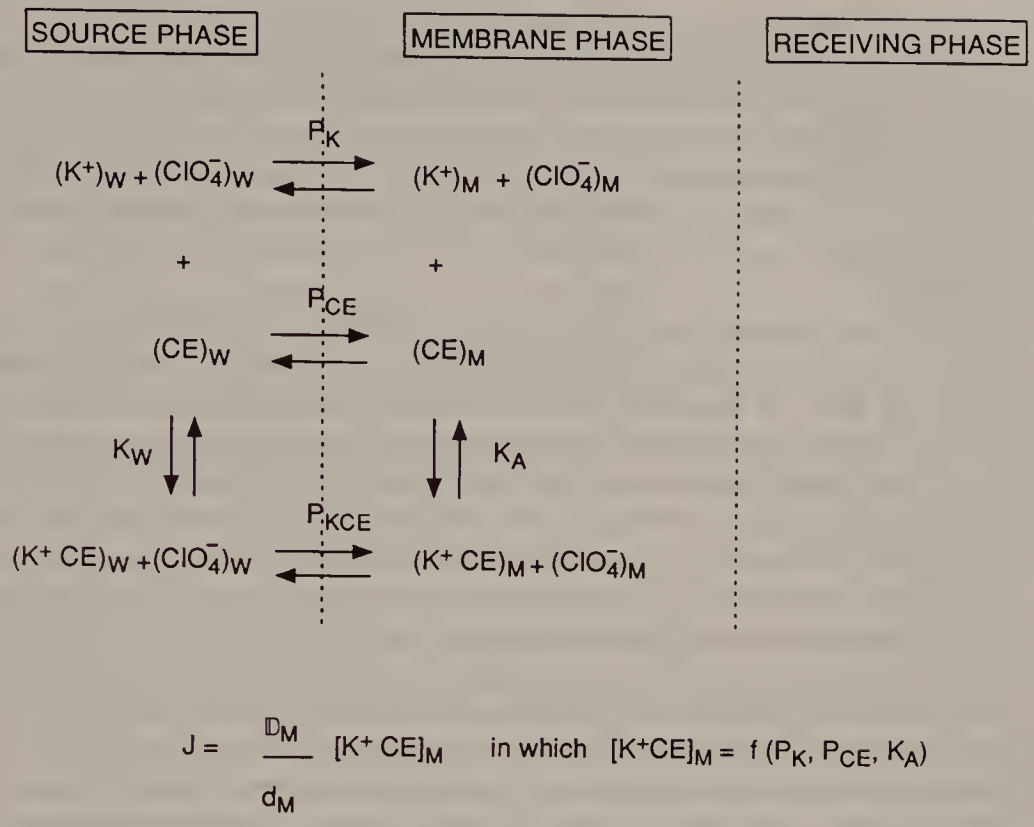


FIG. 3. Schematic representation of a liquid-immobilized membrane separating two aqueous phases. The source phase represents an aqueous solution of $KClO_4$ and the receiving phase only water. The membrane phase consists of a solution of the host molecule in *o*-nitrophenyl octyl ether, absorbed in a porous polymeric film ($d = 100$ μm) of polypropylene. W, aqueous phase; M, membrane phase; K_A, association constant of the complex formed in the membrane phase; CE, ligand, e.g. calix crown ether; P_{CE}, partition coefficient of ligand (membrane/aqueous); P_K, partition coefficient of K^+; J, flux of salt through the membrane ($Mm^{-2}s^{-1}$); D_M, diffusion coefficient of the complex in the membrane phase; d_M, thickness of the membrane; f, general function of variables (P_K, P_{CE}, K_A).

and replicate the species. Consequently, an important goal in host–guest (supramolecular) chemistry for the next decade will be to generate molecular systems that are able to transduce chemical interactions at the molecular level into observable quantities at the macroscopic level. Far-reaching ideas of extending supramolecular chemistry to larger assemblies with such functions have been outlined by Lehn (1988) in terms of 'nanochemistry' (10^{-9}–10^{-7} m). The application of such systems as molecular electronic components will require extension up to the 'nanochemistry' level and also a way to transduce molecular properties to the macroscopic level.

We have decided to follow a different approach for reaching these molecular systems, by integrating synthetic receptor molecules with a semiconductor chip.

This approach takes advantage of the ability to manufacture devices at the submicron level ($\geqslant 3.5 \times 10^{-7}$ m) (Santo & Wollard 1988). As the transducing element we have chosen a field-effect transistor and for the sensing element we have used a variety of macrocyclic polyethers and calix crown ethers. We have focused our attention on the integration of both the sensing and the transducing elements in one molecular system that can be regarded as a chemical entity with a designed function (chemically sensitive field-effect transistor, or CHEMFET).

A field-effect transistor (FET) is able to register the conductance of a semiconductor as a function of an electric field perpendicular to the SiO_2 gate oxide surface (Fig. 4). Bergveld (1970) demonstrated that the conductance of a FET is influenced by the interface potential at the oxide/aqueous solution because of the (de)protonation of silanol groups at the SiO_2 surface. He then showed that chemical information from the solution was transduced into an electronic signal that can be processed by the microchip.

We decided to use this FET to study direct transduction of complexation reactions between hosts and guests into electronic signals by attaching receptor molecules covalently to the SiO_2 gate surface of a FET. This required first the elimination of an undesired pH dependance by covalent modification of the SiO_2 surface (van den Berg et al 1985). A monolayer was not sufficient for this purpose, so we have focused on the covalent attachment of thin (10–25 μm) polymer layers to the 'gate' surface. Silylation with 3-(methoxysilyl)propyl methacrylate or 3-(triethoxysilyl)propylamine gave reactive surfaces that were subsequently modified by co-polymerization with acrylate monomers and prepolymers of polyurethane, respectively. These covalently attached layers

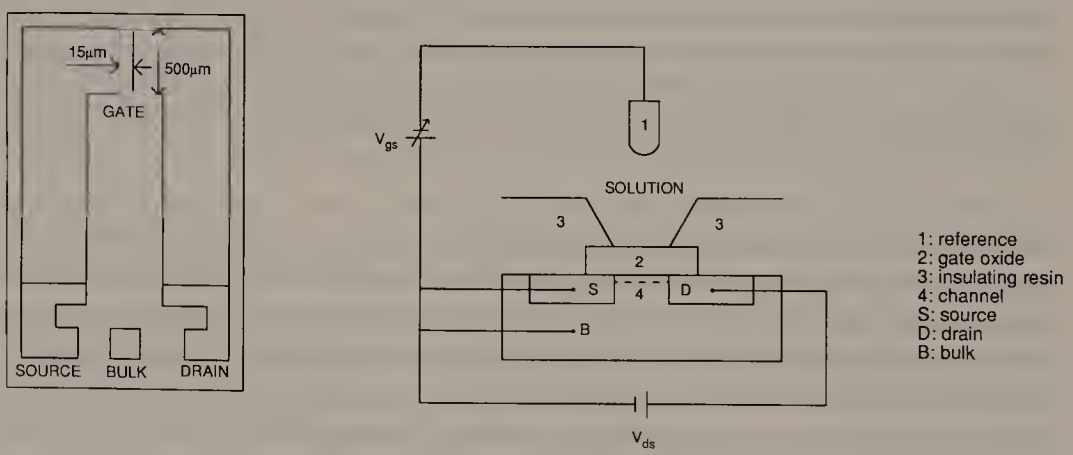


FIG. 4. Schematic representations (top and side views) of a field-effect transistor (FET) in which S, D, and B are the electrical contacts for measuring the relative electrical potentials. (1) electrical reference; (2) the insulating SiO_2 layer (gate); (3) the resin that insulates the devices from the solution; (4) the conducting channel between the two contacts S and D.

almost completely eliminated the pH sensitivity of the microchip. When small amounts of anionic species, always present in these polymers, are neutralized with small amounts of lipophilic cations, these devices are insensitive not only to pH but also to the ionic strength of aqueous solutions. This gives a reference FET (REFET) which is a prerequisite for the potentiometric measurement of chemical interactions (Skowronska-Ptasinska et al 1990).

The complete shielding of the SiOH groups by the covalently attached polymers has one disadvantage. The equilibria at the interface are no longer thermodynamically defined, because a common species (H^+ or OH^-) is lacking. This gives rise to unstable electrical potentials. We have solved this problem in the following way. The SiO_2 surface is first reacted with 3-(methoxysilyl)propyl methacrylate. The bulk of this reagent limits the effective silylation of the SiOH groups to 1.5 functional groups per nm^2 and consequently 3.5 SiOH groups per nm^2 surface will remain. Subsequent photo-co-polymerization with 2-hydroxyethyl methacrylate (HEMA) covers the SiO_2 surface with covalently attached polyHEMA (layer thicknesses between 5 and 15 μm). The polyHEMA was subsequently saturated with a buffered KCl solution to give a hydrogel. This means that at constant local pH the residual SiOH groups fix the interface potential of the FET (Sudhölter et al 1990).

The hydrogel was reacted with methacryloyl chloride in order to attach covalently reactive methacryloyl groups, and a polysiloxane membrane containing methacrylate groups was covalently attached (Fig. 5). The structure of this polysiloxane membrane, which will ultimately contain the receptor molecules, has to meet very distinct requirements for the transduction of host–guest interactions to the FET device.

We have optimized the glass transition temperature (T_g) and the dielectric constant (ϵ) by systematic variation of the ratio of the different siloxane building blocks and we have introduced receptor molecules that selectively complex alkali cations (van der Wal et al 1990) (Fig. 5). Again it proved to be important that the receptor molecules are sufficiently hydrophobic, in order that they should not leach out of the thin membranes that are covalently linked to the chip. These membranes have a distinct upper limit of thickness; too thick layers isolate the FET electrically. Calix crown ethers showed excellent transduction of variable concentrations of K^+ in the range of 10^{-4}–10^{-1} M, even when 10^{-1} M of NaCl was present. These results demonstrate that our first objective—the *covalent* integration of molecular receptors and a FET into a chemical system (CHEMFET)—has been achieved. This molecular system connects the chemical and the electronic domains (Reinhoudt & Sudhölter 1990). Computer simulations of the entire system are in progress.

The first practical applications of our approach to the transduction of complexation reactions into electronic signals are chemoselective sensors. When such a sensor was incorporated into a flow injection device we were able to measure K^+ concentrations in biological fluids in the millimolar range.

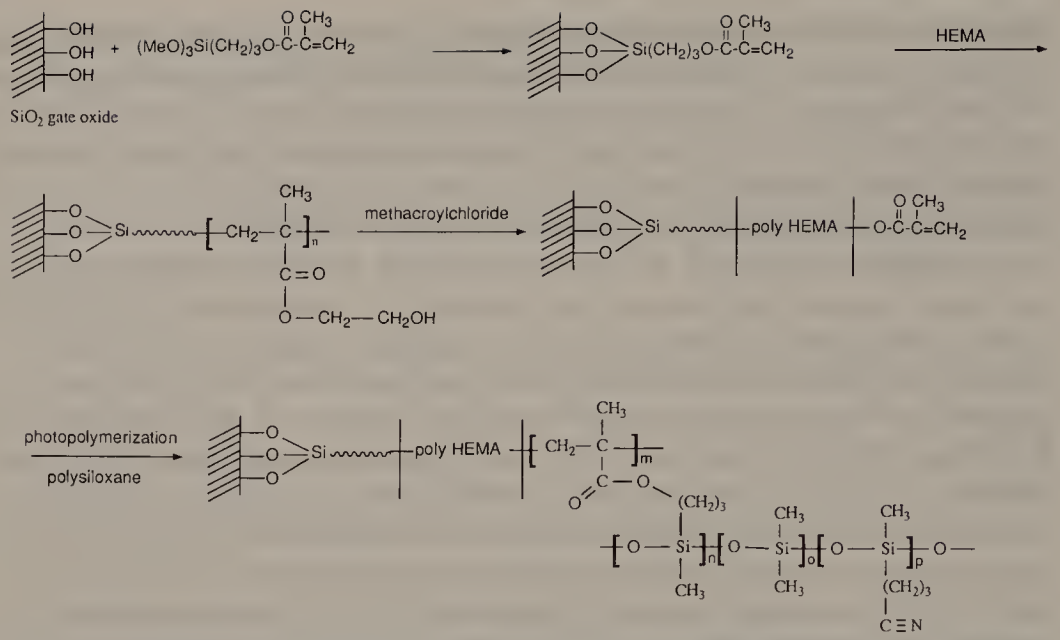


FIG. 5. Schematic representation of the chemical substructures of a membrane covalently attached to the SiO_2 gate surface of a field-effect transistor. HEMA, hydroxyethyl methacrylate.

These two examples of how complexation properties of receptor molecules can be extrapolated to macroscopic properties, (the transport of salts through liquid membranes, and the transduction of chemical into electronic signals) demonstrate that this process requires more than just creating larger assemblies of receptor molecules.

Acknowledgements

The authors would like to thank their colleagues, who are mentioned in the list of references, for their contributions to this work. Financial support for our research was provided by the Centre of Microelectronics (CME) Twente, the Koninklijke/Shell Laboratories in Amsterdam, the Netherlands Technology Foundation (STW), and the EEC Science programme.

References

Bergveld P 1970 Development of an ion-sensitive solid-state device for neurophysiological measurements. IEEE (Inst Electr Electron Eng) Trans Biomed Eng 17:70–71

Cram DJ, Lein GM 1985 Host–guest complexation. 36. Spherand and lithium and sodium ion complexation rates and equilibria. J Am Chem Soc 107:3657–3668

Dijkstra PJ, Den Hertog HJ Jr, Van Eerden J, Harkema S, Reinhoudt DN 1988 4*H*-pyran and pyrylium hemispherands: partly preorganized ionophores with reactive molecular cavities. J Org Chem 53:374–382

Dijkstra PJ, Brunink JAJ, Bugge K-E et al 1989 Kinetically stable complexes of alkali cations with rigidified calix[4]arenes: synthesis, X-ray structures, and complexation of calixcrowns and calixspherands. J Am Chem Soc 111:7567–7575

Ghidini E, Ugozzoli F, Ungaro R, Harkema S, Abu El-Fadl A, Reinhoudt DN 1990 Complexation of alkali metal cations by conformationally rigid, stereoisomeric calix[4]arene crown ethers: a quantitative evaluation of preorganization. J Am Chem Soc 112:6979–6986

Gutsche CD 1989 Calixarenes. Royal Society of Chemistry, Cambridge (Monographs in supramolecular chemistry, vol 1)

Lamb JD, Christensen JJ, Oscarson JL, Nielsen BL, Asay BW, Izatt RM 1980 The relationship between complex stability constants and the rates of cation transport through liquid membranes by macrocyclic carriers. J Am Chem Soc 102:6820–6824

Lehn J-M 1988 Supramolekulare Chemie-Moleküle, Übermoleküle und molekulare Funktionseinheiten (Nobel-Vortrag). Angew Chem 100:91–116

Pedersen CJ 1967 Cyclic polyethers and their complexes with metal salts. J Am Chem Soc 89:7017–7036

Reinhoudt DN, Sudhölter EJR 1990 The transduction of host–guest interactions into electronic signals by molecular systems. Adv Mat 2:23–32

Santo B, Wollard K 1988 The world of silicon: it's a dog eat dog. IEEE (Inst Electr Electron Eng) Spectrum 25:30–39

Skowronska-Ptasinska M, van der Wal PD, van den Berg A, Bergveld P, Sudhölter EJR, Reinhoudt DN 1990 Reference field effect transistors based on chemically modified ISFETs. Anal Chim Acta 230:67–73

Stolwijk TB, Sudhölter EJR, Reinhoudt DN 1987 Crown ether mediated transport: a kinetic study of potassium perchlorate transport through a supported liquid membrane containing dibenzo-18-crown-6. J Am Chem Soc 109:7042–7047

Stolwijk TB, Sudhölter EJR, Reinhoudt DN 1989a Effect of crown ether lipophilicity on the facilitated transport of guanidinium thiocyanate through an immobilized liquid membrane. J Am Chem Soc 111:6321–6329

Stolwijk TB, Vos LC, Sudhölter EJR, Reinhoudt DN 1989b Partition coefficients of crown ethers. Recl Trav Chim Pays-Bas 108:103–108

Sudhölter EJR, van der Wal PD, Skowronska-Ptasinska M, van den Berg A, Bergveld P, Reinhoudt DN 1990 Modification of ISFETs by covalent anchoring of poly(hydroxyethyl methacrylate) hydrogel. Introduction of a thermodynamically defined semiconductor-sensing membrane interface. Anal Chim Acta 230:59–64

van den Berg A, Bergveld P, Reinhoudt DN, Sudhölter EJR 1985 Sensitivity control of ISFETs by chemical surface modification. Sensors & Actuators 8:129–148

van der Wal PD, Skowronska-Ptasinska M, van den Berg A, Bergveld P, Sudhölter EJR, Reinhoudt DN 1990 New membrane materials for potassium-selective ion-sensitive field-effect transistors. Anal Chim Acta 231:41–52

Van Herk G, De Zeeuw D 1978 Unilateral kidney blood flow measurement using the $^{81}Rb/^{81m}Kr$ ratio. Contrib Nephrol 11:67–72

Wienk MM, Stolwijk TB, Sudhölter EJR, Reinhoudt DN 1990 Stabilization of crown ether containing supported liquid membranes. J Am Chem Soc 112:797–801

DISCUSSION

Albery: You showed a diagram of your MOSFET (metal oxide field-effect transistor) device (Fig. 4) in which the metal contact was replaced by the solution. In that diagram the device was very big but you illustrated the reference electrode as very small. I imagine that the sizes are the opposite way around, with the

reference electrode being quite large; or have you miniaturized the reference electrode?

Reinhoudt: For the electronics field, that is the important question. We were not the first to use FETs, but I think we were the first chemists to integrate the whole system covalently so that it has a long life-time. The reference electrode is made in exactly the same way as the CHEMFET. We remove the pH sensitivity of the original silicon dioxide in exactly the same way as we do for a CHEMFET, by covering it with a series of polymer layers, but leaving out the ionophore. On top of that there is a delicately balanced equilibrium between hydrophobic anions and hydrophobic cations. It has been recognized only recently that in most polymers there are some anionic sites that are introduced during the synthetic process by an emulsifier or a starter, for example. By high resolution secondary ion mass spectrometry we have shown that anionic sites are present in all polymers that we have investigated so far, even PVC (van den Berg et al 1987). Because of this, there will be a response to changes in ionic strength. You can eliminate that effect almost completely using a small amount of a quaternary ammonium cation and a long alkyl chain which has a high partition coefficient.

Albery: What determines the potential of that reference electrode?

Reinhoudt: That's a mystery. Both the anions and the cations will determine the potential, but when you combine their effects they compensate each other to give a potential of almost zero over a useful range of ionic strength.

Albery: That is most mysterious.

Reinhoudt: This, for us, is the most difficult thing to understand. When you model the whole system in order to find out how the potential is built up at the interface, you find that on changing from an immobilized anion to an immobilized cation there should be a sharp change from cation to anion selectivity. I think the easiest way to view the reference electrode is as a combination of a cation- and an anion-selective electrode, with the same number of anionic and cationic sites.

Stoddart: You described the anchorage as being of a covalent nature. Did I misunderstand you? Surely, the calix crown was non-covalently attached in some way to the polysiloxane. If it was, what is the nature of that interaction? Does the difficulty of attaching the calix crown covalently limit the device's life-time? Do you plan to attach it covalently?

Reinhoudt: The covalent attachment I referred to is of the membrane system (the hydrophilic membrane, the intermediate layer and the hydrophobic membrane) to the ISFET chip. There has been much debate about the possibility of linking the carrier molecules to the hydrophobic membrane. You can model the potential build-up at the interface because the diffusion potential is a result of the mobility of all the species involved, assuming there has to be a thermodynamic equilibrium. If you immobilized the ionophore it would have zero mobility and, theoretically, there would no longer be a potential. But what

is zero mobility? Could you attach the ionophore via a long chain and still maintain some mobility? We have done some experiments involving covalent attachment not of the calixarene, but of a calcium-selective ionophore. We see a normal Nernstian response of 30 mV per decade when this ionophore is covalently attached to the membrane. The attachment is made by co-polymerization in the final stage of preparation of the hydrophobic membrane.

Although the calix crown is not covalently attached, our K^+ sensor works for at least six months because the partition coefficient of the calix crown is about 10^{14}, so it is retained in the membrane. 18-Crown-6 or a cryptand would be retained in the membrane no longer than minutes.

Gokel: You said that your molecules were highly selective for K^+ over Na^+, but that the selectivity assessed in *o*-nitrophenyl octyl ether did not correspond to that determined in chloroform. Do you have any idea how the selectivity of your molecules compares with that of valinomycin, the selectivity of which is known in a number of media? Also, why did you choose to use a calixarene rather than valinomycin, when valinomycin is probably more selective and probably equally hydrophobic (although it might prove quite hard to attach covalently)?

Reinhoudt: Valinomycin is about three times more selective than the calix crown. My estimate of valinomycin's selectivity in chloroform is about 10 000:1. We have also tested valinomycin in a liquid-immobilized membrane transport experiment (Stolwijk et al 1989). There the K^+/Na^+ selectivity is about 60:1.

Gokel: What was the selectivity of the calixarene there?

Reinhoudt: It was 23:1.

Gokel: So valinomycin is about three times more selective in that situation also.

Reinhoudt: The difference in selectivity is related to the transfer energy of both ligands (K^+ and Na^+) between one solvent and another.

Gokel: Why do you not use valinomycin?

Reinhoudt: We began this work with the aim of making a completely stable, covalently attached system. It's not easy to covalently attach valinomycin to the hydrophobic membrane. I wouldn't say that it would be impossible to construct a device with valinomycin that has almost the same response as one with a calixarene, but although the initial response of valinomycin at very low K^+ concentrations is a little better than that of the calixarenes, in the relevant concentration range (10^{-3}–10^{-2}) it is not.

Ron Breslow: In your experiment you were changing a DC signal to an AC signal. There are a number of things that you can do with AC that you can't do with DC, such as studying phase relationships and non-linear responses. Do you think you will go on to use an AC rather than a DC signal?

Reinhoudt: I am not a solid-state physicist but I understand more or less how the ISFET works. The ISFET measures a change of the membrane potential and that change is transduced as a difference in the current in the channel, which

is immediately compensated for by a change in the potential. You keep the current constant by changing the potential and that's what you actually record (see Fig. 4).

Albery: If you go over to using an AC rather than a DC technique, the kinetics of the system will become significant. At the moment, you don't need to bother about your rather curious thermodynamics. I think the problems that this would cause would outweigh the advantages.

Ron Breslow: Problems are a good thing!

Reinhoudt: We have tried to measure the response rate of the CHEMFET, but we could not record the build-up of the signal quickly enough to see whether the rate of complexation is the rate-determining step. We have never been able to follow the process of signal build-up; it reaches the maximal level almost immediately.

Albery: Does it happen over a microsecond time-scale?

Reinhoudt: I don't think we can measure this in membranes, but it's faster than 0.1 second.

References

Stolwijk TB, Sudhölter EJR, Reinhoudt DN 1989 Effect of crown ether lipophilicity on the facilitated transport of guanidinium thiocyanate through an immobilized liquid membrane. J Am Chem Soc 111:6321–6329

van den Berg A, van der Wal P, Skowronska-Ptasinska M, Sudhölter EJR, Reinhoudt DN, Bergveld P 1987 Nature of anionic sites in plasticized poly(vinyl chloride) membranes. Anal Chem 59:2827–2829

Molecular recognition and molecular sensors

W. John Albery

Molecular Sensors Unit, University College, Oxford OX1 4BH, UK

Abstract. Enzyme–substrate recognition provides a convenient and powerful basis on which to construct molecular sensors. In direct enzyme electrodes the rate of the enzyme reaction is transduced into a current using an electrode made of a conducting organic salt. *In vivo* microelectrodes designed to measure glucose have been constructed and used in the brain of the freely moving rat. Another strategy is to use enzymes that operate with NADH; the NADH can readily be oxidized on conducting organic salt electrodes. Results for the measurement of micellar equilibria involving bile acids are presented. The packed-bed wall-jet electrode provides a device of greater sensitivity; results for the measurement of femtomoles of acetylcholine obtained by microdialysis from cerebral fluid demonstrate the power of this method. The wall-jet ring-disc electrode can be used in an electrochemical immunoassay again at the femtomole level. Finally, enzyme inhibition can be used to make a sensor for toxic substances such as H_2S at the p.p.m. level.

1991 Host–guest molecular interactions: from chemistry to biology. Wiley, Chichester (Ciba Foundation Symposium 158) p 55–72

A molecular sensor is a device which uses molecular recognition to measure the concentration of a target analyte in the presence of other substances and in real time. Conventional analytical procedures involve the separation of the target analyte from its original environment. Such procedures, for instance chromatography, require relatively elaborate equipment and take time. By contrast, the molecular sensor can be dipped straight into fluids such as blood, raw sewage or fermenting beer and will give a reading of the concentration in a matter of seconds. A further advantage of the direct-reading molecular sensor is that in most cases it does not perturb the chemical equilibria in the sample under investigation. Separation procedures inevitably do so. Hence, using molecular sensors one can measure the activity of target species *in situ*. Our work on bile acids described below shows how this can be done.

One strategy designed to achieve recognition without separation is to build into the device an act of molecular binding. Ion-selective electrodes are a well-established example where crown ethers, for instance, have been designed and synthesized to have cavities of the right size to trap the target ion. It might be

thought that it is a relatively simple matter to design a molecule with the right-sized spherical cavity but, in practice, even this simple task has led to more misses than hits. When one turns to synthesizing molecules that recognize more complicated shapes, such as a glucose molecule or a hormone, then the problem can be described as challenging or well-nigh impossible, depending on whether you are an optimist or a pessimist. The advances described in this symposium may point the way to the future but, up to now, we have had to rely on chemical evolution, which has provided us with a range of molecules that recognize their substrates—the enzymes. And of course the substrates that enzymes recognize are those very molecules that are important in biochemistry, clinical medicine, neurophysiology and the life sciences in general. Hence, the use of enzymes to provide the act of molecular recognition has much to commend it.

In a molecular sensor, in order to obtain a signal one has to monitor in some way the act of molecular recognition. If one is using an enzyme the simplest method is to measure the rate of the reaction of the enzyme with its substrate. One of the most fruitful advances of the last decade has been the development of enzyme electrodes where the rate of the enzymic reaction is followed electrochemically. It had been thought that enzymes and electrodes were incompatible, with the biochemist complaining that 'electrodes denature my lovely proteins' and the electrochemist wingeing that 'proteins poison my beautiful electrode'. But several strategies have now evolved by which the enzyme chemistry and the electrochemistry can be happily married. For instance, H. A. O. Hill and I. J. Higgins (Cass et al 1984) have successfully pioneered the use of inorganic mediators, such as different sorts of ferrocene. In our work we have developed the use of conducting organic salts as electrode materials (Albery et al 1985a, 1987a,b,c). In this paper I shall describe various types of enzyme electrode and their applications.

Direct enzyme electrodes

The advantage of an electrode made of a conducting organic salt is that one can achieve the oxidation of the active site of the enzyme on the surface of the electrode (Albery et al 1987a). One does not have to add a mediator to act as an electron shuttle. Writing the active site of a flavoenzyme as FAD, the reaction scheme is as shown in (1),

$$\begin{array}{ll} \text{Solution} & SH_2 + FAD \rightarrow S + FADH_2 \\ \text{TTF-TCNQ electrode} & FADH_2 \rightarrow FAD + 2H^+ + 2e \end{array} \qquad (1)$$

where TTF (tetrathiafulvalene) and TCNQ (tetracyanoquinone) are as shown in Fig. 1.

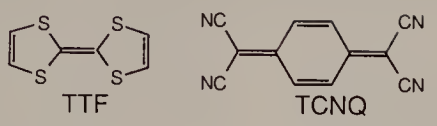


FIG. 1. Structures of tetrathiafulvalene, TTF, and tetracyanoquinone, TCNQ.

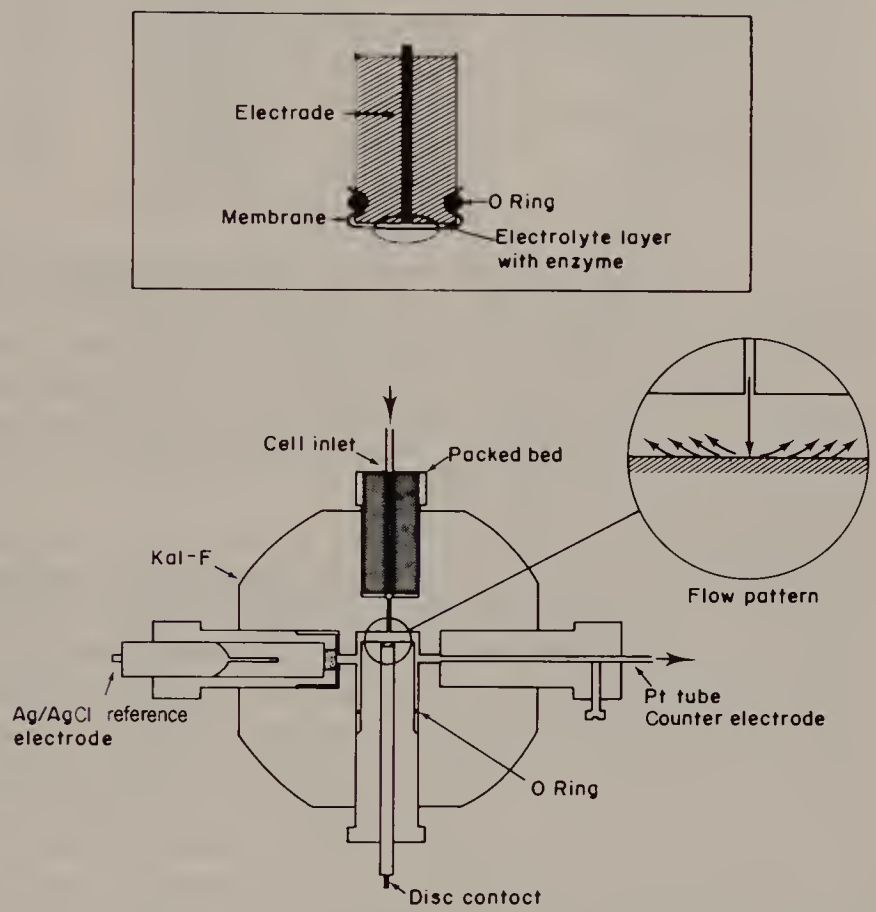


FIG. 2. Two different types of enzyme electrode: a direct (membrane) electrode (*top*) and the packed-bed wall-jet electrode (*bottom*).

In Table 1 we list combinations of enzymes and substrates which we have shown can be used as enzyme electrodes according to this scheme. A typical enzyme electrode is illustrated in Fig. 2 (upper diagram). The enzyme is retained in a thin film of solution behind a dialysis membrane. In general we find linear calibration plots between micromolar and millimolar concentrations of substrate. In many cases the rates of the enzymic and electrochemical reactions are so fast that the response is controlled by the diffusion of substrate through the

TABLE 1 Enzymes and substrates for direct enzyme electrodes

Enzyme	*Number*	*Substrate*
L-Amino-acid oxidase	EC 1.4.3.2	L-Amino acids
D-Amino-acid oxidase	EC 1.4.3.3	D-Amino acids
Choline oxidase	EC 1.1.3.17	Betaine aldehyde Choline
Glucose oxidase	EC 1.1.3.4	Glucose
(*S*)-2-Hydroxy-acid oxidase	EC 1.1.3.15	Glycolate
Lactate 2-monooxidase	EC 1.13.12.4	Lactate
Pyruvate oxidase	EC 1.2.3.3	Pyruvate
Salicylate 1-monooxygenase	EC 1.14.13.1	Salicylate
Sulphite oxidase	EC 1.8.3.1	Sulphite
Xanthine oxidase	EC 1.1.3.22	Xanthine

membrane. This is desirable in the operation of the sensor because it means that the response is not affected by a decline in the absolute activity of the enzyme. We have run a glucose electrode continuously for 28 days (Albery et al 1987a). At the end of the 28 days the response was only 10% less than that at the start of the experiment. Kinetic analysis showed that this slight decline was due to deterioration of the membrane. This type of result is very encouraging for the development of *in vivo* monitors.

Another feature of the system which is helpful in making *in vivo* electrodes is that an enzyme like glucose oxidase is very strongly adsorbed onto the electrode. Once an electrode has been exposed to glucose oxidase, the only way the enzyme can be removed is by polishing the electrode. This means that in making microelectrodes one can simply mix the enzyme and the organic salt together. There is no need to trap the enzyme behind a membrane. Typical results (Boutelle et al 1986) for the variation of glucose concentration in the brain of the freely moving rat are shown in Fig. 3. They are compared with the blood glucose concentration determined in the conventional way, which involves using a centrifuge, a column and a glucose analyser. It can be seen that there is a good correlation between the two techniques, both showing the drop in glucose on the injection of insulin and the rise in glucose level when the rat was given a biscuit. Such electrodes can continue to give results for over a month.

NADH enzyme electrodes

Over 250 enzymes use the ubiquitous cofactor NADH. We have found that NADH can be readily oxidized on an electrode made of the conducting organic salt NMP-TCNQ, where NMP is the *N*-methylphenazinium ion. This allows us a second strategy by which to construct enzyme electrodes. In this strategy instead of oxidizing the active site of the enzyme we oxidize the NADH, as shown

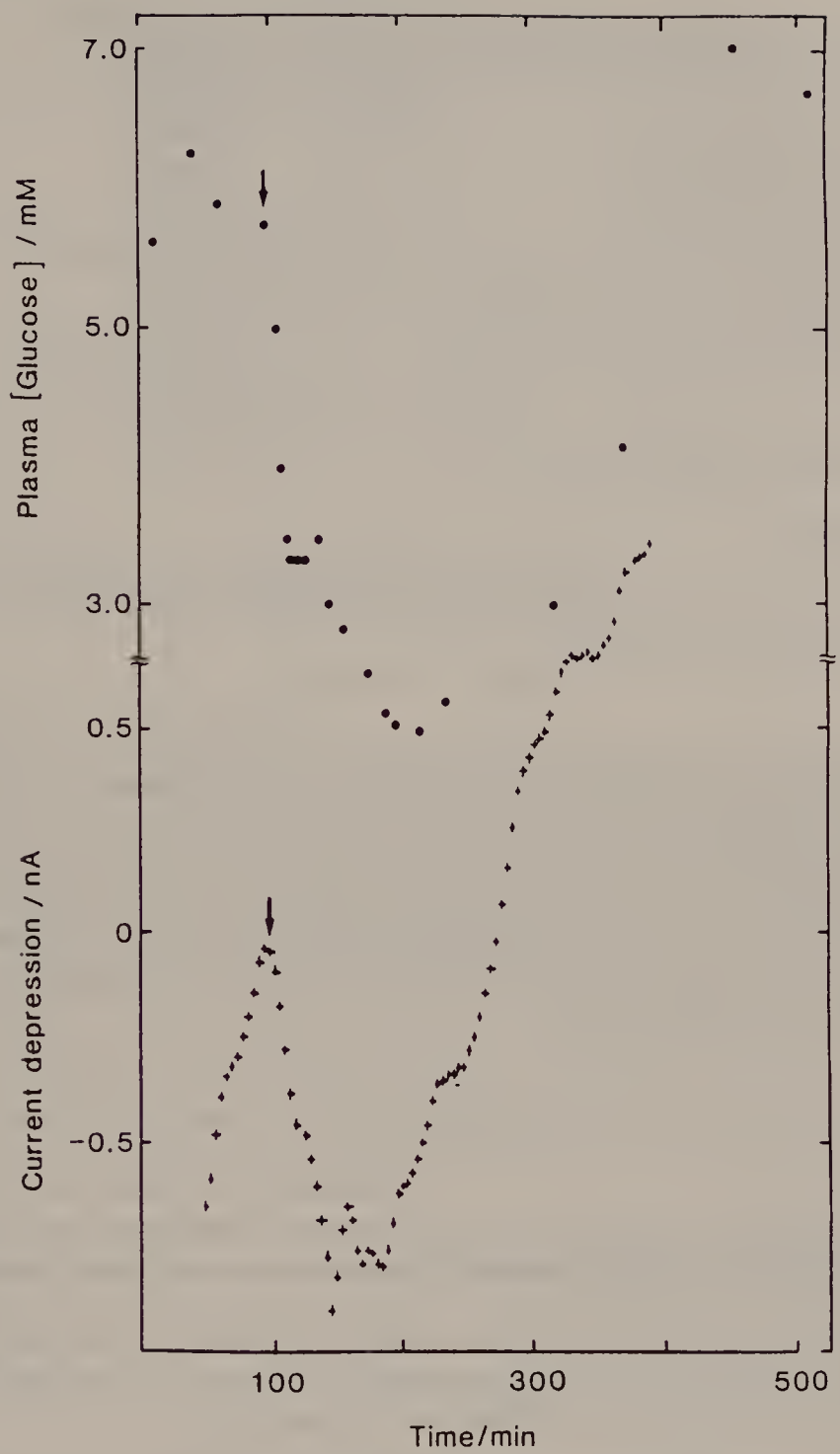


FIG. 3. A comparison of the traditional measurement of blood glucose (●) with the current (+) from an enzyme electrode implanted in the brain of a freely moving rat. The arrow shows the time when insulin was injected. The glucose levels rose when the rat was given a biscuit.

TABLE 2 Enzymes and substrates for NADH enzyme electrodes

Enzyme	*Number*	*Substrate*
Alcohol dehydrogenase	EC 1.1.1.1	Ethanol
Glutamate dehydrogenase	EC 1.4.1.2	Glutamate
Glycerol dehydrogenase	EC 1.1.1.6	Glycerol
3-Hydroxybutyrate dehydrogenase	EC 1.1.1.30	3-Hydroxybutyrate
3α-Hydroxysteroid dehydrogenase	EC 1.1.1.50	Bile acids
Monoethylene glycol dehydrogenase	—	Monoethylene glycol

in (2). In Table 2 we collect together examples of substrates and enzymes we can use in this reaction scheme.

Solution

$$SH_2 + NAD^+ \xrightarrow{\text{Dehydrogenase}} S + NADH + H^+ \qquad (2)$$

NMP-TCNQ electrode

$$NADH \longrightarrow NAD^+ + H^+ + e$$

In our studies on bile acids we have used the enzyme electrode to investigate the equilibrium between monomers of the bile acid and the amount of bile acid present as micelles. Conventional analytical procedures do not discriminate between the two forms but measure the total amount of bile acid present. By fitting the electrode with dialysis membrane of a porosity that excludes the micelles we can measure the concentration of the free monomer (M). The biosensor uses so little material that it does not perturb the equilibrium. We can write for the equilibrium:

$$nM \overset{K}{\rightleftarrows} M_n$$

Note that this is a true equilibrium and we are not assuming the usual phase-separation model to describe micelles. The total concentration of bile acid, c, is given by $c = m + nKm^n$, where m is the concentration of free monomer measured by the sensor. It is helpful to define a 'critical micelle concentration', c_*, as that concentration where the amount of monomer, m, is equal to the amount in the micelles, $c - m$. Then we find that

$$\log(y) = \log[(c-m)/m] = (n-1)[\log(m) - \log(c_*/2)]$$

In Fig. 4 we display some typical experimental results for taurodeoxycholic acid. The linear plot shows the success of the model. In Table 3 we

TABLE 3 Values obtained for the number of monomers in the micelle (n) and for the critical micelle concentration (c_*) for different bile acids using an NADH enzyme electrode

Bile acid	*Conjugate*	*n*	$c_*(mM)$
Ursodeoxycholic acid	None	22	1.52
Chenodeoxycholic acid	None	40	1.24
Ursodeoxycholic acid	Glycine	6.5	1.72
Chenodeoxycholic acid	Glycine	27.5	1.56
Deoxycholic acid	Glycine	101	1.47
Chenodeoxycholic acid	Taurine	8.3	1.73
Deoxycholic acid	Taurine	31	1.53

collect some values of n (the number of monomers in the micelle) and c_* for different bile acids. This work shows the advantage of being able to measure *in situ* concentrations without disturbing the equilibrium.

Packed-bed wall-jet electrodes

An alternative strategy to using the membrane electrode is to immobilize the enzyme on a packed bed. As shown in Fig. 2, the analyte flows through the packed bed and then onto a wall-jet electrode (Albery et al 1985b). The wall-jet system is chosen because it has known hydrodynamics, a small dead space and high sensitivity. The advantage of this system is that the contact time of the analyte with the bed can be long enough for complete conversion of substrate into product by the immobilized enzyme. This leads to higher sensitivity than the membrane electrode (Albery et al 1990a). The natural level of acetylcholine in the rat's brain is approximately nanomolar; a membrane electrode would not be sufficiently sensitive. We therefore use microdialysis (Ungerstedt 1986) to sample the brain fluid. We find that collection over 25 minutes will typically yield 12 femtomoles of acetylcholine.

Our assay (Albery et al 1988) then uses three coupled enzyme reactions on the packed bed. First we use acetylcholinesterase to turn acetylcholine into choline:

$$\text{Acetylcholine} \xrightarrow{\text{Acetylcholinesterase}} \text{Choline}$$

The choline is oxidized by choline oxidase to betaine:

$$2O_2 + \text{Choline} \xrightarrow{\text{Choline oxidase}} \text{Betaine} + 2H_2O_2$$

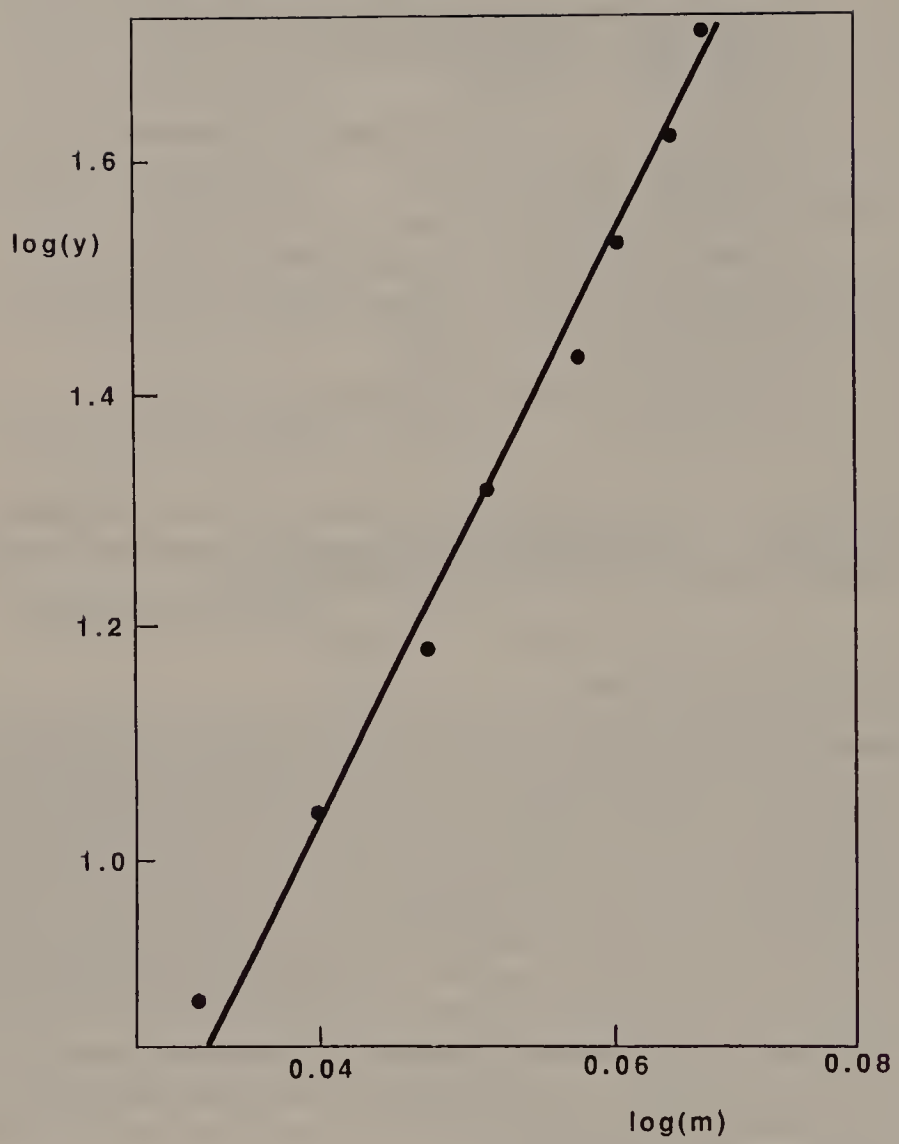


FIG. 4. Typical plot of log(y) against log (m) to test the equilibrium model for micelle formation for taurodeoxycholic acid.

Many classical assays would then detect H_2O_2 electrochemically. This electrochemical reaction is badly behaved, so it is better to follow Hill (Frew et al 1986) and use a third enzyme, horseradish peroxidase, to convert the badly behaved H_2O_2 into the well-behaved ferricinium ion, Fcp^+:

$$H_2O_2 + 2Fcp \xrightarrow{\text{Horseradish peroxidase}} H_2O + 2Fcp^+$$

The ferricinium ion can then be easily reduced at a mild potential on a glassy carbon wall-jet electrode. We have shown that we can measure amounts as small as 0.1 femtomoles. Using this technique we have monitored the effects of atropine on acetylcholine levels in the brain of the freely moving rat.

Electrochemical immunoassays

The wall-jet electrode can also be used in electrochemical immunoassays. In these assays the molecular recognition is provided not by an enzyme binding its substrate but by the antigen–antibody reaction. We have developed a sensitive technique for measuring human immunoglobin G (IgG) (Fig. 5). Goat anti-human IgG is immobilized on the glassy carbon disc of a wall-jet assembly. The usual sandwich procedure is then employed. The wall-jet electrode is dipped into the analyte containing the IgG, which binds specifically to the anti-IgG. The electrode is then exposed to goat anti-human IgG that had been conjugated with horseradish peroxidase. The wall-jet electrode is then fitted into its assembly. In this case the disc electrode is surrounded with a concentric ring electrode, giving a wall-jet ring-disc electrode (Albery & Brett 1983). The horseradish peroxidase is 'switched on' by supplying it with its two substrates,

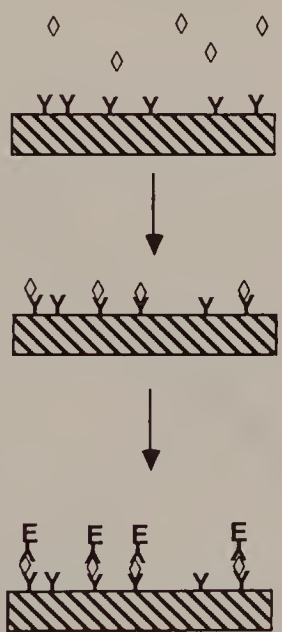


FIG. 5. The electrochemical immunoassay of immunoglobulin G. Goat anti-human IgG (Y) is immobilized on the glassy carbon disc of a wall-jet assembly (*top*), which is then dipped into the target analyte containing human IgG (◇) (*middle*). The electrode is then exposed to goat anti-human IgG coupled to horseradish peroxidase (E) (*bottom*). Horseradish peroxidase activity is thus a measure of IgG in the sample.

ferrocene in the flowing buffer and H_2O_2. The H_2O_2 is generated electrochemically on an upstream packed-bed electrode. The immobilized enzyme on the disc generates ferricinium ion which is then detected downstream on the concentric ring electrode.

The method has very high sensitivity for a number of reasons. First, each target molecule is labelled with an enzyme turning over a thousand times a second. Secondly, using the ring-disc hydrodynamics, the ferricinium ion is measured before it becomes diluted in the bulk of the solution. Thirdly, the electrochemical generation of the H_2O_2 substrate means that one can discriminate against unwanted background current. It is the difference in current with and without H_2O_2 that measures the Fcp^+ produced by the enzyme. With a ring current of 1 nA corresponding to a flux of 10^{-14} mol s^{-1} and an enzyme turning over at 10^3 s^{-1} it is possible to measure amounts as small as 10^{-17} mol of material on the disc. A typical calibration plot is shown in Fig. 6. In this experiment a straight line is obtained in the nanomolar range.

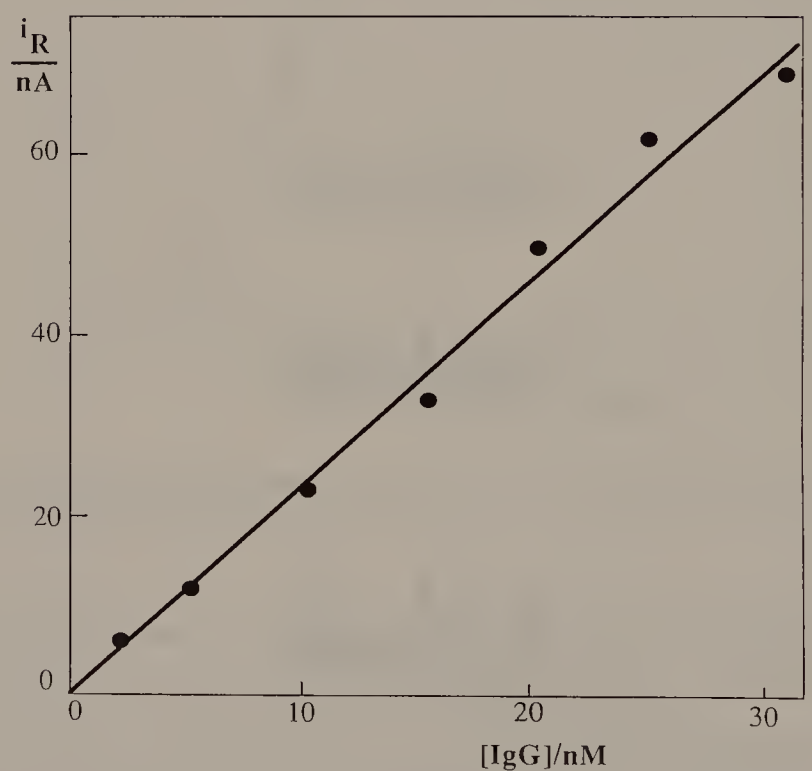


FIG. 6. A typical calibration plot for the electrochemical immunoassay of human immunoglobulin G.

Inhibition enzyme electrodes

In the immunoassay just described we can see how sensitivity can be achieved by switching on the enzyme horseradish peroxidase. For the same reasons we can also achieve great sensitivity by switching off an enzyme with an inhibitor. In classical amperometry one would measure, say, two electrons per analyte molecule; an inhibitor that stops an enzyme turning over a thousand times per second has a much more dramatic effect. We have developed a sensor for toxic gases such as H_2S. The sensor uses the respiratory enzyme cytochrome oxidase (E), which undergoes a cycle of reactions with oxygen and cytochrome *c*(C), as shown in (3).

$$\begin{aligned} \mathrm{E(red)} + \mathrm{O_2} &\rightarrow \mathrm{E(ox)} + \mathrm{H_2O_2} \\ \mathrm{E(ox)} + \mathrm{C(red)} &\rightarrow \mathrm{E(red)} + \mathrm{C(ox)} \end{aligned} \qquad (3)$$

We then use a modified gold electrode (Albery et al 1990b) to reduce C(ox) to C(red), thereby measuring the rate of enzyme turnover. An inhibitor such as H_2S binds to the enzyme, rendering it inactive:

$$E + I \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} EI$$

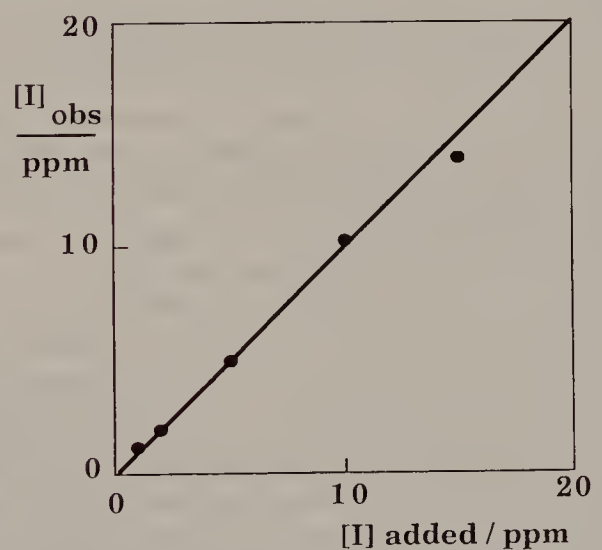


FIG. 7. A typical calibration plot for the determination of H_2S using the inhibition enzyme electrode.

Kinetic analysis shows that at any time the concentration of inhibitor is given by

$$[I] = (k_{-I}/k_I)(i_0/i - 1) + (1/k_I)[-\mathrm{dln}(i)/\mathrm{dt}]$$

where i is the current and i_0 the current before the arrival of any inhibitor. The arrival of inhibitor causes d^2i/dt^2 to go sharply negative. We use this feature to measure i_0 and to start the measurement of [I]. We have developed the necessary software to do this and within a matter of seconds a value of [I] is obtained. A calibration plot is shown in Fig. 7. It can be seen that levels of H_2S in the parts per million range can be measured. It is interesting that the more toxic the substance is, the more sensitive is the sensor. We have dubbed the device 'the chemical canary'. But, unlike the miner's canary, the feathers, the claws and the little red beak have been eliminated, and we have kept only the crucial respiratory chemistry.

Acknowledgements

I thank the following colleagues and co-workers: Dr Marianne Fillenz, Dr Martyn Boutelle, Dr Sally Durant, Dr Andrew Hopkins, Dr Edmund Magner, Bernard Mangold and Girish Rao. We are grateful to the Science and Engineering Research Council for financial support.

References

Albery WJ, Brett CMA 1983 The wall-jet ring-disc electrode. Part 1. Theory. J Electroanal Chem 148:201–210

Albery WJ, Bartlett PN, Craston DH 1985a Amperometric enzyme electrodes. Part 2. Conducting salts as electrode materials for the oxidation of glucose oxidase. J Electroanal Chem 194:223–245

Albery WJ, Haggett BGD, Jones CP, Pritchard MJ, Svanberg LR 1985b The packed bed wall-jet electrode and the determination of NO_3^-. J Electroanal Chem 188:257–263

Albery WJ, Bartlett PN, Cass AEG 1987a Amperometric enzyme electrodes. Philos Trans R Soc Lond B Biol Sci 316:107–119

Albery WJ, Bartlett PN, Bycroft M, Craston DH, Driscoll BJ 1987b Amperometric enzyme electrodes. Part 3. A conducting salt electrode for the oxidation of four different flavoenzymes. J Electroanal Chem 218:119–126

Albery WJ, Bartlett PN, Cass AEG, Sim KW 1987c Amperometric enzyme electrodes. Part 4. An enzyme electrode for ethanol. J Electroanal Chem 218:127–134

Albery WJ, Boutelle MG, Cass AEG, Durrant SLT, Fillenz M, Zetterstrom T 1988 A new assay for measuring basal acetylcholine in rat brain dialysate. Br J Pharmacol C133

Albery WJ, Boutelle MG, Durrant SLT, Fillenz M, Hopkins AR, Mangold BP 1990a Sensitive enzyme electrodes. Philos Trans R Soc Lond A Math Phys Sci 333:49–61

Albery WJ, Cass AEG, Mangold BP, Zhi Xin S 1990b Inhibited enzyme electrodes. Part 3. A sensor for low levels of H_2S and HCN. Biosensors Bioelectronics 5:397–413

Boutelle MG, Stanford C, Fillenz M, Albery WJ, Bartlett PN 1986 An amperometric enzyme electrode for monitoring brain glucose in the brain of the freely moving rat. Neurosci Lett 72:283–286

Cass AEG, Davies G, Francis GD et al 1984 Ferrocene mediated enzyme electrode for amperometric determination of glucose. Anal Chem 57:667–671

Frew JE, Harmer MA, Hill HAO, Libor SI 1986 A method of estimation of hydrogen peroxide based on mediated electron transfers of peroxidases at electrodes. J Electroanal Chem 201:1–12

Ungerstedt U 1986 Microdialysis—a new bioanalytical sensing technique. Curr Sep 7:43–46

DISCUSSION

Crumpton: In the electrodes with a separating dialysis membrane, the membrane obviously acts as a barrier. In some instances, particularly in neuronal responses, the speed of production of the analyte (e.g. acetylcholine) will be faster than its rate of diffusion across the membrane. In fact, you won't be measuring the actual rate of production in terms of response time.

Albery: There are various strategies that we can employ. The fastest-responding electrode that we have for neurophysiology is the carbon paste electrode. This measures ascorbate, because the electrode isn't selective and ascorbate is the electroactive compound present at highest concentration in the brain. We have recently had some very exciting results using the ascorbate electrode. We pinched a rat's tail for a second and saw the ascorbate concentration leap up; as soon as we removed the forceps, the ascorbate concentration fell again. The time resolution of that electrode is down to half a second and, for the first time, we have been able to follow chemical changes resulting from a behavioural stimulus. The glucose electrode doesn't have a membrane—glucose oxidase is adsorbed straight onto the conducting organic salt electrode—so it shows a reasonably fast response time. In microdialysis, the analyte has to diffuse through the dialysis membrane, so the response time will be of the order of tens of seconds.

Using the NAD 'trick', we have developed a sensor for glutamate. We have a hypothesis, which is physiologically sound but chemically untested, as yet, that when glutamate is released the ascorbate level also goes up, because of the way the glutamate re-uptake mechanism works. Our current approach is to use the glutamate enzyme electrode to validate the link between glutamate and ascorbate; then we shall be able to use the fast-reacting ascorbate electrode to look at the rapid events.

Crumpton: Over what orders of magnitude can you get straight line relationships between concentration and response?

Albery: Potentiometric sensors, such as the those described by Professor Reinhoudt, work over many orders of magnitude of concentration, but the problem with these is that a noise of 10 mV causes significant errors in the measurement of concentration. Amperometric sensors tend to show a linear relationship between current and concentration and you can tailor the membrane

to suit your needs. For example, if you wanted to measure high concentrations of glucose that would normally saturate the enzyme, you could use a thick membrane that would reduce sensitivity but increase the linearity. Any particular device probably measures reliably over two orders of magnitude. You can tailor the device to suit your concentration range, but the dynamic range will not be as good.

Brändén: Do endogenous inhibitors interfere with sensors when they are implanted in an animal?

Albery: Our results with the glucose electrode, although it did lose some sensitivity, are encouraging, because everyone said that proteases would inactivate it. I think we have avoided this by using large amounts of the enzyme. One thing that we want to do is to combine the microdialysis technique with the implanted electrode, so that the electrode can be calibrated continually; decreasing sensitivity would then not matter. You could implant a microdialysis probe and an electrode in the same place, use microdialysis with a fresh electrode in the laboratory to calibrate the implanted electrode, then use the implanted electrode to look at the rapid changes that are happening.

Esther Breslow: What degree of spatial resolution do you think will be possible, given that the brain is made up of multiple cell types?

Albery: There are two schools of thought. Dr Fillenz and I have tended to use electrodes with a diameter of, say, 1 mm; such electrodes are suitable for sampling a particular brain region. Other people have implanted carbon fibre electrodes, which can have a point sharp enough to allow you to sample single neuronal events. The problem with carbon fibre electrodes is, at least at the moment, that it is very difficult to keep them working for longer than a day or two. The approach depends on what you want to study. Our rats are operated on, then they recover; we want to measure things in normal, freely moving rats. If you use a carbon fibre electrode, the electrode stops working before the rats have recovered. It appears that it's the healing process that actually damages the carbon fibre electrode. In our rats, glial cells grow around the end of the electrode, providing a sampling compartment. This takes a couple of days. If you pull a carbon fibre electrode out after two days it appears as if that process has damaged the tip of the electrode.

Esther Breslow: That raises the question of the biochemistry of stroke and the mechanism by which the reintroduction of oxygen into the injured part of the brain is thought to cause injury. One hypothesis is that the enzyme xanthine dehydrogenase, which normally uses the cofactor NAD, switches to using FAD. It then uses oxygen as the electron acceptor, thereby producing hydrogen peroxide. Could you use this kind of electrode system to establish which cofactor was being used after injury?

Albery: That is a very interesting problem and it is the sort of thing one could test. We haven't thought about that particular problem. We have talked about getting fast responses from implanted electrodes. The advantage of the

microdialysis technique is that once you have got the dialysate, you can sample it with three or four different electrodes. You could, for example, use the horseradish peroxidase reaction to determine whether hydrogen peroxide was indeed being produced. The dialysis technique doesn't give such good time resolution, but it allows you to look at the correlation between three or four different analytes.

Ron Breslow: With these electrochemical methods calibration and reproducibility can be problems. Some of your systems are reversible, so this should not be a problem, but your sandwich immunoassay is irreversible—once the antigen is bound to the antibody it cannot be removed. Do you find calibration or reproducibility a problem?

Albery: The biggest problem with the immunoassay is not the electrochemistry, which is beautiful, but the antigen–antibody reaction. The kinetics of that process is probably the factor over which we have least control. Once you have made the sandwich, you can sample the amount of enzyme three or four times.

Ron Breslow: But do you get the same response to a particular level of antigen when you use different electrodes?

Albery: Yes, but you have to control the antigen–antibody reaction carefully by getting the temperature and incubation time right. George Wilson tells me that we don't take enough trouble in the immobilization of the antibody onto the carbon surface; the orientation of the antibody on the surface is important, so we would probably get better results if we took more trouble at that stage.

Crumpton: I don't understand why temperature should be a problem, because immunological reactions—that is, the direct interaction of antigen with antibody—classically have a negligible temperature coefficient, at least between 0 and 37 °C.

Albery: That's very interesting—why is that?

Crumpton: It's a fact of life!

Stoddart: Has anyone used electrochemical technology in conjunction with catalytic antibodies?

Albery: We had a little foray into this, but I don't know if anyone else has.

Sutherland: Could you tell us a little more about the nature of your conducting organic salt?

Albery: Dr Bartlett started my group off on this. We were working with modified electrodes and we were worried about there not being enough reaction centres; he suggested using *N*-methylphenazinium, which was traditionally used by biochemists to oxidize NADH. If you want lots of *N*-methylphenazinium, why not have it as a cation in an organic salt? We then discovered that Kulys had actually had the same idea and had shown that you can oxidize NADH on the NMP-TCNQ salt (Kulys et al 1982).

We were surprised when we found we could oxidize so many enzyme active sites. We have spent a lot of time looking at the mechanism of this oxidation.

The model that we favour at the moment to explain why these salts are so good at oxidizing the active sites is that when the enzyme sits on the conducting salt electrode, the constituents of the salt in the hydrophobic region where the enzyme is can actually shuffle around a bit, although not in the solution. There is a sort of mobile carrier by which one or other of the constituents of the salt can actually go into the active site and shuffle backwards and forwards in a sort of hydrophobic pocket. There is an interesting question here about potential distribution. An electrochemist normally thinks about the Helmholtz layer with a thickness of 3 Å; the potential drops sharply over this thin layer. When you put a protein onto a salt like ours, the potential becomes somewhat distributed all over the glob of protein, so you have made a funny local region. The model that we find fits best at the moment involves the TTF^+ coming out of the salt, shuttling into the active site and then shuttling back out again; locally, there is a high concentration of mediator.

Sutherland: There are not many candidates for the salt.

Albery: We tried a few different salts and found that TTF-TCNQ was best for most of the flavoenzymes whereas NMP-TCNQ was best for reactions involving NADH. The others we tried weren't as good, but they all showed some effect.

Brändén: I am a little puzzled by your point about the reoxidation of the coenzyme within the enzyme active site by the electrode. Most of the enzymes you discussed have non-covalently bound coenzymes, therefore one would expect the coenzyme to dissociate after the reaction and consequently the free coenzyme would be reoxidized on the electrode surface.

Ron Breslow: That's not true of all flavoenzymes.

Albery: Most flavoenzymes have some free FAD, but we did an experiment with glucose oxidase in which we dialysed out all the free FAD so there was only bound FAD, and we saw no difference. I don't think the system works by FAD shuttling.

Brändén: There could easily be a slight diffusion of the coenzyme.

Albery: I used to think that was what happened, that the enzyme was strongly adsorbed with the active site up against the surface of the electrode; but a moment's thought tells you that that isn't a very good strategy, because the glucose needs to be able to get in and out. I now think it more likely that the glucose is finding the active site on the outside of the enzyme–electrode complex, and maybe the electron relay is taking place from the active site backwards, through the enzyme.

Sutherland: You have been measuring organic substrates using enzymes; perhaps some of the organic chemists might like to tell you when you will give up enzymes in favour of organic hosts, or is that too far into the future to visualize?

Albery: That's a very interesting question. What about a host for ethanol, for example? We can measure ethanol quite easily using alcohol dehydrogenase.

Sutherland: You are giving us a particularly awkward challenge. We're fairly happy with cations, but not so happy with organic molecules.

Dunitz: Surely, alcohol dehydrogenase isn't specific for ethanol?

Albery: You are quite correct; it will also react with butanol. We are busy at the moment trying to make a sugar analyser in which there will be three or four different enzymes, all of which have some response to a number of different sugars. If you have a mixture of four sugars, then with results from four different enzymes one can solve the simultaneous equations to find the constituents of the mixture.

Stoddart: When you introduce enzymes into organic solvents, which are unnatural environments, the substrate specificities often change quite remarkably and sometimes become broader. To take glucose oxidase as an example, are other sugar substrates accepted or is the enzyme absolutely specific for glucose? Does it accept galactose or mannose, for example?

Albery: No; as I said, these enzymes aren't entirely specific.

Stoddart: I am raising the point that you are putting the enzymes in an unnatural environment. Can we extrapolate from what is known about their action under physiological conditions to their action in such unnatural environments?

Albery: In the first electrode I described, the enzyme is largely in the electrolyte layer (some will be absorbed but most of it is in the electrolyte layer). With the implanted electrode, in which we mix up the enzyme with a conducting organic salt, the enzyme is on the electrode surface. We have no evidence of markedly different selectivities between enzymes in these two electrode types, but it's not something we have investigated systematically.

Sutherland: For some purposes a loss of selectivity might be quite useful because it extends your range of substrates!

Albery: We are working on an interesting problem for a brewing company at the moment; they want to know when it is best to stop the brewing process. We are trying to see whether we can find what particular mixture of amino acids produced by the yeast corresponds to 'best bitter'. We do want to have some selectivity, but it can also be useful to be able to analyse mixtures.

Hamilton: You mentioned that activity will vary depending on the orientation of the enzyme on the electrode surface. Do you have ways of controlling the surface organization? Could site-directed mutagenesis techniques be used to specifically orient the enzyme?

Albery: That's an interesting idea.

Crumpton: I am not convinced that when the enzyme is adsorbed onto the surface of the electrode there is necessarily occlusion of the enzyme active site. It's a characteristic feature of antibodies against enzymes that they are not inhibitory when the substrate is small (for example, urease and catalase). As you increase the size of the substrate (for example, with oligonucleotides of increasing size and ribonuclease), you can actually titrate the substrate size to

a point at which the antibody inhibits the enzyme. In this case, the limiting substrate size is a measure of access to the enzyme active site in the presence of antibody.

Albery: Are you saying that if we took our electrode with the adsorbed enzyme on it, we could titrate to see how open the active site is?

Crumpton: Exactly.

Albery: That's a good idea.

Reference

Kulys JJ, Cenas NK, Svirmickas GJS, Svirmickiene VP 1982 Chronoamperometric stripping analysis following biocatalytic pre-concentration. Anal Chim Acta 138:19–26

The natural design of vancomycin family antibiotics to bind their target peptides

Jonathan P. Waltho* and Dudley H. Williams†

*Department of Molecular Biology and Biotechnology, University of Sheffield, Sheffield S10 2TN, UK and †University Chemical Laboratory, Lensfield Road, Cambridge CB2 1EW, UK

Abstract. The vancomycin family of antibiotics provide a rare opportunity among natural systems to study a molecular recognition process in which both the 'receptor' and the 'ligand' are relatively small molecules. Unlike the vast majority of antibiotics, in the vancomycin family the antibiotic performs the role of the receptor. All members of the family are covalently cross-linked heptapeptides that contain a variety of glycosidic modifications. Their site of action in bacterial cell walls is modelled by simple dipeptides and tripeptides. NMR experiments have been used to characterize the binding of these species through the study of both the complex and the free components. In unbound antibiotics conformational freedom is observed in regions of the molecule not severely restricted by covalent linkages. On binding of the ligand much of this conformational freedom is lost and the hydrophobic side chains of the antibiotics reside close to the intermolecular hydrogen-bonding interactions, thus shielding these interactions from the solvent. The charged amino groups of the N-terminus and disaccharide region of vancomycin are orientated not to optimize intermolecular electrostatic interactions but rather to retain solvation. This causes further hydrophobic faces to be presented to the ligand. Removal of saccharide units from the antibiotics leads to small losses in binding energy but may have considerable influence on the selectivity of the antibiotics. Specific dimerization through the non-ligand-binding faces of ristocetin is observed at millimolar concentrations. The geometry of the dimeric complex enables a close approach of the ligand carboxylate anion and the charged amino group of the novel sugar, ristosamine.

1991 Host–guest molecular interactions: from chemistry to biology. Wiley, Chichester (Ciba Foundation Symposium 158) p 73–91

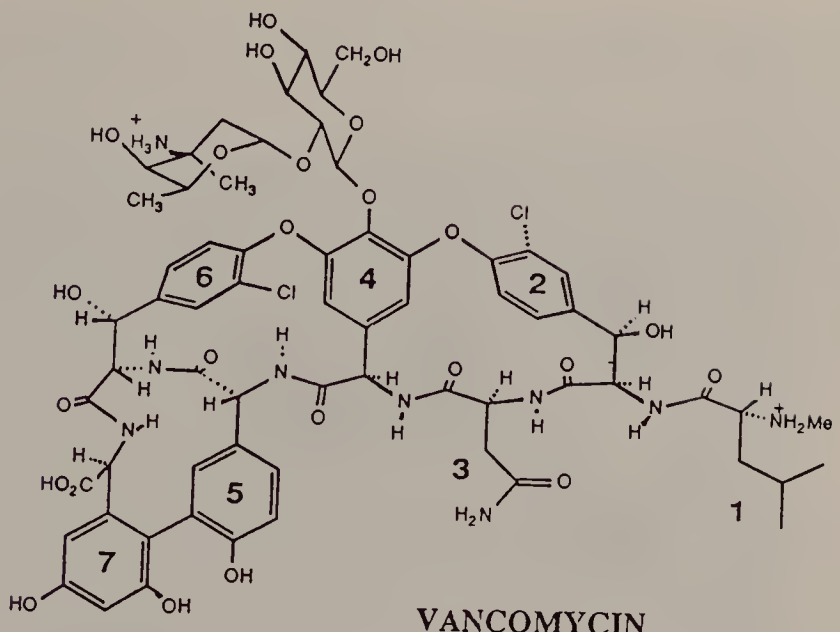
Unlike most intermolecular recognition processes in Nature, the interaction between the vancomycin family of antibiotics and models of the site of action involves two relatively small molecules in a highly selective, high binding affinity complex. Whereas other antibiotics act as the ligand, those in the vancomycin family perform the role of the receptor in the complex and the bacterial cell wall provides the ligand. The molecular masses of the complexes fall in the range 1500–3000 daltons. This not only facilitates the study of these systems, but also

imposes unusual requirements on the design of the antibiotic so that a high binding affinity for its target ligand is achieved while the appropriate selectivity is retained. In this paper we wish to discuss some aspects of the 'natural design' of the antibiotic systems which achieves this interaction. Our discussion supports the view that the many complex features of this and other natural products have evolved to serve one or several specific purposes in the 'life-cycle' of the molecule, and are not the result of non-specific variation (Williams et al 1989).

As a result of the considerable clinical and agricultural success of the vancomycin family of antibiotics, extensive efforts made by several pharmaceutical companies have led to the discovery of numerous variants of the family. For a number of years our group and others have been involved in the structural elucidation of members of this family (reviewed in Barna & Williams 1984) and in defining the molecular basis for their mode of action (see below). The structures of all known antibiotics of the vancomycin family are based on a heptapeptide backbone, the side chains of which are extensively modified through covalent cross-linkage (see Fig. 1). Further modifications are widespread, such as the methylation of amines, chlorination of aromatic rings, sulphonation of phenolic groups, and (of major concern in this paper) glycosylation of hydroxyl groups—involving both commonly occurring monosaccharides and novel amino sugars.

This family of antibiotics can be broadly subdivided into two according to whether or not residues 1 and 3 are covalently cross-linked. We shall limit our discussion to two members of the family, vancomycin and ristocetin A, the covalent structures of which are shown in Fig. 1. These molecules typify the two subfamilies. Furthermore, both species contain novel amino sugars which, as we shall discuss, appear to perform quite different functions.

Elucidation of the mode of action of these antibiotics stemmed from the observation that both vancomycin and ristocetin bind to cell wall mucopeptide precursors terminating in the dipeptide sequence -D-Ala-D-Ala (Perkins 1969). By observing changes in chemical shifts of the nuclear magnetic resonances, and their nuclear Overhauser effects (discussed in more detail in Methods, below) for ristocetin in the presence and absence of the cell wall analogue *N*-acetyl-D-Ala-D-Ala, the orientation of this analogue with respect to the antibiotic was elucidated (Kalman & Williams 1980a,b). Figure 2 illustrates the geometry of the complex. The dashed lines indicate hydrogen bond formation between the carbonyl groups of one component and the NH groups of the other component. It can be seen that the carboxyl group at the C-terminus of the peptide ligand (upper right) forms three hydrogen bonds to three NH groups which lie in a pocket at the N-terminal end of the antibiotic structure (lower right). Additional hydrogen bonds occur between the carbonyl group of residue 4 of the antibiotic and the NH group of the C-terminal alanine, and between the NH group of residue 7 and the *N*-acetyl carbonyl group of the peptide ligand.



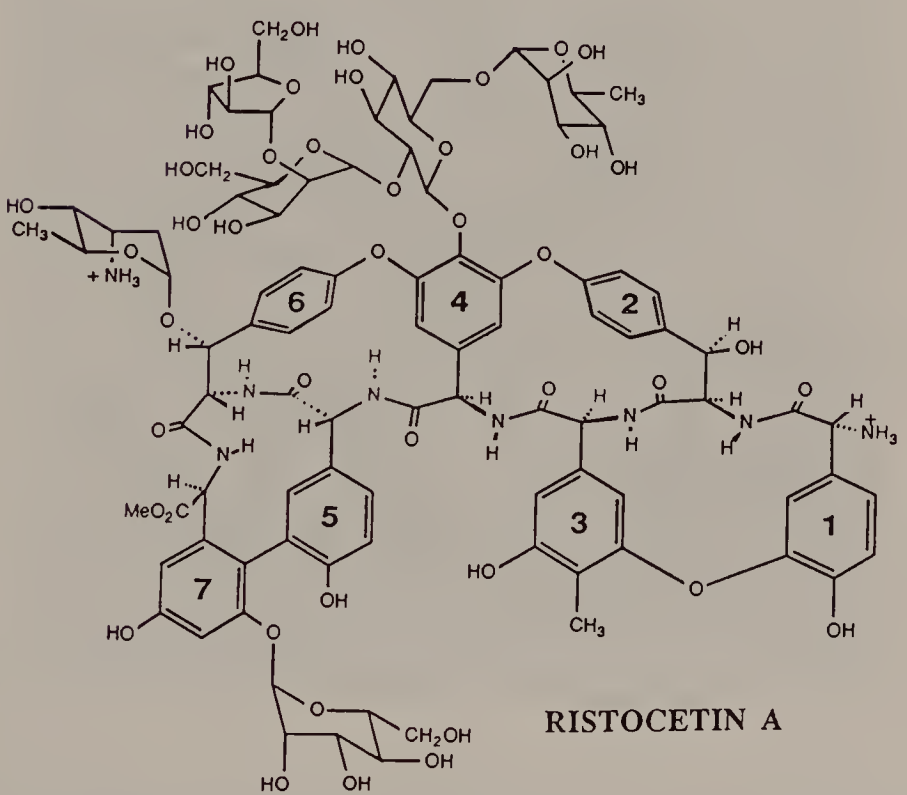


FIG. 1. The covalent structures of the antibiotics vancomycin (*upper*) and ristocetin A (*lower*) and the numbering scheme for their constituent amino residues.

FIG. 2. An exploded view (Corey–Pauling–Koltun model) of the geometry of the complex formed between ristocetin A and the cell wall model peptide di-*N*-Ac-L-Lys-D-Ala-D-Ala. The antibiotic is oriented as in Fig. 1. The intermolecular hydrogen bonds are indicated by the broken lines.

On the basis of this information, we have been able to explore some of the more subtle factors contributing to the stability and selectivity of this intermolecular recognition process. This paper is concerned with three aspects of this work: (1) the conformations adopted by the relatively flexible regions of the antibiotic binding cleft and the consequent hydrophobic shielding of intermolecular interactions; (2) the role of some of the natural glycosidic modifications of antibiotics of the vancomycin family in increasing binding affinity and in determining specificity; (3) the dimerization of ristocetin and the involvement of the protonated amino group of ristosamine.

Methods

The following discussions rely strongly on the observation of a nuclear magnetic resonance (NMR) property, called the nuclear Overhauser effect (NOE), of the hydrogen atoms within the molecules under study. This property provides a means with which to determine the spatial proximity of hydrogen nuclei through observation of their mutual relaxation towards equilibrium from an excited state created within the NMR experiment. The details of the NOE and experimental procedure for its measurement are fully discussed in Neuhaus & Williamson (1989) and Wüthrich (1986), and the application of these techniques to vancomycin family systems is discussed in detail in Waltho et al (1987) and Kannan et al (1988). Studies on unbound vancomycin were performed in water, water/acetonitrile and dimethyl sulphoxide/water mixtures (Waltho et al 1988a). Studies on vancomycin bound to the peptide ligand di-*N*-acetyl-L-Lys-D-Ala-D-Ala were performed in dimethyl sulphoxide/chloroform/water mixtures (Kannan et al 1988). Studies on both unbound and bound ristocetin were performed in water, water/acetonitrile and dimethyl sulphoxide/water mixtures (Williamson & Williams 1985, Waltho & Williams 1989).

Conformations of the binding cleft

The extensive cross-linkage of side chains to form a series of small cyclophane rings within the peptide region of the antibiotics imposes severe conformational restriction on the various parts of the molecule. Free rotation of the cross-linked aromatic rings is prohibited at room temperature and fluctuation of the backbone is extremely limited. In addition to making a significant contribution in reducing the entropy loss on complexation, the rigidity of a majority of the antibiotic provides a framework against which to measure the properties of those parts of the molecule with greater conformational mobility. The most conformationally free region of the peptide portion of the antibiotics is the side chains of residues 1 and 3.

FIG. 3. A CPK model of the geometry of the complex formed between vancomycin and di-*N*-Ac-L-Lys-D-Ala-D-Ala (hatched protons). The orientation of the antibiotic is as shown in Fig. 1. Indicated is the binding pocket formed by the ring 6 chlorine atom (A), the vancosamine 6-methyl group (B), and the aromatic ring of residue 2 (C) for the C-terminal alanine residue of the ligand (D). Also indicated for the ligand are the carboxylate anion (E) that binds to the amide protons of residues 2, 3 and 4 of the antibiotic (as shown in Fig. 2) and the N-terminal alanine methyl group (F). The positional relationship between the methyl group F, the C-terminal alanine α-proton (G) and the isopropyl terminus of residue 1 of vancomycin (H) defines the hydrophobic binding pocket of the ligand carboxylate anion. This pocket is extended by the methyl group of the N-terminus (J). The charged amino groups of the N-terminus and vancosamine (K) are exposed for maximum solvation.

Vancomycin

In vancomycin, residues 1 and 3 are *N*-methyl-D-leucine and L-asparagine respectively (see Fig. 1). Thus, there are no cross-linkages, and considerable flexibility of the side chains is observed in the unbound antibiotic. The intra-side chain NOEs of the *N*-methylleucine residue graduate from negative to positive, moving to the side chain termini from the α-proton (Waltho et al 1988a), which indicates the considerably isotropic nature of the internal

reorientations of this side chain. Furthermore, in the asparagine residue, the intra-side chain NOEs show a similar trend, with the primary amide to β-proton NOEs at the zero cross-over point. On addition of the peptide ligand, gradation of the NOE build-up is unobservable and these residues take on the overall tumbling properties of the complex. NOEs indicate that in the predominantly populated (most prevalent) conformations, the terminal methyl groups of the *N*-methylleucine residue are close to the α-proton of the C-terminal alanine residue of the peptide ligand and the methyl group of the N-terminal alanine residue (see Fig. 3) (Waltho et al 1988b). Thus, the electrostatic interactions between the carboxylate of the peptide ligand and the amide-rich carboxylate binding pocket of the antibiotic receptor are shielded from solvation by the hydrophobic side chain of residue 1.

Ristocetin

In ristocetin, residues 1 and 3 are D- and L-hydroxyphenylglycine derivatives respectively, and are joined by a phenoxy coupling (see Fig. 1). The cross-linkage imparts more conformational restriction to this region than that found in the analogous region of vancomycin, but it remains the most conformationally flexible cyclophane ring in the molecule. In unbound ristocetin, the NOEs between the α-proton of residue 1 and the *ortho* protons of ring 1 indicate either an 'open face' conformation for ring 1 or conformational exchange involving both 'open' and 'closed face' conformations. (The 'closed face' conformation is as shown in Fig. 2. In an 'open face' conformation, the bond connecting the α-carbon and the phenyl ring of residue 1 is rotated to further expose the upper face of ring 1—as shown in Fig. 2—to the solvent.) The NOEs observed in the complex indicate overwhelming population of a 'closed face' conformation. The protons of ring 1 are close in space to both the α-proton of the C-terminal alanine, and the methyl group of the N-terminal alanine residue of the peptide ligand. In water, water/acetonitrile and dimethyl sulphoxide/water mixtures, the conformation of residue 3 is not measurably changed on binding. In methanol however, some conformational flexibility is observed in the unbound state (Fesik et al 1986).

Hydrophobic shielding of intermolecular interactions

The results of the conformational changes occurring on binding of the peptide ligand to ristocetin are closely analogous to those observed for vancomycin. In both systems, the intermolecular electrostatic interactions of the carboxylate binding pocket are shielded from solvation by the side chain of residue 1, and to a lesser extent by that of residue 3. Thus, these interactions occur in a relatively low dielectric environment, away from the depolarizing influences of the solvent. This is closely analogous to the increased association constants of ionic species in solvents with decreasing polarity.

The role of the antibiotic sugars

From the early work (Kalman & Williams 1980a,b, Williams et al 1983) it is evident that the major basis for the interaction between antibiotic and cell wall analogue lies in the aglycone portion of the antibiotics. However, the antibiotics are known to carry one to six sugars (see Fig. 1). While it always appeared likely that these sugars would promote antibiotic activity *in vivo* by promoting aqueous solubility, it is now becoming evident that at least some of them aid binding in subtle and interesting ways.

Vancosamine in vancomycin

Vancosamine is a novel amino sugar unique to a few members of the vancomycin family of antibiotics. In vancomycin it is bonded via a glucose molecule to residue 4 of the antibiotic (see Fig. 1). When vancosamine is selectively removed, the constant of binding to the peptide ligand drops by a factor of approximately three (see Table 1 and Kannan et al 1988). An identical reduction in binding constant is observed on *N*-acetylation of vancosamine. For vancomycin in the unbound state, rotation of the disaccharide unit is observed through NOEs between the methyl groups of vancosamine and both the front and back faces of rings 2 and 6. Similar NOEs link the anomeric proton of glucose to both faces of the antibiotic. On complexation with the peptide ligand, evidence for such a rotation is lost, and NOEs show that the predominantly populated conformation has the 6-position methyl group of vancosamine in close proximity to an aromatic proton of ring 2 and the methyl group of the C-terminal alanine residue of the peptide ligand. Such a conformation is shown in Fig. 3. Note that the charged amino group of vancosamine and the carboxylate anion of the peptide ligand are far removed from their position of closest approach. On *N*-acetylation of vancosamine, the conformational ensemble with rotation of the disaccharide returns. Hence, the interaction between the disaccharide and the ligand–aglycone complex is subtly modified by the presence of the charged amino group.

The position of vancosamine in the antibiotic–peptide complex leads us to conclude that it is the hydrophobic extension of the binding pocket for the methyl

TABLE 1 Binding to di-*N*-Ac-L-Lys-D-Ala-D-Ala of vancomycin and derivatives characterized by the removal and *N*-acetylation of its amino sugar vancosamine

Antibiotic	*Binding constant* ($\times 10^{-3}$)
Vancomycin	1500
Desvancosaminylvancomycin	490
N'-Acetylvancomycin	440

Adapted from Kannan et al 1988.

group of the C-terminal alanine residue of the ligand by the vancosamine 6-methyl group (rather than any direct electrostatic interaction between the amino group of vancosamine and the peptide ligand carboxylate anion) that is the predominant interaction increasing the stability of the complex. The loss of any significant contribution to stability on *N*-acetylation of vancosamine is consistent with the charge group on the sugar effecting an intramolecular 'salting out' of the hydrophobic interactions binding the C-terminal alanine methyl group.

Mannose on ring 7 of ristocetin

One remarkable similarity between the mannose unit bound to ring 7 of ristocetin (see Fig. 1) and vancosamine on vancomycin is the NOEs observable between the sugar protons and the methyl groups of the peptide ligands. In the case of mannose on ristocetin, NOEs are observed between the sugar and the N-terminal alanine residue of the ligand. No loss in binding energy is observed on removal of the ring 7 mannose (along with the ring 4 tetrasaccharide) with the tripeptide ligand, although these sugars appear to make a large contribution to the stability of the dipeptide complex (as discussed in Williams & Waltho 1988).

Selectivity as a role for vancosamine and mannose

Although the above results indicate that the sugars make only small or insignificant contributions to binding affinity between the antibiotics and the tripeptide ligand, it appears that their presence has considerable influence on the selectivity of the binding site, and thus on the effectiveness of these agents against specific bacterial types. Before any structural details were known for the antibiotics it was shown that vancomycin strongly disfavours a large side chain in the C-terminal residue of the ligand (Nieto & Perkins 1971). Table 2 contains examples of this specificity. Although such specificity had been observed, it is only now that the roles of the sugar substituents in the natural design of this specificity are coming to light. It appears that it is the energy

TABLE 2 Binding constants ($\times 10^{-3}$) of vancomycin and ristocetin to peptides that model antibiotic binding sites within mature cell walls of some Gram-positive bacteria

Peptide	*Vancomycin*	*Ristocetin*
Di-*N*-Ac-L-Lys-D-Ala-D-Ala	1500	590
Di-*N*-Ac-L-Lys-D-Ala-D-Lys	14	100
Di-*N*-Ac-L-Lys-D-Ala-D-Leu	9	610
N-Ac-L-Ala-D-Glu-Gly	480	0.7

Adapted from Nieto & Perkins 1971.

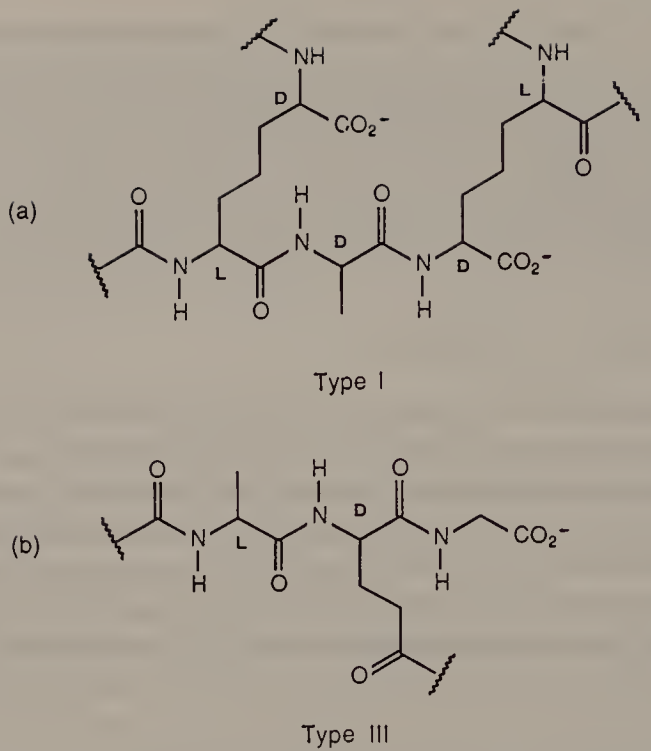


FIG. 4. Peptide portions of the mature cell walls of (a) type I (e.g. bacilli) and (b) type III (e.g. micrococci) Gram-positive bacteria, that have the correct motif for binding antibiotics of the vancomycin family.

penalty involved in displacing the sugars to accommodate a side chain larger than that of alanine that is the most significant feature of the sugars in binding (Williams & Waltho 1988).

The requirement for such high selectivity may well be the result of the nature of the mature cell walls surrounding the target bacteria. Bacterial cell walls are biosynthesized from the inside; thus to inhibit biosynthesis it is necessary for the antibiotic to pass through the mature wall. The mature walls of Gram-positive bacteria of type I (e.g. bacilli) contain within *every* peptidoglycan cross-linkage the amino acid sequence shown in Fig. 4a. Any antibiotic lacking specificity for alanine as the C-terminal residue of the target ligand (for example, ristocetin) will bind equally as strongly to the massive excess of 'decoy' sites in the mature wall as to the intended target. Thus, the antibiotic action is effectively buffered by the mature wall. In contrast, Gram-positive bacteria with type III cell walls (e.g. micrococci) contain within their mature walls the sequence shown in Fig. 4b. Inspection of the binding constants for a model of this peptide, *N*-acetyl-L-Ala-D-Glu-Gly (see Table 2) reveals that vancomycin, which lacks selectivity for large side chains at the D-Glu position of this peptide, binds to the model peptide with only a small loss in affinity relative to di-*N*-acetyl-L-Lys-D-Ala-D-Ala, whereas the affinity of the type III cell wall model for ristocetin is over three orders of magnitude lower.

Experimentally (Sinha & Neuhaus 1968), a 50% inhibition of peptidoglycan synthesis in *Micrococcus lysodeikticus* occurred with 7.5 μg/ml of ristocetin but required 20 μg/ml of vancomycin. More significantly, in a *M. lysodeikticus* cell-free system that synthesizes peptidoglycan, the addition of mature cell walls from the same bacterium completely reversed the inhibition of the synthesis caused by vancomycin. Under the same conditions, only a 30% reversal of the inhibition caused by ristocetin was observed.

Dimerization of ristocetin

At millimolar concentrations, the binary complex between ristocetin and di-*N*-acetyl-L-Lys-D-Ala-D-Ala begins to undergo self-association. The interconversion between associated and non-associated forms is sufficiently slow for separate NMR resonances to be observable for the various species. At 1.5 mM, the ratio of the three observable antibiotic–ligand complexes is approximately 4:1:1. This changes to 1:5:5 by 15 mM, indicating an association constant of *ca* 2×10^3 M^{-1} (Waltho & Williams 1989). The two sets of resonances for the associated forms may be satisfied by either an asymmetrical dimer, or the presence of two distinct dimeric species. The geometry of the dimerization process is evident from the intermolecular NOEs and chemical shift changes occurring in dimerization. It appears that the novel amino sugar of ristocetin, ristosamine, that is attached to residue 6 of the antibiotic (see Fig. 1) and resides mostly on the non-bonding face, is intimately involved in the dimerization process. NOEs link the 2 and 3 positions of ristosamine with aromatic protons of ring 2 and the α-proton of residue 3, and resonance 6f with 4b (for nomenclature of protons, see Fig. 5). Large chemical shift changes ($\Delta\delta > 0.4$ p.p.m.) are observed for the 'back face' of ring 6 (i.e. the back face according to the orientation depicted in Fig. 2), the α-proton of residue 4 and aromatic proton 4f (see Fig. 5).

From an inspection of molecular models it is proposed that the arrangement of intermolecular hydrogen bonds shown in Fig. 5 would satisfy all the NOE and chemical shift perturbation data. In such a dimeric form, the stacking of the aromatic rings, in particular those of residues 4 and 6, is responsible for the unusual chemical shifts of the NMR resonances. It is noteworthy that the dimerization process brings the charged amino group of ristosamine to a position where it may influence electrostatically the functional groups of the carboxylate binding pocket. This may well be linked to the unusual axial rather than equatorial configuration of the amino group on ristosamine and implicates ristosamine in a role quite unlike those discussed for vancosamine and the ring 7 mannose above.

The difference between the two forms of dimer is as yet unsolved. The chemical shift differences are reflected most prominently in the tetrasaccharide portion of the molecule; the largest are observed in the glucose and rhamnose

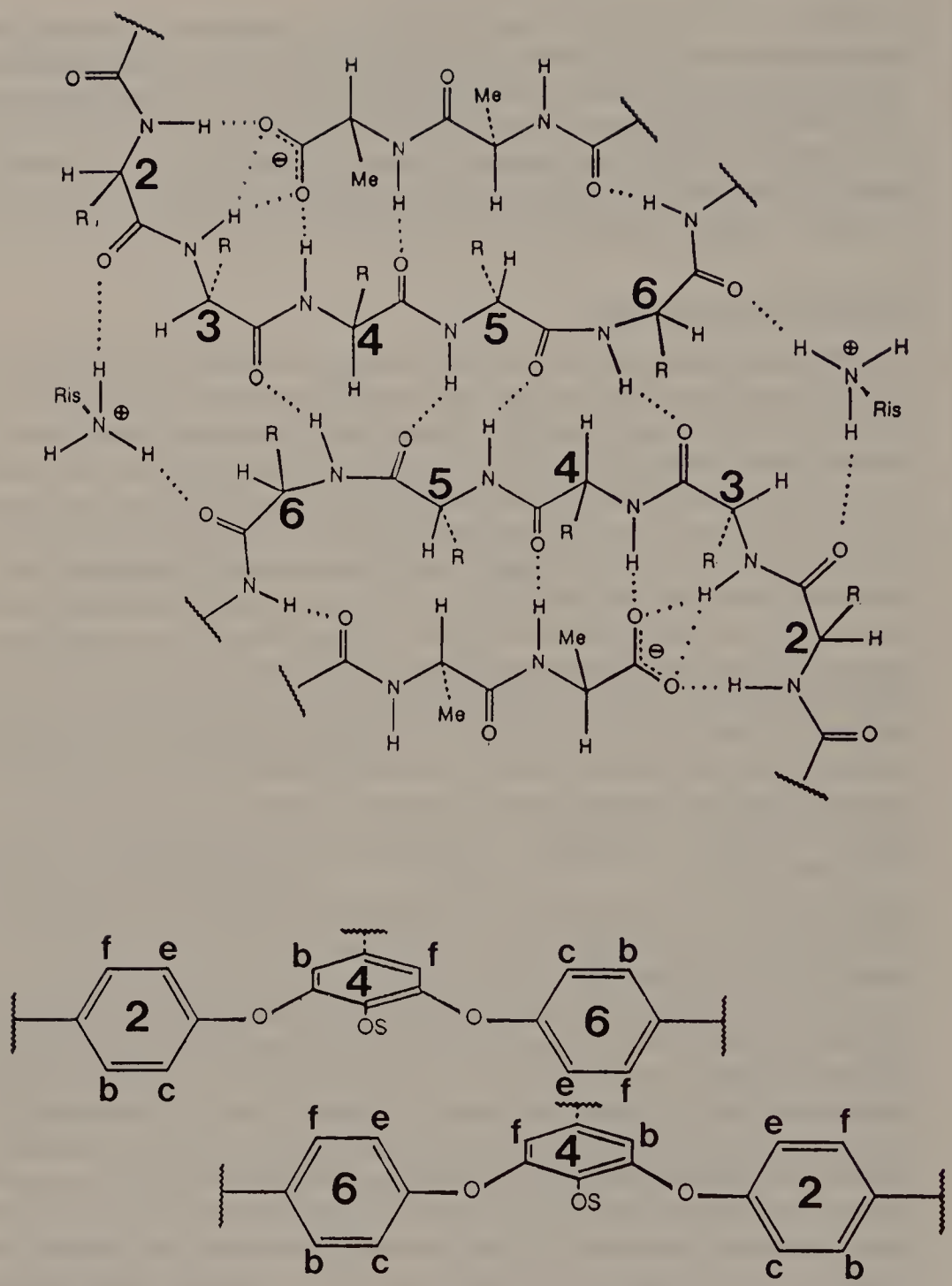


FIG. 5. The orientation of the dimer formed by ristocetin A binding to the cell wall peptide model di-*N*-Ac-L-Lys-D-Ala-D-Ala, showing (*upper*) the proposed arrangement of intermolecular hydrogen bonds (including the charged amino group of ristosamine) and (*lower*) the corresponding arrangement of the aromatic rings of residues 2, 4 and 6 (with proton nomenclature).

portions of the tetrasaccharide, and the resonances surrounding the site of attachment to ring 4 (i.e. 2c, 2e, and 6c all have $\Delta\delta > 0.3$ p.p.m.). NOEs from the anomeric proton of the glucose of the tetrasaccharide suggest a difference in rotamers about the ring 4–tetrasaccharide linkage.

At present one can merely speculate whether this very specific dimerization process has a role in antibiotic activity *in vivo*. However, recent experiments (J. P. Scott, R. R. Montgomery & G. S. Retzinger, unpublished work 1990) have shown that two very significant medical properties of ristocetin, namely its ability to flocculate various proteins and to agglutinate platelets in the presence of von Willebrand factor, are stimulated only by the dimeric form of the molecule. It is proposed that its mode of action under these circumstances involves the bifunctionality of the ristocetin dimer cross-linking multiple copies of a recognition site in these proteins.

Conclusion

Against the background of a severely conformationally restricted antibiotic it is possible to observe the changes that occur on binding of the peptide ligand in the few less restricted regions. In both vancomycin and ristocetin the side chains of residues 1 and 3 populate conformations in the presence of the peptide ligand that create a hydrophobic environment around the carboxylate binding pocket, and prevent solvation of the intermolecular electrostatic interactions. Vancosamine on vancomycin also serves to increase the size of the hydrophobic binding pocket—in this case in the region where the methyl group of the C-terminal alanine residue of the peptide ligand is bound. This interaction is significant to the overall binding constant only when it is 'salted out' by the charged amino group on the sugar. However, its main role does not appear to be its contribution to the binding affinity, but rather the size selectivity it imparts on the binding pocket. A similar role is proposed for the mannose attached to ring 7 of ristocetin. Such selectivity may have a role in reducing the affinity of the antibiotics for unproductive sites in the mature cell walls of the target bacteria. Ristosamine, attached to residue 6 of ristocetin, is strongly involved in the dimerization of ristocetin. Its axially substituted charged amino group is positioned close to the amide groups involved in the binding of the peptide ligand on the dimer partner. Dimerization appears to be important in the protein-precipitating properties of ristocetin (J. P. Scott, R. R. Montgomery & G. S. Retzinger, unpublished work 1990).

Acknowledgements

The authors are deeply indebted to the considerable number of co-workers that have contributed to our understanding of antibiotics of the vancomycin family and their mode of action. In particular we would like to thank Dr Rajamoorthi Kannan and Dr Nicholas

Skelton for their contribution to the present work. The authors are also grateful to Dr Retzinger for making available a manuscript prior to publication.

References

Barna JCJ, Williams DH 1984 The structure and mode of action of glycopeptide antibiotics of the vancomycin group. Annu Rev Microbiol 38:339–357

Fesik SW, O'Donnell RT, Gampe RT, Olejniczak ET 1986 Determining the structure of a glycopeptide–Ac_2-Lys-D-Ala-D-Ala complex using NMR parameters and molecular modeling. J Am Chem Soc 108:3165–3170

Kalman JR, Williams DH 1980a An NMR study of the structure of the antibiotic ristocetin A. The negative nuclear Overhauser effect in structure elucidation. J Am Chem Soc 102:897–905

Kalman JR, Williams DH 1980b An NMR study of the interaction between the antibiotic ristocetin A and a cell wall peptide analogue. Negative nuclear Overhauser effects in the investigation of drug binding sites. J Am Chem Soc 102:906–912

Kannan R, Harris CM, Harris TM, Waltho JP, Skelton NJ, Williams DH 1988 Function of the amino sugar and N-terminal amino acid of the antibiotic vancomycin in its complexation with cell wall peptides. J Am Chem Soc 110:2946–2953

Neuhaus D, Williamson MP 1989 The nuclear Overhauser effect in structural and conformational analysis. VCH Publishers, New York

Nieto M, Perkins HR 1971 The specificity of combination between ristocetins and peptides related to bacterial cell wall mucopeptide precursors. Biochem J 124:845–852

Perkins HR 1969 Specificity of combination between mucopeptide precursors and vancomycin or ristocetin. Biochem J 111:195–205

Sinha RK, Neuhaus FC 1968 Reversal of the vancomycin inhibition of peptidoglycan synthesis by cell walls. J Bacteriol 96:374–382

Waltho JP, Williams DH 1989 Aspects of molecular recognition: solvent exclusion and dimerization of the antibiotic ristocetin when bound to a model bacterial cell-wall precursor. J Am Chem Soc 111:2475–2480

Waltho JP, Williams DH, Selva E, Ferrari P 1987 Structure elucidation of the glycopeptide antibiotic complex A40926. J Chem Soc Perkin Trans I 2103–2107

Waltho JP, Williams DH, Stone DJM, Skelton NJ 1988a Intramolecular determinants of conformation and mobility within the antibiotic vancomycin. J Am Chem Soc 110:5638–5643

Waltho JP, Cavanag J, Williams DH 1988b Aspects of molecular recognition: use of a truncated drived pseudo-NOESY experiment to elucidate the environment of intermolecular electrostatic interactions in vancomycin. J Chem Soc Chem Commun 707–709

Williams DH, Waltho JP 1988 Molecular basis of the activity of antibiotics of the vancomycin group. Biochem Pharmacol 37:133–141

Williams DH, Williamson MP, Butcher DW, Hammond SJ 1983 Detailed binding sites of the antibiotics vancomycin and ristocetin A: determination of intermolecular distances in antibiotic/substrate complexes by use of the time-dependent NOE. J Am Chem Soc 105:1332–1339

Williams DH, Stone MJ, Hauck PR, Rahman SK 1989 Why are secondary metabolites (natural products) biosynthesised? J Nat Prod (Lloydia) 52:1189–1208

Williamson MP, Williams DH 1985 ^{1}H NMR studies of the structure of ristocetin A and of its complexes with bacterial cell wall analogues in aqueous solution. J Chem Soc Perkin Trans I 949–956

Wüthrich K 1986 NMR of proteins and nucleic acids. Wiley, New York

DISCUSSION

Waring: If there's a significant electrostatic contribution to the binding energy between *N*-acetyl-D-Ala-D-Ala and vancomycin, presumably there should be an ionic strength dependence of the binding, which could probably be quite easily measured *in vitro*, and maybe even an ionic strength dependence of the antibacterial activity.

Waltho: As far as I know, the ionic strength dependence of binding or antibacterial activity hasn't been measured. There is a strong electrostatic component to the interaction; *N*-acetylation of the N-terminus of vancomycin, for example, causes a 20-fold drop in binding energy (Kannan et al 1988). A similar situation exists for ristocetin (Herrin et al 1985). Of course, hydrophobic interactions will also be affected by ionic strength changes.

Rebek: If I interpret Fig. 2 correctly, in ristocetin the primary amino group is on the exo-face of the molecule, and can't be in contact with the carboxylate.

Waltho: That's correct.

Rebek: There is some distance between the amine and the carboxylate.

Waltho: Thomas and co-workers (Herrin et al 1985) have prepared a derivative of ristocetin in which the stereochemistry of residue 1 is reversed, which positions the protonated N-terminal amine adjacent to the carboxylate anion of the peptide ligand. They found there was a reduction in binding energy. It appears that the design of ristocetin is such that its N-terminal amino group is kept fully solvated as opposed to being used in a direct salt bridge. Of course, an intermolecular electrostatic interaction is retained, as discussed above.

Roberts: It is interesting that the sugar of vancomycin spins around on the top of the molecule. You postulated that the amino group on the sugar contributes to the binding in some way. Does the sugar spin around in the complex? When you make the complex, does the interaction with the amino group, whatever it is, slow the spinning to hold the sugar in one place?

Waltho: We see evidence for population of a single conformation only in the complex in which the methyl group of the 6-position of vancosamine is adjacent to the C-terminal alanine methyl group of the peptide ligand (see Fig. 3). Therefore, the protonated amine is on the non-binding face of vancomycin. We don't see population of the other rotamer that has the protonated amine on the binding face. If we *N*-acetylate the sugar selectively, we see a return of the rotation in the complex.

Roberts: One of the reasons that the interaction with the amino group of the sugar contributes relatively little to the overall binding energy may be the decrease in entropy resulting from immobilization of the sugar on binding.

Lemieux: Is the rotation of the sugar around the aglyconic bond only? That is, does it involve changes in only the ψ-angle?

Waltho: We can isolate it only to the linkage between the glucose and ring 4. We have no evidence for the population of multiple conformations about the vancosamine–glucose glycosidic linkage.

Lemieux: Have you tried molecular modelling calculations to see what the difference in energy may be between the various allowed conformers?

Waltho: We did a simulation in which we forced a rotation around the ϕ-angle of the glucose–aglycone glycosidic linkage (Waltho et al 1988). We had problems with the electrostatic terms, but if we eliminated, or at least suppressed, the electrostatic terms we came up with a barrier to rotation that would have given a rotational frequency of the order of 10^9–10^{10} Hz; in other words, there's very little steric restriction to rotation.

Esther Breslow: Do you have any evidence for a difference in dimerization of ristocetin between the bound and unbound states? Do both members of the dimer bind peptide?

Waltho: Unfortunately, the u.v. difference binding constants were all measured at concentrations at which the monomeric state predominates. We can estimate the dimerization constants for both free and bound ristocetin from their NMR spectra. It appears that the dimerization is stronger in the presence of the ligand peptide, but it is difficult to quantify accurately.

Esther Breslow: Then the question would arise as to what the mechanism of that is: why does the dimer bind more strongly? Is it due to a conformational change in the antibiotic?

Waltho: If the constant of binding to the peptide ligand were stronger for ristocetin dimers than monomers, that might be a reflection of the stabilization imparted by the striking complementarity of hydrogen-bonding groups of dimerization and those involved in peptide ligand binding. In particular, the close proximity of the amino group of ristosamine to the amide groups binding the peptide carboxylate anion may well account for the increased stability (see Fig. 5).

Esther Breslow: Are there other potential mechanisms? For example, might the peptide bind across the monomer–monomer interface? This is all hypothetical, but do you know if both peptides are bound in the dimeric case?

Waltho: Yes. We can detect binding of a peptide to both ristocetin molecules in the dimer, because the dimer is asymmetrical. We are unsure of the cause of the asymmetry in the dimer. NMR chemical shift changes suggest that the region showing the greatest asymmetry is the tetrasaccharide unit. We don't know whether there is some sort of specific, asymmetrical interaction between the sugars, or whether there is asymmetry in the aglycone.

Vinter: Modelling studies suggest that there is a mutual inversion movement about the two chlorinated rings centred on the connecting ring, such that the right side is forward when the left side is back and vice versa. These two complementary shapes can come together, back-to-back, to form a reasonable dimeric structure that is consistent with the experimental evidence so far; this

idea is still speculative and at the modelling stage. There seems to be a small, subtle shift in conformation of the ring region, which is not as rigid as one would imagine.

Waltho: There is evidence from NMR spectroscopy for that kind of motion. If you try to quantify distances between the two protons of ring 4 and the α-proton of residue 4 using NOE build-ups, you find that these distances are both too short to satisfy the covalent structure; that is, there is some motional averaging process around the Cα—Cβ bond of residue 4 that brings the ring protons closer to the α-proton.

Hamilton: I am interested in the origin of the stereoselectivity in vancomycin binding for the C-terminal residue of the dipeptide substrate. Your models with the C-terminal D-alanine suggested that the α-proton was directed away from the binding cavity, so it didn't appear that there would be a steric problem involved in binding the opposite enantiomer in that C-terminal position; all one would lose is a rather weak interaction between the methyl group and the hydrophobic pocket provided by the diphenyl ether.

Waltho: I agree that losing the interaction between residue 4 and the methyl group is not crippling, because binding is relatively little affected if the C-terminal residue of the peptide is glycine.

Hamilton: I believe there is some binding to L-amino acids. Popieniek & Pratt (1987) have some recent results using a fluorescent probe that suggest binding occurs but that it's weak.

Waltho: In the original work, Nieto & Perkins (1971) found that the binding of vancomycin to peptides containing L-amino acids was below the threshold that they could measure. That indicates a difference between D- and L-amino acids of least four orders of magnitude. However, if you make the molecular model you can twist it in a reasonable fashion to accommodate L-amino acids, so the origin of this specificity remains a puzzle. The most likely candidate for the part of the molecule that imposes the stereochemical restriction over the C-terminal residue of the peptide is the side chain of residue 1, the methylleucine. Dudley Williams' group are attempting to change the N-terminal residue to investigate this. So far, they have succeeded in using the Edman degradation technique to remove the N-terminal residue (Booth et al 1987).

Ron Breslow: About eight years ago we considered the question of whether this strong binding of *N*-acetylated dipeptides to vancomycin might be used to drive the synthesis of *N*-acetylated dipeptides; the monomers don't bind very well but dipeptides bind extremely strongly. In 1982, A. W. Schwabacher in our laboratory looked at a lot of versions of what might come together to make the dipeptide, such as a methyl ester with the free amino acid. We saw no evidence for any promotion of dipeptide synthesis by vancomycin when we were trying to make the dipeptide that should bind; there are a number of possible explanations for that, one of which is that we perhaps didn't try hard enough!

Rebek: That anticipates much of what the catalytic antibody researchers want to do, to use high binding constants to drive peptide synthesis.

Ron Breslow: The question is, does the transition state look enough like the product to enable you to get the driving force? In some cases, such as in first-order reactions, the answer is yes, but we were unable to make binding to vancomycin force synthesis.

Sutherland: Stabilization of charges that are developed in the transition state would be important; that appears to be absent from a vancomycin-bound transition state.

Ron Breslow: Yes and no. There is a hydrogen bond between the NH of the peptide you are trying to make and vancomycin, and an ammonium ion that might well make a hydrogen bond. The only other thing to worry about is the developing O^- of the carbonyl group, and we have no control over that.

Brändén: It's interesting to note that the mode of binding of the peptide to the antibiotic resembles the mode of binding of serine proteases to their peptide substrates; there is one extended chain on the antibiotic and another on the peptide. These extended chains are aligned in an antiparallel fashion with hydrogen bonds between them. There are also specificity pockets for the side chains, exactly as is found in the serine proteases. Of course, there are not the sort of catalytic groups, such as the oxyanion hole, that would participate in the stabilization of the transition state for peptide formation.

Rebek: One of the really charming things about this class of antibiotics is that they are, as far as I know, the only ones that have a concave binding surface. The other antibiotics have convex functionality and they are directed towards the folds of biological macromolecules. This one is unique in that it is directed towards a small piece or detail of the target molecule.

Hamilton: But the strategy is also completely different; vancomycin antibiotics attack the substrate rather than attacking the catalyst like the penicillins do. That is a rather inefficient way to design an antibiotic, but obviously it has its attractions; for example, it's more difficult to develop resistance towards a molecule of this type.

Waring: I promise to show Julius Rebek another antibiotic with a splendid concave binding surface!

No one really knows why microorganisms make these fiendishly complicated organic molecules, but this one presumably went to some lengths to activate chlorines and put them into this amazingly complex structure. Have you any idea why? Do you know what effect those chlorine atoms have?

Waltho: The only relevant experimental work that I know of was carried out in Tom Harris's laboratory at the Vanderbilt University (Harris et al 1985). They replaced the chlorine atoms of rings 2 and 6 with hydrogens and observed small losses in binding energy, which they attributed to the increased entropy of the unbound state of the hydrogenated form.

Ron Breslow: Chlorine might protect against oxidative degradation, as it should deactivate the ring.

Kennard: I should like to point out that the molecular structure of vancomycin that forms the basis of much of this work was determined from the X-ray diffraction analysis of CDP-I, a degradation product obtained from vancomycin with loss of ammonia (Sheldrick et al 1978). It was a very difficult structure to solve and inspired the development of many new direct methods for the X-ray analysis of compounds of similar complexity. There is no other crystal structure in this area, as far as I know.

Sutherland: To finish, I should like to comment that if we had to design a synthetic molecule to bind peptides that had to function in aqueous solution and we were allowed only a limited number of binding sites, I think some of us would feel pretty pessimistic about the outcome. This work has demonstrated that we could try; there's something there that we can learn from.

References

Booth PM, Stone DJM, Williams DH 1987 The Edman degradation of vancomycin: preparation of vancomycin hexapeptide. J Chem Soc Chem Commun 1694–1695

Harris CM, Kannan R, Kopecka H, Harris TM 1985 The role of the chlorine substitutents in the antibiotic vancomycin: preparation and characterization of mono- and didechlorovancomycin. J Am Chem Soc 107:6652–6658

Herrin TR, Thomas AM, Perun TJ, Mao JC, Fesik SW 1985 Preparation of biologically active ristocetin derivatives: replacements of the 1′-amino group. J Med Chem 28:1371–1375

Kannan R, Harris CM, Harris TM, Waltho JP, Skelton NJ, Williams DH 1988 Function of the amino sugar and N-terminal amino acid of the antibiotic vancomycin in its complexation with cell wall peptides. J Am Chem Soc 110:2946–2953

Nieto M, Perkins HR 1971 Modifications of the acyl-D-alanyl-D-alanine terminus affecting complex-formation with vancomycin. Biochem J 123:789–803

Popieniek PH, Pratt RF 1987 A fluorescent ligand for binding studies with glycopeptide antibitoics of the vancomycin class. Anal Biochem 165:108–113

Sheldrick GM, Jones PG, Kennard O, Williams DH, Smith GA 1978 Structure of vancomycin and its complex with acetyl-D-alanyl-D-alanine. Nature (Lond) 271:223–225

Waltho JP, Vinter JG, Davis A, Williams DH 1988 Forces in molecular recognition: comparison of experimental data and molecular mechanics calculations. J Comput Aided Mol Des 2:31–41

General discussion II

Crystals as supramolecules

Dunitz: Crystalline solids are not specifically mentioned in the titles of any of the lectures delivered at this symposium, nor, apart from protein crystals, in the introductory words of our Chairman. The only allusion to them has been as devices for yielding structural information. I would like to remind you all that a crystal is, in a sense, the supramolecule *par excellence*—a lump of matter, of macroscopic dimensions, millions of molecules long, held together in a periodic arrangement by just the same kind of non-bonded interactions as those that are responsible for molecular recognition and complexation at all levels. Indeed, crystallization itself is an impressive display of supramolecular self-assembly, involving specific molecular recognition at an amazing level of precision.

In view of this, it may seem all the more surprising that most compounds can exist in more than one crystalline form. The crystal structure of a compound is generally not unique, but depends on the conditions of crystallization. Polymorphism is ubiquitous. Indeed, it has been said that the number of polymorphs of a given substance is approximately proportional to the amount of time spent in trying to crystallize that substance. Most crystallographers examine one crystal and determine its structure, then the matter is finished. If they were to look at two or three crystals, they would often find structural differences, mainly differences in molecular packing, but often also in conformation. Free energy differences between polymorphs are usually quite small, a few kilocalories per mole at most, and may have different temperature dependences. This means that over a quite small range of temperature, and particularly between room temperature and the melting point, one polymorph or another can be the thermodynamically stable form. A metastable form can persist for years, or it can undergo spontaneous transformation to the stable form.

Polymorphism can be thought of as a kind of higher form of isomerism; it is in fact isomerism plus cooperativity. A polymorphic transformation is, then, a kind of isomerization reaction. Once we know the crystal structure adopted by a given molecule, we can easily analyse all the factors that contribute to its stability. Thanks to the Cambridge Structural Database (Cambridge Crystallographic Data Centre, University of Cambridge, UK; Allen et al 1983), we have a store of nearly 100 000 crystal structures to practise on. We can reassure ourselves that they all correspond to potential energy minima. What we cannot do, in spite of all the computational power that is available today,

is to solve the reverse problem—we still cannot predict the crystal structure or structures that a given molecule will adopt. This failure testifies that we do not understand sufficiently well these processes of self-assembly. It's a good challenge for anybody who thinks they understand anything about self-assembly to design the crystal structure of a given molecule. More attention should be paid to these phenomena by chemists, especially those interested in supramolecular chemistry. Apart from their interest as devices for yielding molecular structures, crystals have a lot to tell us about the details of intermolecular interactions. I am convinced that the study of polymorphism and of phase transformations in molecular crystals can teach us much about fundamental aspects of molecular assemblies.

Sutherland: There are a number of papers appearing on clathrates at the moment. The clathrate designers have gone a little way towards applying intuitive methods for designing compounds that will crystallize with cavities in the structure; however, this is done by analogy with similar structures.

Kennard: That's quite different from taking any one molecule and predicting the crystal structure.

Sutherland: I am not claiming they have done that. I was merely saying that we are making a little progress in the direction that Jack Dunitz had indicated.

Ron Breslow: Is it possible to predict the crystal structure of a really symmetrical molecule like cubane, or is even that too much of a challenge? With a highly symmetrical molecule, you have a feeling about how it might pack.

Dunitz: Well, let's take benzene as an example.

Ron Breslow: No, I deliberately didn't chose benzene because of this question of edge–face packing discussed earlier (p 18). Cubane doesn't have quite that problem.

Dunitz: I cannot remember what the structure of cubane is, but I suspect that at room temperature it forms a plastic crystal, like adamantane, because of its high symmetry. In the crystal, such molecules behave as if they were practically spherical. In a plastic crystal, orientational order is lost, although translational order is maintained; a plastic crystal is, in a sense, intermediate between a proper crystal and a liquid. (In a liquid crystal it is the other way round; orientational order is maintained and translational order is lost.)

Anyway, even for simple molecules of that type, we cannot predict crystal structures with any confidence.

Vinter: It depends on how precise you want to be and on the degree of intermolecular interference encountered in the crystal structure. With moderately unperturbed systems we can simulate many crystal structures to within 0.001 Å or better. More subtle features are difficult to reproduce and I am sure that's because we still lack a full understanding of the potentials involved. We are not too far away, and are getting closer as molecular mechanics and wave mechanics slowly merge, but we are not as close as we should be, considering the extent to which we use modelling methods blindly.

Kennard: I do not think we are close. If we were, we could solve the crystal structure from knowledge of the intermolecular interactions, but, in fact, the situation is just the other way round—we deduce the precise interactions from the crystal structure. I hope that the Cambridge Structural Database (Allen et al 1983) will one day be replaced by just a set of rules, but this is still far into the future.

As Jack Dunitz said, the Cambridge Structural Database could be used much more extensively to examine problems of this type. We have recently released a new graphics version (Version 4) which is particularly 'chemist-friendly', and is as easy to use as an Apple Macintosh. By drawing a chemical structure on the screen you can ask quite elaborate questions and analyse the three-dimensional data stored in the database with a variety of software tools. The Cambridge Structural Database System is widely distributed, and I urge all of you to try out the facilities it offers for new insight into the kind of questions we are discussing at this meeting.

Dunitz: If one is given the unit cell dimensions, the problem is much more tractable. By bringing the molecules in from infinity and putting them into a box of the right size, we have implicitly disposed of all the attractive interactions. What is left are the repulsive interactions, and the problem now consists of finding the structure where these are minimized.

Vinter: I accept that, but we are making progress.

Kennard: Is that conceptual progress?

Vinter: Yes, I think so, particularly in the realms of electrostatics.

Stoddart: In the context of the earlier discussion, I would like to say something about our tetracationic macrocycle, which is composed of two paraquat residues and bridging paraxylylene units. We have four crystal structures with different included molecules that are isostructural. They are quite different from those that I described for the catenanes. The cyclophanes form a kind of chimney-stack and placed between the tetracations there are four hexafluorophosphate counter ions. Included inside the tetracations you can have 1,2-dimethoxybenzene, 1,4-dimethoxybenzene, or 1,5-dimethoxynaphthalene, or you can have solvent in the form of acetonitrile. These complexes are all isostructural. We are using the very principle that Jack Dunitz has just enunciated to build supramolecular structures; we string together these neutral molecules covalently as suggested by the X-ray crystal structures, to produce complexes with [*n*]rotaxane-like structural features to them.

Albery: I was interested in Jack Dunitz's point about polymorphism. Does anyone have a hunch as to whether the kinetics is controlled by a nucleation process that actually leads to all these variations? We are interested in measuring kinetics of dissolution processes using the wall-jet electrode system. Quite a lot more work could now be done on the kinetics of both crystal formation and crystal dissolution, in order to see the different pathways that open up. Is my feeling that the nucleation kinetics is important correct?

Dunitz: Absolutely; the problem in liquid–solid phase transitions, while still very difficult, has a chance of being solved in the next decade by molecular dynamics calculations. The problem of nucleation in solid–solid phase transitions is much more intractable and will hold out, I would guess, for a further decade or two. One difficulty is that the study of phase transitions in solids has fallen into the realm of physics. Mainly because of the influence of those great physicists, Landau and Lifschitz, there is a tremendous amount of mythology about solid–solid phase transitions, but I have the impression that these authors (and many of their followers) rarely looked at the actual phenomena. A lot of the information in the textbooks needs to be re-written—or ignored.

Molecular devices and sensors

Stoddart: The building of molecular devices has so far happened successfully from the top down (miniaturization), and we have heard some elegant examples of this. I wonder if John Albery or David Reinhoudt, from their experience, could give a lead to those of us here who want to build such devices from the molecule up, by a process of 'maxiturization', if I might coin the term. John Albery had a very elegant acetylcholine sensor made up of three enzymes in a row. Do you think this is the way forward, putting catalytic systems on top of one other to get one particular system to talk to another at the molecular level?

Albery: People who work on molecular sensors divide into different camps on this matter. Some spend their time trying to develop molecular electronics. My philosophy has been rather different: I have always been in favour of the separation of function. I don't want to put my electrolyte into the amplifer—I would rather keep them separate. In my experience, you sometimes need to use a miniature device; if you want to look at a single neuron, for example, then you need a very small device. On the whole, we make our sensors at a macroscopic level because we think they work better. Separation of function is what I believe in; if I understand the kinetics of transport through the membrane or the enzyme kinetics, then I can leave the physicists to do the amplification.

In the particular case of the acetylcholine sensor, to detect low levels you need 100% conversion on the packed bed. That puts constraints on the kinetics of the enzyme system, on flow rate and on the dimensions of the bed. It's best to get 100% conversion, because you are then not dependent on the activity of the enzyme on the bed. I believe that we will see more serial reaction chains of that type; that is a sensible way to develop things.

Stoddart: Imagine that, rather than simply sensing a material, you wanted to go further and to be able to store information in the device. Do you think information could be stored within devices that have evolved through the miniaturization technique, or do you think that technique will not be capable of handling the storage?

Albery: I think that if you want to store a lot of information, particularly if you want fast switching, you have to go for miniaturization. I am greatly interested in what we may be able to achieve in this area with conducting polymers. These are interesting materials, because you can lay them down electrochemically and may be able to make thin plastic 'wires'.

Reinhoudt: There are different ways of trying to connect the chemical and the electronic worlds. In one of the camps are the analytical chemists, who want to construct sensors. We entered this field with a different philosophy, wanting to to apply molecular recognition and to communicate this to the outside world. We ended up making sensors and now everybody focuses on this application. The other approach would be to synthesize molecular receptors and to store information by complexation, for example. However, I haven't been able to see a way to communicate with individual molecules. We should bear in mind that Nature doesn't do this; Nature organizes molecules in higher aggregates and builds up membrane potentials. In our work, communication is also achieved simply through differences in concentration, and thus in electrical potential, on a field-effect transistor (FET). That's why we chose this way to start, not because we think it's ultimately the best way. The method of communication is an intriguing problem. Although I have been to several meetings on molecular electronics, nobody has ever made it clear to me how best to approach the problem. It's easy to say that you can take a matrix of organic material, to store information, and then do molecular computing, but, to me, that's science fiction, not science.

Stoddart: At a meeting earlier this year Professor H. Kuhn spoke about the need for the aggregation of molecular components if this communication problem is going to be addressed successfully. When pressed to speculate about the numbers of molecular components involved, he suggested in the region of 10^4.

Sutherland: That is close to the level of detection that John Albery can achieve.

Albery: The level of sensitivity at the moment is impressive, but if one wants to pass information, I suspect that photons would be better than electrons.

Sutherland: The level of 10^4 molecules was mentioned in terms of photons.

Reinhoudt: Even 10^4 molecules would form a two-dimensional network of only 2500 nm^2, and one must be able to address such an assembly individually with a light beam to read the stored information, so the wavelength of light will be a problem.

Sutherland: Yes, there is a problem associated with the wavelength of light.

Reinhoudt: The limit of miniaturization that I pointed to is simply the limit of the photolithography, which is limited by the wavelength of light. If you went down to X-ray wavelength you could store a lot more information. I think the problem is that you have to use so much energy at this small surface area that you destroy organic molecules.

Dunitz: Does the problem not depend on the temperature at which you want your device to work? If it's going to work at 1 K, you can probably manage with one molecule.

Reinhoudt: That's correct. Professor E.T.H. Wild has the technology to make a compact disc containing enough music for a lifetime—80 years of music on one compact disc—but it works only at 5 K.

Reference

Allen FH, Kennard O, Taylor R 1983 Systematic analysis of structural data as a research technique in organic chemistry. Acc Chem Res 16:146– 153

Clefts as receptor and enzyme analogues

Julius Rebek Jr

Department of Chemistry, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

Abstract. Synthetic receptors for small biological targets have become a popular research topic in molecular recognition. This paper discusses the optimal functional group complements and the scaffolds that are ideal for such purposes. Specifically, remote steric barriers are used to control the conformation of (that is, to preorganize) hosts derived from acridine skeletons, triaryl benzenes and related systems. These structures separate entropic effects from enthalpic effects and show that entropy is an important contributor to high affinity. In a comparative study lactams are shown to be superior to imides in their capacity for self-association. Imides are shown to have higher affinity than lactams for adenine derivatives because of the presence of an unconventional hydrogen bond. Finally, preorganization in the context of chemical catalysis is demonstrated in two systems, one involving hemiacetal cleavage and a second involving a self-replicating system.

1991 Host–guest molecular interactions: from chemistry to biology. Wiley, Chichester (Ciba Foundation Symposium 158) p 98–114

Recent developments in molecular recognition now make possible the synthesis of receptors with a variety of sizes, shapes and functional linings. The questions are not so much how to build something, but what to build and why. There are some technical questions as well. How rigid or flexible should these synthetic receptors be? Is it possible, or even desirable, to control their sizes to tenths of ångströms? What can be learned from model systems that cannot be tackled in the naturally occurring systems through genetic engineering? In this chapter we report on the progress that we have made in addressing some of these issues.

Preorganization

How well things fit together depends on their predisposition to do so, a matter frequently referred to as 'preorganization' (Cram 1988). While the term conjures up entropic effects, it is hard to separate these from enthalpic effects. Take for example the comparison of 2,2′-bipyridyl and *o*-phenanthroline as metal-chelating ligands. The latter has a higher affinity for metals; its ground state

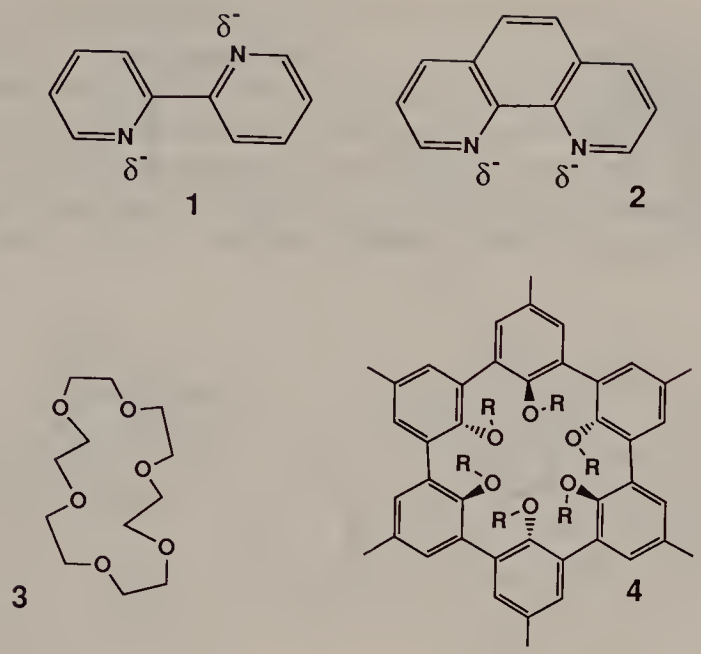


FIG. 1. Entropy or enthalpy in preorganized chelating agents: 2,2′-bipyridyl vs. *o*-phenanthroline and 18-crown-6 vs. a spherand.

conformation resembles that of the complex. Is the enhanced affinity due to the loss of rotation between the 2-pyridyl nuclei or the destabilized ground state in which the dipoles (lone pairs) converge, or both? A similar question can be posed for the crown ether, 18-crown-6, compared to a spherand (Fig. 1), and recent work in conformationally restricted crown ethers (Iimori & Still 1989, Iimori et al 1989) has made important progress on this question.

Our own approach was somewhat different and began with the premise that the enthalpic and entropic effects were more likely to be distinguished in large structures rather than from further investigations of small molecules. Our initial studies on this issue involved the acridine system shown in Fig. 2. In **5**, the condensation product of Kemp's triacid and proflavin, the *ortho* hydrogens permit rotation about the $C_{aryl}—N_{imide}$ bond. As a result, three conformations exist: the 'in–out' one shown, a divergent one, and a convergent one. The *ortho* methyl groups of **6**, however, lock the system into the convergent mode—that is, the OH bonds of the acid converge towards the centre of the structure.

Titrations with triethylenediamine (diazabicyclo[2.2.2]octane, DABCO) gave the chelated complexes **7**, which showed that **6** had a higher affinity than did **5**; the ratio of association constants was 12 : 1 or about 1.5 kcal/mol difference in affinities (Rebek et al 1987). This value is larger than might be expected from that based on statistics, that value being a factor of three. Rather surprisingly, when quinoxaline was the substrate, **5** bound it somewhat better than did **6**, the value being 2.9×10^4 as compared to 2.3×10^4. This may be due to the

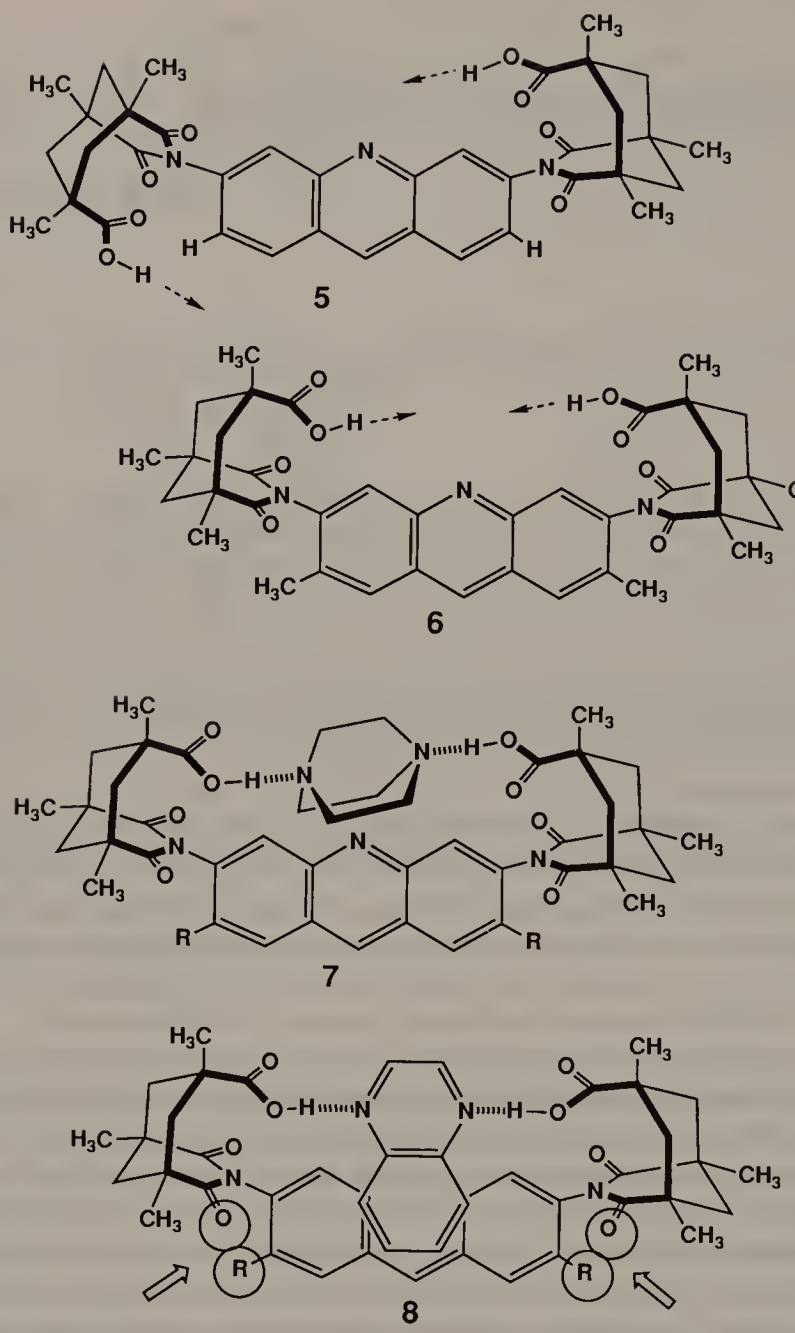


FIG. 2. Conformation and complexes of acridine-derived diacids. An in–out conformation, **5**; a convergent conformation, **6**; diazabicyclo[2.2.2]octane (DABCO) complex, **7**; and quinoxaline complex with aryl stacking, **8**.

greater N–N separation of the pyrazine ring compared to DABCO. As Jorgensen et al (1989) have pointed out, a pyrazine nucleus is slightly too large for the cleft bearing the *ortho* methyls and prefers to float about slightly above the two diacids. The *ortho* hydrogens would permit rotations and freedoms from

buttressing effects that could now accommodate the heterocycle in a more chelated manner, as indicated in **8**.

The message seemed to be that even larger structures were required, lest the conformational restrictions that are introduced mask the issue by inadvertently changing the size or the shape of the receptor. Accordingly, we moved to the triaryl benzenes as spacers to hold apart three carboxyl groups (Fig. 3), each provided by Kemp triacid subunits (Kemp & Petrakis 1981). The molecules are

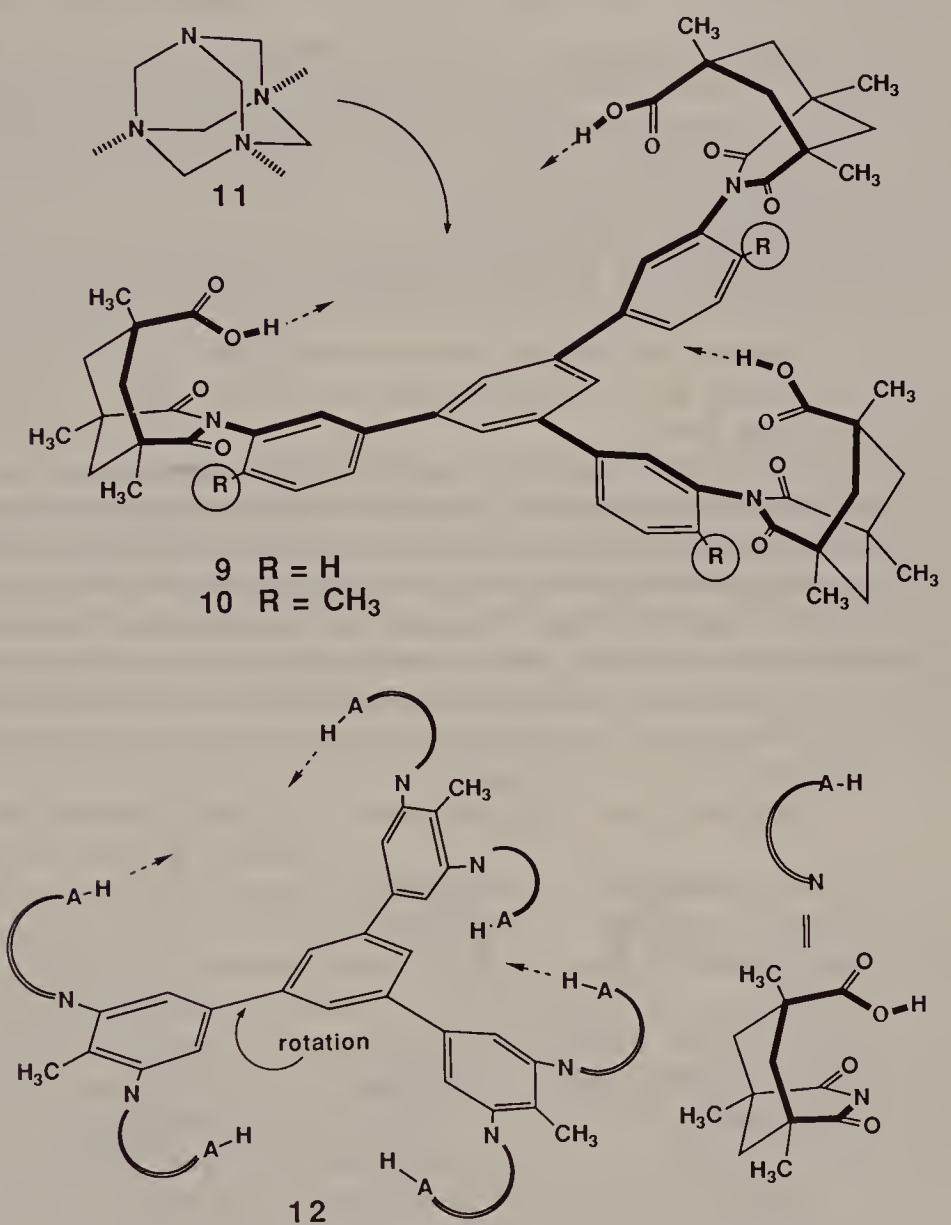


FIG. 3. Receptors (**9,10**) derived from triaryl benzenes complex with hexamethylene tetramine (**11**) and related ligands.

readily prepared by acid-catalysed trimerization (Ebmeyer & Vögtle 1989) of the appropriate nitroacetophenones followed by their catalytic reduction. The amines are then condensed with the anhydride acid chloride derivative of Kemp's triacid.

Direct comparisons could be made in this system between *ortho* hydrogen and *ortho* methyl groups—for example, structures **9** and **10** (Fig. 3). Complexation of hexamethylene tetramine (**11**) titrated in CD_3CN showed the *ortho* methyl compound to be some 1.5 to 3.5 times more effective than the *ortho* hydrogen compound (J. Huff & N. Horiuchi, unpublished work 1990). Rotation between the aryl–aryl bonds permits conformations in which all three carboxyls are on the same side, or another in which two are on one side and one is on the other. Thus **10** has two conformations whereas the *ortho* hydrogen molecule **9** has eight. Thus the seemingly passive methyl group has the four-fold effect anticipated on purely entropic grounds.

We also examined complex formation between hexaacid **12** (Fig. 3) and other azaadamantane derivatives. Its preorganization is as complete as we can make it. Here we see an *allosteric* effect that is expressed as positive binding cooperativity. Binding on one face fixes the dihedral angles in a conformation optimal for binding at the remote site on the opposite face (N. Horiuchi, unpublished work 1990).

The allosteric effect could be magnified in a system based on the tetraarylethylene skeleton (Fig. 4). The results of the binding studies with **13** and several diamines are shown in Table 1 (F. Ebmeyer, unpublished work 1990). Very large cooperativity is observed—for example, nearly 200-fold for the smaller guest. Presumably some intramolecular hydrogen bonds exist in the uncomplexed host species but, on the formation of a 1 : 1 complex, these are destroyed, forcing the two non-participating carboxyls to the opposite face of

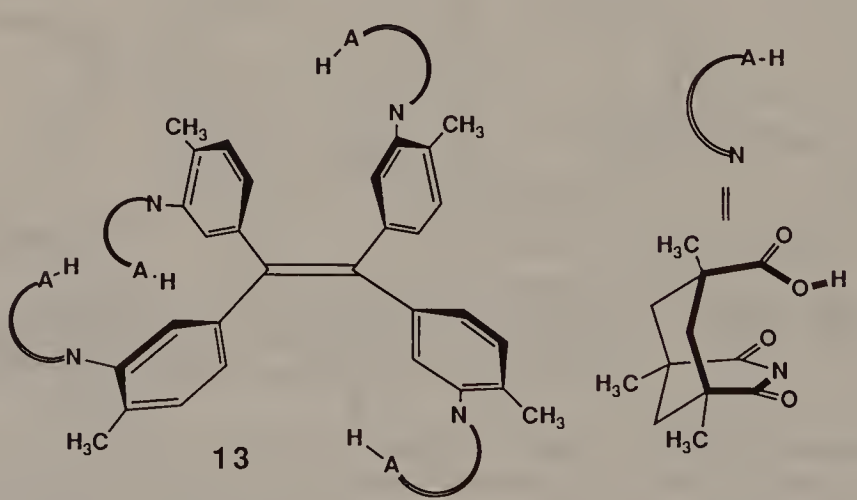


FIG. 4. Receptor (**13**) based on a tetraarylethylene skeleton.

TABLE 1 Association constants for the complexation of various amines with tetraacid 13 in $CDCl_3$ at 25 °C

Guest	$K_{1:1}$	$K_{1:2}$
Hydrazine	630 (16.0)	35000 (25.9)
Ethylenediamine	9 (5.4)	1700 (18.4)
Triethylenediamine (DABCO)	150 (12.4)	360 (14.6)
2-Aminopyrimidine	1400 (18.0)	90 (11.2)

the structure. These are now exposed for amine binding, having been deprived of the stabilizing influence of hydrogen bonds.

Allosteric effects may be one means by which Nature addresses the problem of preorganization. Binding at one site is used to organize (or disorganize) the remote site and turn on or off the enzyme or receptor. The model systems shown here demonstrate that this can be quite effective, and the enhancements observed for the carboxyl functional groups are quite large, suggesting that there is more than entropy at work here.

Hydrogen bonding

We have already reported extensively on the use of these model compounds to study base pairing (Askew et al 1989, Williams et al 1989). The advantage they offer is that both hydrogen bonding and aryl stacking interactions can be observed in $CDCl_3$. This solvent acts as a magnifying glass and permits associations to be observed and mapped out that would not otherwise be accessible in a more competitive environment, such as H_2O. From an analysis of the binding in synthetic receptors has emerged a profoundly useful mnemonic concerning secondary interactions (Jorgensen & Pranata 1990): secondary intermolecular interactions alter the nature of the hydrogen bonds. These interactions can be stabilizing or destabilizing and can therefore affect association constants quite dramatically.

We have examined these secondary effects in two systems. In the first, the relative hydrogen-bonding affinities of imides and lactams were examined (Jeong et al 1990). Such functional groups have been studied in the past but their self-association or dimerization is quite small in $CDCl_3$. However, with the U-turn inherent in Kemp triacid derivatives, it was possible to arrange the acidic and basic sites to converge in an intramolecular manner, as shown in Fig. 5. Three sets of structures were prepared: the diimides **14**, the racemic dilactams **15**, and the hybrid imide lactam (**16**) structures, separated by various flexible spacers (the $O\text{-}(CH_2)_n\text{-}O$ linkages). Their tendencies to form cyclic arrays of hydrogen bonds of an intramolecular sort are summarized in Table 2.

The lactams consistently showed a higher tendency to self-associate than did imide functions. In accord with Jorgensen and Pranata's analysis, the

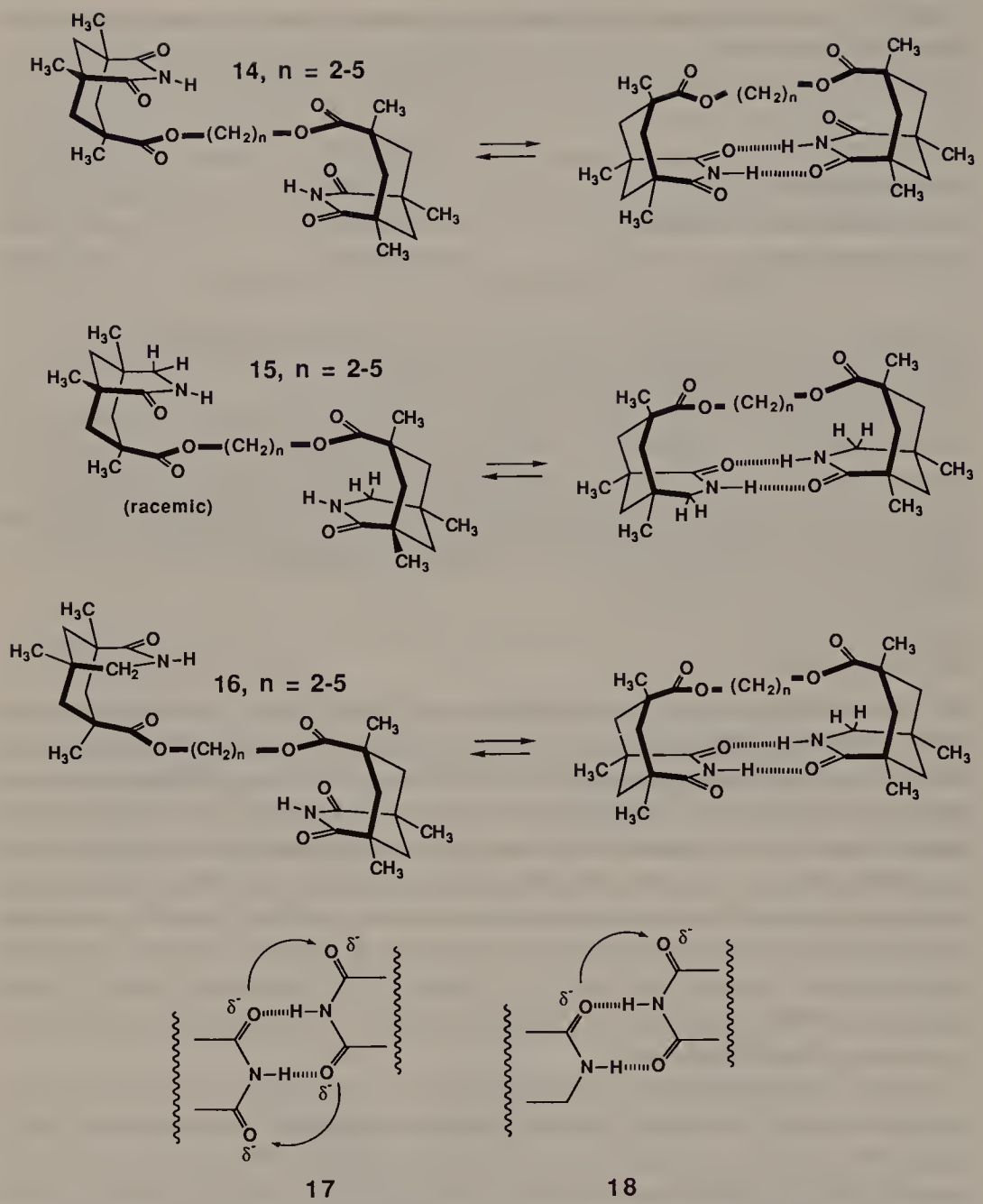


FIG. 5. Hydrogen-bonding affinities of imides and lactams: imide–imide interactions, **14**; lactam–lactam interactions, **15**; imide–lactam interactions, **16**; and destabilizing effects of 'spectator' carbonyls (**17** and **18**).

'spectator' oxygens of the imides (those not involved in the hydrogen bonds) appeared to destabilize nearby hydrogen bonds. The effect appears to involve ≈ 0.4 kcal mol^{-1} in these systems per destabilizing interaction, as shown in **17** and **18** (Fig. 5). The enhanced acidity of the imide is of no advantage in

TABLE 2 Cyclization data obtained for the imide and lactam systems shown in Fig. 6, in $CDCl_3$ at 25 °C

	Imide–imide (14)			*Lactam–lactam (15)*		*Lactam–imide (16)*	
n	*% cyclic form*	K_{obs}	$K_{c(corr)}$	*% cyclic form*	K_c	*% cyclic form*	K_c
2	63	1.7	0.85	86	6.1	75	3
3	83	5.2	2.6	96	24	87–90	8
4	32	0.45	0.22	48	0.92	27–44	0.43–0.78
5	39	0.64	0.32	55	1.2	39–45	0.66–0.82

K_c, equilibrium constant for cyclization.

hydrogen bonding to other imides or lactams; however, it does appear to be an important factor in base pairing to adenine and our second system involved the use of adenine derivatives as probes.

In direct comparisons, the specific cases involved were the naphthyl derivatives shown in Fig. 6. The imide **19** showed *higher* affinity than the lactam **20** toward 9-ethyladenine (**21**) and the imide was approximately 1.5 kcal mol^{-1} more effective in its contact with adenine, despite the presence of the spectator carbonyl (K. Jeong et al, unpublished work 1990). In a second comparison, the flexible diimide **14** with a five-carbon spacer was compared to the corresponding (*meso*) dilactam **15**. Again, the imides outperformed the dilactams by a very large margin in the formation of complexes **22** and **23**.

That Nature chose imides rather than primary amides to base pair with adenine must carry with it some message, but what is it? It might be found in a recent calculation (W. Jorgensen, unpublished work 1990) in which the calculated structure of the base pair shows a great deal of polarization of the hydrogen at C-2. Thus a sort of hydrogen bond exists between the 'spectator' carbonyl and this polarized hydrogen, as shown in **24** (Fig. 6). Presumably a similar polarization occurs at the hydrogen (at C-8) in the chelated derivatives. The interpretation draws support from the unconventional hydrogen bonds to CH hydrogens in caffeine observed previously in the crystalline state (Donohue 1968). Thus model systems provide a means by which such subtle but intrinsic factors in hydrogen bonding can be assessed at the level of less than 1 kcal mol^{-1}.

Catalysis

One of the historical goals in biomimetic chemistry has been the design and synthesis of enzyme models (Hosseini et al 1989, Anslyn & Breslow 1989). Molecular recognition plays a key role in the binding step—that is, formation of the reactive complex. But how then is one to arrange functional groups in the complex to promote actual chemical reactions? A convergent distribution of functional groups here again appears to be the key, in that it can 'focus' activity on a highly localized site. The ultimate goal is to merge the recognition

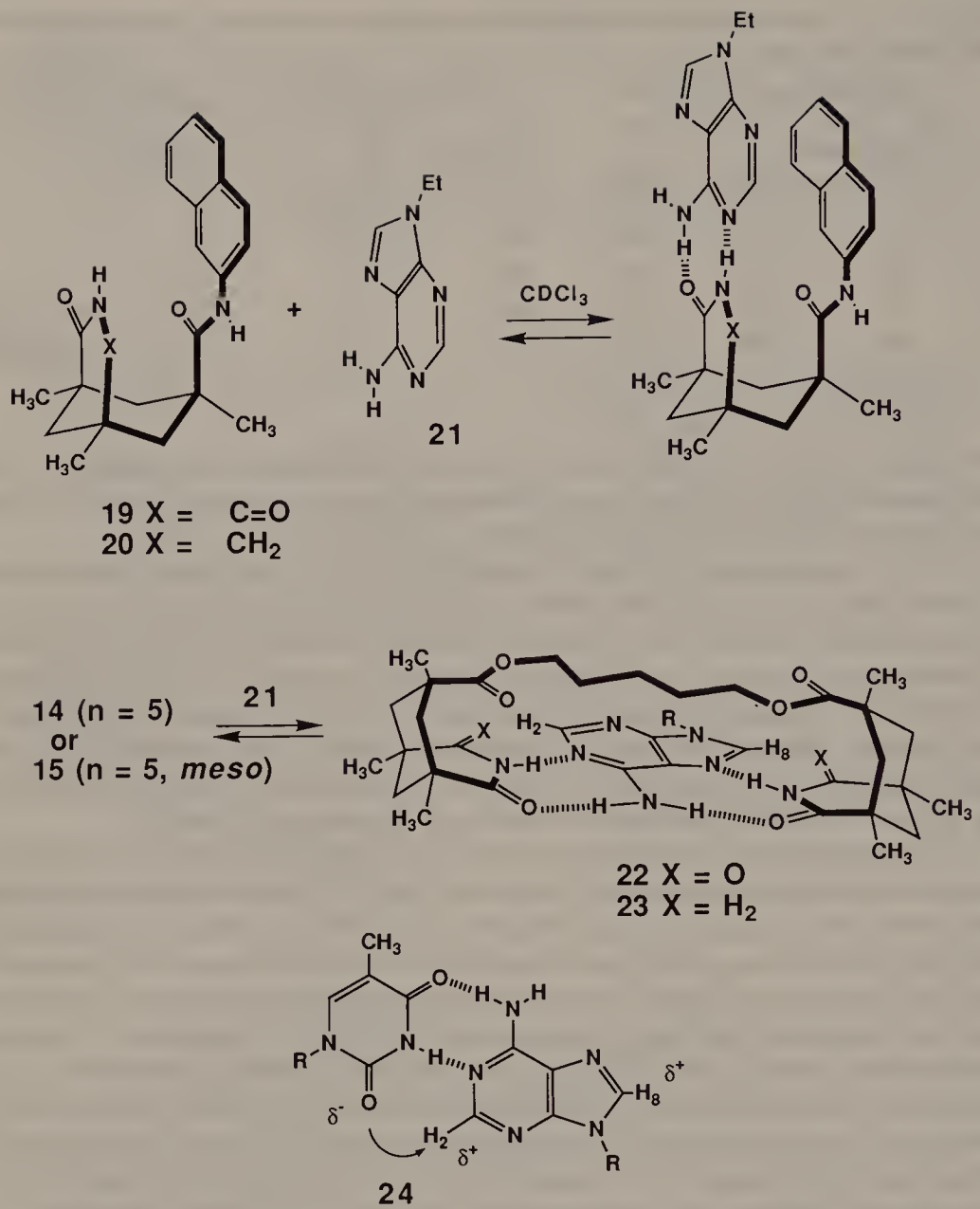


FIG. 6. Superior affinity of imides over lactams for adenine derivatives: binding to rigid receptors, **19** and **20**; binding to flexible receptors, **14** or **15**, which gives **22** and **23**; and unconventional hydrogen bonding in **24**.

and reaction steps in space and in time so that *transition states* would be recognized best, as was anticipated by Pauling so many years ago (Pauling 1946).

We have made progress in two areas. The first involves the cleavage of hemiacetals, specifically, the dihydroxy dioxane **25** shown in Fig. 7. Conventional acids and bases, and even the much-admired 2-pyridone, are quite poor in

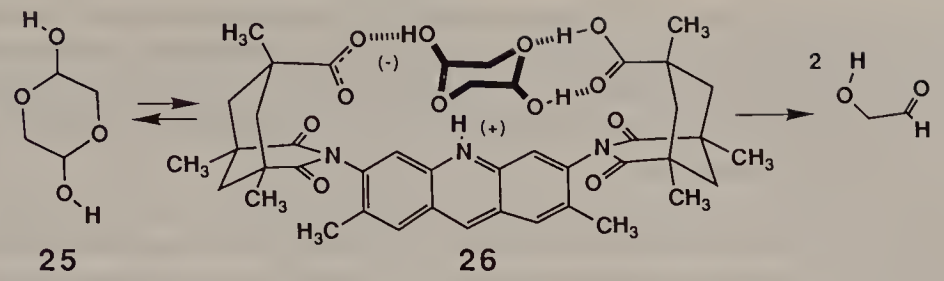


FIG. 7. Cleavage of a hemiacetal (**25**) through concerted catalysis by an enzyme mimic, an acridine diacid (**26**).

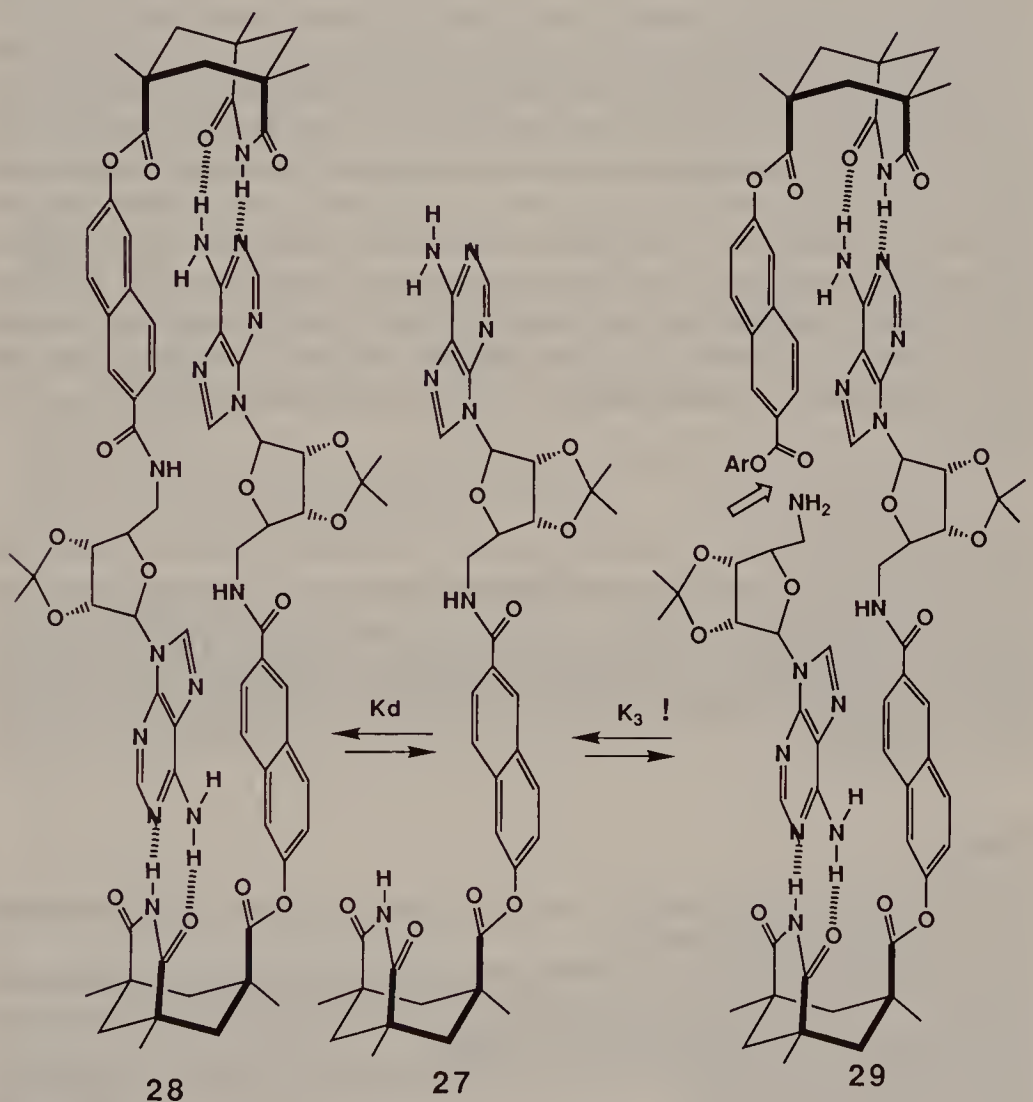


FIG. 8. Dimerization (K_d) and autocatalysis (K_3) in a self-replicating system.

catalysing the dissociation of this material into hydroxyacetaldehyde. However, the acridine diacid is excellent in this regard (Wolfe et al 1988). The structure fits nicely within the cleft (**26**), where a constellation of functional groups is poised to effect the bond-breaking steps. The result is a 10^4 rate enhancement of the dissociation reaction.

The second system involves an intriguing structure capable of *self-replication* (Tjivikua et al 1990). This structure **27** (Fig. 8) is self-complementary; that is, an adenine is covalently bound to a receptor for adenine, and whereas it can dimerize to **28**, it is unable to fold up upon itself in an intramolecular manner. However, it can gather the two components from which it is made on its surface and thereby promote an otherwise bimolecular reaction to a unimolecular one. Adding the template product to a mixture of the active ester and the amine nucleophile enhances the coupling rate. From the known association constants, the relative concentrations of the various species can be calculated; the termolecular complex **29** is typically there as only a few per cent of the total concentration, but its calculated effective molarity is quite high, >100 M. This system shows the unusual feature of autocatalysis. Whereas this has been observed in nucleic acid systems, both enzymically and non-enzymically (Zielinski & Orgel 1987, von Kiedrowski et al 1989), the case at hand involves the formation of an amide or peptide bond, enhanced by base-pairing events elsewhere in the molecule. Thus it provides a very tenuous bridge between the nucleic acid and peptide worlds (Orgel 1987). We are currently pursuing the construction of even sturdier bridges between these two systems.

Acknowledgements

I am most grateful for the capable experimental assistance of my co-workers at the University of Pittsburgh and M.I.T., whose names appear in the original literature citations. I thank Professor W. Jorgensen for helpful discussions and for sharing his results with us prior to their publication. Financial support was provided by the National Science Foundation through their Special Creativity Grant programme, and the National Institutes of Health.

References

Anslyn E, Breslow R 1989 Geometric evidence on the ribonuclease model mechanism. J Am Chem Soc 111:5972–5973

Askew B, Ballester P, Buhr C et al 1989 Molecular recognition with convergent functional groups. VI. Synthetic and structural studies with a model receptor for nucleic acid components. J Am Chem Soc 111:1082–1090

Cram DJ 1988 The design of molecular hosts, guests and their complexes. Angew Chem Int Ed Engl 27:1009–1020

Donohue J 1968 Selected topics in hydrogen bonding. In: Rich A, Davidson N (eds) Structural chemistry and molecular biology. W H Freeman, San Francisco, p 443–465

Ebmeyer F, Vögtle F 1989 Selective molecular recognition of trihydroxy benzenes. Angew Chem Int Ed Engl 28:79–81

Hosseini MW, Lehn JM, Jones KC, Plute KE, Mertes KB, Mertes MP 1989 Supramolecular catalysis: polyammonium macrocycles as enzyme mimics for phosphoryl transfer in ATP hydrolysis. J Am Chem Soc 111:6330–6335

Iimori T, Still WC 1989 Podand ionophores. A new class of nonmacrocyclic yet preorganized hosts for cations. J Am Chem Soc 111:3439–3440

Iimori T, Erickson SD, Rheingold AL, Still WC 1989 Enantioselective complexation with a conformationally homogeneous C_2 podand ionophore. Tetrahedron Lett 30:6947–6950

Jeong KS, Tjivikua T, Rebek J Jr 1990 Relative hydrogen-bonding affinities of imides and lactams. J Am Chem Soc 112:3215–3217

Jorgensen W, Pranata J 1990 Importance of secondary interactions in triply hydrogen bonded complexes: guanine-cytosine vs uracil 2,6-diaminopyridine. J Am Chem Soc 112:2008–2010

Jorgensen WL, Boudon S, Nguyen TB 1989 Structure and binding for Rebek's diacid in chloroform. A demure host for pyrazine. J Am Chem Soc 111:755–757

Kemp DS, Petrakis SK 1981 Synthesis and conformational analysis of *cis,cis*-1,3,5-trimethylcyclohexane-1,3,5-tricarboxylic acid. J Org Chem 46:5140–5143

Orgel LE 1987 Evolution of the genetic apparatus: a review. Cold Spring Harbor Symp Quant Biol 52:9–16

Pauling L 1946 Molecular architecture and biological reactions. Chem Eng News 24:1375–1377

Rebek J Jr, Askew B, Killoran M, Nemeth D, Lin FT 1987 Convergent functional groups 3. A molecular cleft recognizes substrates of complementary size, shape and functionality. J Am Chem Soc 109:2426–2431

Tjivikua T, Ballester P, Rebek J Jr 1990 A self-replicating system. J Am Chem Soc 112:1249–1250

von Kiedrowski G, Woltzka B, Helbing B 1989 Sequence dependence of template-directed syntheses of hexadeoxynucleotide derivatives with 3′-5′ pyrophosphate linkage. Angew Chem Int Ed Engl 28:1235–1237

Williams K, Askew B, Ballester P et al 1989 Molecular recognition with convergent functional groups VII: energetics of adenine binding with model receptors. J Am Chem Soc 111:1090–1094.

Wolfe J, Nemeth D, Costero A, Rebek J Jr 1988 Convergent functional groups: catalysis of hemiacetal cleavage in a synthetic molecular cleft. J Am Chem Soc 110:983–984

Zielinski WS, Orgel L 1987 Autocatalytic synthesis of a tetranucleotide analogue. Nature (Lond) 327:346–347

DISCUSSION

Kennard: Dr Rebek, how many of the complexes you described were actually solved by X-ray analysis?

Rebek: None of them; the receptors themselves have lower solubility than the complexes.

Kennard: It will be interesting to see if there are any surprises when you have the crystal structures.

Rebek: We are fairly confident about most of the structures in solution, because they show the expected intermolecular NOEs. I think it's useful for organic chemists to look at hydrogen-bonding patterns, because they help in the selection of the right complement for a given situation or functional group. Moreover, they explain a lot of mysterious data.

Ron Breslow: I would like to ask about the difference in binding energy between the lactam and the imide. Of course, the basicity of the carbonyl differs in these two groups, so you might worry about whether that explains the difference. That's not what Jorgensen invoked.

Rebek: Jorgensen & Severance (1990) have now done the calculations and have come to the same conclusion as we have; the spectator oxygen destabilizes the hydrogen bond between the imides.

Ron Breslow: What kind of calculation did Jorgensen do—SCF-CI?

Rebek: You are asking the wrong person.

Ron Breslow: But the point is that these calculations may not be valid.

Rebek: He reproduces the binding data well and he didn't get the data before he did the calculation.

Dunitz: But that only means that his calculation has reproduced your result. The interpretation of it is still open to endless discussion, because there is no rigorous way of separating out that little bit of stabilization into so much of A and so much of B.

Rebek: I don't know if I should be Jorgensen's apologist here. All I can do is present you with the best data I can obtain.

Ron Breslow: Jorgensen does a calculation that gives an energy for this structure. To establish which part of the interaction leads to a particular piece of that energy you have to do another calculation in which you remove one of the atoms; when you do that, many other effects come into play, and I don't know how you would distinguish between them.

Sutherland: Doesn't it depend on how you look at your hydrogen bonds? If you consider them as point charge interactions you can identify individual contributions. If you use a quantum mechanical approach I agree that it would be very difficult to separate out individual interactions.

Ron Breslow: Jorgensen does a quantum calculation on one system; to say which interaction is the one that matters you have to do another calculation in which you change the basic structure.

Dunitz: You do a metacalculation.

Rebek: I believe so; you mutate parts of the structure.

Ron Breslow: This is a general problem. Quantum mechanics is useful, but it cannot always tell you which is the interaction that matters. The experimental finding that lactams are better at self-assembly than imides is the fundamental datum; whether the Jorgensen calculation adds anything to that is another question. If you had not done the experiment, I certainly wouldn't take the calculation as my guide.

Dunitz: Let's imagine Jorgensen's calculation was so accurate that it could serve as an exact analogue computer for your molecule. It would then produce the same result as your experiment. But this perfect agreement would not tell you how to factorize the extra stabilization of the dimer into so much of this hydrogen bond, so much of that one, and so much from the spectator oxygen. There is no rigorous way to use quantum mechanics to partition the energy; to do that you need philosophy—or poetry!

Ron Breslow: You do the second calculation hoping that you know what you have erased. Was the solvent included in the calculations?

Rebek: Yes.

Dunitz: If I remember correctly, it was Planck who said that the facts are the only knowledge we have; the rest is imagination, poetry.

Rebek: I have to have some poetry!

Ron Breslow: There's a theoretical physicist who said that when the experiments do not agree with the calculations, the experiments should be re-done!

Reinhoudt: If you dissect the different energy contributions to the complexation process for a series of compounds, the only thing that matters is whether the experimental result corresponds to the predicted value.

Dunitz: You are proposing that additivity is the most important principle in science, but it's not; we know of many situations where things are not additive.

Reinhoudt: If the contributions are additive they provide good guidance for the synthetic chemist. Cram tried to assess the energy contributions of hydrogen bonding and the positive charge on the nitrogen in *tert*-butyl ammonium complexes (Timko et al 1977). Additivity works as a general rule in flexible systems but it does not apply if there are steric constraints.

Sutherland: What you are really saying is that we can use a simplified model if it makes useful predictions. This must be true; we don't always have to use the quantum mechanical model.

Reinhoudt: Dr Rebek, have you ever used thioamides? The closest you could get to removing the oxygen would be to replace it with another heteroatom.

Rebek: We have made thioamides, but it never occurred to us to do any binding studies with them.

Reinhoudt: There would be a larger negative charge and the C=S group is also more polarizable.

Rebek: The interesting question is why Nature chose imides, not lactams, to base pair with adenine. We have now compared these two systems, and have found that imides bind better and that there's a good reason for that. Crystallography rescues us here through some observations made by Donohue (1968) and others. We have compared imides and lactams with a simple naphthalene surface, with flexible spacers and with a more rigid one (**19–24**). The association constants for the imides are invariably higher, by up to two orders of magnitude. According to Jorgensen's calculations (Jorgensen &

Severance 1990) there seems to be some unconventional hydrogen bonding between the highly polarized hydrogen at C-2 and the otherwise spectating oxygen. This can take place on both the Watson–Crick edge and the Hoogsteen edge (see structure **24**). In certain crystal structures, such as those of caffeine and other heterocycles, you frequently see what is often referred to as a hydrogen bond to those hydrogens; heteroatoms are at the right sort of distance and orientation for hydrogen bonding to the polarized C—H.

Ron Breslow: But there's another possible explanation for that. Adenine is a base, so you want the more acidic partner for it; that's what you are looking at.

Kellogg: I gather that you have used both the 1,3-dicarboxylic acids and the structurally more rigid 1,5-dicarboxylic acids. Can you estimate how much is gained from the preorganization?

Rebek: I think that if you don't build in enthalpic destabilizations, then counting the number of different conformations that the flexible molecule can have compared to those that are required for binding will allow you to calculate the enhancement obtained through preorganization. But in 1,5-dicarboxylic acids the two carboxyl groups communicate with each other, so you can't isolate or separate the enthalpic and entropic contributions to preorganization.

Kellogg: That is correct. As long ago as 1939 Dippy noted that pK_a values suggest that there is some communication between carboxyl groups even in 1,5-di-substituted acids.

Hamilton: There was a *cis–trans* isomerization of the amide in the self-replicating system that you described (Fig. 8). I wondered whether rotation might occur not at the *cis*-amide stage, but at the tetrahedral intermediate that precedes it. The tetrahedral intermediate that is formed by the reaction of the amine with the *p*-nitrophenyl ester will have a C—N single bond and so rotation should be easier. One of the proposed mechanisms for proline isomerase enzymes involves formation of a tetrahedral intermediate which lowers the energy for rotation of the C—N bond. Harrison & Stein (1990) have shown that this is probably not the mechanism; nevertheless, it's an interesting possibility.

Rebek: I think that suggestion is a good one. It may be that at the tetrahedral intermediate the two components can drift partially away from one another. I don't see what advantage that would have over the alternative since it would break up the template effect.

Hamilton: I suppose this is a question of the transition state energies. What is the difference between the transition state energy for amide bond rotation at the *cis* amide and that at the tetrahedral intermediate? Intuitively, one would think that it would be lower for the tetrahedral intermediate, because in the *cis* amide a partial carbon–nitrogen double bond has to be broken.

Sutherland: For the *cis* amide, the barrier for rotation is only 14 or 15 kcal.

Rebek: I hoped that the *cis*-amide system would be stabilized by the hydrogen bonds, because the *cis* amide is supposed to be only 2–3 kcal less stable than the *trans* amide. With the *cis* amide there can be hydrogen bonding in an

intramolecular sense, but the *trans* hydrogen bonds by dimerization, so there is no real change in the number of hydrogen bonds in the isomerization.

Gutsche: In an autocatalysed reaction, one would expect the rate to increase with time. Did you observe that?

Rebek: We studied the initial coupling rates; if you add an inhibitor the rate drops, but if you add some of the product the rate increases. The rate of change is really a function of how large the autocatalytic component is in comparison to the background coupling rate. If the autocatalytic component is large, you will see sigmoidal reaction rate curves. Also, you work in a concentration window that will alter the shape of that curve; for example, the background coupling is a bimolecular reaction but the autocatalytic component is a termolecular one, so at very high concentrations the autocatalytic component dominates, regardless of the relative contributions of the two components. The result will be a sigmoidal curve. In our experiments, with the concentration regime used and rate ratios observed, the plots were almost perfectly linear and we can simulate that by computation.

Ron Breslow: You said that the termolecular complex (**29**) formed only a small percentage of the total concentration and that adding the template product increased the coupling rate. By how much was the rate increased under those circumstances?

Rebek: The coupling rate was increased by 50%.

Ron Breslow: Can you get a larger rate increase using other concentration regimes, or is that the optimum? You have the problems of dimerization and inhibition by the product to worry about and if you lower the concentrations there will be other problems.

Rebek: The dimerization of the template is a problem, but because the system is termolecular you can reach a concentration at which you get larger proportions of the components in the termolecular complex. In a practical sense, solubility prevents you from getting higher concentrations.

Stoddart: With the 1,3,5-triarylbenzene-based receptor in which there were six identical arms (**12**) there was, as expected I think, no cooperativity in binding two substrates. The next obvious experiment to do is to desymmetrize the receptor and introduce three arms identical to each other and another three arms identical to each other but different from the first three. You could then test to see if there is cooperativity in the binding of two different substrates. Have you done that?

Rebek: No; unfortunately, my Japanese co-worker Nobu Horiuchi went back home.

Stoddart: Would you predict that there would be cooperativity?

Rebek: Yes, but of the heterotropic sort. You could use the binding of one substrate on one face to preorganize the opposite binding face for a different substrate. It's a nice idea; we have been trying to mono-functionalize this beast, but it's quite difficult.

References

Dippy JFJ 1939 The dissociation constants of monocarboxylic acids; their measurement and their significance in theoretical organic chemistry. Chem Rev 25:151–211

Donohue J 1968 Selected topics in hydrogen bonding. In: Rich A, Davidson N (eds) Structural chemistry and molecular biology. WH Freeman, San Francisco, CA, p 443–465

Jorgensen WL, Severance DL 1991 Chemical chameleons: hydrogen bonding with imides and lactams in chloroform. J Am Chem Soc 113:209–216

Harrison RK, Stein RL 1990 Mechanistic studies of peptidyl prolyl cis-trans isomerase: evidence for catalysis by distortion. Biochemistry 29:1684–1689

Timko JM, Moore SS, Walba DM, Hiberty PC, Cram DJ 1977 Host-guest complexation. 2. Structural units that control association constants between polyethers and *tert*-butylammonium salts. J Am Chem Soc 99:4207–4219

Enzyme mimics

Ronald Breslow

Department of Chemistry, Columbia University, New York, NY 10027, USA

Abstract. 'Two is better than one' in cooperating systems. Synthetic ditopic host molecules have been constructed from cyclodextrins and from synthetic hydrophobic cavities bridged with simple connectors or with catalytic groups. These host molecules show very strong binding of appropriate substrates, and selective catalysis. A cyclodextrin bis-imidazole shows bifunctional catalysis of the cleavage of a cyclic phosphate that is simultaneous, judged from isotopic studies. The best geometry is consistent with mechanistic evidence on the preferred mechanism for such bifunctional catalysis. Certain polar compounds bind to a still unidentified cellular receptor and induce differentiation of malignant cells. Ditopic drugs that have been designed and synthesized bind strongly to two receptor sites and induce cancer cell differentiation at low concentrations with few side-effects. Clinical results with these compounds are already promising in a few cancer patients.

1991 Host–guest molecular interactions: from chemistry to biology. Wiley, Chichester (Ciba Foundation Symposium 158) p 115–127

Dimeric hosts showing very strong binding of appropriate substrates

The hydrophobic binding of substrates into cyclodextrins in water has proved to be useful in the design of enzyme mimics. Some synthetic hydrophobic cavities are also useful, and have been compared with the cyclodextrins in a few cases (Breslow et al 1986). In the best cases the binding constants for the cyclodextrin complexes are of the order of $10^4\,M^{-1}$, a value similar to many enzyme–substrate binding constants, but the binding is not as strong as that of antibody–substrate complexes. We have prepared a number of cyclodextrin dimers to see whether we could achieve strong binding with appropriate substrates and could use that binding to achieve effective and selective catalysis. Early results are extremely promising.

With a single linkage between two cyclodextrin rings we can see binding constants of the order of $10^9\,M^{-1}$ with some substrates (Breslow et al 1989a), but the systems are sensitive to the precise geometries of the two components. When there are two links between the rings, a more complicated situation arises (Fig. 1); the rings can be locked either in an occlusive geometry (which we call the 'clam shell' arrangement) or in an aversive geometry in which the two rings cannot cooperate in binding a substrate (which we call the 'love seat'

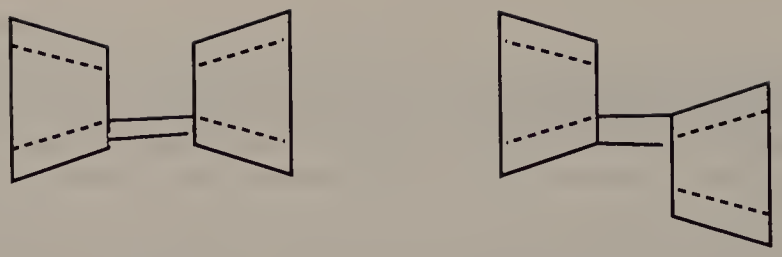


FIG. 1. When two cyclodextrins are doubly linked, the resulting dimers can have either the strongly binding, occlusive geometry that we call the 'clam shell' configuration or the weakly binding, aversive geometry that we call the 'love seat' geometry.

arrangement). Normally both isomers are formed and we must therefore separate them.

We have prepared such doubly linked cyclodextrin dimers by the use of cyclodextrin diiodides, whose uses we have described elsewhere (Anslyn & Breslow 1989a). Two covalent links can be made directly (Breslow & Chin 1990) or, alternatively, the cyclodextrin diiodides can be converted to cyclodextrin bis-imidazoles. We shall describe the catalytic properties of these bis-imidazoles later, but now point out that they can serve as metal ligands, producing dimers with an intervening metal ion. Several such systems have been explored (R. Breslow & R. Zarzycki, unpublished work 1990).

We have also prepared cyclodextrin dimers with a single metal ligand linkage between the two cyclodextrin binding groups. In this case we see an interesting metal-catalysed cleavage of substrates that bind with ester groups directly above the metal ions (R. Breslow & S. Halfon, unpublished work 1990).

We have designed dimers with geometric constraints such that substrates are bound with distortion in the geometric direction of the transition states for their reactions. Such use of binding energy is typical of enzymic reactions.

Bifunctional catalysis of phosphate ester cleavage by imidazole groups in ribonuclease mimics

The enzyme ribonuclease catalyses two steps in the cleavage of RNA (ribonucleic acid). First there is an ester interchange reaction, in which the 3′,5′ phosphate diester is converted to a 2′,3′-cyclic phosphate, with cleavage of the RNA chain. The enzyme then catalyses the hydrolysis of the 2′,3′-cyclic phosphate to regenerate the 2′ hydroxyl group and form a 3′ phosphate monoester where the chain had been broken. We have studied mimics of both reactions.

The ester interchange cleavage of RNA has been examined (Breslow & Labelle 1986), as has the simple cyclization-cleavage of a dinucleotide, UpU (Anslyn & Breslow 1989b). The catalyst used was simply imidazole buffer; we wanted to know whether the freely moving imidazoles of the buffer would indeed act

as catalysts, just as the imidazole rings do in the enzyme. More importantly, we asked what mechanism they would use. The mechanisms proposed for the enzymic reaction are supported by little direct evidence, but rely on analogy with simple model reactions. We wanted to determine the preferred mechanism for catalysis by the enzymic catalytic groups.

We found that the cleavage of RNA was indeed catalysed by imidazole buffer, and with kinetic behaviour that pointed to an interesting mechanism (Fig. 2). Both the basic imidazole and the acidic imidazolium ion components of the buffer played a role in the catalysis, but they acted sequentially. First one component converted the phosphate diester to a phosphorane monoanion intermediate, then the other buffer component catalysed the fragmentation of this phosphorane to form the cyclic phosphate product.

Simple kinetics could not tell us whether the first catalyst was the basic imidazole or the acidic imidazolium ion, but a study of the rearrangement that accompanies fragmentation indicated that the first catalyst was the acid, the imidazolium ion. This was the only catalyst for the isomerization of 3′,5′-UpU to 2′,5′-UpU, a process that shares the first step with the cleavage process but then does not require the second catalyst. In fact, basic imidazole was a 'negative

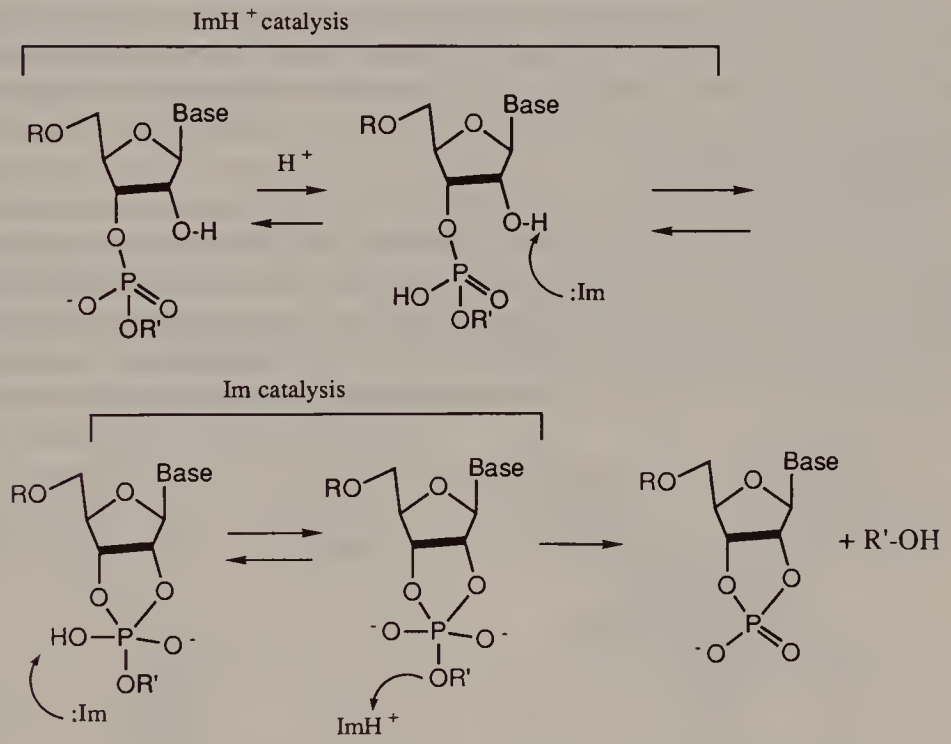


FIG. 2. Cleavage of RNA or of the dinucleotide UpU by imidazole buffer involves sequential bifunctional catalysis, first by the buffer acid and then by the base. The phosphate group is protonated to promote cyclization, but the proton is then removed to promote fragmentation of the intermediate to the cyclic phosphate product.

catalyst' for the isomerization, as would be expected if it diverts a common intermediate away from isomerization and toward fragmentation. We have seen similar kinetic behaviour with the dinucleotide ApA (R. Breslow & D.-L. Huang, unpublished work 1990).

Because the acidic catalyst acts first, it must be protonating the substrate phosphate ion. This is not the role usually assigned to the imidazolium ion in ribonuclease, but an examination of the enzymic data convinced us that this is actually the role that it first plays in the enzyme as well. Interestingly, the proton of the imidazolium ion can be replaced by a Zn^{2+} ion bound to imidazole (Breslow et al 1989b). The Zn^{2+} ion is a more effective catalyst than is the proton and it changes the mechanism in sensible ways.

With this mechanistic information in hand we re-examined an enzyme mimic we had first made some years ago (Breslow et al 1978). Cyclodextrin bis-imidazole can bind a cyclic phosphate ester into the cavity and the two imidazole rings can then cooperate in catalysing the hydrolysis of this phosphate ester. This is the same type of reaction as that catalysed by the enzyme, although our substrate is not the same and is chemically more reactive than is a 2′,3′-cyclic phosphate ester of ribose. We had originally designed our catalysts to use the previously accepted mechanism, in which the base delivers a water molecule to the phosphorus while the acidic imidazolium ion protonates the leaving group. With this mechanism, the two catalytic groups should be on opposite sides of the cyclodextrin cavity. However, with our new mechanism the imidazolium ion should be much closer to the imidazole, so that it can protonate the phosphate anion oxygen.

The appropriate cyclodextrin diiodide allowed us to make a new cyclodextrin bis-imidazole (Fig. 3), with the two groups on neighbouring glucose residues of the cyclodextrin (Anslyn & Breslow 1989a). This was actually the best catalyst of all. It cannot use the previously accepted mechanism; apparently even in a bifunctional catalyst the acidic imidazolium ion preferentially acts to protonate the phosphate ion. However, when both catalytic groups are in the same

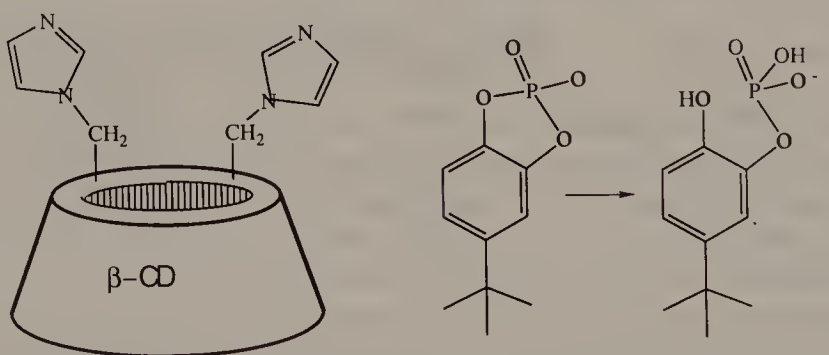


FIG. 3. When the cyclodextrin carries imidazole groups on neighbouring glucose residues (β-cyclodextrin A,B bis-imidazole) it is the best catalyst for the hydrolysis of a cyclic phosphate substrate.

molecule, it is not as likely that they act sequentially. The entropy factors that make simultaneous bifunctional catalysis unlikely when the catalytic imidazoles are freely moving buffer components are not the same when everything is connected, as in an enzyme or in our cyclodextrin-based catalysts. Thus, simple kinetic methods cannot address the question of whether catalysis is sequential or simultaneous, as they can and did in our buffer studies. However, the technique of *proton inventory*, involving reactions in various mixtures of H_2O and D_2O, was used. With the enzyme ribonuclease, proton inventory had shown the type of curve indicative of simultaneous two-proton transfer in the rate-determining step. With our catalyst, the same technique also showed that we were dealing with simultaneous bifunctional catalysis, not sequential catalysis (Anslyn & Breslow 1989c). The observed isotope effects were similar to those seen for the enzyme, but a little larger.

Although this technique is not new, there is still concern about whether it is a solid guide to mechanism. Luckily, in our case we could easily check it. The cyclodextrin carrying only *one* imidazole also catalyses the cleavage of our cyclic phosphate ester. It is less efficient, and shows only base catalysis. The proton inventory method showed only a single proton moving. Thus, bifunctional catalysis is probably truly simultaneous.

How can several things happen at the same time in a catalytic process? What is meant by 'simultaneous' in this case? The imidazolium ion is hydrogen bonded to the substrate phosphate anion, and the imidazole base group carries a hydrogen-bonded water molecule (Fig. 4). When that water attacks the phosphorus the pK values change such that the two hydrogen-bonding protons should now shift. They still vibrate along the internuclear line, but the energy minima are now such that the imidazolium ion will have donated its proton to the new phosphorane anion, even while retaining some hydrogen bonding to it. Similarly, the imidazole will have 'captured' a proton from water while it still retains a hydrogen bond. This movement of the energy minima along a hydrogen bond corresponds to proton transfer, whatever the timing of motion of the vibrating protons themselves.

Our mimics of the enzyme ribonuclease imitate the mechanisms, the pH dependences and even the isotope effects seen for the enzyme itself. With the correct geometry our bifunctional catalyst is quite efficient, although not yet

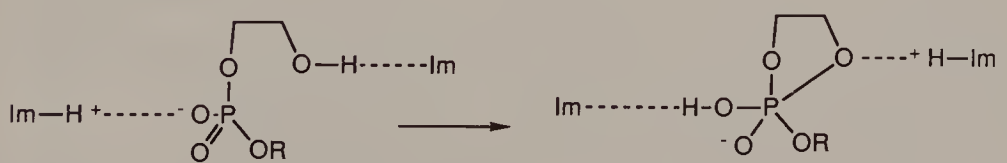


FIG. 4. The two hydrogen-bonding protons both move as the new P—O bond is made. The process is simultaneous in the sense that the positions of the energy minima both move, even though the protons are vibrating around those minima.

up to the catalytic level of the enzyme. We believe that achieving enzymic catalysis levels is just a matter of time.

Ditopic cytodifferentiating agents

About 15 years ago we started a collaboration with Paul Marks and Richard Rifkind, now of the Sloan-Kettering Institute for Cancer Research. We were following up an observation made by Charlotte Friend, that when pre-erythrocytes infected with murine erythroleukaemia virus were treated with dimethyl sulphoxide (DMSO), some of them differentiated to form apparently normal erythrocytes. Without this treatment, they did not differentiate and expressed a type of leukaemia. We established that many other polar organic solvents had similar effects, but the concentrations required were rather high (Tanaka et al 1975). I thought that, with luck, we might be able to find a simple way to improve the efficacy of such compounds.

There are many reasons why a high concentration of a polar compound might be needed to induce differentiation; one possible reason is that more than one such molecule has to bind. If so, a high concentration would be needed to increase the probability that two or more solvent molecules bind at the same time. Although we had no evidence that this was the case, we decided to design some molecules that might take advantage of this hypothetical requirement for multiple binding. We made molecules with *two* polar solvent units linked by a chain. If both solvent units have to bind for the biological effect to occur, and if the chain is the correct length to permit this double binding, such compounds might be particularly effective.

We were very lucky. Our first such compound (Fig. 5), hexamethylene-bis-acetamide (HMBA), was considerably more effective than were any of the monomeric polar solvents (Reuben et al 1976, 1978). A related compound with the amide groups turned around, suberic acid bis-dimethylamide (SBDA) (Fig. 5), was almost identical in activity. Since that time many biological studies have been performed on these early ditopic cytodifferentiating agents (Marks et al 1989). They have been shown to induce the differentiation of many cancer cell types in tissue culture, producing cells with normal behaviour. On the basis of this work HMBA was put into Phase I clinical trials a few years ago and is about to enter Phase II trials. There have been one or two remarkable reversals of diagnosed terminal cancer in patients, with one patient with non-small-cell lung cancer now apparently functioning normally five years after she had been diagnosed as a hopeless case. However, it is much too early to know whether these compounds will be generally useful in cancer treatment.

After this early success we spent many years trying to create drugs with much better activity. In humans, doses of HMBA of about 20 grams per day are required to achieve the blood level (5 mM or so) that we believe is needed for effectiveness; this is an annoyingly large dose but is not impossible to administer.

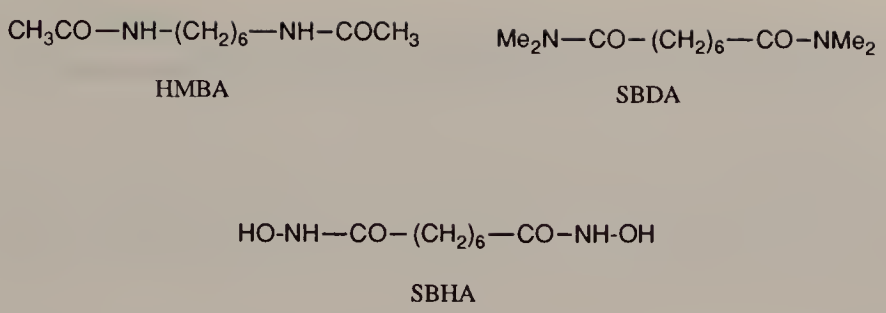


FIG. 5. Three cytodifferentiating agents that utilize double interaction with a biological receptor. HMBA, hexamethylene-bis-acetamide; SBDA, suberic acid bis-dimethylamide; SBHA, suberic bis-hydroxamic acid.

Many structural variations were tried until finally chemical reasoning was brought to bear. The dependence of effectiveness on chemical structure indicated strongly that the two polar groups of HMBA are binding to a receptor of some sort, not just sitting in water solvent. Thus we decided to create stronger binding, even though we did not know the exact nature of the receptor.

A polar amide group must be binding either to a metal ion or to hydrogen-bonding groups. In either case, an amide is not the optimal functionality for such binding. Hydroxamic acids are *N*-hydroxyamides; they are much better ligands to metal ions, and can make more hydrogen bonds, and perhaps stronger ones, than a simple amide can make. Our first example was suberic bis-hydroxyamic acid (SBHA) (Fig. 5), and it was very effective. In our cytodifferentiation assay it was over 100 times more active than HMBA and other bis-hydroxamic acids are even better. The dose of such compounds needed in humans should be only a few milligrams per day, rather than the grams per day of HMBA. Furthermore, the compounds do not seem to have any unusual toxicity.

Much additional work, now under way, will be needed to see whether these new compounds are truly promising leads as cytodifferentiating agents for human treatment. However, from the structural requirements for activity, including the geometries imposed in rigid bis-hydroxamic acids, we already know a lot about the nature of the binding sites in the receptors for these molecules. In this case, as well as in the binding and catalysis studies described in the first part of this paper, two has proved to be better than one, provided the geometries are correct.

In this real-life drug design problem we have had to fit the geometry of the drug to the unknown geometry of a still unidentified receptor molecule, but this simply meant that a little more cutting and fitting was required to optimize the system. In drug design one hardly ever knows the nature and shape of the host when trying to design an appropriate drug guest. Chemical intuition, and some luck, can overcome this seemingly daunting problem. Now the tailored guests can be used to learn more about the host, and about the process of

cytodifferentiation that is central not just to a new approach to cancer chemotherapy, but to understanding normal cell development as well.

Acknowledgements

The work described here has been supported by the National Institutes of Health and the Office of Naval Research. I wish to thank my collaborators and co-workers, who are named in the references.

References

Anslyn E, Breslow R 1989a On the mechanism of catalysis by ribonuclease: cleavage and isomerization of the dinucleotide UpU by imidazole buffers. J Am Chem Soc 111:4473–4482

Anslyn E, Breslow R 1989b Geometric evidence on the ribonuclease model mechanism. J Am Chem Soc 111:5972–5973

Anslyn E, Breslow R 1989c Proton inventory of a bifunctional ribonuclease model. J Am Chem Soc 111:8931–8932

Breslow R, Chin S 1990 Strong binding of ditopic substrates by a doubly linked occlusive C_1 'clamshell' as distinguished from an aversive C_2 'loveseat' cyclodextrin dimer. J Am Chem Soc 112:9659–9660

Breslow R, Labelle M 1986 Sequential general base-acid catalysis in the hydrolysis of RNA by imidazole. J Am Chem Soc 108:2655–2659

Breslow R, Doherty JB, Guillot G, Lipsey C 1978 β-Cyclodextrinylbisimidazole, a model for ribonuclease. J Am Chem Soc 100:3227–3229

Breslow R, Czarnik AW, Lauer M, Leppkes R, Winkler J, Zimmerman S 1986 Mimics of transaminase enzymes. J Am Chem Soc 108:1969–1978

Breslow R, Greenspoon N, Guo T, Zarzycki R 1989a Very strong binding of appropriate substrates by cyclodextrin dimers. J Am Chem Soc 111:8296–8297

Breslow R, Huang D-L, Anslyn E 1989b On the mechanism of action of ribonucleases: dinucleotide cleavage catalyzed by imidazole and Zn^{2+}. Proc Natl Acad Sci USA 86:1746–1750

Marks PA, Breslow R, Rifkind RA, Ngo L, Singh R 1989 Polar/apolar chemical inducers of differentiation of transformed cells: strategies to improve therapeutic potential. Proc Natl Acad Sci USA 86:6358–6362

Reuben RC, Wife RL, Breslow R, Rifkind RA, Marks PA 1976 A new group of potent inducers of differentiation in murine erythroleukemia cells. Proc Natl Acad Sci USA 73:862–866

Reuben RC, Khanna PL, Gazitt Y, Breslow R, Rifkind RA, Marks PA 1978 Inducers of erythroleukemic differentiation; relationship of structure to activity among planar-polar compounds. J Biol Chem 253:4214–4218

Tanaka M, Levy J, Terada M, Breslow R, Rifkind RA, Marks PA 1975 Induction of erythroid differentiation in murine virus infected erythroleukemia cells by highly polar compounds. Proc Natl Acad Sci USA 72:1003–1007

DISCUSSION

Rebek: Do you ever see any unusual kinetics of binding of substrates to these bis-cyclodextrins, such as a rapid rate for one end of the substrate and a slower rate for the other end?

Ron Breslow: We haven't seen that yet. Our systems have flexibility in them and I don't think we will see slow binding. If we bridged them *meta* or *para*, that is, A–C or A–D, so they were 'locked' closed, and if we put polar groups at the ends of the substrate so that it had trouble threading through, we might well see some interesting kinetic effects. Normally, binding of materials to cyclodextrins is not on the stopped-flow scale, but on the pulse-relaxation scale, so you would have to do fancy experiments to see such kinetic effects. If we made the host more rigid and imposed some kinetic barriers, we might see the kind of effect you are talking about. There are certainly complexes formed that have the wrong geometry as well as those with the right geometry, but all this happens quickly.

Rebek: Have you done bimolecular reactions? You could take a phenylalanine bound to one cyclodextrin and an activated phenylalanine bound to the other part of the dimer and try to catalyse peptide formation. Presumably the peptide product would have a geometry different from that which is optimal for the starting materials; it would dissociate and the system could turn over.

Ron Breslow: The advantage of fragmentation reactions is that the products bind more weakly than the starting material, so you regenerate the catalyst. Synthesis is more interesting for chemists, but the disadvantage of synthesis is that the product often binds more strongly than the starting material, leading to product inhibition. We have not done a study on this in this case, although peptide synthesis is one of the things that the metal ions are there for; the metal ion is present not only to cleave an amide, but also to make an amide. We expect that unless we do some interesting geometry, we will not solve the problem of product inhibition. One of the reasons that we built this molecule with a torque in it is that if we are lucky, the product amide, which wants to be linear, will dissociate easily because it won't fit well in the twisted cavity. We did Diels–Alder reactions in cyclodextrin cavities with a diene and dienophile which both fit into the cavity and were essentially pushed together (Rideout & Breslow 1980). The problem was that the product was strongly bound, so we didn't see turnover. People working with catalytic antibodies picked up on this and also did Diels–Alder reactions. There were two different tricks that were used; my former student Don Hilvert used a diene in which the Diels–Alder product underwent a spontaneous fragmentation and changed its geometry so that it could dissociate. Peter Schultz used a somewhat different technique, but in both cases the problem of product inhibition had to be addressed.

Stoddart: If you do an asymmetry analysis of the cyclodextrin dimers you find that the 'clam shell' arrangement is an asymmetrical one. The 'love seat' arrangement has two-fold symmetry. The implications are that the two binding sites differ from one another in the clam shell, whereas in the love seat arrangement the two binding sites are the same. Is the rate of exchange of a substrate slow enough to allow you to identify the two different sites in the clam shell by NMR? Could you bind different substrates in these two different sites for subsequent studies of enzyme mimicry?

Ron Breslow: We have not detected *that* asymmetry, although we know it's there. The A–B and B–A arrangements are essentially diastereoisomers because there is also a chirality in the glucose residue, so going clockwise and anticlockwise around these cavities is not the same. That means that in the clam shell, each link is from an A to a B, whereas in the love seat there's an A–A and a B–B linkage; that essentially gives the extra asymmetry that you referred to.

Stoddart: In the clam shell every proton must be in a different environment, whereas there is a symmetrical matching of pairs of protons in the love seat.

Ron Breslow: In NMR spectra we haven't yet seen all the expected chirality. We do see the chirality of the bridging groups themselves; in one arrangement (clam shell) the C-2 protons of the two imidazolium bridging groups have the same chemical shift, but they do not in the other arrangement. We haven't yet seen any difference by NMR in the two halves of a bound substrate, although it should be present. The most interesting thing we might be able to do is to bind two different substrates or to make the two linkages differ in such a way that one side could interact with something while the other cannot.

Stoddart: The linkages could be different, but they don't need to be different in the love seat to provide two different binding cavities for substrates.

Ron Breslow: There's a difference of chirality in principle, but whether that's chemically useful is another question. There isn't much chiral distinction in simple binding to cyclodextrins even though the cavity is made up of chiral glucoses. If chirality isn't expressed well in the monomer, it won't be expressed well in the dimer either.

We have also made dimers of synthetic cavities which don't have this problem of chirality. Some of the synthetic cavities have higher binding constants than cyclodextrin.

Crumpton: You said that the binding constants for the cyclodextrin complexes are not as high as those for antibody–antigen complexes. One could extend the comparison with antibodies to make the point that a bivalent interaction gives an increase in affinity over monovalent interaction. As you have said, if you increase the valency further, you should get additional increases in affinity of interaction. Of course, in the case of antibodies there's a very good example of that, namely multivalent IgM compared with bivalent IgG. Why not have a cyclodextrin trimer or tetramer? To a biologist, that seems an obvious way to increase affinity of interaction.

Ron Breslow: This is an interesting point. You cannot join three cyclodextrins around a central point so that they are close to one another (or to the central point). You could put cyclodextrins in a line to make a tetramer in a string; then the binding sites would be rather far apart, though they could fold around a substrate. An alternative, which I think is perhaps a simpler way to achieve what you are invoking, is not to use only hydrophobic binding—you eventually have to make some peculiar substrates to keep all the interactions hydrophobic. Our system of two cyclodextrins with a metal ion in the middle ought to have

very high affinity for a substrate with two hydrophobic binding groups and an ionic binding ligand or a metal binding ligand in the middle. Catalytic studies show that it will cleave phosphate esters, but we don't yet know the binding constant. I think we could reach very high binding constants more easily by having a third binding interaction that could be ionic or metal coordination.

Sutherland: Anderson & Sanders (1989) have described a tris-metalloporphyrin in which there are three binding sites for 2,4,6-tri-4-pyridyl-*s*-triazine; two of them functioned well, but the third didn't quite match the requirements of the guest ligand.

Crumpton: Isn't it possible that this is a reflection of steric considerations, rather than lack of matching with the guest ligand?

Ron Breslow: There's no question that three is better than two, and that four is better still, but you have to construct the molecules without leaving so much flexibility that it costs you too much entropy. As I said, some flexibility leads to slightly decreased affinity; if there is too much flexibility the molecule can collapse into a totally inactive conformation. When we had enough flexibility to give us different conformations, some of which were active, binding affinity was not as good. We didn't get affinity of 10^{10} until we introduced the second binding interaction and 'locked' the conformation to lose the rotational motion.

Crumpton: Many people have tried to induce cellular differentiation using drugs and to model antitumour drugs on known inducers of differentiation. Some of these work well in both *in vivo* and *in vitro* systems, but, surely, the situation is more complicated than simply inducing differentiation to an end cell that doesn't undergo further division. Nevertheless, it's a good idea, but we are a long way from achieving success in a general sense. Isn't the fundamental problem the complexity of the induction of tumorigenicity, which requires a number of separate and different genetic changes?

Dimerization of your primordial drug gave an increased binding. While this is indeed true, I am not convinced that the increased binding is *solely* the basis for the drug's increased effectiveness. In many receptor systems dimerization is a prerequisite for induction of a biological response. This is true for extracellular receptors where cross-linkage of the receptor initiates the biological response, and for intracellular receptors such as the oestrogen receptors, where dimerization provides the mechanism by which the biological response is induced. Although you are seeing an increased affinity of binding to a receptor (whatever that receptor is), cross-linkage of the receptor, as a consequence of using a dimeric ligand, is likely to be the inducing biological effect.

Ron Breslow: We have considered that. According to that argument, one receptor binds to each end of the ligand and then something is cross-linked, such as two proteins on the surface. The argument against that is that the dimeric

compounds are not all that much more effective than monomeric amides, which cannot cross-link receptors. The dimeric amides were about 100-fold more effective than the monomer. Also, the cytodifferentiating ability is sensitive to chain length and to the nature of the chain. A chain length of six carbons is better than five or seven. If only cross-linkage was required, why would the response be so sensitive to chain length? Surely, linking would be sufficient and you wouldn't need any 'fine tuning'. I think the strongest argument is that the dimer shows the same behaviour as monomeric material, which cannot cross-link proteins.

Crumpton: I don't agree that the monomers cannot effect cross-linkage. In the direct sense they obviously cannot, because they are monovalent. However, theoretically a monomeric ligand can, on binding, induce a conformational change in the receptor that promotes receptor dimerization. It appears that oestrogen acts in this way.

Ron Breslow: That's not cross-linking by the drug, it's changing the geometry of the protein so as to induce a new geometry which is active; that does not require a physical linkage.

Your other point was whether differentiating agents are generally useful—most aren't, but we have never found a solid tumour in culture that didn't respond to HMBA. In Phase I clinical trials HMBA has had only limited exposure. We have tested it in a series of patients with leukaemia and some responded. We have also used it in some patients with solid tumours, the most strikingly successful of which I described to you. I don't want to claim that we always get success with this compound; on the other hand, the clinicians tell me that in Phase I, many of the patients are fundamentally moribund.

Crumpton: That's a common argument.

Ron Breslow: The induction of differentiation is not instantaneous. It takes a few months to reach the full response and many of these people don't live for more than a couple of weeks after the treatment begins. I'm not claiming that this drug is the magic bullet, but there is no evidence to say that it isn't going to be useful. We now have a more effective compound, which I can't tell you about for legal reasons, that has a general effect. There are two classes of these drugs. Some bis-hydroxamic acids seem to have general effectiveness, whereas others are quite specific. There may be two kinds of receptor site, one general, found everywhere, and the other specific to the mouse. We can tune the geometry of these compounds so that they will cure only a mouse, not a human. Perhaps we could do it in the other direction. We can alter effectiveness and selectivity by changing geometry. We are putting some of these more effective compounds into human clinical trials as fast as we can.

Crumpton: There are other compounds, like retinoids for example, which induce cellular differentiation *in vitro*, but which do not appear to be particularly effective *in vivo*. Thus, those who have a high retinoid intake don't have fewer tumours than those with a low intake (Wald 1987).

References

Anderson HL, Sanders JKM 1989 Synthesis of a cyclic porphyrin trimer with a semi-rigid cavity. J Chem Soc Chem Commun 1714–1715

Rideout D, Breslow R 1980 Hydrophobic acceleration of Diels-Alder reactions. J Am Chem Soc 102:7816–7818

Wald NJ 1987 Retinol beta-carotene and cancer. Cancer Surv 6:635–651

Binding of antibiotics to DNA

Michael Waring

Department of Pharmacology, University of Cambridge, Tennis Court Road, Cambridge, CB2 1QJ, UK

Abstract. The DNA molecule can serve as host to numerous guest ligands, some of which are antibiotics, and almost all of which are endowed with anticancer or antimicrobial activity. Many guest ligands are quite large and complex in structure, and an array of intermolecular contacts underlie their complementarity to their macromolecular receptor. Often the process of molecular recognition involves conformational adjustments on the part of the interacting species, but the lion's share of the adjusting is demanded of the DNA helix which commonly ends up considerably distorted. Generally the lock must bend to accommodate the key. Two fundamentally different modes of binding can be identified: intercalation and minor groove-binding. The former mode is exemplified by daunomycin or actinomycin and the latter by netropsin or distamycin. Intercalation is associated with substantial extension and unwinding of the helix whereas groove binding is characterized by replacement of the spine of hydration and by lesser effects on helix geometry such as local bending. Bifunctional (or bis-) intercalation, as seen with echinomycin, causes the most far-reaching perturbations in the structure of DNA. It might even involve altered base pairing, which has been observed in complexes between echinomycin and oligonucleotides but has not yet been detected with macromolecular DNA in solution.

1991 Host–guest molecular interactions: from chemistry to biology. Wiley, Chichester (Ciba Foundation Symposium 158) p 128–146

Drugs which bind to DNA so as to distort its structure and disrupt its function have long been recognized as important tools for chemotherapy (Fig. 1). They include a number of useful antimicrobials, particularly antiprotozoal drugs, as well as the majority of the agents in current use for the treatment of cancer. Among the latter are several antibiotics—substances of natural origin—with which this chapter will be chiefly concerned. They are extremely diverse in chemical structure (Fig. 2). This diversity reflects the variety of the means by which they interact with DNA, but, thanks to recent advances in chemistry and molecular biology, we are well placed to identify common motifs in drug–DNA recognition processes. Using this knowledge we can make sensible predictions that will guide research. The impetus to pursue this work comes largely from the well-founded belief that a rational approach to chemotherapy and drug design will emerge from our efforts to exploit the immense specificity inherent in the structure of DNA as it relates to such matters as the regulation of gene expression.

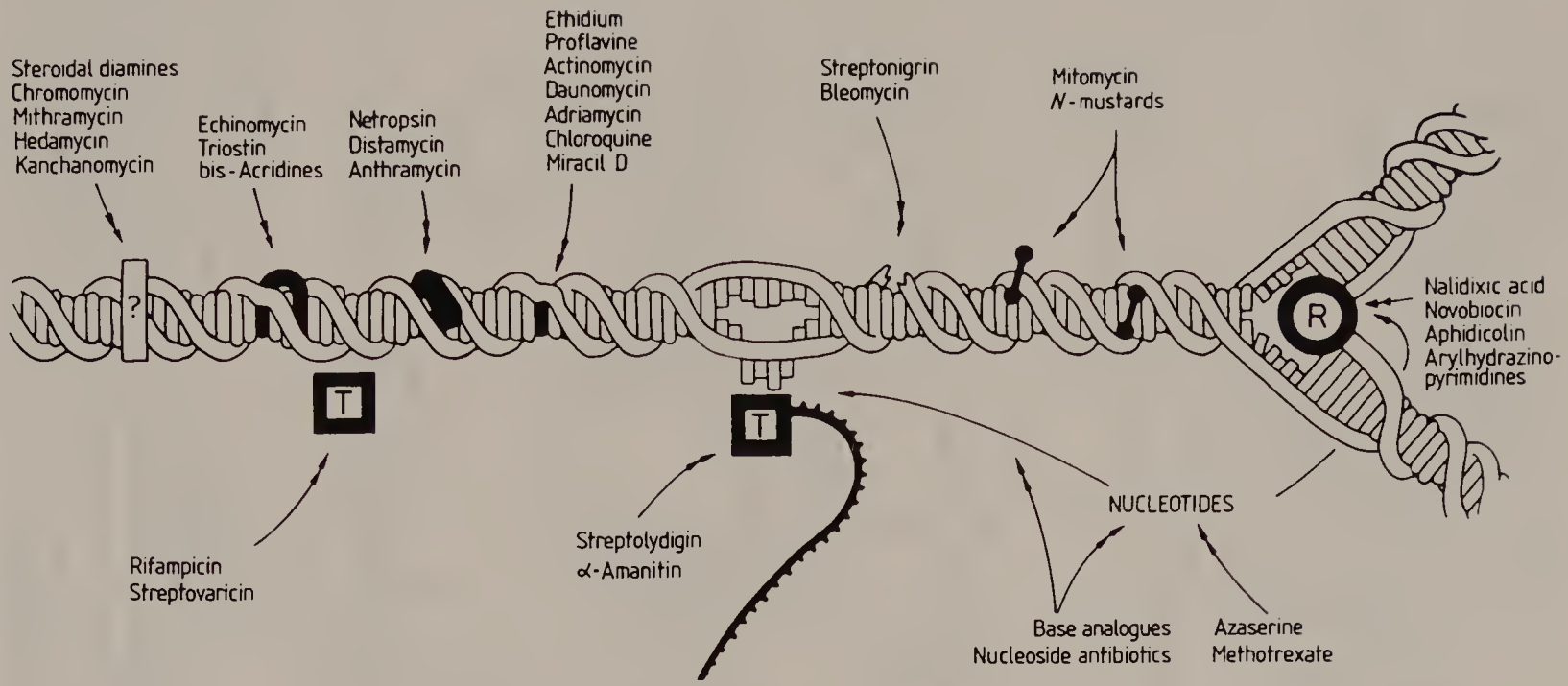


FIG. 1. Sites of action of antibiotics and drugs that inhibit nucleic acid synthesis. The diagram shows a portion of a double-helical DNA molecule in process of replication at the right-hand end and being transcribed into RNA at a region near the middle. R represents the replicating enzyme, T the transcribing enzyme (RNA polymerase). Towards the left-hand end an RNA polymerase molecule is about to form an initiation complex with the DNA and start the synthesis of a new RNA strand. Actions of inhibitors are represented by double-headed arrows and are purely diagrammatic; it is not intended to imply that the sites of action represented here account for the whole mode of action of any drug. (From Gale et al 1981.)

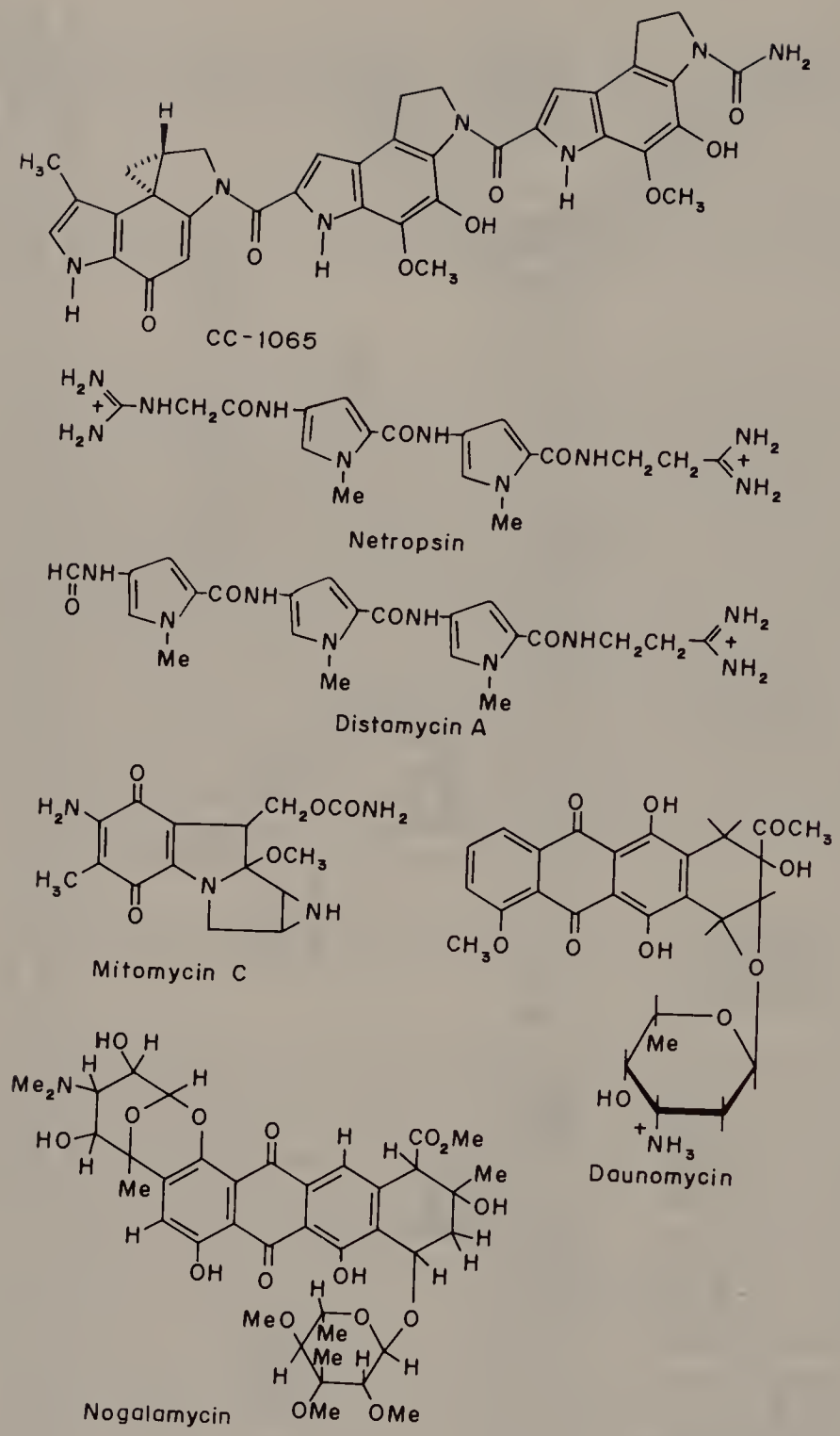


FIG. 2. (*see legend on facing page*)

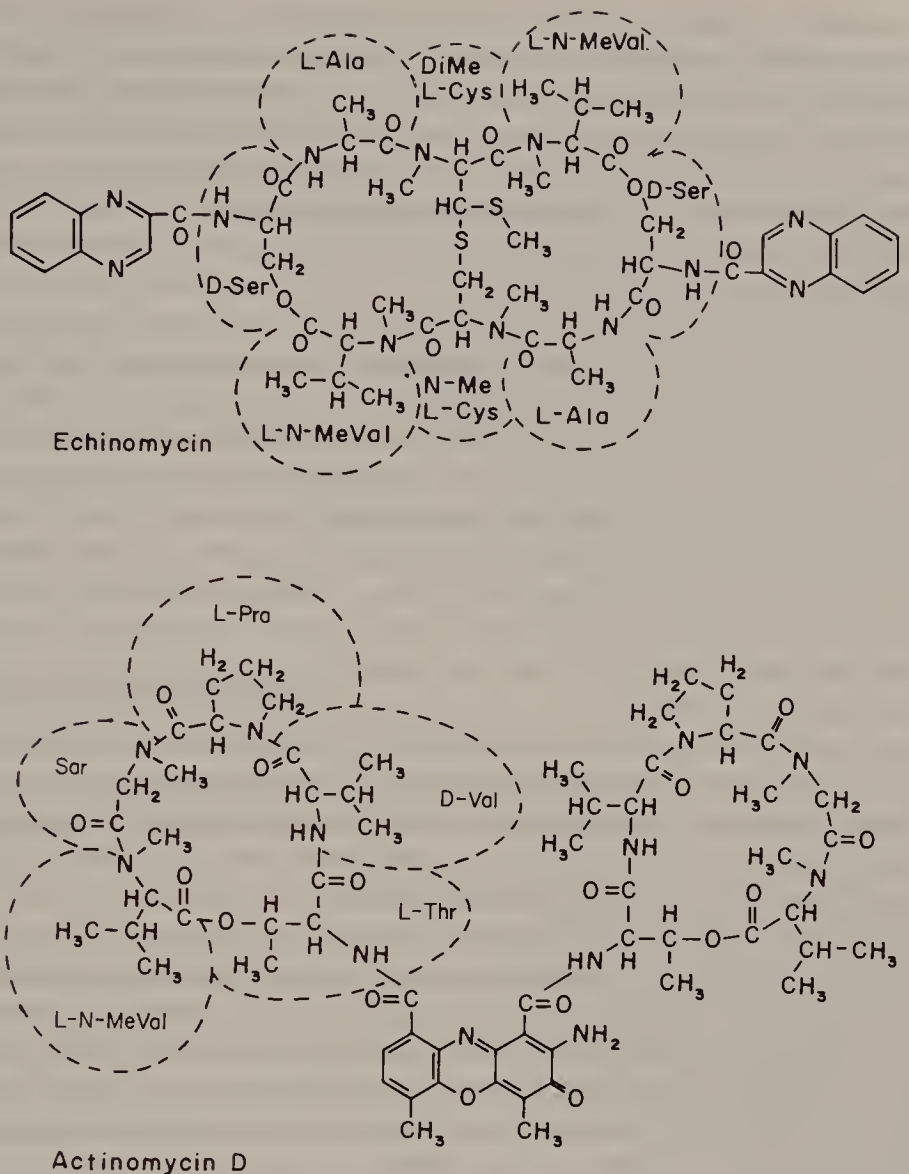


FIG. 2. Structural formulae of antibiotics which bind to DNA.

Modes of binding to DNA

Outstanding among techniques which have shed light on the exact details of antibiotic–DNA recognition is X-ray crystallography. Its successful application is evident in the rash of publications on the structure of antibiotic–oligonucleotide complexes that have appeared during the past ten years (Wang 1987, Kopka et al 1987). But the identification of different modes of binding to DNA and the broad outlines of the forces which drive them originate from ten or twenty

years of earlier research using techniques drawn from a range of scientific disciplines. Broadly speaking, we can identify four principal modes of binding of small ligands to DNA: electrostatic interaction, covalent bonding, intercalation and reversible groove-binding (almost always in the minor groove of the B-form DNA helix). Each is associated with some degree of perturbation of DNA structure, and the nature of that perturbation typically leads to the assignment of criteria by which the mode can be identified. That in turn prescribes the diagnostic techniques to be used.

Purely electrostatic binding is rare, and probably never happens with antibiotics. Its nearest equivalent is probably the formation of self-stacked aggregates of positively charged acridines in the vicinity of a polyanion, which need not be DNA (Armstrong et al 1970, Gale et al 1981). Covalent bonding, on the other hand, is now known to be at the bottom of the action of several antitumour antibiotics of exquisite potency and represents a major area of interest for the development of novel anticancer drugs (Warpehoski & Hurley 1988). (It is also, paradoxically, at the bottom of the cancer-producing process initiated by a range of powerful carcinogens such as the polycyclic aromatic hydrocarbons and aflatoxins.) A sterling example of such novel anticancer drugs is the antibiotic CC-1065 (Fig. 2), which reacts highly selectively with the N-3 of adenine residues in AT-rich tracts of DNA, forming a strong complex in the minor groove of the helix (Warpehoski & Hurley 1988). Another is the antibiotic mitomycin C, which reacts with the 2-amino groups of neighbouring guanine residues so as to cross-link the antiparallel strands of DNA, again in the minor groove of the helix (Tomasz 1990).

The two other modes of interaction with DNA, namely intercalation and groove-binding, together constitute the major modes of interaction, which have led to an explosion of research on molecular interactions. Interest in intercalation dates from the early 1960s when it was first postulated as an explanation for the action of acridines and ethidium (reviewed by Gale et al 1981). The first antibiotics identified as intercalators were the anthracyclines daunomycin and nogalamycin, soon followed by actinomycin. A cardinal feature of the intercalation model is the local unwinding and extension of the DNA helix that is required to permit sandwiching of the intercalative ligand chromophore between the base pairs. These two characteristic distortions have become accepted as the diagnostic criteria for intercalation and they can be measured by hydrodynamic experiments (such as viscosity measurements) using closed circular duplex DNA or sonicated rod-like DNA fragments respectively (Gale et al 1981). Conversely, antibiotics capable of binding tightly to DNA without eliciting the phenomena attributable to helix unwinding and extension are identified *prima facie* as non-intercalators. They have all turned out to be groove-binders and to lodge in the minor groove of the helix. Examples include the *N*-methylpyrrole peptides netropsin and distamycin. Structural distortions associated with their binding to DNA are minor compared to those associated

with the intercalators; there may be a tendency for the minor groove to close up around the entrapped antibiotic and some bending of the helix axis may ensue, but the conformational adjustments are much more equally shared between the ligand and the receptor than is the case with the intercalators, where the lion's share of the adjusting is manifestly on the part of the DNA.

On the whole, the distinction between intercalators and groove-binders is clear so far as antibiotics are concerned, but that is by no means so for small synthetic DNA-binding molecules. Recent evidence indicates that some ligands can be both intercalators and groove binders; for example, the fluorescent chromosome stain diamidino-2-phenylindole (DAPI) appears to intercalate adjacent to G·C base pairs but to bind in the minor groove at AT-containing clusters (Wilson et al 1990). More mysterious are such oddities as steroidal diamines (Waring & Henley 1975, Gale et al 1981) or triphenylmethane dyes like crystal violet (Wakelin et al 1981), which clearly unwind the DNA helix but cannot intercalate in the normal sense. Much remains to be learned about non-classical binding to DNA by small molecules.

Conformational adjustments

As well as the macroscopic changes in DNA structure caused by intercalation, many microscopic changes can be provoked by binding of antibiotics, frequently in a markedly sequence-selective fashion. These microscopic changes reveal how exquisitely sensitively the DNA molecule can adapt itself to accommodate a small molecule—indeed, how intrinsically flexible the double-helical structure of DNA is. Sugar–phosphate backbone torsion angles are readily susceptible to deformation within allowed ranges, leading to altered sugar puckering as well as displacement of phosphate groups; Watson–Crick base pairs can buckle, shear, propeller twist, tilt and roll. Exact details of these antibiotic-induced distortions have emerged from a series of elegant crystallographic studies on antibiotic–oligonucleotide complexes, backed up by NMR investigations which have used the increased sophistication of two-dimensional methods to confirm (and not infrequently to discover) the existence of conformational changes in solution. Among intercalators the anthracyclines have proved particularly informative (Searle et al 1988, Williams et al 1990, Frederick et al 1990). Among non-intercalators, the narrow groove-binders netropsin and distamycin have been the most thoroughly investigated (Kopka et al 1985, 1987, Coll et al 1987, 1989).

The distamycin–oligonucleotide and netropsin-oligonucleotide complexes, as expected, share many common features (Fig. 3). The antibiotic molecules adopt a crescent shape that is not very different from their conformation when free in solution, and lodge neatly in the minor groove of the DNA helix at AT-rich sites, where they effectively replace the natural spine of hydration and form bifurcated hydrogen bonds simultaneously with adenine N-3 and thymine O-2

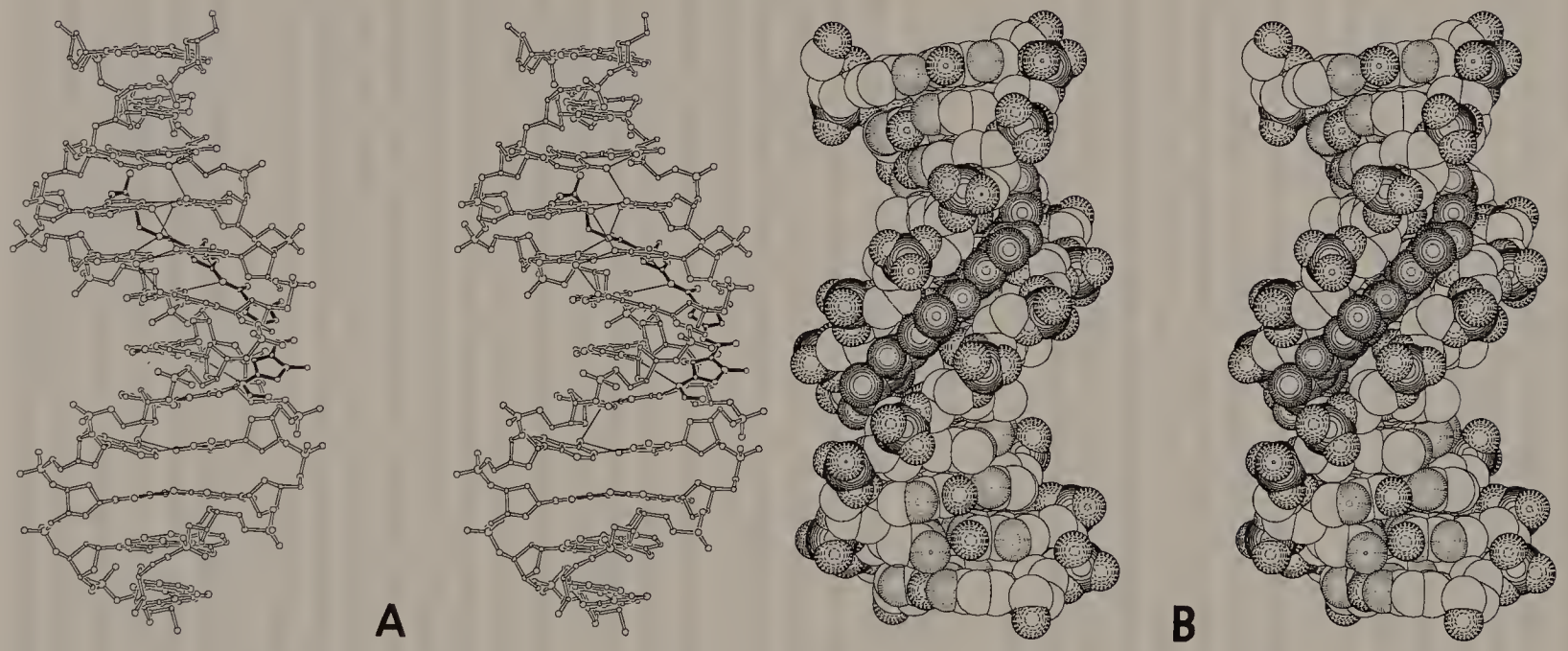


FIG. 3. Stereo diagrams of a 1 : 1 complex between distamycin and the self-complementary dodecanucleotide d(CGCAAATTTGCG) which forms a mini-DNA helix of twelve base pairs. (A) Skeletal diagram: the bonds of distamycin are solid, whereas those of the DNA are open. The distamycin molecule can be seen wrapping around the side and continuing onto the back of the DNA helix. Bifurcated hydrogen bonds that connect the antibiotic molecule to the narrow minor groove are shown as thin lines. Likewise, bifurcated interactions in the major groove due to the propeller twisting of the base pairs are also drawn as thin lines. The propeller twisting of the base pairs is readily evident in this diagram. (B) Van der Waals diagram: the complex is rotated 90° relative to A. All atoms of distamycin are shaded with heavy black circles, whereas in the DNA nitrogens are drawn as dotted spheres, oxygens as dashed spheres, phosphorus as concentric circles, and carbons are unshaded. Illustration from Coll et al (1987) kindly provided by Dr Loren Williams.

atoms. In the process they nestle up closely to the floor and walls of the groove, making numerous van der Waals contacts that serve to stabilize the whole structure and glue the two strands of the double helix together. Bending of the helix can thereby be strongly affected, as happens with kinetoplast DNA (Wu & Crothers 1984). Most interestingly, it has recently been discovered that it is also possible for *two* distamycin molecules to bind simultaneously in the minor groove of oligonucleotides containing runs of five or six A·T base pairs (Pelton & Wemmer 1989, 1990). The two antibiotic molecules lie side-by-side in antiparallel orientation, making numerous electrostatic hydrogen-bonding and stacking interactions with each other and with the walls of the minor groove, which is significantly expanded by comparison with the 1 : 1 antibiotic : helix complex. Thus, the flexibility of DNA structure is neatly underlined by the characteristics of the various distamycin complexes, in which the width of the minor groove can be narrowed to around 3.4 Å to encompass a single antibiotic molecule or expanded to at least 6.8 Å to accommodate a 'stack' of two.

The overriding message of the intercalator–oligonucleotide complexes is one of conformational adaptation. Depending upon the exact orientation of the intercalated chromophore, which may vary from a 'spearing' mode, in which its long axis is nearly perpendicular to the adjacent base pairs, to one in which it is practically parallel to them, the base pairs may be buckled, twisted, or flattened to optimize van der Waals contacts (alias stacking interactions). Appended substituents, such as the sugars attached to anthracyclines (Fig. 2), will tend to pack more or less neatly into the groove(s) so as to provide additional stabilization through more van der Waals interactions, hydrogen bonds and electrostatic forces. Sensitive responses of the receptor molecule are everywhere in evidence and are well exemplified by detailed structural comparison of adriamycin- and daunomycin–oligonucleotide complexes (Frederick et al 1990) in which subtle effects attributable to the presence or absence of a hydroxyl group on the antibiotic appear magnified, and may be related to the significantly different anticancer clinical activities of these important drugs.

Understandably, structural distortions of DNA provoked by antibiotic binding have far-reaching consequences for nucleoprotein interactions and biological effects. Thus, the rotational orientation of DNA on the surface of nucleosome core particles is disturbed by antibiotic binding (Portugal & Waring 1987), and interference by antibiotics with the operation of topoisomerase II leads to the appearance of strand breaks (Hertzberg 1990). These observations are likely to be highly relevant to piecing together the chain of events that lead to the killing of cancer cells by antibiotics.

Bifunctional and hybrid molecules

In parallel with the explosive growth of knowledge about mechanisms of drug–DNA interaction there have been intense efforts on the part of chemists

and biochemists to construct all manner of structurally complex molecules, often directly exploiting themes perceived with the antibiotics referred to above, with the goal of enhancing the selectivity of ligand–DNA interaction. The concept of 'isohelical fit' to the minor groove of DNA, prompted by the discoveries made with netropsin and similar molecules, with or without recognition of the critical 2-amino group of guanine, has inspired the search for further 'lexitropsins' (Kopka et al 1987). Hybrid molecules consisting of linked moieties (often antibiotic derivatives) known to have differing binding properties have also been made, and a number of them do indeed bind selectively to particular DNA nucleotide sequences. Exciting developments along similar lines, employing oligonucleotides to recognize biologically important sequences in DNA, are described by Hélène and co-workers (Montenay–Garestier et al 1990, this volume).

Among true antibiotics, the quinoxaline depsipeptides echinomycin (Fig. 2) and its disulphide cross-bridged precursor triostin A have served to provide a paradigm for bifunctional interaction with DNA—in this case bis-intercalation. These antibiotics were the first bis-intercalators to be discovered (Waring & Wakelin 1974, Waring 1979, 1990). They have a rigid, staple-like structure (Fig. 4)

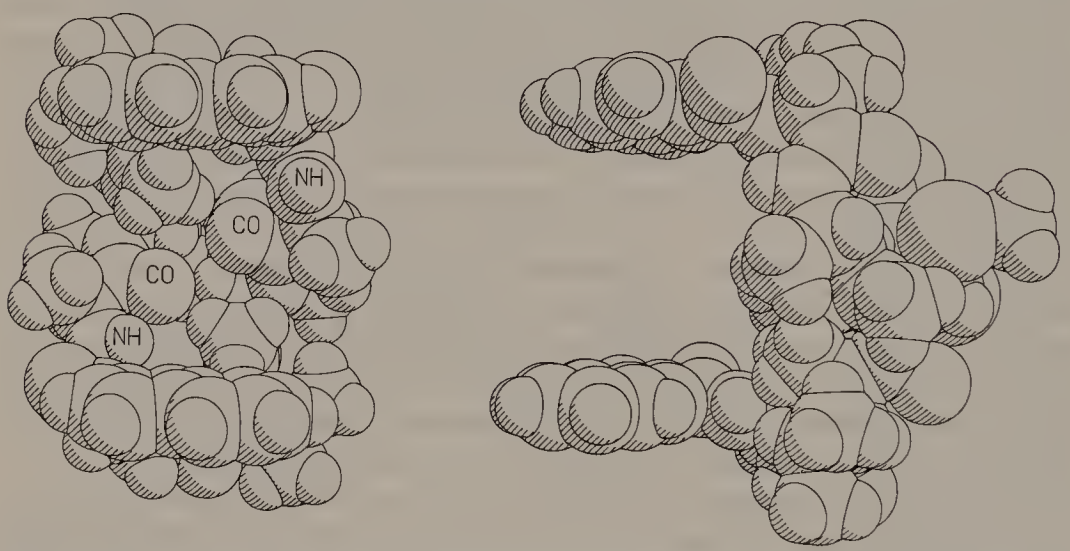


FIG. 4. Crystal structure of a bis-quinoline analogue of echinomycin (2QN) prepared by directed biosynthesis (Gauvreau & Waring 1984). In the illustration on the left the molecule is viewed down its quasi-dyad axis with the aromatic chromophores projecting out towards the viewer; this reveals the face of the peptide ring presented towards the DNA base pairs with the NH and CO groups of the two L-alanine residues lying in a diagonal array from lower left to upper right. These are the groups which mediate the recognition of G·C base pairs via hydrogen bonding to the 2-amino and N-3 of the guanine bases. In the illustration on the right the molecule is viewed from the 'side', having been turned through 90° about a vertical axis, so as to show its staple-like arrangement which is ideally suited for bis-intercalation. Small rotations have been applied to the points of attachment of the aromatic rings so as to bring their planes exactly parallel. The illustration is based on unpublished work of G. M. Sheldrick, P. G. Jones, E. F. Paulus & M. J. Waring. For a description of the comparable crystal structure of triostin A see Sheldrick et al 1984.

which enables them to recognize the CpG dinucleotide sequence in DNA. They bis-intercalate so as to sandwich two G·C base pairs between their quinoxaline chromophores, in the process unwinding the helix by about 48° and extending it by nearly 6.8 Å, nominally the theoretical maximum for perfect stacking between chromophores and base pairs. As the crystallographic results reveal, however, the stacking is by no means perfect and the DNA structure undergoes numerous modifications to permit the formation of a large number of van der Waals contacts (21 in an echinomycin–hexanucleotide complex) which are mainly responsible for stabilizing the structure (Wang et al 1984, Ughetto et al 1985, Quigley et al 1986, Wang 1987). But the most dramatic modification is the rearrangement of both A·T and G·C base pairs flanking the bound antibiotic from Watson–Crick to Hoogsteen form; this requires rotation of the purine nucleosides to the unnatural *syn* conformation as well as protonation of the cytosine bases (Fig. 5).

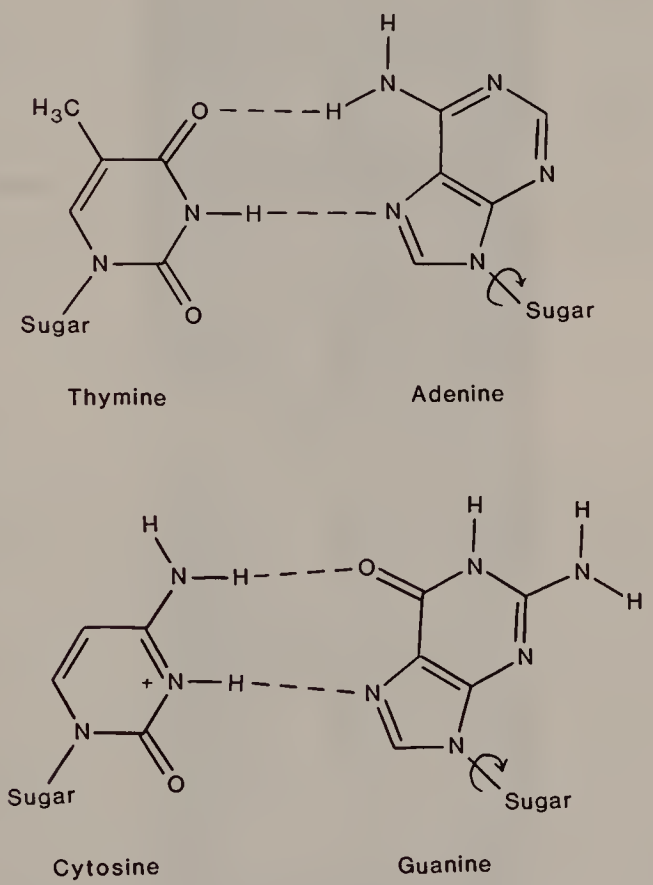
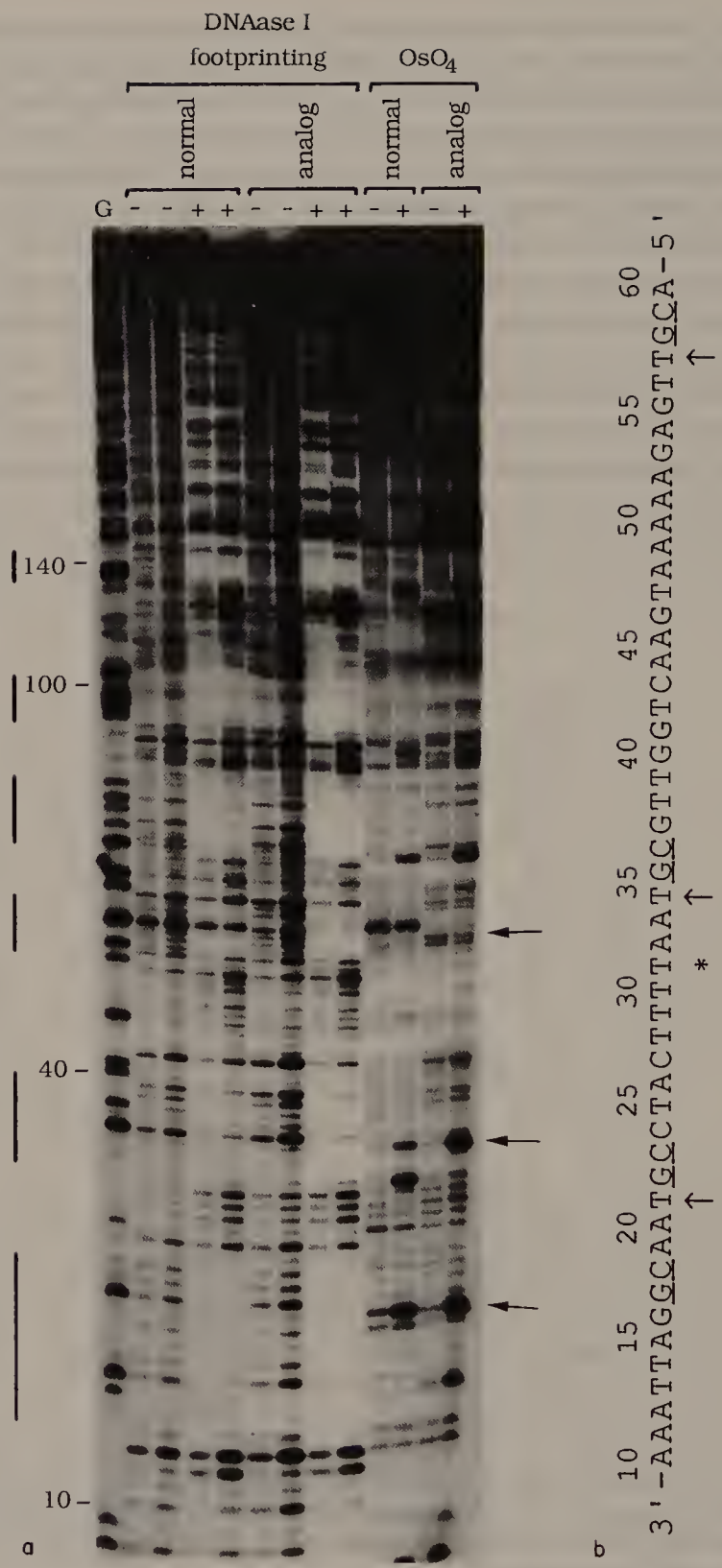


FIG. 5. Purine–pyrimidine base pairing according to the Hoogsteen scheme. The placement of the pyrimidines is much the same as in the Watson–Crick scheme, but the purines are turned over. Note that the C·G Hoogsteen base pairing requires protonation of the cytosine at N-3.


DNAase I
footprinting
OsO4
normal
analog
normal
analog
G
- - + + - - + + - + - +
140 –
100 –
40 –
10 –
a
b
3'-AAATTAGGCAATGCCTACTTTTAATGCGTTGGTCAAGTAAAAAGAGTTGCA-5'
10 15 20 25 30 35 40 45 50 55 60
*

Careful NMR experiments have revealed that echinomycin can, in certain circumstances, cause nucleotide bases to pair according to the Hoogsteen scheme in oligonucleotides in solution (Gao & Patel 1989, Gilbert et al 1989). However, independent attempts to detect Hoogsteen pairs in echinomycin complexes with macromolecular DNA, using a variety of probes, have failed (Mendel & Dervan 1987, Portugal et al 1988, Jeppesen & Nielsen 1988, McLean & Waring 1988).

Thus far only one experiment has directly addressed the question of whether echinomycin forces base pairs in large DNA molecules to flip and adopt Hoogsteen pairing (McLean et al 1989). This was done by synthesizing DNA containing deoxytubercidin (7-deazadeoxyadenosine) residues in place of the natural A residues, thus preventing A·T pairs forming Hoogsteen hydrogen bonds because of the lack of the essential purine N-7 atom (cf. Fig. 5). The results were clear cut: no differences were found when the 7-deaza-A-substituted DNA was compared with normal DNA, neither in respect of the footprinting pattern, which detects the physical binding sites of echinomycin, nor in the reactivity towards osmium tetroxide, a probe that sensitively detects antibiotic-induced changes in the accessibility of the 5,6-double bond of thymidine residues (Fig. 6). It was concluded that Hoogsteen base pairing cannot be the cause of echinomycin-induced hyperreactivity of DNA towards osmium tetroxide, which most probably results from the large local unwinding of the helix. Moreover, preventing the possibility of Hoogsteen base pairing did not preclude echinomycin from binding to the DNA.

FIG. 6. (*opposite*) (a) Autoradiograph showing the results of DNAse I cleavage of DNA fragments containing deoxyadenosine in both strands (lanes marked normal) or containing 7-deazadeoxyadenosine in one strand (lanes marked analog) in the absence (lanes −) or presence (lanes +) of 4 μM echinomycin. For each experiment, two samples corresponding to 1 min and 5 min digestions with the enzyme were run in adjacent lanes. Numbers on the left refer to nucleotide coordinates in the original tyrT fragment (for a full description of the construction of this fragment, which contains the upstream regulatory elements for tyrosine tRNA transcription in *E. coli*, see McLean et al 1989). Bars highlight footprints or antibiotic-binding sites—regions of diminished cleavage by DNAse I in the presence of echinomycin. Also shown are the results of OsO_4 reactions with these two fragments in the absence (lanes −) or presence (lanes +) of 4 μM echinomycin. The three thymidine residues immediately 3′ to echinomycin-binding sites in this strand are marked with arrows. Lane G, dimethyl sulphate/piperidine marker specific for guanine. (b) Sequence of the lower (Crick) strand of the tyrT DNA fragment from positions 10–65 for comparison with the autoradiograph in (a). Echinomycin binding sites, centred on CpG steps, are underlined and the same three thymidine nucleotides are marked with arrows. For full experimental details and discussion see McLean et al 1989.

Conclusions

The conformational diversity of DNA under the influence of ligand binding is well illustrated by its interaction with antibiotics. Almost every parameter of the Watson–Crick B-form DNA helix can be modified by binding of one antibiotic or another. In general, the intercalating agents provoke more far-reaching perturbations of structure than do agents which bind in the minor groove. Future research will increasingly focus on how structural modifications of DNA affect its participation in macromolecular assemblies and interactions with proteins, and on the biological implications.

Acknowledgements

Research in the author's laboratory was supported by grants from the Cancer Research Campaign, the Medical Research Council and the Royal Society. Many collaborators and other colleagues participated in the experimental work and discussions which led to the present state of understanding; their contributions are recorded with deep appreciation. Dean Gentle provided expert and willing technical assistance.

References

Armstrong RW, Kurucsev T, Strauss UP 1970 The interaction between acridine dyes and deoxyribonucleic acid. J Am Chem Soc 92:3174–3181

Coll M, Frederick CA, Wang AHJ, Rich A 1987 A bifurcated hydrogen-bonded conformation in the d(A.T) base pairs of the DNA dodecamer d(CGCAAATTTGCG) and its complex with distamycin. Proc Natl Acad Sci USA 84:8385–8389

Coll M, Aymami J, van der Marel GA, van Boom JH, Rich A, Wang AHJ 1989 Molecular structure of the netropsin–d(CGCGATATCGCG) complex: DNA conformation in an alternating AT segment. Biochemistry 28:310–320

Frederick CA, Williams LD, Ughetto G et al 1990 Structural comparison of anticancer drug–DNA complexes: adriamycin and daunomycin. Biochemistry 29:2538–2549

Gale EF, Cundliffe E, Reynolds PE, Richmond MH, Waring MJ 1981 The molecular basis of antibiotic action, 2nd edn. Wiley, Chichester

Gao X, Patel DJ 1989 Antitumour drug–DNA interactions: NMR studies of echinomycin and chromomycin complexes. Q Rev Biophys 22:93–138

Gauvreau D, Waring MJ 1984 Directed biosynthesis of novel derivatives of echinomycin. Part II. Purification and structure elucidation. Can J Microbiol 30:730–738

Gilbert DE, van der Marel GA, van Boom JH, Feigon J 1989 Unstable Hoogsteen base pairs adjacent to echinomycin binding sites within a DNA duplex. Proc Natl Acad Sci USA 86:3006–3010

Hertzberg RP 1990 Agents interfering with DNA enzymes. In: Hansch C, Sammes PG, Taylor JB (eds) Comprehensive medicinal chemistry. Pergamon Press, Oxford vol 2:753–791

Jeppesen C, Nielsen PE 1988 Detection of intercalation-induced changes in DNA structure by reaction with diethyl pyrocarbonate or potassium permanganate. Evidence against the induction of Hoogsteen base pairing by echinomycin. FEBS (Fed Eur Biochem Soc) Lett 231:172–176

Kopka ML, Yoon C, Goodsell D, Pjura P, Dickerson RE 1985 The molecular origin of DNA-drug specificity in netropsin and distamycin. Proc Natl Acad Sci USA 82:1376–1380

Kopka ML, Pjura PE, Goodsell DS, Dickerson RE 1987 Drugs and minor groove binding in B-DNA: netropsin and Hoechst 33258. In: Eckstein F, Lilley DMJ (eds) Nucleic acids and molecular biology. Springer-Verlag, Berlin, vol 1:1–24

McLean MJ, Waring MJ 1988 Chemical probes reveal no evidence of Hoogsteen base-pairing in complexes formed between echinomycin and DNA in solution. J Molecular Recognition: 1:38–151

McLean MJ, Seela F, Waring MJ 1989 Echinomycin-induced hypersensitivity to osmium tetroxide of DNA fragments incapable of forming Hoogsteen base pairs. Proc Natl Acad Sci USA 86:9687–9691

Mendel D, Dervan PB 1987 Hoogsteen base pairs proximal and distal to echinomycin binding sites on DNA. Proc Natl Acad Sci USA 84:910–914

Montenay-Garestier T, Hélène C, Thuong NT 1990 Design of sequence-specific bifunctional nucleic acid ligands. In: Host–guest molecular interactions: from chemistry to biology. Wiley, Chichester (Ciba Found Symp 158) p 147–157

Pelton JG, Wemmer DE 1989 Structural characterization of a 2:1 distamycin A.d(CGCAAATTGGC) complex by two-dimensional NMR. Proc Natl Acad Sci USA 86:5723–5727

Pelton JG, Wemmer DE 1990 Binding modes of distamycin A with $d(CGCAAATTTGCG)_2$ determined by two-dimensional NMR. J Am Chem Soc 112:1393–1399

Portugal J, Waring MJ 1987 Analysis of the effects of antibiotics on the structure of nucleosome core particles determined by DNAse I cleavage. Biochimie 69:825–840

Portugal J, Fox KR, McLean MJ, Richenberg JL, Waring MJ 1988 Diethyl pyrocarbonate can detect a modified DNA structure induced by the binding of quinoxaline antibiotics. Nucleic Acids Res 16:3655–3670

Quigley GJ, Ughetto G, van der Marel GA, van Boom JH, Wang AHJ, Rich A 1986 Non-Watson–Crick GC and AT base pairs in a DNA–antibiotic complex. Science (Wash DC) 232:1255–1258

Searle MS, Hall JG, Denny WA, Wakelin LPG 1988 NMR studies of the interaction of the antibiotic nogalamycin with the hexadeoxyribonucleotide duplex $d(5'GCATGC)_2$. Biochemistry 27:4340–4349

Sheldrick GM, Guy JJ, Kennard O, Rivera V, Waring MJ 1984 Crystal and molecular structure of the DNA-binding antitumour antibiotic triostin A. J Chem Soc Perkin Trans II 10:1601–1605

Tomasz M, Borowy-Borowski H, McGuinness BF 1990 Course of recognition and covalent reactions between mitomycin C and DNA. Sequence selectivity of a cross-linking drug. In: Pullman B, Jortner J (eds) Molecular basis of specificity in nucleic acid–drug interactions. Kluwer, Dordrecht (23rd Jerusalem Symposium on Quantum Chemistry & Biochemistry) p 551–564

Ughetto G, Wang AH, Quigley GJ, van der Marel GA, van Boom JH, Rich A 1985 A comparison of the structure of echinomycin and triostin A complexed to a DNA fragment. Nucleic Acids Res 13:2305–2323

Wakelin LPG, Adams A, Hunter C, Waring MJ 1981 Interaction of crystal violet with nucleic acids. Biochemistry 20:5779–5787

Wang AHJ 1987 Interactions between antitumor drugs and DNA. In: Eckstein F, Lilley DMJ (eds) Nucleic acids and molecular biology. Springer-Verlag, Berlin vol 1:53–69

Wang AH, Ughetto G, Quigley GJ et al 1984 Molecular structure of the DNA–triostin complex. Science (Wash DC) 225:1115–1121

Waring MJ 1979 Echinomycin, triostin and related antibiotics. In: Hahn FE (ed) Antibiotics, vol 5/part 2: Mechanism of action of antieukaryotic and antiviral compounds. Springer-Verlag, Heidelberg, p 173–194

Waring MJ 1990 The molecular basis of specific recognition between echinomycin and DNA. In: Pullman B, Jortner J (eds) Molecular basis of specificity in nucleic acid–drug interactions. Kluwer, Dordrecht (23rd Jerusalem Symposium on Quantum Chemistry & Biochemistry) p 225–245
Waring MJ, Henley SM 1975 Stereochemical aspects of the interaction between steroidal diamines and DNA. Nucleic Acids Res 2:567–586
Waring MJ, Wakelin LPG 1974 Echinomycin: a bifunctional intercalating antibiotic. Nature (Lond) 252:653–657
Warpehoski MA, Hurley LH 1988 Sequence selectivity of DNA covalent modification. Chem Res Toxicol 1:315–333
Williams LD, Egli M, Gao Q et al 1990 Structure of nogalamycin bound to a DNA hexamer. Proc Natl Acad Sci USA 87:2225–2229
Wilson WD, Tanious FA, Barton HJ et al 1990 The interaction of unfused polyaromatic heterocycles with DNA: intercalation, groove-binding and bleomycin amplification. Anti-Cancer Drug Des 5:31–42
Wu HM, Crothers DM 1984 The locus of sequence-directed and protein-induced DNA bending. Nature (Lond) 308:509–513

DISCUSSION

Kennard: The startling thing about the interaction of echinomycin with DNA is that nobody predicted the complex which was found in the crystal structure (Ughetto et al 1985). There was a tremendous amount of theoretical work published on the binding of bifunctional intercalators to DNA. However, no one included in the calculations a radical change in the DNA structure such as the presence of Hoogsteen base pairs. Although in the latest experiments you reported you did not see Hoogsteen pairing, I think that we are now sufficiently open-minded to postulate a number of different hypotheses. Hoogsteen pairing does not necessarily have to be involved—it could be almost any other type of base pairing.

One of the important things to realize is that classical B-DNA is not the only form of DNA which can bind ligands. We are far from understanding exactly how these drugs interact with DNA. Although a fair number of X-ray structures have been reported for oligonucleotide–drug complexes, each class of drugs was complexed to virtually the same or very similar oligomers. The groove binders have been studied with the Dickerson dodecamer (Wing et al 1980) or variations on this sequence, all of which have a narrow AATT region. Similarly, daunomycin has been studied with a hexamer d(CGATCG) and variants on this sequence (for a recent review see Kennard & Hunter 1989). There is not yet enough information for us to be confident that we fully understand the rules governing DNA–drug interactions. The main facts which have emerged are that the van der Waals contacts are more important than anticipated, that there is a balance between these contacts and the hydrogen bonding between the drug and the DNA, and that three-centred hydrogen bonds play an important role.

I do not think we could, as yet, predict how any particular drug would bind to a randomly chosen base sequence. We should exercise caution.

Waring: That is a succinct summary of the state of play. There's no doubt that we have had our eyes opened and we've already had lots of surprises. But life is full of surprises and I am sure that we are in for more.

Esther Breslow: There is obviously some energy cost when the drug binds to and distorts the DNA, and at the same time there is presumably some sort of chelate effect, because the intact echinomycin binds better than it would if it were cut in half. Is anything known about how strongly the 'half molecule' would bind and what would happen to DNA conformation?

Waring: There is a fair amount of information on structure–activity relationships for both biological activity and binding to DNA (Katagiri et al 1975, Waring 1979). Stabilization of the echinomycin–DNA complex is mainly due to van der Waals forces. A large number of van der Waals contacts are formed between the peptide ring system and the minor groove of the DNA. There is probably relatively little contribution to the binding energy from intercalation of the quinoxaline chromophores, which we have long suspected to be rather poor intercalators. Indeed, the available crystal structures reveal that in each case the intercalating chromophore is rather poorly stacked over the nucleotides.

Years ago we did all sorts of terrible things to echinomycin, such as trying to knock bits off the molecule (Lee & Waring 1978). Almost anything that you do to the peptide ring abolishes the molecule's biological activity and its ability to bind to DNA. I have been to many meetings of this sort where chemists spot the thioacetal group and tell me that I could knock it off with x, y or z; we tried x, y or z, not very successfully, because the cross-bridge seems to be buried in the structure and remarkably refractory to chemical attack, usually ending up with about 1 or 2% of a product in which the thioacetal bridge is split. In general, we saw no binding of these products to DNA. Richard Olsen has synthesized a compound which lacks a cross-bridge across the middle (Fox et al 1980). In this molecule there are actually two alanines in place of the cysteines, so it presumably has a more 'floppy' peptide ring. That molecule did bind to DNA, but with a low binding constant.

Esther Breslow: Is it the edge of the staple, rather than the two prongs, which distorts the DNA?

Ron Breslow: What happens if one of the prongs is missing?

Waring: Richard Olsen gave us one of his precursors for the total synthesis of triostin and we knocked off the benzyloxycarbonyl groups. The peptide which remained showed only very weak interaction with DNA, so prongs of some sort are essential.

Esther Breslow: Did that molecule still alter the DNA structure?

Waring: Not much, so far as we could tell. I should perhaps add that anything that alters the orientation of the chromophore with respect to the peptide ring,

or modifies it radically, abolishes or severely reduces binding to DNA. Richard Olsen synthesized an echinomycin analogue with an L centre instead of a D centre at the serine holding on one of the chromophores. This molecule looks nine-tenths normal, but one of the chromophores is disposed at a crazy angle. The molecule interacted only very weakly with DNA (Fox et al 1980). A similar analogue with both D-serines replaced by L-amino acids did not bind detectably to natural DNA, though a feeble effect on the melting temperature of poly(dA-dT) was observed (Lee & Waring 1978).

Esther Breslow: Can't you fool the system by forcing more ligand into solution and thereby get weakly binding ligands to interact with the DNA?

Waring: No; these antibiotics tend to have rather low solubility. The saturation limit in water of echinomycin is 5 μM, which leads to endless experimental difficulties because the near-u.v. absorption coefficient is low and it's not fluorescent.

Esther Breslow: You can't get an absolute value because of the solubility problems, but how much energy do you estimate that you get from the extra prong, relative to the total binding energy?

Waring: We did attempt some thermodynamic analysis, but it was based on an inappropriate method of handling binding isotherms, now superseded, and I wouldn't place any reliance on the numbers we got (Waring & Wakelin 1974, Wakelin & Waring 1976). I think the time may be right for someone to get the energetics sorted out.

Brändén: Do the non-intercalating antibiotics show any sequence specificity and, if so, what is responsible for the sequence specificity in the minor groove of DNA?

Waring: The short answer is that nearly all the non-intercalators that bind in the minor groove of the helix do so at AT-rich sequences; they do not like the 2-amino group of guanine, which is the one hydrogen bond donor that sits in the minor groove. The AT-rich sequences are typically associated with a relatively narrow minor groove, and that can be narrowed further or widened to accommodate the ligand. The minor groove binders classically strip off the ordered layer of water in the minor groove and replace it with hydrogen-bonding groups of their own. Attempts are now being made in various laboratories, notably those of Richard Dickerson in Los Angeles and Bill Lown in Edmonton, to make compounds called 'lexitropsins'; these are variations on the structure of compounds like netropsin and distamycin that will recognize a G·C pair or at least tolerate a G·C pair in the binding site. I think they are having some success.

Kennard: I am not sure if the groove binders are necessarily all that specific. In the case of one oligomer, d(CGCGAATTCGCG), complexed to the groove binder Hoechst 33258, two independent determinations gave different results. In one determination (Teng et al 1988) the drug was found to bind to the central four AATT base pairs, whereas in the other (Pjura et al 1987) the drug bound asymmetrically to the AATT region.

Rebek: In molecules like echinomycin, are both the chromophores on the same face in the absence of the DNA?

Waring: Some very nice NMR work has been done on the conformation of echinomycin in solution (Cheung et al 1978, Williamson & Williams 1981, Williamson et al 1982). As far as we can tell, the conformation of the antibiotic in solution is practically identical to its conformation when bound to DNA.

Kennard: The same is true of triostin A. The crystal structures with and without DNA are very similar, with two almost parallel chromophores in each case (Sheldrick et al 1984).

Waring: It must be said that in the crystal structures it is rare for the chromophores of the antibiotics to be placed exactly parallel. In one particular antibiotic crystal structure, that of the semi-biosynthetic bis-quinoline analogue of echinomycin called 2QN (unpublished), one chromophore is disposed in almost exactly the optimum position for intercalation, but the other one is bent off a bit. We are not daunted by this because it's clear that the point of attachment of the chromophores to the peptide enjoys relatively free rotation, so that presumably with minimum energy costs the two chromophores could be brought into an exactly parallel arrangement. In that arrangement, the spacing between their points of attachment is generally just about right to accommodate two DNA base pairs or slightly more.

Kennard: The structure of actinomycin hasn't been solved yet, although Jain & Sobell (1990) have reported the crystal data for a complex. It would be extremely interesting to see how well the current molecular models could predict that structure.

Vinter: That's a challenge I will attempt to take on!

Caruthers: Another way to address this problem is to use different oligonucleotides with echinomycin to see if the Hoogsteen-type base pairing is still formed. Have you done that?

Waring: We have done very little oligonucleotide work in my laboratory. I agree that by careful analysis of interactions between echinomycin and oligonucleotides it should be possible, at least in those model systems, to observe what happens between two judiciously spaced antibiotic binding sites in an oligonucleotide. One would of course still worry about 'end effects', distal to the antibiotic binding sites, but I would feel quite confident about observations on internal base pairs—that is, those located between two filled binding sites. Such studies are underway in at least two laboratories in the USA. (Gao & Patel 1988, 1989, Gilbert et al 1989).

References

Cheung HT, Feeney J, Roberts GCK, Williams DH, Ughetto G, Waring MJ 1978 The conformation of echinomycin in solution. J Am Chem Soc 100:46–54

Fox KR, Olsen RK, Waring MJ 1980 Interaction between synthetic analogues of quinoxaline antibiotics and nucleic acids: role of the disulphide cross-bridge and D-amino acid centres in des-*N*-tetramethyl-triostin A. Br J Pharmacol 70:25–40

Gao X, Patel DJ 1988 NMR studies of echinomycin bisintercalation complexes with d(A1-C2-G3-T4) and d(T1-C2-G3-A4) duplexes in aqueous solution: sequence-dependent formation of Hoogsteen A1.T4 and Watson-Crick T1.A4 base pairs flanking the bisintercalation site. Biochemistry 27:1744–1751

Gao X, Patel DJ 1989 Antitumour drug–DNA interactions: NMR studies of echinomycin and chromomycin complexes. Q Rev Biophys 22:93–138

Gilbert DE, van der Marel GA, van Boom JH, Feigon J 1989 Unstable Hoogsteen base pairs adjacent to echinomycin binding sites within a DNA duplex. Proc Natl Acad Sci USA 86:3006–3010

Jain SC, Sobell HM 1990 Actinomycin: molecular mechanism of action; structure of crystalline complex of 7-aminoactinomycin with DNA like dodecamer. Abstracts of the American Crystallographic Association Meeting (1990) 86 PGO3

Katagiri K, Yoshida T, Sato K 1975 Quinoxaline antibiotics. In: Corcoran JW, Hahn FE (eds) Antibiotics, vol 3: Mechanism of action of antimicrobial and antitumour agents. Springer-Verlag, Heidelberg, p 234–251

Kennard O, Hunter WN 1989 Oligonucleotide structure: a decade of results from single crystal X-ray diffraction studies. Q Rev Biophys 22:327–379

Lee JS, Waring MJ 1978 Interaction between synthetic analogues of quinoxaline antibiotics and nucleic acids: changes in mechanism and specificity related to structural alterations. Biochem J 173:129–144

Pjura PE, Grzeskowiak K, Dickerson RE 1987 Binding of Hoechst 33258 to the minor groove of B-DNA. J Mol Biol 197:257–271

Sheldrick GM, Guy JJ, Kennard O, Rivera V, Waring MJ 1984 Crystal and molecular structure of the DNA-binding antitumour antibiotic, triostin A. J Chem Soc Perkin Trans II 10:1601–1605

Teng MK, Usman N, Frederick CA, Wang A H-J 1988 The molecular structure of the complex of Hoechst 33258 and the DNA dodecamer d(CGCGAATTCGCG). Nucleic Acids Res 16:2671–2690

Ughetto G, Wang AHJ, Quigley JG, van der Marel GA, van Boom JH, Rich A 1985 A comparison of the structure of echinomycin and triostin A complexed to a DNA fragment. Nucleic Acids Res 13:2305–2323

Wakelin LPG, Waring MJ 1976 The binding of echinomycin to DNA. Biochem J 157:721–740

Waring MJ 1979 Echinomycin, triostin and related antibiotics. In: Hahn FE (ed) Antibiotics, vol 5 part 2: Mechanism of action of antieukaryotic and activiral compounds. Springer-Verlag, Heidelberg, p 173–194

Waring MJ, Wakelin LPG 1974 Echinomycin: a bifunctional intercalating antibiotic. Nature (Lond) 252:653–657

Williamson MP, Williams DH 1981 Manipulation of the nuclear Overhauser effect by the use of a viscous solvent: the solution conformation of the antibiotic echinomycin. J Chem Soc Chem Commun 165–166

Williamson MP, Gauvreau D, Williams DH, Waring MJ 1982 Structure and conformation of fourteen different antibiotics of the quinoxaline group determined by 1H NMR. J Antibiotics (Tokyo) 32:62–66

Wing R, Drew H, Takano T, Broka C, Tanaka S, Itakura K, Dickerson RE 1980 Crystal structure analysis of a complete turn of B-DNA. Nature (Lond) 287:755–758

Design of sequence-specific bifunctional nucleic acid ligands

T. Montenay-Garestier†, C. Hélène† and N. T. Thuong*

*†Laboratoire de Biophysique, INSERM U201, CNRS UA481, Muséum National d'Histoire Naturelle, 43 rue Cuvier, 75005 Paris and *Centre de Biophysique Moléculaire, CNRS, 45071, Orléans Cedex 02, France*

Abstract. Homopyrimidine oligodeoxynucleotides have been covalently linked to intercalating agents. These bifunctional nucleic acid ligands bind to the major groove of DNA at homopurine·homopyrimidine sequences, where they form triple helices. The homopyrimidine oligonucleotide binds parallel to the purine strand of the double helix. Two hydrogen bonds are formed between bases of the oligonucleotide and the purines engaged in Watson–Crick base pairs. The intercalating agent inserts its aromatic ring at the triplex–duplex junction, resulting in a strong stabilization of the triple helical structure. Bifunctional oligonucleotide–intercalator conjugates provide new tools for a selective control of gene expression. In addition, irreversible reactions can be targeted to the oligonucleotide recognition sequence. Cleavage reactions can be induced by a copper–phenanthroline chelate or an ellipticine derivative covalently linked to the triple helix-forming oligonucleotide.

1991 Host–guest molecular interactions: from chemistry to biology. Wiley, Chichester (Ciba Foundation Symposium 158) p 147–157

Regulation of gene expression in living organisms involves the selective recognition of nucleic acid base sequences by regulatory proteins and also by nucleic acids themselves. Artificial control of gene expression can be achieved using oligodeoxynucleotides and their derivatives (Toulmé & Hélène 1988). The more popular targets have been messenger ribonucleic acids (mRNA), whose translation is blocked by binding of an antiparallel complementary oligonucleotide (the so-called 'antisense strategy'). The efficiency of the biological effect depends on the activity of an endogenous enzyme, RNase H, which cleaves the mRNA in the mRNA–oligodeoxynucleotide hybrid (Cazenave et al 1987a). When the oligonucleotide target is located upstream of the initiation codon, translation inhibition might also involve an RNase H independent inhibition of the translation machinery. We have shown previously that covalent attachment of an intercalating agent to one or both ends of an oligonucleotide strongly increases its inhibitory effect (Cazenave et al 1987a, Toulmé et al 1986).

In an *in vitro* translation assay, this increase was ascribed to enhanced binding of the oligonucleotide to its target sequence resulting from additional binding energy provided by the intercalating agent (Asseline et al 1983, 1984). In cell culture, two additional effects contribute to the enhanced biological activity: (i) addition of an intercalating agent to the 3′ end protected the oligonucleotide against degradation by 3′-exonucleases (Cazenave et al 1987b, Verspieren et al 1987) and (ii) the intercalating agent facilitated uptake of the oligonucleotide by the cells (Verspieren et al 1987, Zérial et al 1987). Such oligonucleotide–intercalator conjugates were shown to selectively kill trypanosomes (Verspieren et al 1987) and to inhibit the cytopathic effect of influenza virus in MDCK cells (Zérial et al 1987).

Double-stranded DNA can also be recognized by oligonucleotides via Hoogsteen base pairing of homopyrimidine oligonucleotides to homopurine·homopyrimidine sequences in DNA. This is the basis of the 'anti-gene strategy' for the selective control of gene expression at the level of gene transcription or replication. Thymine and protonated cytosine form two hydrogen bonds with Watson–Crick A·T and G·C base pairs respectively (Fig. 1). Attachment of a photoactive group to the oligonucleotide allowed us to photo-induce cross-linking of the oligonucleotide to each strand of duplex DNA and to cleave the DNA strands at the cross-linked sites by piperidine treatment (Le Doan et al 1987, Praseuth et al 1988). DNA-cleaving reagents such as Fe–EDTA (Moser & Dervan 1987, Strobel et al 1988) and Cu–phenanthroline (François et al 1988, 1989a,b) were shown to induce site-specific double-strand cleavage upon addition of a reducing agent. Sequence-specific intercalation occurs when an intercalating agent is covalently linked to a homopyrimidine oligonucleotide (Sun et al 1989). When phenanthroline is chosen as the intercalating agent, the cleavage reaction induced upon addition of a reducing agent to the copper chelate occurs in the minor groove, even though oligonucleotide binding takes place in the major groove (François et al 1989b). Cleavage can be photo-induced by an ellipticine derivative bound to the triple helix-forming oligonucleotide (Perrouault et al 1990).

Oligonucleotide–intercalator conjugates

Homopyrimidine oligonucleotides bind to homopurine·homopyrimidine sequences of double-helical DNA. Thymine and protonated cytosine form two hydrogen bonds with A·T and G·C base pairs respectively (Fig. 1). These hydrogen bonds involve the purine bases of Watson–Crick base pairs. They are located in the major groove of the double helix. In such triple helices the third strand has a parallel orientation with respect to the homopurine sequence.

Figure 2 shows the structure of an 11mer homopyrimidine oligonucleotide that binds to DNA at the indicated homopurine·homopyrimidine sequence, to form a triple helix. As shown in Table 1, the stability of the triple helix is

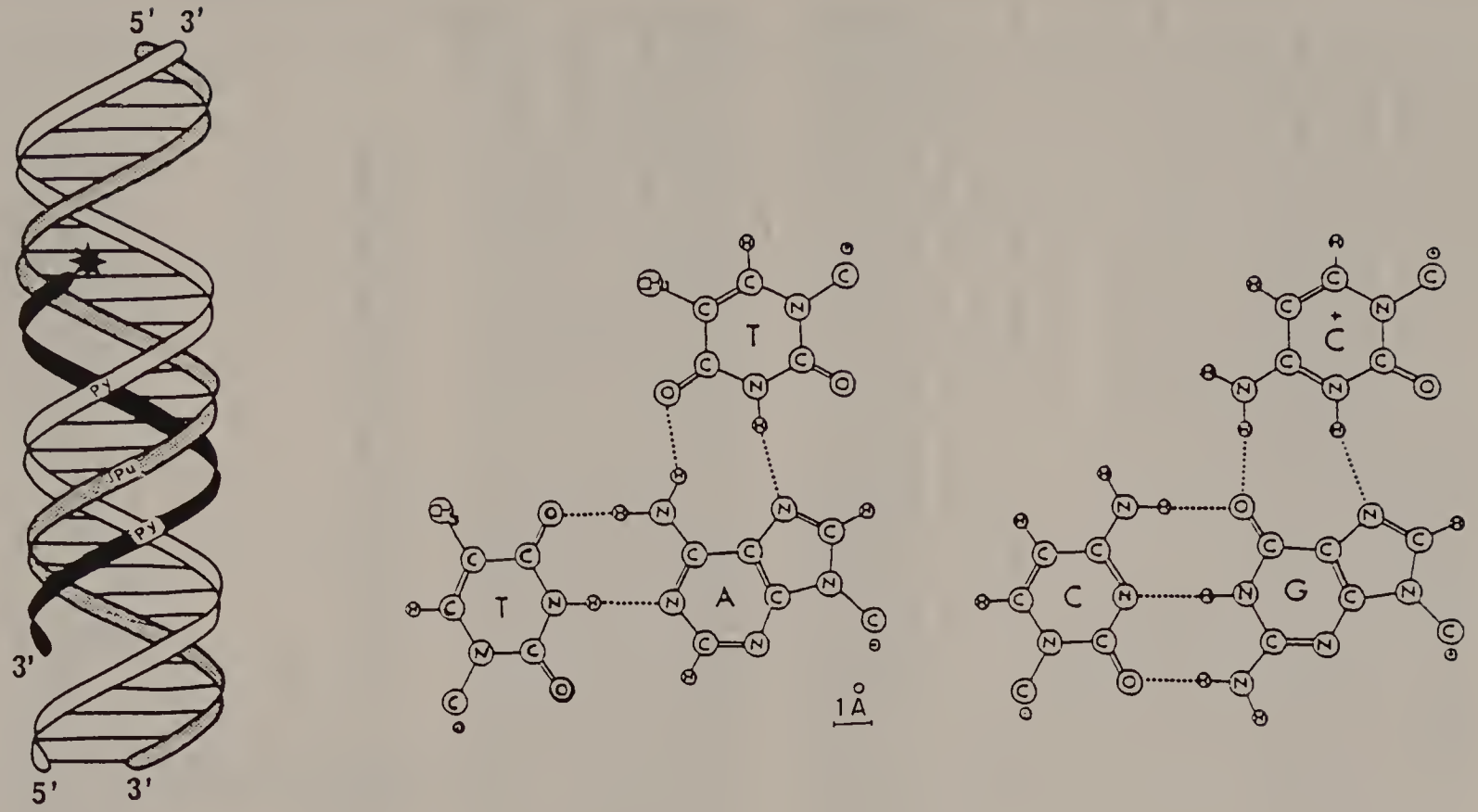


FIG. 1. Oligonucleotide binding to duplex DNA. *Left*: the third strand homopyrimidine oligonucleotide (black) binds to the major groove of duplex DNA at a homopurine·homopyrimidine sequence. The star represents a reactive group that can induce irreversible reactions involving one or both strands of DNA. *Right*: Hoogsteen hydrogen bonding of T and protonated C to A·T and G·C Watson–Crick base pairs respectively. Py, pyrimidine; Pu, purine.

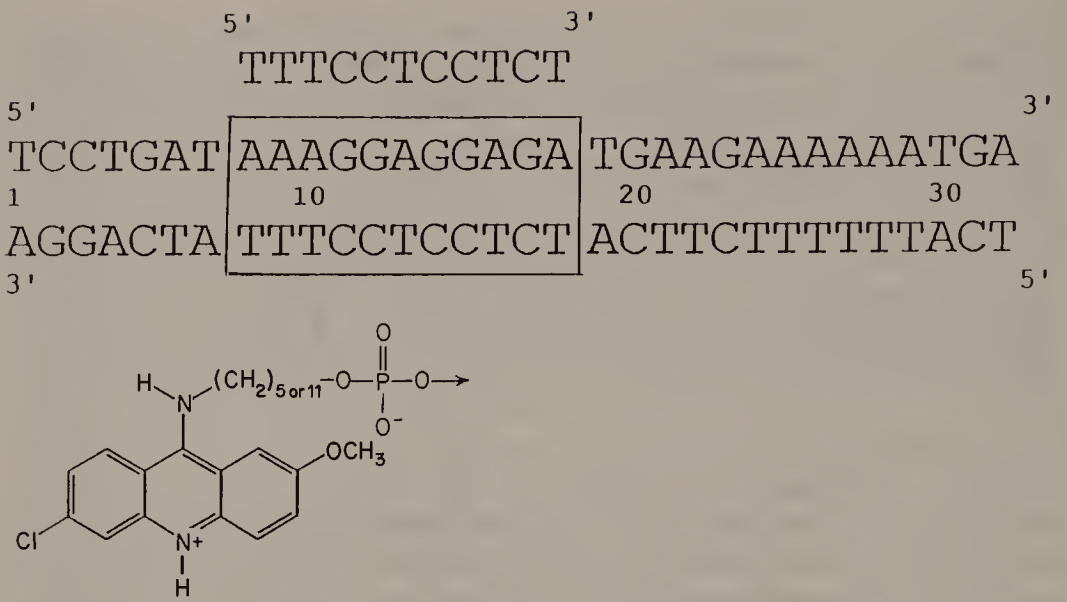


FIG. 2. Sequence of the duplex DNA fragment used in the studies described in the text with the triple helix-forming 11mer oligonucleotide. The acridine derivative shown below (2-methoxy-6-chloro-9-aminoacridine) was linked to either the 5′- or 3′-phosphate of the oligonucleotide via a linker containing either 5 (m_5) or 11 (m_{11}) methylene groups.

enhanced when the oligonucleotide is attached to an intercalating agent, acridine. A previous report has described the property of the 5′-substituted 11mer under different experimental conditions (Sun et al 1989). The gain in stability is higher when acridine is linked to the 5′ end of the oligonucleotide. Changing the length of the linker on the 3′ end influences complex stability. These differences can be explained in structural terms by taking into account the distortion that occurs at the 5′ and 3′ junctions between triplex and duplex structures.

By means of molecular modelling using the AMBER 3 software, the structure of the 11mer bound to a 32 base pair double helix and the triplex–duplex

TABLE 1 Temperatures of half-dissociation of triple helices formed between a 32mer double helix and different substituted 11mers

Oligonucleotide	*Tm (°C)*
11mer	19.3
Acr-m_5-5′11mer	37.4
11mer3′-m_5-Acr	25.5
11mer3′-m_{11}-Acr	30.0

See Fig. 2 for abbreviations of the intercalating agent and linker structures. Measurements were carried out in a 10 mM sodium cacodylate buffer (pH 6.0) in the presence of 0.1 M NaCl and 1 mM spermine. All oligonucleotide concentrations (32mer and 11mer) were 1 μM.

junctions have been studied. The complex formed between the oligonucleotide and its target was split in two parts. One, denoted 5′ junction, is composed of six triplets and six base pairs. It corresponds to the 5′ end of the pyrimidine strand bound by Hoogsteen hydrogen bonds to the target. The other, the 3′ junction, is also composed of six base pairs and six triplets, but at the 3′ end of the Hoogsteen-bonded strand. The coordinates of the atoms used to construct both junctions were taken from Arnott X-ray diffraction data obtained on triple helix fibres (Arnott et al 1976) for all atoms in the triplets, and from the canonical B-DNA data (Arnott et al 1980) for atoms in the base pairs. Molecules were generated with the JUMNA program (Lavery et al 1986), which uses helicoidal parameters (Lavery & Sklenar 1989) especially suited for the generation of nucleic acids.

Minimization of these structures was performed with AMBER (Weiner et al 1986). The computed energies for intercalation of the acridine derivative (2-methoxy-6-chloro-9-aminoacridine) either free or covalently attached to the 11mer oligonucleotide are shown in Table 2. It can be seen that the strongest stabilization occurs when the acridine ring is intercalated at the triplex–duplex junction. The m_5 (five methylene groups) and m_{11} (eleven methylene groups) linkers are nearly equivalent on the 5′ end, whereas intercalation with the m_5 linker is unlikely to occur on the 3′ end because the linker is too short.

TABLE 2 Computed energies at the 5′ and 3′ junctions for various sites of intercalation of an acridine derivative

	Energy of intercalation (kcal mol^{-1})	
Acridine derivative intercalation	*5′ junction*	*3′ junction*
Free acridine derivative		
DH	−28.5	−26.0
DH/TH	−30.0	−24.2
TH	−20.7	−10.8
Acridine/m_5 linker		
DH	−4.5	−3.9
DH/TH	−15.0	−1.4
TH	−4.0	No convergence
Acridine/m_{11} linker		
DH	−8.1	−18.6
DH/TH	−13.3	−20.2
TH	−12.1	+3.0

Three intercalation sites were investigated: in one (TH), the acridine derivative (2-methoxy-6-chloro-9-aminoacridine) is placed between the last two triplets of the triple helix; in the second (DH/TH), it is placed at the junction between the triple and the double helix; in the third (DH), intercalation is between the first two base pairs of the double helix. The acridine derivative was linked to the triple helix-forming oligonucleotide via 5 (m_5) or 11 (m_{11}) methylene groups. See text for explanation of 5′ and 3′ junction.

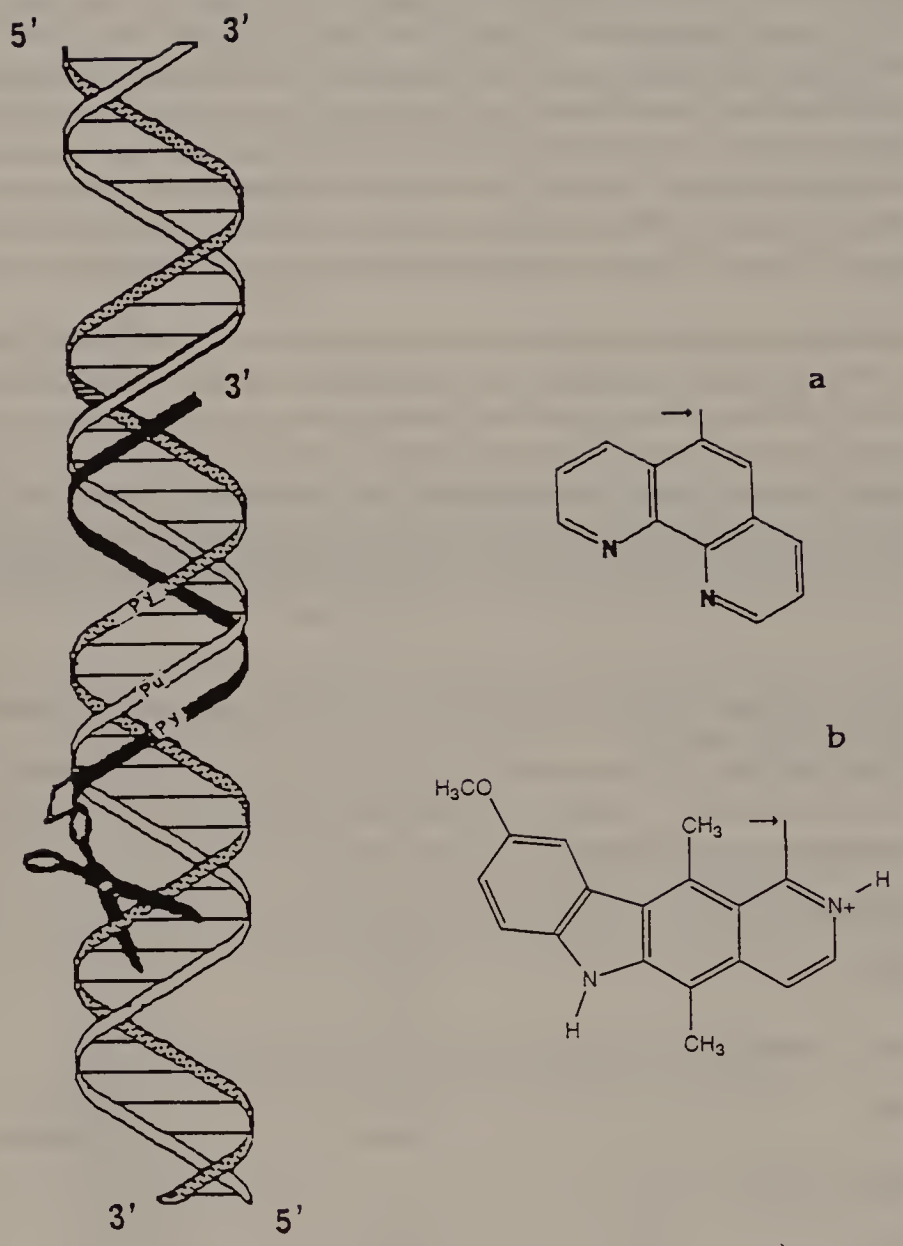


FIG. 3. Sequence-specific artificial nucleases. Oligonucleotide-intercalator conjugates recognize the major groove of duplex DNA at specific sequences (see Fig. 1) and induce cleavage reactions. If phenanthroline (a) is used as an intercalator, cleavage of the two strands of DNA is observed (François et al 1989c). Cleavage can also be photo-induced by an ellipticine derivative (b) Perrouault et al 1990. The arrow indicates the site of attachment to the oligonucleotide.

Recognition and cleavage of double helical DNA at specific sites by Cu–phenanthroline tethered to an oligonucleotide

The same 11 base pair homopurine·homopyrimidine target sequence as that shown in Fig. 2 was chosen as a target for an oligonucleotide–phenanthroline conjugate. Phenanthroline was attached to the 5′ end of the 11mer homopyrimidine oligonucleotide. Various linkers were used to tether the 5-amino group of 5-amino-1,10-phenanthroline to a 5′-thiophosphate group of the oligonucleotide. In the presence of Cu^{2+} and a reducing agent (β-mercaptopropionic acid) cleavage was observed of the two strands of a 32 base pair duplex DNA fragment containing the 11 base pair target sequence (François et al 1989b). The most efficient cleavage was obtained with a pentamethylene carboxamide linker. The cleavage sites on the two strands had an asymmetrical distribution. They were shifted towards the 3′ end, indicating that the cleavage reaction had taken place in the minor groove. The species responsible for cleavage are diffusible $OH^{\cdot}$ radicals which react within a restricted domain, as indicated by the limited number of bonds which undergo cleavage.

The asymmetrical distribution of the cleavage sites observed with Cu–phenanthroline tethered to the 11mer oligonucleotide indicated that cleavage occurred in the minor groove even though phenanthroline was brought into the major groove by oligonucleotide binding. This result can be explained if phenanthroline intercalates at the triplex–duplex junction as previously described for an acridine derivative tethered to the 11mer oligonucleotide (Sun et al 1989). The linker was attached to the C-5 position of the phenanthroline ring. Therefore, intercalation brings the nitrogen atoms (N-1 and N-10) into the minor groove, where copper chelation occurs, followed by oxidative attack of the deoxyribose and strand cleavage (Fig. 3).

The target sequence of the 11mer oligopyrimidine is contained within the DNA of simian virus SV40. There is a single copy of the 11 base pair sequence among the 5243 base pairs of the viral DNA. The phenanthroline-substituted oligonucleotide was shown to cleave SV40 DNA at a single site on both linear and circular DNA (François et al 1988, 1989b). Double-strand cleavage was observed by agarose gel electrophoresis. The cleaved phosphodiester bonds were then identified by polyacrylamide gel electrophoresis of denatured SV40 fragments obtained after cleavage by the phenanthroline–oligonucleotide conjugate. The cleavage sites were identical to those observed on the 32mer duplex DNA described above. The efficiency of double-strand cleavage reached about 70% at 20 °C in a pH 7.4 buffer containing 10 mM Na phosphate, 0.1 M NaCl and 1 mM spermine. Cleavage was measured as a function of temperature, and was found to be reduced by 50% at 30 °C. The oligonucleotide–phenanthroline conjugate formed a much more stable triple helix than the unsubstituted 11mer because as well as phenanthroline intercalation there was copper chelation in the minor groove, which locks the complex in place. At low

temperatures (< 10 °C) a secondary cleavage site was observed, which corresponded to binding of the oligonucleotide to a mismatched sequence (François et al 1989b).

Photo-induced cleavage of duplex DNA by an oligonucleotide–ellipticine conjugate

An ellipticine derivative (1-amino-5,11-dimethyl-9-methoxy ellipticine) was covalently attached via its 1-amino group to the 3′-phosphate of the 11mer oligonucleotide. This oligonucleotide was bound to a 32 base pair DNA fragment containing the 11 base pair homopurine·homopyrimidine target sequence where a triple helix is formed. Upon u.v. irradiation ($\lambda > 310$ nm) cleavage of the two strands of duplex DNA was observed where expected if the 11mer oligonucleotide were bound in a parallel orientation with respect to the homopurine strand (Perrouault et al 1990). Photo-induced cross-linking of the oligonucleotide–ellipticine conjugate to the target sequence was also observed upon irradiation. Neither cleavage nor cross-linking was observed when the triplex was dissociated at high temperature.

Conclusions

Homopyrimidine oligonucleotides bind to the major groove of DNA at homopurine·homopyrimidine sequences, forming local triple helices. An aromatic ring (acridine, phenanthroline or ellipticine) attached to the oligonucleotide via an appropriate linker may intercalate at the junction between the double helix and the triple helix. The complex is stabilized by the additional binding energy that is provided by this interaction. In the case of phenanthroline, intercalation provides the required structure for double-strand cleavage by the copper complex in the minor groove, even though the reaction is targeted to a specific sequence by recognition of the major groove. The negative charges of the oligonucleotide prevent intercalation of acridine, phenanthroline or ellipticine at DNA sequences where the oligonucleotide does not find its binding sequence. In other words, non-specific binding of the intercalating agent to any DNA sequence is lost. Intercalation takes place only at sequences which are recognized by the oligonucleotide. Therefore, oligonucleotide–intercalator conjugates are highly sequence-specific DNA intercalating agents that could be used to control gene expression at the DNA level. We (François et al 1989c) and others (Maher et al 1989, Hanvey et al 1990) have shown that local triple helix formation inhibits cleavage of DNA by restriction enzymes and binding of transcription factors (Maher et al 1989). Preliminary data also indicate that DNA replication might be inhibited (Birg et al 1990). Recognition of homopurine·homopyrimidine sequences of DNA by homopyrimidine oligonucleotides might not be the only recognition code for triple helix

formation. There is evidence that such sequences can also be recognized by homopurine oligonucleotides that are able to block transcription (Cooney et al 1988). Further applications of artificial sequence-specific double-strand nucleases can be envisaged, for example, in gene mapping experiments on long DNA fragments or for site-directed mutagenesis. The cleavage reaction can be chemically induced (e.g. using Cu–phenanthroline 'activated' by a reducing agent) or photochemically induced (e.g. using ellipticine). These reactions can take place in the low-melting agarose that is used to prepare long DNA macromolecules such as chromosomes. Pulse-field electrophoresis can then be used to separate the cleaved fragments. Site-directed mutagenesis can also be contemplated after induction of sequence-specific modifications in DNA *in vitro* or *in vivo*, using nuclease-resistant oligonucleotides such as oligo-[α]-deoxynucleotides (Le Doan et al 1987, Praseuth et al 1988).

Acknowledgements

This work was supported by INSERM, CNRS, the Muséum National d'Histoire Naturelle and Rhône-Poulenc Santé. The Ligue Nationale Française contre le Cancer and the Fondation pour la Recherche Médicale are also acknowledged for their support. We wish to thank our collaborators U. Asseline, J. Chomilier, J. C. François, C. Giovannangeli, T. Le Doan, L. Perrouault, T. Saison-Behmoaras and J. S. Sun who have been involved in the work summarized in this paper.

References

Arnott S, Bond PJ, Selsing E, Smith PJC 1976 Models of triple stranded polynucleotides with optimised stereochemistry. Nucleic Acids Res 3:2459–2470

Arnott S, Chandrasekaran R, Birdsall DL, Leslie AGW, Ratliff RL 1980 Left handed DNA helices. Nature (Lond) 283:743–746

Asseline U, Thuong NT, Hélène C 1983 Nouvelles substances à forte affinité spécifique pour des séquences d'acides nucléiques: oligodésoxynucléotides liés de façon covalente à un agent intercalant. C R Acad Sci Paris 297:369–372

Asseline U, Delarue M, Lancelot G, et al 1984 Nucleic acid binding molecules with high affinity and base sequence specificity: intercalating agents covalently linked to oligodeoxynucleotides. Proc Natl Acad Sci USA 81:3297–3301

Birg F, Praseuth D, Zerial A et al 1990 Inhibition of simian virus 40 DNA replication in CV1 cells by an oligonucleotide covalently linked to an intercalating agent. Nucleic Acids Res 18:2901–2908

Cazenave C, Loreau N, Thuong NT, Toulmé JJ, Hélène C 1987a Enzymatic amplification of translation inhibition of rabbit globin mRNA mediated by antimessenger oligodeoxynucleotides covalently linked to intercalating agents. Nucleic Acids Res 15:4717–4736

Cazenave C, Chevrier M, Thuong NT, Hélène C 1987b Rate of degradation of α and β oligodeoxynucleotides in *Xenopus* oocytes. Implications for antimessenger strategies. Nucleic Acids Res 15:10507–10521

Cooney M, Czernuszewicz G, Postel EH, Flint SJ, Hogan ME 1988 Site specific oligonucleotide binding represses transcription of human c-myc gene in vitro. Science (Wash DC) 241:456–459

François JC, Saison-Behmoaras T, Chassignol M, Thuong NT, Hélène C 1988 Restriction-like artificial endonucleases: oligonucleotides covalently linked to a copper-phenanthroline chelate recognize the major groove of DNA and cleave at specific sequences. C R Acad Sci Paris 307:849–854

François JC, Saison-Behmoaras T, Chassignol M, Thuong NT, Hélène C 1989a Sequence-targeted cleavage of single and double-stranded DNA by oligothymidylates covalently linked to 1,10-phenanthroline. J Biol Chem 264:5891–5898

François JC, Saison-Behmoaras T, Thuong NT, Hélène C 1989b Inhibition of restriction endonuclease cleavage via triple helix formation by homopyrimidine oligonucleotides. Biochemistry 28:9617–9619

François JC, Saison-Behmoaras T, Barbier C, Chassignol M, Thuong NT, Hélène C 1989c Sequence-specific recognition and cleavage of duplex DNA via triple helix formation by oligonucleotides covalently linked to phenanthroline-copper chelate. Proc Natl Acad Sci USA 86:9702–9706

Hanvey JC, Shimizu M, Wells RD 1990 Site-specific inhibition of EcoRI restriction/modification enzymes by a DNA triple helix. Nucleic Acids Res 18:157–161

Lavery R, Sklenar H 1989 Defining the structure of irregular nucleic acids: conventions and principles. J Biomol Struct & Dyn 6:655–667

Lavery R, Sklenar H, Zakrzewska K, Pullman B 1986 The flexibility of the nucleic acids. II. The calculation of internal energy and applications to mononucleotide repeat DNA. J Biomol Struct & Dyn 3:989–1014

Le Doan T, Perrouault L, Praseuth D, et al 1987 Sequence specific recognition, photocrosslinking and cleavage of the DNA double helix by an oligo α thymidylate covalently linked to an azidoproflavine derivative. Nucleic Acids Res 15:7749–7760

Maher III JL, Wold B, Dervan PB 1989 Inhibition of DNA binding proteins by oligonucleotide-directed triple helix formation. Science (Wash DC) 245:725–730

Moser HE, Dervan PB 1987 Sequence specific cleavage of double helical DNA by triple helix formation. Science (Wash DC) 238:645–650

Perrouault L, Asseline U, Rivalle C et al 1990 Sequence specific artificial photoinduced endonucleases based on triple helix forming oligonucleotides. Nature (Lond) 344:358–360

Praseuth D, Perrouault L, Le Doan T, Chassignol M, Thuong NT, Hélène C 1988 Sequence specific binding and photocrosslinking of an α and β oligo deoxynucleotide to the major groove of DNA via triple helix formation. Proc Natl Acad Sci USA 85:1349–1353

Strobel SA, Moser HE, Dervan PB 1988 Double-stranded cleavage of genomic DNA at a single site by triple-helix formation. J Am Chem Soc 110:7927–7929

Sun JS, François JC, Montenay-Garestier T et al 1989 Sequence-specific intercalating agents. Intercalation at specific sequences on duplex DNA via major-groove recognition by oligonucleotide intercalator conjugates. Proc Natl Acad Sci 86: 9198–9202

Toulmé JJ, Hélène C 1988 Antimessenger oligodeoxynucleotides: an alternative to antisense RNA for artificial regulation of gene expression—a review. Gene 72: 51–58

Toulmé JJ, Krisch HM, Loreau N, Thuong NT, Hélène C 1986 Specific inhibition of mRNA translation by complementary oligonucleotides covalently linked to intercalating agents. Proc Natl Acad Sci USA 83:1227–1231

Verspieren P, Cornelissen AWCA, Thuong NT, Hélène C, Toulmé JJ 1987 An acridine-linked oligodeoxynucleotide targeted to the common 5′ end of trypanosome mRNAs kills cultured parasites. Gene 61:307–315

Weiner SJ, Kollman PA, Nguyen DT, Case DA 1986 An all atom force field for simulations of proteins and nucleic acids. J Comput Chem 7:230–252

Zérial A, Thuong NT, Hélène C 1987 Selective inhibition of the cytopathic effect of type A influenza viruses by oligonucleotides covalently linked to an intercalating agent. Nucleic Acids Res 15:9909–9919

Synthesis and biochemical studies of dithioate DNA

Marvin H. Caruthers, Graham Beaton, Lendell Cummins, Darla Graff, Yun-Xi Ma, William S. Marshall, Henri Sasmor, Paula Norris and Eric K. Yau

Department of Chemistry and Biochemistry, University of Colorado, Boulder, CO 80309-0215, USA

Abstract. Dithioate DNA was synthesized and used for various biochemical studies. Results from these studies indicate that dithioate DNA is a potent inhibitor of HIV Reverse Transcriptase, activates endogenous RNase H in HeLa cell nuclear extracts, and is a useful probe for studying protein–DNA interactions.

1991 Host–guest molecular interactions: from chemistry to biology. Wiley, Chichester (Ciba Foundation Symposium 158) p 158–168

Although synthetic DNA having the natural 3′–5′ phosphodiester linkage has proven to be extremely valuable for a large number of biochemical and biological applications (Caruthers 1985, Caruthers 1989), certain key uses require modified synthetic polynucleotides. For example, for many potential applications DNA needs to be nuclease resistant, tagged with reporter groups, selectively linked to supports, or valuable as a therapeutic agent. For some time now, our research has focused on the development of derivatives that meet these criteria. Among the more promising is a new compound called dithioate DNA. This derivative, which contains an internucleotide phosphodiester group having sulphur at the two non-linking positions, is a mimic of natural DNA with many possibly unique properties. For example, it is achiral and ionic like natural DNA, but it can be derivatized easily and is stable toward nucleases. In this chapter, we shall outline the chemistry for synthesizing dithioate DNA, then present results that indicate how this derivative may prove useful in a large array of biochemical and biological applications.

Synthesis of phosphorodithioate DNA

For dithioate DNA to be used for many different applications, chemical synthesis methodologies must be rapid, high yielding and compatible with current approaches for preparing natural DNA. If these criteria are met, then this derivative can be incorporated into normal DNA in a pre-selected manner to

generate a polynucleotide with the desired combination of natural and dithioate internucleotide linkages.

The approach that meets these criteria is outlined in Fig. 1 (Brill et al 1989). Briefly, deoxynucleoside 3′-phosphorothioamidite synthons (**3**) are condensed in any desired sequence with a growing polynucleotide on a polymer support. These compounds (**3**) are prepared by condensing a suitably protected deoxynucleoside with bis(dimethylamino)chlorophosphine and triethylamine to yield the diamidite, which is converted, without isolation, to the phosphorothioamidite by addition of 2,4-dichlorobenzylmercaptan. After aqueous extraction and precipitation, these synthons can be stored without decomposition as dry powders under inert gas atmospheres. Synthesis of DNA begins by coupling the deoxynucleoside 3′-phosphorothioamidite with a deoxynucleoside linked to a silica support (**2**) to yield a thiophosphite triester (**4**). The activating agent is tetrazole. This step is followed by sulphurization using elementary sulphur to yield the protected phosphorodithioate derivative (**5**), capping or acylation of the unreactive silica-linked nucleoside with acetic anhydride (**6**) and detritylation with dichloroacetic acid. Further repetitions of this cycle using either dexoynucleoside phosphorothioamidites or deoxynucleoside phosphoramidites as synthons and tetrazole as an activator yield oligodeoxynucleotides with normal phosphate diester and phosphorodithioate diester linkages in the desired combination. So far, segments containing up to 28 dithioate linkages have been prepared by this approach. Yields of 96–98% internucleotide bond formation per cycle (based upon measurement of dimethoxytrityl cation released during detritylation) and 97–98% dithioate per linkage have been obtained (the remaining 2–3% internucleotide linkages are phosphorothioate). Similar results can be obtained with the deoxynucleoside 3′-pyrrolidinothioamidite, but use of

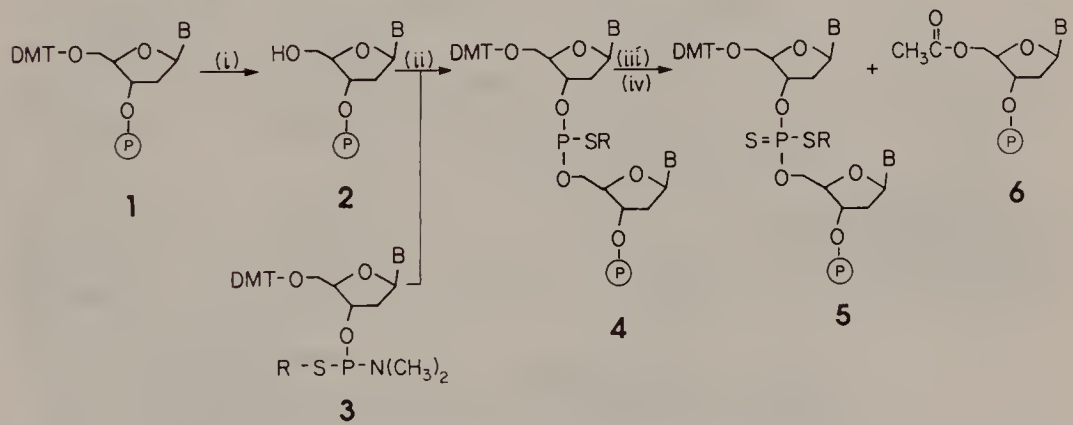


FIG. 1. Synthesis of deoxyoligonucleotide dithioates from deoxynucleoside phosphorothioamidites. (i) 2% dichloroacetic acid (ii) tetrazole (iii) sulphur (iv) dimethylaminopyridine + acetic anhydride. Abbreviations: R, 2,3-dichlorobenzyl; DMT, di-*p*-methoxytrityl; B, *N*-benzoyladenine, *N*-benzoylcytosine, thymine, or *N*-isobutyrylguanine; Ⓟ, silica gel support.

sulphur-protecting groups other than 2,4-dichlorobenzyl, such as β-cyanoethyl, leads to higher yields of the phosphorothioate (5–10%).

Although the method outlined in Fig. 1 allows the rapid synthesis of a large variety of phosphorodithioate-containing oligonucleotides, it is primarily useful only for preparing small quantities of DNA (1 μmole or less). For certain applications, such as therapeutic or biophysical (NMR and X-ray crystallography) studies, considerably larger quantities of phosphorodithioate DNA are needed. These considerations led to the development of a non-polymer support, phosphate triester approach (Yau et al 1990) for synthesizing this derivative (Fig. 2). The first step was to prepare the fully protected phosphorodithioate triesters (**7a,b**) from commercially available deoxynucleoside phosphoramidites by treatment first with 4-chloro- or 2,4-dichlorobenzyl-mercaptan and tetrazole and then, without isolation, with a saturated sulphur solution. Intermediates from which DNA can be synthesized (**8a,b**; **9b**) can be generated from **7a,b** by removal of the β-cyanoethyl group (Et_3N:CH_3CN, 1:1, at room temperature for 2–3 h) or the dimethoxytrityl group (3% trichloroacetic acid at 0 °C for 30 min). Further condensation of the resulting two synthons using triisopropylbenzenesulphonyl chloride in CH_2Cl_2 yields the completely protected dinucleoside phosphorodithioate (**10a,b,c**), contaminated with 1% of **11a,b,c**, the dinucleoside phosphorothioate. The latter compound can be removed by silica gel column chromatography. This chemistry has been used to synthesize a deoxyoctathymidine containing phosphorodithioate internucleotide linkages (approximately 200 mg) and a large assortment of di- and trinucleotide dithioates (0.5 to 1.0 mmoles). These results suggest that this approach will prove useful for the large-scale synthesis of dithioate oligonucleotides.

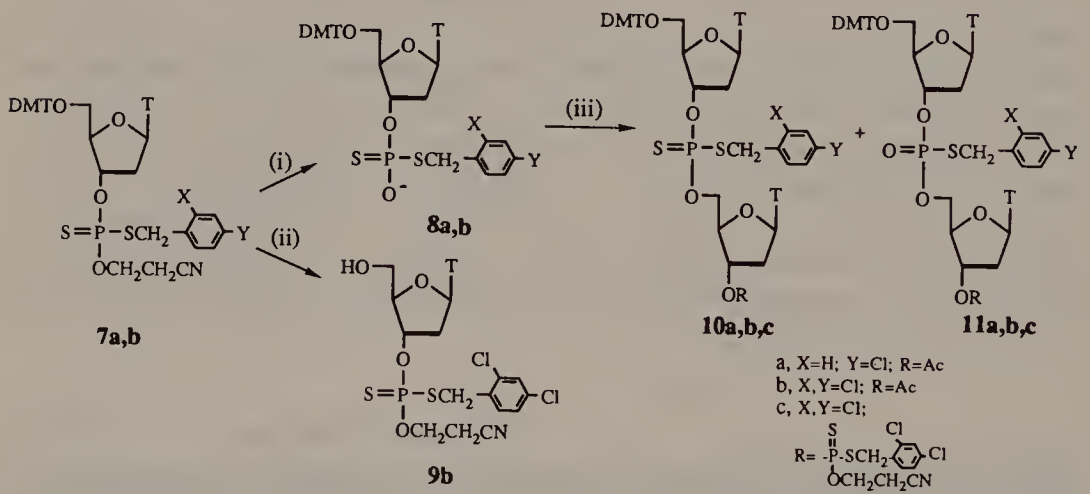


FIG. 2. Synthesis of deoxydinucleotide dithioates from a solution phosphotriester approach. (i) Et_3N:CH_3CN (1:1, v:v) (ii) 2–3% trichloroacetic acid (iii) triisopropylbenzenesulphonyl chloride. Ac, acetyl; DMT, di-*p*-methoxytrityl; T, thymine.

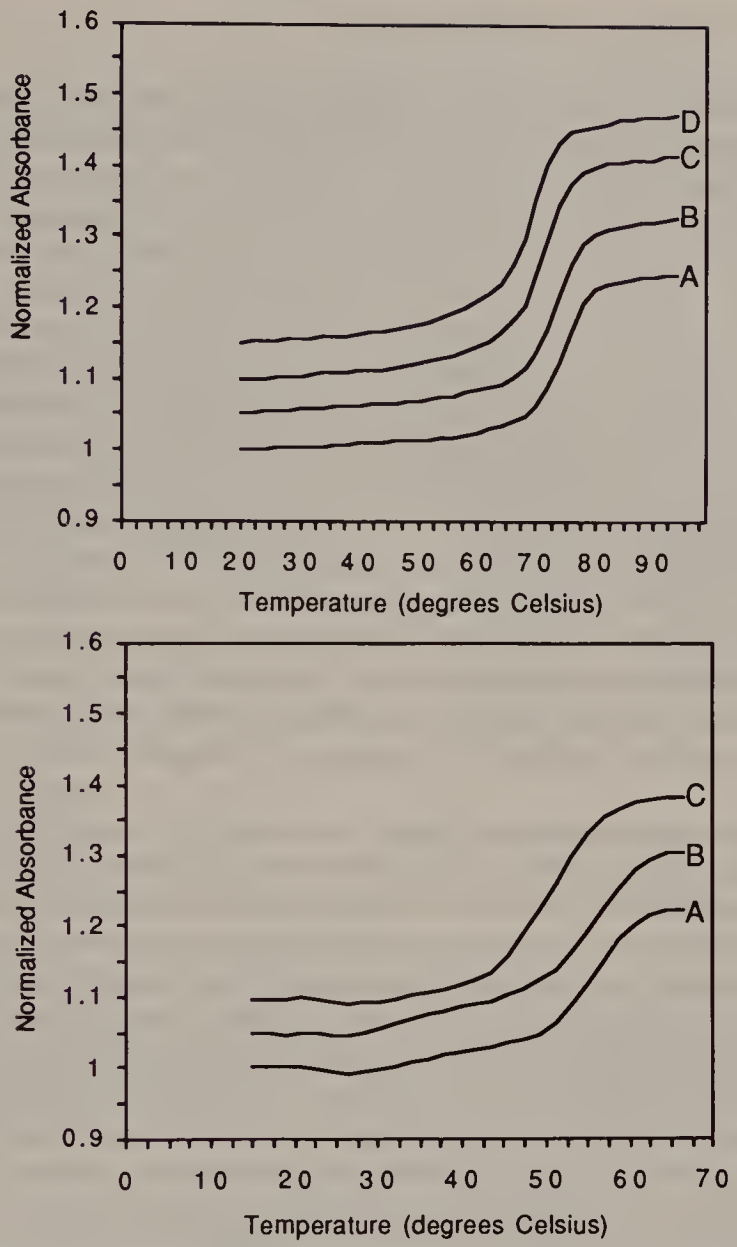


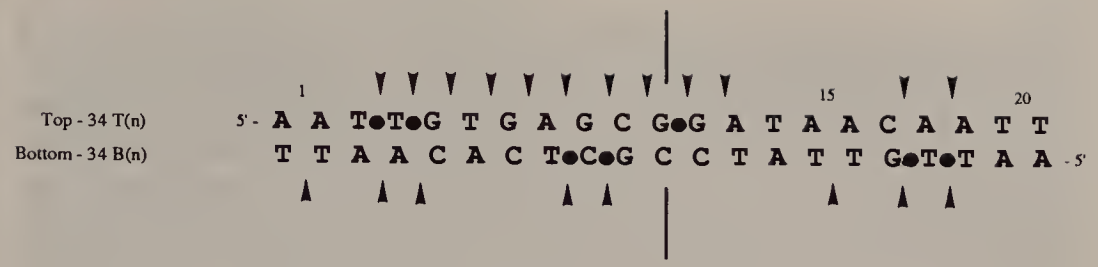
FIG. 3. Studies of dithioate DNA duplexes. Oligonucleotides containing dithioate linkages were annealed to the complementary, unmodified DNA segments and the absorbance was monitored with temperature. x, phosphorodithioate; p, phosphate.
Top panel:
A, d(CpCpApApApApGpGpCpCpGpApGpApApGpCpGpApT);
B, d(CxCpApApApApGxGpCpCpGpApGxApApGpCpGpAxT);
C, d(CxCpApAxApApGxGpCpCxGpApGxApApGxCpGpAxT);
D, d(CxCpAxApAxApGxGpCxCpGxApGxApAxGpCxGpAxT).
Bottom panel:
A, d(GpApTpTpCpApGpCpTpApGpTpCpCpA);
B, d(GpApTpTpCxApGpCpTpAxGpTpCpCpA);
C, d(GpApTxTpCpAxGpCpTxApGpTxCpCpA).

Biochemical studies with dithioate DNA

It is of particular interest whether dithioate DNA is resistant to nucleases and capable of forming duplexes. Results with snake venom phosphodiesterase I and calf spleen phosphodiesterase II (Grandas et al 1989) and nuclease P_1 (Porritt & Reese 1990) indicate that this derivative is not degraded by nucleases. More recent results have also shown that dithioate DNA cannot be degraded by the 3′–5′ exonuclease activity of bacteriophage T4 DNA polymerase (L. Gold & M. Caruthers, unpublished results) or the nucleases present in HeLa cell nuclear extracts (L. Cummins & M. Caruthers, unpublished results). Recent data (Fig. 3) has also demonstrated that oligonucleotides containing dithioate linkages readily form stable duplexes, with somewhat reduced melting points (m.p.) relative to unmodified DNA. In a 20mer with 53% dithioate linkages (10 of 19), the m.p. is reduced by 5 °C. With smaller numbers of these linkages (4 of 19, 21%; 7 of 19, 37%), m.p. reductions of 1 °C and 3 °C respectively are observed. Similarly, with a 15mer having two or four dithioates per segment, m.p. reductions of 0.4 °C and 4.4 °C respectively suggest that the same trend continues with perhaps a more significant change in duplex stability for shorter segments. With oligonucleotides containing only dithioate linkages, duplexes are observed only with excess of the analogue oligonucleotide (m.p. studies are in progress).

Oligonucleotide dithioates have also been shown to induce RNase H activity in HeLa cell nuclear extracts. The 20mer with 10 dithioate linkages listed in Fig. 3 is complementary to the 5′-oligonucleotide sequence of U2 snRNA, an essential component of a cell's RNA-splicing machinery. When placed in a HeLa cell nuclear extract, this compound stimulates endogenous RNase H to degrade U2 RNA (L. Cummins, data not shown) in a manner identical to earlier results with the same DNA sequence with natural internucleotide linkages (Black et al 1985, Krainer & Maniatis 1985). However, unlike the normal oligonucleotide, the dithioate derivative stimulates complete degradation of U2 RNA. This is because the natural segment is degraded by endogenous nuclease, whereas the dithioate oligomer, even with only 10 modified linkages, appears to be resistant to degradation. These results are very encouraging for experiments in which the objective is to control RNA synthesis.

Also of considerable interest is whether oligonucleotides containing phosphorodithioate linkages significantly alter the conformation of DNA duplexes. In order to test this possibility and also to gain new insights on how these compounds may be used for various biochemical studies, the *lac* operator–*lac* repressor system was examined. The basic approach was to prepare a series of *lac* operators in which each duplex contained one phosphorodithioate linkage and then to test how this modification altered the ability of the *lac* repressor to recognize the operator and form a stable complex. A summary of these results is presented in Fig. 4. Generally, minor reductions in binding affinity were



Duplex	Relative affinity	Duplex	Relative affinity
34 T(3)	0.7	34 B(1)	0.9
T(4)	does not bind	B(3)	does not bind
T(5)	0.5	B(4)	0.9
T(6)	0.7	B(8)	1.2
T(7)	0.5	B(9)	0.3
T(8)	0.8	B(15)	0.8
T(9)	0.5	B(17)	does not bind
T(10)	0.5	B(18)	0.5
T(11)	0.4		
T(12)	1.1		
T(17)	1.0		
T(18)	0.8		

FIG. 4. Studies of the *lac* repressor–*lac* operator interaction. Assays of repressor–operator interactions were completed on nitrocellulose filters using standard procedures. The symbols T and B refer to the top and bottom strands. Relative affinity is the ratio of equilibrium binding constants for modified, dithioate-containing operators to unmodified operators. Although each oligonucleotide used in the binding assay was 34 nucleotides in length, only the central portion corresponding to the *lac* operator is shown. The numbers 1–20 refer to internucleotide phosphate or phosphorodithioate linkages. The duplexes assayed are defined by the position of the phosphorodithioate linkage in each duplex. For example T(3) and B(1) refer to two duplexes containing dithioate internucleotide linkages at positions 3 and 1 on the top (T) and bottom (B) strands, respectively. Both strands are numbered 1–20 left to right as shown. The axis of symmetry of the operator is designated with a vertical line and the sites in operators that contain dithioate linkages are indicated by arrow heads.

observed (0–70%), although at two sites (T12 and B8) a slight enhancement in complex stability was seen. This information can then be used to introduce probes, reporter groups and cross-linking agents such as fluorescein or heavy metals without altering the ability of the *lac* repressor to recognize the *lac* operator (H. Sasmor and G. Beaton, unpublished results). This approach should be quite general and provide new insights into protein–nucleic acid complexes.

There are, however, three sites (T4, B3 and B17) at which insertion of a phosphorodithioate linkage leads to a complete loss in affinity of repressor for operator, as assayed by binding of the complex to nitrocellulose filters. These sites are either symmetrical with one another (T4 and B17) or with the known nuclease-sensitive site (Galas & Schmitz 1978) that is associated with the formation of a stable *lac* repressor–*lac* operator complex (B3 and T18). This data may be telling us something about conformational changes associated with formation of a *lac* repressor–*lac* operator complex. Perhaps the conformation the *lac* operator has in the absence of the *lac* repressor is altered in some way when it forms a complex with the repressor. In this model, the key DNA sites involved in conformational changes would be T4, B3, T18 and B17. Insertion of a phosphorodithioate at sites T4, B3 or B17 would either prevent the desired conformational change or alter the operator structure in such a way that recognition by the repressor is reduced. Because repressor–operator affinity is unaltered when a dithioate linkage is inserted at T18, this model would argue that a dithioate at T18 aids the interaction. The nuclease sensitivity of this site (but not the symmetrical B3 position) argues for different protein–DNA conformations with the two operator half-sites.

Because oligonucleotide analogues with modifications in the phosphate diester linkage have potential therapeutic applications (Zon 1989), the phosphorodithioate DNA derivative has been investigated in this regard. In our first experiments, the inhibition of retroviral reverse transcriptases by oligodeoxynucleotide phosphorodithioates was examined. The results are summarized in Table 1. The assay involves carrying out repair synthesis with

TABLE 1 A summary of ID_{50}[a] values for phosphorodithioate DNA inhibition of human immunodeficiency virus reverse transcriptase

	Compound[b]	*ID_{50}*
I	d(CpCpCpCpCpCpCpCpCpCpCpC)	36 μM
II	d(C–C–C–C–C–C–C–C–C–C–C–C–C–C–C–C)	1.7 μM
III	d(CxCxCxCxCxCxCxCxCxCxCxC)	60 nM
IV	d(CxCpCxCpCxCpCxCpCxCpCxCpC)	2.0 μM
V	d(TxTxTxTxTxTxTxTxTxTxTxT)	30 nM
VI	d(AxAxAxAxAxAxAxAxAxAxAxA)	75 nM
VII	d(CxCxCxCxCxCxCxCxCxCxCxCxCxCxCxCxCxC)	10 nM
VIII	d(CxCxCxCxCxCxCxCxC)	220 nM
IX	d(CxCxCxCxCxCxC)	1.2 μM
X	d(CxCxCxC)	20 μM

[a]Concentration of oligodeoxynucleotide that causes 50% inhibition of reverse transcriptase repair activity (see text for details).

[b]Internucleotide linkages are represented by x for phosphorodithioate, p for phosphate, and – for phosphorothioate.

human immunodeficiency virus (HIV) reverse transcriptase using a primer (15mer) in complex with a template (30mer) in the absence or presence of the oligodeoxynucleotide phosphorodithioate at increasing concentration. The assay was completed under saturating conditions of primer-template and deoxynucleotide triphosphates and the results were analysed by denaturing polyacrylamide-gel electrophoresis. The concentration of oligodeoxynucleotide dithioate that inhibits this repair reaction to the 50% level is called the ID_{50} and is used to compare data.

The results indicate that phosphorodithioate DNA strongly inhibits HIV reverse transcriptase. An oligodeoxycytidine 14mer containing only phosphorodithioate linkages (III) has an ID_{50} value (50% inhibition) with HIV reverse transcriptase of 60 nM. The corresponding normal oligonucleotide (I) is 600-fold less inhibitory ($ID_{50} = 36\,\mu M$), whereas the oligodeoxycytidine phosphorothioate 15mer (II) is only 28-fold less effective ($ID_{50} = 1.7\,\mu M$) than the phosphorodithioate (III). Also, the extent of inhibition correlates with the content of phosphorodithioate. For example, compound IV with 50% phosphorodithioate and 50% phosphate diester linkages is 33-fold less inhibitory ($ID_{50} = 2\,\mu M$) than the phosphorodithioate 14mer (III). Equally of interest were the observations that inhibition correlates with length (longer oligodeoxynucleotide dithioates are more inhibitory) and with the sequence of the homopolynucleotide (dT > dC > dA). Although the potential of these reagents as anti-HIV therapeutics is clear, future research must address the biochemistry (kinetics and the mechanism of complex formation), cell biology, and pharmacological utility of dithioate DNA.

Conclusions

These results demonstrate that dithioate DNA can be readily synthesized, is stable toward nucleases and forms duplexes with normal DNA and that it has several potential biochemical applications. The discovery of this attractive new DNA derivative should stimulate new research in the polynucleotide field and lead to the synthesis of even more exciting analogues in the years ahead.

Acknowledgements

This research was supported by NIH (GM21120 and GM25680).

References

Black DL, Chabot B, Steitz JA 1985 U2 as well as U1 small nuclear ribonucleoproteins are involved in premessenger RNA splicing. Cell 42:737–750

Brill WKD, Tang JY, Ma YX, Caruthers MH 1989 Synthesis of oligodeoxynucleoside phosphorodithioates via thioamidites. J Am Chem Soc 111:2321–2322

Caruthers MH 1985 Gene synthesis machines: DNA chemistry and its uses. Science (Wash DC) 230:281–285

Caruthers MH 1989 Synthesis of oligonucleotides and oligonucleotide analogues. In: Cohen JS (ed) Oligonucleotides: antisense inhibitors of gene expression, vol 12 of Topics in molecular and structural biology. Macmillan, Basingstoke p 7–22

Galas DJ, Schmitz A 1978 DNAase footprinting: a simple method for the detection of protein–DNA binding specificity. Nucleic Acids Res 5:3157–3170

Grandas A, Marshall WS, Nielsen J, Caruthers MH 1989 Synthesis of deoxycytidine oligomers containing phosphorodithioate linkages. Tetrahedron Lett 30:543–546

Krainer AR, Maniatis T 1985 Multiple factors including the small nuclear ribonucleoproteins U1 and U2 are necessary for pre-mRNA splicing in vitro. Cell 42:725–736

Porritt GM, Reese CB 1990 Use of the 2,4-dinitrobenzyl protecting group in the synthesis of phosphorodithioate analogues of oligodeoxyribonucleotides. Tetrahedron Lett 31:1319–1322

Yau EK, Ma YX, Caruthers MH 1990 Synthesis of dinucleoside and dinucleotide phosphorodithioates via a phosphotriester approach. Tetrahedron Lett 31:1953–1956

Zon G 1989 Pharmaceutical considerations. In: Cohen JS (ed) Oligonucleotides: antisense inhibitors of gene expression, vol 12 of Topics in molecular and structural biology. Macmillan, Basingstoke p 233–247

DISCUSSION

Roberts: You said that the dithioate linkage is resistant to nucleases. In your experiment involving RNA cleavage, the oligonucleotide that you used had alternating dithioate and ordinary linkages. Are the ordinary linkages in those oligonucleotides protected from nucleases by their dithioate neighbours?

Caruthers: Yes, the ordinary linkages do appear to be protected against nuclease degradation. We have found that exposure to HeLa cell nuclear extracts for 21 hours leaves approximately 85% of the 20mer with alternating ordinary and dithioate linkages undegraded. We have no explanation for the protection mechanism. I suspect that if we looked more carefully with large excesses of nucleases, rather than using the HeLa cell extract, we would see complete cutting at the normal linkages.

Ron Breslow: Another way to protect against nucleases is to include either phosphonate or phosphinate linkages. What effect do these have? Is ribonuclease stimulated by phosphinates?

Caruthers: I would guess that phosphinates, especially those in which the 3′-oxygen is replaced by carbon, would be stable towards most nucleases. These compounds have not yet been synthesized. The neutral analogues, such as the methyl phosphonates, do not stimulate ribonuclease H.

Ron Breslow: I was thinking about anionic phosphonic acids with a negative charge in which carbon replaces a neutral internucleotide oxygen.

Caruthers: These would be very interesting analogues but they have yet to be synthesized.

Ron Breslow: If anionic phosphinate-linked DNA were stable to nucleases and base-paired with RNA, it would presumably activate ribonuclease H.

You have some wonderful compounds, so I don't know that one needs others; I am just curious whether instead of solving the problem using sulphur one could solve it with carbon.

Brändén: Are eukaryotic and prokaryotic DNA and RNA polymerases inhibited by dithioate DNA to the same extent as HIV reverse transcriptase?

Caruthers: They are inhibited, but to a somewhat lesser extent. We have looked at α-polymerase and the Klenow fragment of *E. coli* DNA polymerase. Both are inhibited about 20-fold less than HIV reverse transcriptase under the same assay conditions. This inhibition, however, does not appear to have a toxic effect on cells in the experiments we have done so far. Perhaps these compounds don't get into the nucleus, or perhaps the α-polymerase is tied up as part of the replisome (DNA-replicating complex) and is unavailable for binding to these oligonucleotides.

Crumpton: How do you think a 14mer gets into cells? The lack of effect of dithioate oligonucleotides on viability would fit in with a failure to get into the cells. Was inhibition of HIV reverse transcriptase assayed in a cell-free assay system?

Caruthers: Yes, that was done with purified reverse transcriptase. We have not studied the transport of oligonucleotides into cells, but others have. Currently, three different methods to get anionic oligonucleotides into cells are being investigated. One method involves packing them into liposomes. Loke et al (1989) have shown that there is a protein which binds and transports oligonucleotides. Phosphorothioate DNA segments are transported to a much lesser extent than normal oligomers by this system and we don't know whether it will transport dithioate oligonucleotides into cells. The third method involves derivatizing DNA with cholesterol (Letsinger et al 1989). Presumably, oligonucleotides could be transported into most cells using this approach.

Several people have investigated the transport of methylphosphonate DNA, which is uncharged. Such compounds seem to diffuse into cells, but, of course, they also diffuse out. Therefore you have to maintain extremely high concentrations to see a result (Kulka et al 1989).

Rebek: What is the p*K* of the dithioate linkage?

Caruthers: We haven't measured it in our oligonucleotides. The p*K* values of *o,o*-dialkylphosphorodithioate esters are about half a p*K* unit more acidic than the oxygen analogues.

We don't know why these dithioate oligonucleotides bind more tightly to polymerases than normal DNA of the same sequence. Perhaps it's an ionic effect. Alternatively, sulphur might be more polarizable than oxygen and thus interact more strongly through van der Waals contacts. Additional studies using different salt and solvent conditions may help us address this question.

Waring: There's a great deal of literature about the effects of ions on the melting temperature of DNA. Your dithioate compounds would presumably react differently from normal phosphates to ions as regards site binding (as

opposed to non-specific electrostatic interactions). Have you looked for such differences?

Caruthers: We have done two relevant studies. Using the *lac* operator with dithioate linkages shown in Fig. 4 we have measured the stability of the *lac* repressor–*lac* operator interaction in relation to various heavy metals. Mercury and silver block the ability of the *lac* repressor to recognize dithioate-containing operators. This suggests that such metals bind very tightly to phosphorodithioate internucleotide linkages.

In the other study we measured the rate of DNA synthesis by HIV reverse transcriptase and the Klenow fragment of *E. coli* DNA polymerase I using different primers. Polymerization rates with primers containing dithioate internucleotide linkages were faster than with the same unmodified primer–template system. As these enzymes require magnesium ions, and the Klenow fragment perhaps requires zinc, our results suggest that dithioate DNA does not inactivate polymerases by chelating enzyme active site metal ions.

References

Kulka M, Smith CC, Aurelian L et al 1989 Site specificity of the inhibitory effects of oligo(nucleoside methylphosphonate)s complementary to the acceptor splice junction of herpes simplex virus type 1 immediate early mRNA 4. Proc Natl Acad Sci USA 86:6868–6872

Letsinger RL, Zhang GR, Sun DK, Ikeuchi I, Sarin PS 1989 Cholesteryl-conjugated oligonucleotides: synthesis, properties, and activity as inhibitors of replication of human immunodeficiency virus in cell culture. Proc Natl Acad Sci USA 86:6553–6556

Loke SL, Stein CA, Zhang XH et al 1989 Characterization of oligonucleotide transport into living cells. Proc Natl Acad Sci USA 86:3474–3478

Conformational flexibility and protein specificity

G. C. K. Roberts

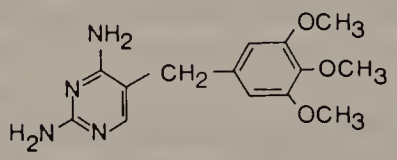
Department of Biochemistry and Biological NMR Centre, University of Leicester, University Road, Leicester LE1 7RH, UK

Abstract. Dihydrofolate reductase is of pharmacological interest as the target for a number of useful drugs including methotrexate, trimethoprim and pyrimethamine. The binding of substrates, inhibitors and coenzymes to the enzyme has been studied by NMR spectroscopy in conjunction with site-directed mutagenesis; changes in the chemical structure of the protein or the ligand are found to have a variety of effects on both the time-average conformation and its fluctuations. These experiments have revealed a number of instances where the specificity is influenced by conformational fluctuations or equilibria. Both the substrate, folate, and the coenzyme, $NADP^+$, have alternative modes of binding to the enzyme. The pteridine ring of folate can bind in two distinct orientations, depending on the ionization state of Asp-26. In the enzyme–trimethoprim–$NADP^+$ complex, the nicotinamide ring of the coenzyme exists in two states, either bound to the enzyme or hanging free in solution. Analogues of the inhibitor pyrimethamine exist in two slowly interconverting conformations, both of which are able to bind to the enzyme. In all these cases the different modes of ligand binding have very similar binding energies and play a role in determining the specificity of the enzyme.

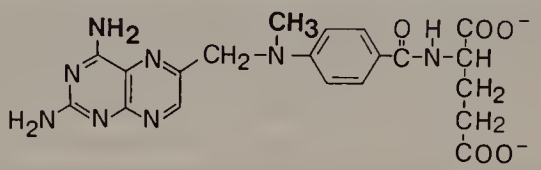
1991 Host–guest molecular interactions: from chemistry to biology. Wiley, Chichester (Ciba Foundation Symposium 158) p 169–186

Emil Fischer's well-known 'lock and key' analogy for the interaction of substrates with enzymes (and, by extension, for all kinds of host–guest interactions) has provided successive generations of scientists with their mental picture of molecular recognition processes for almost a century. With the increase in available structural information over the last thirty years, principally from X-ray crystallography, it has become clear that this picture is not only over-simplified but also, in some respects, misleading. First, like all molecules, proteins and their complexes with ligands have structures which fluctuate at a wide variety of rates, and molecular recognition is, by its nature, a dynamic process. Structural fluctuations, in addition to time-average conformations, help to determine specificity. Second, while there is only one way in which a key will fit a lock, protein–ligand interactions are often more permissive. Not only can

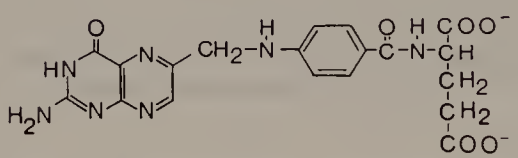
closely related molecules bind in distinctly different ways, but sometimes even a single molecule can bind in two (or more) conformations or orientations with closely similar binding energies. In discussing these effects, I shall take examples from recent work on dihydrofolate reductase (dhfr) from *Lactobacillus casei*.



Trimethoprim



Methotrexate



Folate

Dihydrofolate reductase (EC 1.5.1.3) is the 'target' for the important 'anti-folate' drugs such as methotrexate, trimethoprim and pyrimethamine. Its pharmacological importance, coupled with its conveniently small size (M_r 18 000–25 000 in most species), has led to intensive studies of this enzyme over the last 10–12 years (Blakley 1985, Roberts 1989, 1990). High resolution crystal structures have been determined for dhfr from *L. casei* and other organisms in a number of ligand complexes (e.g. Bolin et al 1982, Volz et al 1982, Oefner et al 1988), and crystal structures are also available for mutants of the *E. coli* enzyme (Howell et al 1986). ^{1}H resonances from more than 75% of the residues of *L. casei* dhfr have been assigned (Hammond et al 1986, Birdsall et al 1990a, Carr et al 1990), and, as a result, 'reporter groups' are available throughout the protein structure. An enormous range of inhibitors of this enzyme has been

synthesized over the last 40 years (Roth & Cheng 1982), and much more recently a number of studies of dhfr by site-directed mutagenesis have been reported (e.g. Howell et al 1986, Mayer et al 1986, Benkovic et al 1988, Prendergast et al 1989, Birdsall et al 1989a, Jimenez et al 1989). One can thus make defined changes in the chemical structure of either partner in the complex and then ask questions about the roles of individual groups in the interaction.

Structural comparisons of mutants and complexes

The functional (kinetic and binding) effects of amino acid substitutions or of alterations in the chemical structure of the ligand can be understood only in the context of studies of their conformational effects. Comparisons of the structures of mutant and wild-type enzyme and of complexes with different ligands, both by crystallography and by NMR, have revealed considerable variation in the conformational consequences of chemical change. Crystallographic studies of the Asp-27→Asn (D27N) mutant* of the *E. coli* enzyme (Howell et al 1986) and NMR studies of the corresponding *L. casei* mutant, D26N (Jimenez et al 1989), show that any structural changes in the enzyme–methotrexate complex are very small and local (and may result from the displacement of a bound water molecule). Sometimes, however, the structural effects extend considerably further through the protein structure. In the case of the T63Q substitution, which involves a residue that normally makes a hydrogen bond to the 2′-phosphate of the coenzyme, NMR analysis shows that the conformational changes are again small, but they affect residues in the substrate binding site and are thus transmitted as far as 25 Å through the protein (J. A. Thomas, J. R. P. Arnold, G. C. K. Roberts, B. Birdsall, J. Feeney & J. Andrews, unpublished work).

Many of the structural comparisons by NMR of mutant dihydrofolate reductases and of complexes with different ligands have been made on the basis of chemical shift comparisons. These provide a simple and sensitive method for detecting structural differences, but it is not generally possible to use observed shift differences to describe the precise *nature* of the difference in conformation. More recently, we have been using nuclear Overhauser effects (NOEs) to obtain a description of these (generally) small structural differences (I. Barsukov, J. R. P. Arnold, J. A. Thomas, G. C. K. Roberts, B. Birdsall & J. Feeney, unpublished work). Further work on the necessary methods is in progress, but the preliminary results are encouraging and hold out the prospect of detailed structural comparisons of complexes in solution. For example, in the case of the mutant T63Q, detailed analysis of NOEs suggests that a likely mechanism for

*The nomenclature used to describe mutants with single amino acid substitutions is based on the single-letter code for the amino acids; thus a mutant in which aspartic acid at position 27 of the wild-type sequence has been replaced by asparagine is denoted D27N. Unless otherwise specified, mutants discussed are those of the *L. casei* enzyme.

the transmission of the conformational effects from the site of substitution to the substrate site is by a slight movement of the α-helix formed by residues 42–49.

Similar comparisons can illuminate the effects of modification of the ligand molecule. For example, the role of the two carboxylate groups of methotrexate, which form ion pairs with groups on the protein (His-28-γ-carboxylate and Arg-57-α-carboxylate), have been studied by using amide analogues (Antonjuk et al 1984, Hammond et al 1987). The γ-amide of methotrexate binds only nine-fold less tightly to the enzyme than methotrexate itself does; NMR evidence indicates that these two complexes are essentially isostructural, and hence that the difference in binding constant is a reasonable measure of the modest contribution of the His-28-γ-carboxylate ion pair to the overall interaction energy. The K_i value for the α-amide, in contrast, is 100-fold less than that of methotrexate itself, apparently indicating that the Arg-57-α-carboxylate ion pair makes a larger contribution to binding. However, the effects of binding of the α-amide on the NMR spectrum indicate that in this case the whole of the *p*-aminobenzylglutamate moiety of the α-amide analogue binds differently, and the His-28-γ-carboxylate ion pair is also broken, so that the difference in binding energy cannot be interpreted in a simple way.

When changes are made to the same region of the binding site, a convergence of the effects of ligand and protein modification is often seen. For example, both 3′,5′-difluoro- and 3′,5′-dichloromethotrexate bind about 100 times less tightly to the enzyme than methotrexate itself does. Structural comparison of their complexes shows that the binding of the pteridine ring and of the glutamate moiety is unaffected, but that the orientation of the benzoyl ring about its axis of symmetry changes by up to 25° to accommodate the bulky substituents (Hammond et al 1987), the change in binding being attributed to changes in non-polar interactions with this ring. Similarly, methotrexate binds 300-fold less tightly to the *E. coli* mutant enzyme L54G, in which one of the side chains involved in these non-polar contacts has been removed, than to the wild-type enzyme (Mayer et al 1986), although in this case the structural consequences are not known. In studies of NADPH binding to the *L. casei* mutant enzyme W21L (Birdsall et al 1989a) a qualitative parallel was seen between the effects of this amino acid substitution and those of replacing the nicotinamide carboxamide with a thioamide (Birdsall et al 1980a). In both cases, the binding constant of the reduced coenzyme is decreased substantially; the effects probably arise from a perturbation of the non-polar interaction between Trp-21 and the nicotinamide that leads to a change in the orientation of the nicotinamide ring in the binding site (Birdsall et al 1989a).

NMR studies of structural fluctuations

NMR can provide information on molecular motion over a wide time scale, from relaxation effects at $10^9\,s^{-1}$, through effects on the shape of

the resonance signal (lineshape effects) at 10^2–10^3 s^{-1}, to hydrogen exchange at 10^{-1}–10^2 s^{-1}. Structural fluctuations in proteins and their complexes occur at rates that span this whole range, as illustrated by our studies of the conformational fluctuations of the inhibitor trimethoprim (2,4-diamino-5-[3′,4′,5′-trimethoxybenzyl]pyrimidine) bound to dhfr. Motions of the symmetrically substituted benzyl ring of trimethoprim were studied using [*m*-methoxy ^{13}C]trimethoprim (Searle et al 1988). Relaxation measurements revealed fluctuations of ± 25–$35°$ at rates of about 10^9 s^{-1}, while lineshape analysis of the ^{13}C resonances shows that 180° 'flips' about the symmetry axis of the benzyl ring are much rarer, but do occur at 250 s^{-1} at room temperature. A major contribution to the barrier to this 'flipping' comes from steric interactions *within* the trimethoprim molecule, between the benzyl and pyrimidine rings, and molecular mechanics calculations show that the flipping of the benzyl ring must be accompanied by a change of 30–60° in the relative orientation of the two rings in bound trimethoprim. This in turn requires transient movements of the amino acid side chains packed around the ring.

Inhibitors such as trimethoprim bind in the protonated state, making crucial electrostatic and hydrogen-bonding interactions with the carboxylate of Asp-26, which accepts hydrogen bonds from the hydrogen at N-1 and the 2-amino group of the inhibitor, while the second proton of the 2-amino group hydrogen bonds, through a water molecule, to Thr-116. By using ^{15}N-labelled trimethoprim and methotrexate, together with inverse detection experiments, we have located the resonances of the protons involved in all three of these hydrogen bonds; this allows us to measure the rate of exchange of these protons with the solvent and hence the rate of making and breaking of the hydrogen bonds between inhibitor and enzyme. In the case of the trimethoprim N-1-H—Asp-26 hydrogen bond, this occurs at a rate of 34 s^{-1} at room temperature (Searle et al 1988) and clearly results from a fluctuation in the protein structure distinct from that involved in the ring flipping described above. It seems very likely that if both fluctuations happened to occur simultaneously, the dissociation of trimethoprim from the enzyme would result.

The rates of all these fluctuations in the structure of the enzyme–trimethoprim complex are affected by coenzyme binding (Searle et al 1988) and also by amino acid substitutions. The mutant D26E is kinetically and structurally similar to the wild-type enzyme (Andrews et al 1989, Birdsall et al 1989a), but there are clear differences in the dynamics of trimethoprim bound to this mutant enzyme. The rates of the conformational fluctuations discussed above are increased as much as 30-fold in this mutant (Birdsall et al 1989a). It appears that structural fluctuations are significantly more sensitive to the effects of amino acid substitutions than is the time-average structure.

Alternative modes of substrate and coenzyme binding

The dhfr inhibitor methotrexate is a close structural analogue of the substrate, folate. The molecules differ in only two ways: there is a methyl substituent (which does not influence the binding to dhfr) on N-10 in methotrexate but not in folate, and a 4-amino rather than a 4-oxo substituent on the pteridine ring of methotrexate. In spite of this close structural similarity, methotrexate is an inhibitor rather than a substrate, and binds to the enzyme some 10^5 times more tightly than folate. Furthermore, the stereochemistry of reduction implies a difference of 180° in the orientation of the pteridine ring between substrate and inhibitor (Charlton et al 1985 and references therein). NMR experiments have revealed a further complication: both the enzyme–folate and the enzyme–folate–$NADP^+$ complexes exist in solution as a mixture of slowly interconverting conformations whose proportions are pH dependent (Birdsall et al 1989b and references therein). Comparison of chemical shifts between the three conformations of the dhfr–folate–$NADP^+$ complex indicates that the structural differences are localized to the active site region, close to the pteridine and nicotinamide rings. NOE experiments (Fig. 1) have recently shown that one important difference between these conformations is the orientation of the pteridine ring of the substrate. In the conformations denoted I and IIa, NOEs are observed between the proton on C-7 of the pteridine portion of bound folate and the two methyl groups of Leu-27, indicating that the ring has the same orientation in the binding site as that of methotrexate. In conformation IIb, in contrast, no such NOEs are observed; given the structural constraints of the pteridine ring binding site, this can only be explained if the ring has turned over by approximately 180° about an axis along the C-2—NH_2 bond (Birdsall et al 1989b). Only the latter is a catalytically productive mode of binding for the substrate. In this case, therefore, the substrate is able to bind, with very similar affinity, in both productive and non-productive orientations. The complex of folate and $NADP^+$ with the mutant D26N exists in only a single conformation, corresponding to the low pH conformer of the wild-type complex, over the pH range 5.5–7.5 (Fig. 2; Jimenez et al 1989), strongly suggesting that the carboxylate of Asp-26 is the group responsible for the pH dependence of the conformational equilibrium. The inhibitor methotrexate binds *only* in the non-productive orientation, most probably because in that orientation it can form an ion pair with Asp-26 (see above).

The complex between dhfr, trimethoprim and $NADP^+$ also exists in an equilibrium between two conformations; this equilibrium is independent of pH, and involves primarily the nicotinamide ring of the coenzyme. In one conformation, this ring is bound specifically to a site on the enzyme, but, in the other, rotations about the nicotinamide ribose C-5′—O and pyrophosphate P—O bonds alter the conformation of the coenzyme so that the nicotinamide ring is no longer in contact with the enzyme, but hanging free in solution, with

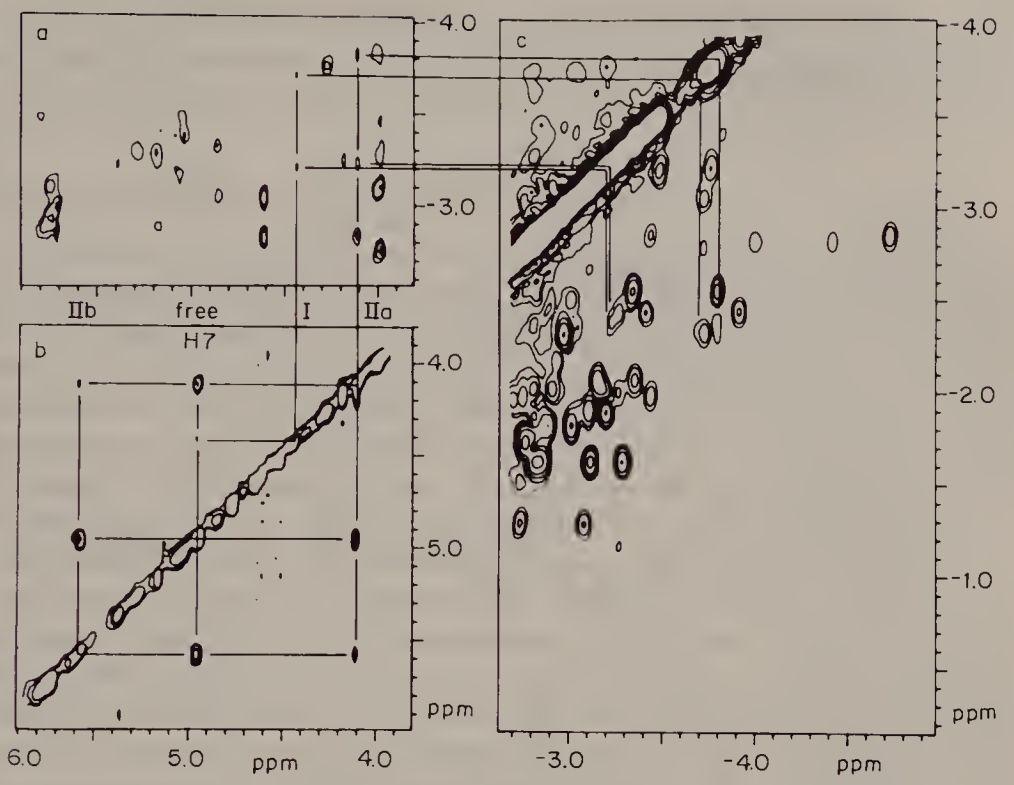


FIG. 1. Nuclear Overhauser effects (NOEs) involving the proton on C-7 of the pteridine portion of folate and methyl groups of the protein in the three conformations of the dihydrofolate reductase–folate–NADP$^+$ complex. Part (b) shows the aromatic region of the two-dimensional exchange spectrum obtained from a sample containing the complex and a two-fold excess of free folate. The exchange cross-peaks linking the 7-proton of free folate at 4.95 p.p.m. to the resonances of the same proton in the three conformations of the complex are connected by lines. Part (a) shows the region of a two-dimensional NOESY (nuclear Overhauser effect spectroscopy) spectrum containing NOE cross-peaks between aromatic and methyl proton resonances. Cross-peaks involving the 7-proton resonance in conformations I and IIa are indicated; in this region of the spectrum there are no cross-peaks involving the 7-proton resonance in conformation IIb. Part (c) shows the high-field region of the COSY (correlated spectroscopy) spectrum of the complex, which helps to identify the methyl protons giving NOEs to the 7-proton as those of Leu-27. From Birdsall et al (1989b) with permission.

the coenzyme being bound solely through its adenosine moiety (Gronenborn et al 1981, Birdsall et al 1984). The fact that these two conformations are almost equally populated shows that the oxidized nicotinamide ring contributes very little to the overall binding energy (for additional evidence see Birdsall et al 1980a). The most likely reason for this is that the favourable interactions between the oxidized nicotinamide ring and the enzyme are almost completely offset by the energetic cost of desolvating the positively charged ring. The relative populations of the two conformations of the enzyme–trimethoprim–NADP$^+$ complex are readily altered by changes in coenzyme or inhibitor structure

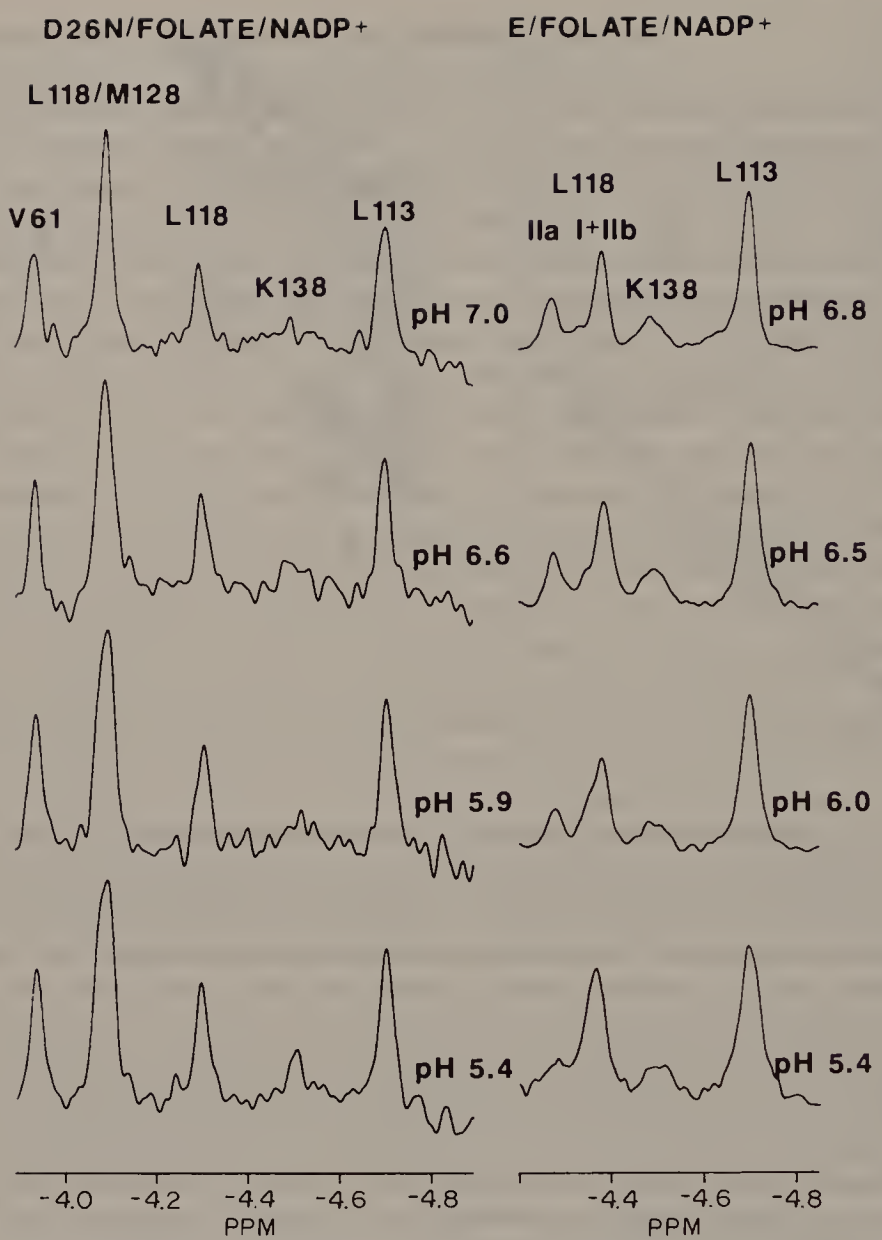


FIG. 2. Comparison of the pH dependence of the high-field region of the ^1H spectra of the complex of folate and $NADP^+$ with, *right*, wild-type dihydrofolate reductase (E) and, *left*, the Asp-26→Asn mutant (D26N). In the wild-type complex the methyl resonance of Leu-118 has two components at high pH, corresponding to conformations IIa and (I + IIb) (see Fig. 1); in the mutant complex, however, only a single, pH independent resonance from Leu-118 is observed. Letters are single-letter codes for amino acids. From Jimenez et al (1989) with permission.

(Gronenborn et al 1981) or by amino acid substitutions in the protein (Birdsall et al 1989a).

These two examples illustrate that, in some circumstances, a given ligand can bind to the same site on a protein in more than one way with very similar binding

energy. There is no reason to believe that dhfr is special in this respect; one must consider the possibility that the problem of fitting two complementary electron distributions together can in general have more than one solution.

Binding of conformational isomers of the ligand

Most small molecules, of course, exist in solution as a mixture of a large number of different conformations, only one of which will exist in the complex formed when the small molecule binds to a protein. (One tends to think of the ligand having 'a' conformation when bound, although, as described above, substantial fluctuations in this conformation can and do occur.) This conformational selection process is obviously an important part of molecular recognition. Under favourable circumstances, when the interconversions between different conformations of the ligand are sufficiently slow, one can study the conformational selection process directly (e.g. Birdsall et al 1980b).

We have recently been studying the binding to dhfr of a number of analogues of the antimalarial pyrimethamine (2,4-diamino-5-[4′-chlorophenyl]-6-ethylpyrimidine) synthesized by Stevens and his colleagues at Aston University (Tendler et al 1988, Birdsall et al 1990b). Rotation about the phenyl–pyrimidine bond of pyrimethamine is, as expected, slow, and four non-equivalent aromatic proton resonances are seen for pyrimethamine bound to the enzyme. In this case, of course, the two rotational isomers about the phenyl–pyrimidine bond are identical. This is, however, not the case in unsymmetrically substituted compounds such as 2,4-diamino-5-[3′-fluorophenyl]-6-ethylpyrimidine (FNP). We have found that this compound binds to the enzyme in two conformational states, A and B, the ratio A : B in the binary complex being 60 : 40. Comparison of the chemical shifts of the protons of the phenyl ring in pyrimethamine and its analogues bound to the enzyme reveals a very similar pattern in each case, as shown in Fig. 3. This indicates that the phenyl ring protons of FNP are experiencing essentially the same environment in each of the two forms, A and B, as that experienced by the corresponding protons in pyrimethamine. This can only be true if A and B represent the two conformational isomers resulting from rotation by 180° about the pyrimidine–phenyl bond of FNP, with the 2,4-diaminopyrimidine ring being bound in the same way in each case.

Measurements of NOEs between an *ortho*-proton of pyrimethamine (which, as shown by its chemical shift, corresponds to the 2′-proton of FNP in conformation A) and the *meta*- and *para*-protons of Phe-30 allow us to distinguish the two isomers of the unsymmetrically substituted compounds. We can thus construct models of the two complexes of FNP; these are depicted in Fig. 4. These models account, in a semi-quantitative way, for the large difference in chemical shift between the protons on either side of the phenyl ring (cf. Fig. 3), suggesting that it results from the magnetic anisotropy of

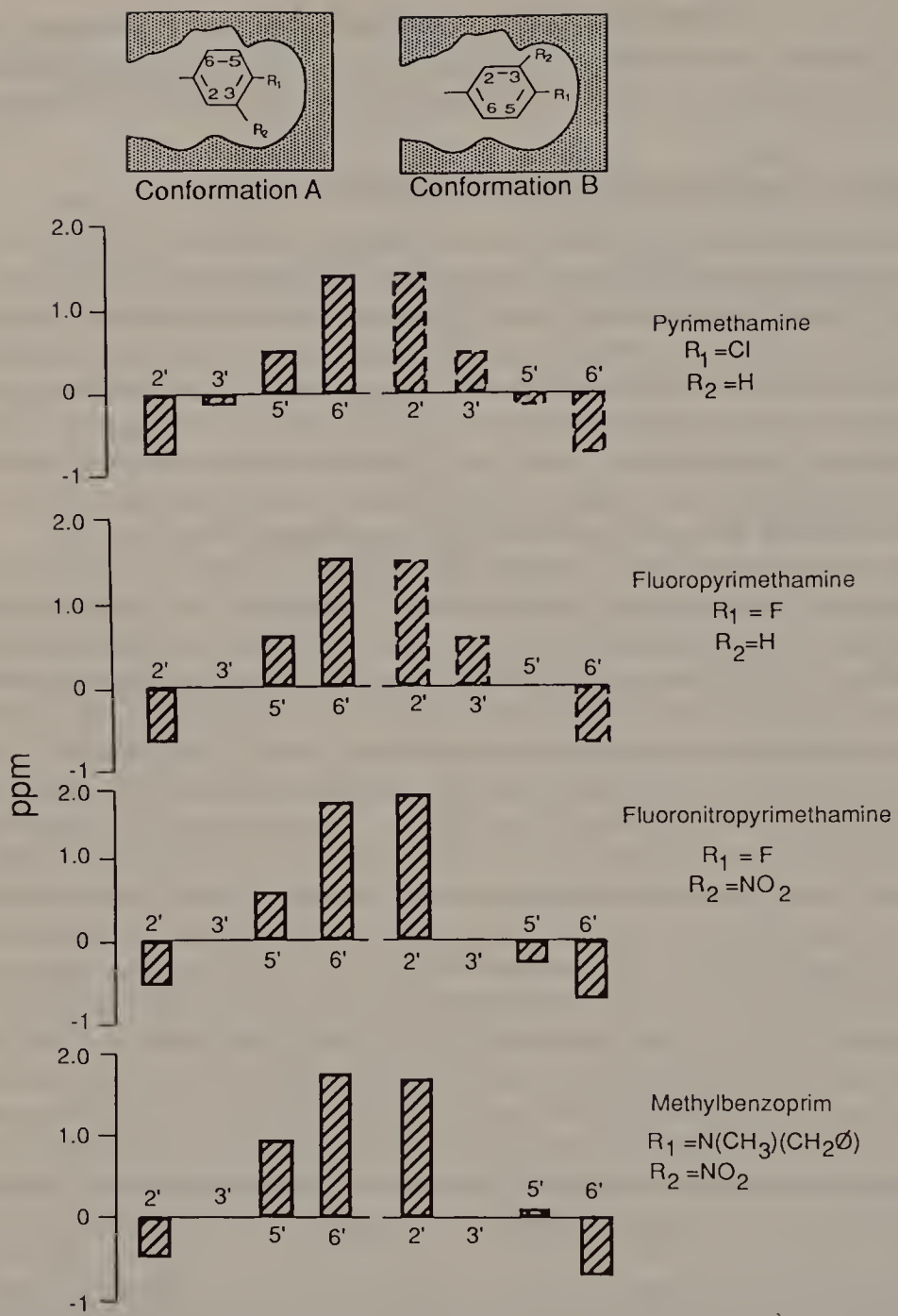


FIG. 3. Schematic representation of the ^1H chemical shift differences between enzyme-bound and free species for pyrimethamine, fluoropyrimethamine, fluoronitropyrimethamine and methylbenzoprim. Note that for the symmetrically substituted compounds pyrimethamine and fluoropyrimethamine the two orientations of the ring are equivalent; the chemical shifts have been indicated (as dashed bars) for conformation B, to facilitate comparison with the data for the other compounds. ϕ represents a phenyl ring. From Birdsall et al (1990b) with permission.

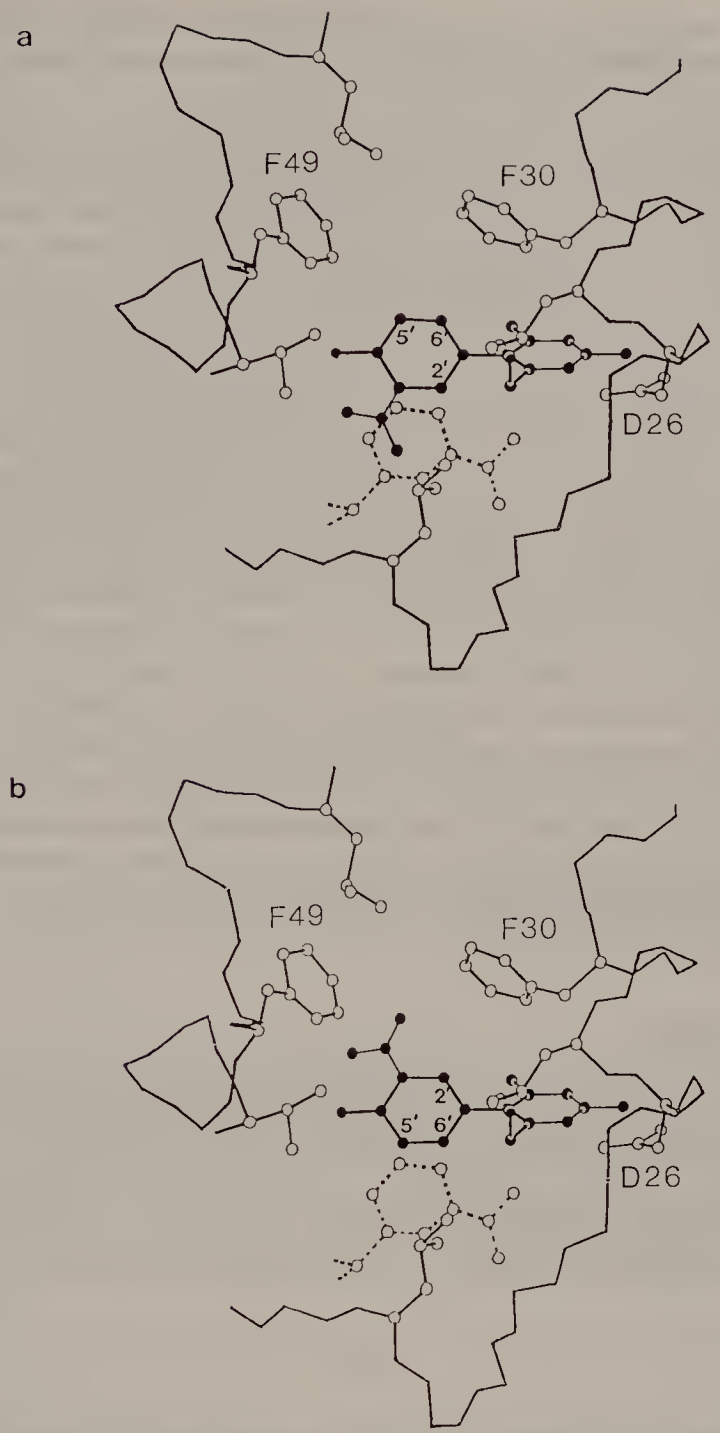


FIG. 4. Model of the binding site in the complex of dihydrofolate reductase with 2,4-diamino-5-[3′-nitro-4′-fluorophenyl]-6-ethylpyrimidine in conformations (a) A and (b) B (see Fig. 3). The position of the nicotinamide ring of the coenzyme is indicated by dashed bonds. The model takes as its starting point the crystallographic data of Bolin et al (1982) for the *L. casei* enzyme–methotrexate–NADPH complex. From Birdsall et al (1990b) with permission.

Phe-30. The models in Fig. 4 are, however, slightly oversimplified in that they represent the two isomers as related by a rotation about the phenyl–pyrimidine bond of exactly 180°. Analysis of the ^1H chemical shifts (e.g. comparison of the *meta*-proton shifts in A and B) and, most obviously, the observation that the chemical shift of the *para*-fluorine differs between the two complexes indicate that this is not entirely true. As yet, we do not know whether these additional effects arise because the angle of rotation is not exactly 180°, or because of small differences in protein conformation between the two complexes. (Any differences in protein conformation must be small, since we have not been able to detect any protein proton which gives separate signals for A and B.)

From Fig. 3 it can be seen that the polar, and potentially hydrogen-bonding, nitro group of FNP is in very different environments in the two isomeric complexes. Nevertheless, the two complexes have binding energies that differ by less than 0.3 kcal mol^{-1}. In conformation A, the nitro group extends partly into the binding site of the nicotinamide ring of the coenzyme and, as one would expect, the slight preference for form A is reversed by addition of coenzyme, an A : B ratio of 0.3 : 0.7 being observed for the ternary enzyme–FNP–$NADP^+$ complex (Tendler et al 1988, Birdsall et al 1990b). Again, the coexistence of these two distinct complexes between unsymmetrically substituted pyrimethamine analogues and the enzyme has important implications for understanding the relation between chemical structure and binding affinity. It is obvious that a given substitution may have different effects on affinity in isomers A and B, and in fact one would need to construct quite separate structure–activity relationships for the two isomers—something which can be done only with the aid of structural information.

Acknowledgements

The work described here is part of a continuing collaboration with Jim Feeney and Berry Birdsall (Mill Hill) and Julie Andrews (Manchester). Many people, whose names appear in the reference list, have made essential contributions to the experiments and their interpretation; my thanks to all of them, and particularly to John Arnold, Janette Thomas, Saul Tendler and Lu-Yun Lian. Work at Leicester has been supported by the Science and Engineering Research Council (Protein Engineering Club) and the Wellcome Trust.

References

Andrews J, Fierke CA, Birdsall B et al 1989 A kinetic study of wild-type and mutant dihydrofolate reductases from *Lactobacillus casei*. Biochemistry 28:5743–5750

Antonjuk DJ, Birdsall B, Cheung HTA et al 1984 A ^1H NMR study of the role of the glutamate moiety in the binding of methotrexate to dihydrofolate reductase. Br J Pharmacol 81:309–315

Benkovic SJ, Fierke CA, Naylor AM 1988 Insights into enzyme function from studies on mutants of dihydrofolate reductase. Science (Wash DC) 239:1105–1110

Birdsall B, Burgen ASV, Roberts GCK 1980a Binding of coenzyme analogues to *Lactobacillus casei* dihydrofolate reductase: binary and ternary complexes. Biochemistry 19:3723–3731

Birdsall B, Feeney J, Roberts GCK, Burgen ASV 1980b The use of saturation transfer NMR experiments to monitor the conformational selection accompanying ligand–protein interactions. FEBS (Fed Eur Biochem Soc) Lett 120:107–111

Birdsall B, Bevan AW, Pascual C et al 1984 Multinuclear NMR characterisation of two coexisting conformational states of the *Lactobacillus casei* dihydrofolate reductase–trimethoprim–$NADP^+$ complex. Biochemistry 23:4733–4742

Birdsall B, Andrews J, Ostler G et al 1989a NMR studies of differences in the conformations and dynamics of ligand complexes formed with mutant dihydrofolate reductases. Biochemistry 28:1353–1362

Birdsall B, Feeney J, Tendler SJB, Hammond SJ, Roberts GCK 1989b Dihydrofolate reductase: multiple conformations and alternative modes of substrate binding. Biochemistry 28:2297–2305

Birdsall B, Arnold JRP, Jimenez-Barbero J et al 1990a The ^{1}H-NMR assignments of the aromatic resonances in complexes of *Lactobacillus casei* dihydrofolate reductase and the origins of their chemical shifts. Eur J Biochem, 191:659–668

Birdsall B, Tendler SJB, Arnold JRP et al 1990b NMR studies of multiple conformations in complexes of *L. casei* dihydrofolate reductase with analogues of pyrimethamine. Biochemistry 29:9660–9667

Blakley RL 1985 Dihydrofolate reductase. In: Blakley RL, Benkovic SJ (eds) Folates and pterins. Wiley, New York, vol 1:191–253

Bolin JT, Filman DJ, Matthews DA, Hamlin RC, Kraut J 1982 Crystal structures of *Escherichia coli* and *Lactobacillus casei* dihydrofolate reductase refined at 1.7 Å resolution. J Biol Chem 257:13650–13662

Carr MD, Birdsall B, Frenkiel TA et al 1991 Dihydrofolate reductase: sequential assignments and secondary structure in solution. Biochemistry, in press

Charlton PA, Young DW, Birdsall B, Feeney J, Roberts GCK 1985 Stereochemistry of reduction of the vitamin folic acid by dihydrofolate reductase. J Chem Soc Perkin Trans I 1349–1353

Gronenborn A, Birdsall B, Hyde EI, Roberts GCK, Feeney J, Burgen ASV 1981 ^{1}H and ^{31}P NMR characterisation of two conformations of the trimethoprim–$NADP^+$–dihydrofolate reductase complex. Mol Pharmacol 20:145–153

Hammond SJ, Birdsall B, Searle MS, Roberts GCK, Feeney J 1986 Dihydrofolate reductase: ^{1}H resonance assignments and coenzyme-induced conformational changes. J Mol Biol 188:81–97

Hammond SJ, Birdsall B, Feeney J, Searle MS, Roberts GCK, Cheung HTA 1987 Structural comparisons of complexes of methotrexate analogues with *Lactobacillus casei* dihydrofolate reductase by two-dimensional ^{1}H NMR at 500 MHz. Biochemistry 26:8585–8590

Howell EE, Villafranca JE, Warren MS, Oatley SJ, Kraut J 1986 Functional role of aspartic acid 27 in dihydrofolate reductase revealed by mutagenesis. Science (Wash DC) 231:1123–1128

Jimenez MA, Arnold JRP, Andrews J et al 1989 Dihydrofolate reductase: control of the mode of substrate binding by aspartate 26. Protein Eng 2:627–631

Mayer RJ, Chen J-T, Fierke CA, Benkovic SJ 1986 Importance of a hydrophobic residue in binding and catalysis by dihydrofolate reductase. Proc Natl Acad Sci USA 83:7718–7720

Oefner C, D'Arcy A, Winkler FK 1988 Crystal structure of human dihydrofolate reductase complexed with folate. Eur J Biochem 174:377–385

Prendergast NJ, Appleman JR, Delcamp TJ, Blakley RL, Freisheim JH 1989 Effects of conversion of phenylalanine-31 to leucine on the function of human dihydrofolate reductase. Biochemistry 28:4645–4650

Roberts GCK 1989 Understanding the specificity of the dihydrofolate reductase binding site. In: Jardetzky O (ed) Protein structure and engineering. Plenum Publishing Corporation, New York (Proc NATO Advanced Study Institute) p 209–220

Roberts GCK 1990 NMR and mutagenesis studies of dihydrofolate reductase. In: Curtius HC, Ghilsa S, Blau N (eds) Pteridines and folic acid derivatives. De Gruyter, Berlin p 681–693

Roth B, Cheng CC 1982 Recent progress in the medicinal chemistry of 2,4-diamino-pyrimidines. Prog Med Chem 19:270–320

Searle MS, Forster MJ, Birdsall B et al 1988 Dynamics of trimethoprim bound to dihydrofolate reductase. Proc Natl Acad Sci USA 85:3787–3791

Tendler SJB, Griffin RJ, Birdsall B, Stevens MFG, Roberts GCK, Feeney J 1988 Direct ^{19}F NMR observation of the conformational selection of optically active rotamers of the antifolate compound fluoronitropyrimethamine bound to the enzyme dihydrofolate reductase. FEBS (Fed Eur Biochem Soc) Lett 240:201–204

Volz KW, Matthews DA, Alden RA et al 1982 Crystal structure of avian dihydrofolate reductase containing phenyltriazine and NADPH. J Biol Chem 257:2528–2536

DISCUSSION

Brändén: In your folate binding studies, do you have any indication of the exchange rates of the two modes of binding in relation to the catalytic rate? In other words, are half the molecules always inhibited?

Roberts: It's clear from the exchange measurements that the folate cannot turn over while it's bound to the protein. The way in which it interconverts between the two modes of binding is to dissociate, turn over and bind again. With these experiments on the enzyme–folate–$NADP^+$ complex, we are not dealing with a catalytically functional complex, because the wrong coenzyme is present. In this complex, the rate of dissociation of folate is too slow to be relevant to catalysis. I think the story may be different, however, in the presence of the correct coenzyme, NADPH, because one of the curiosities of this enzyme is that there is strong negative cooperativity between the binding of NADPH and the binding of the product of the reaction, tetrahydrofolate. If that's true in the case of folate as well, then it's possible that the dissociation rate might become fast enough to be relevant to catalysis. At present, we are trying to measure this dissociation rate directly by stop-flow.

Esther Breslow: The effects of substitution of glutamine for threonine are striking and they relate very well to Martin Karplus's calculations (Gao et al 1989) on the hidden effects of amino acid substitutions in proteins, in the sense that they can be far more extensive than one would ever imagine from simple criteria. Are there any other indications, apart from the NMR and the pK shift, of the conformational change associated with that substitution? Can you see it by other spectroscopic criteria, such as circular dichroism?

Roberts: We haven't looked at the CD spectrum. I would guess that we wouldn't see any effects in the far u.v. because as far as we can tell from the NMR spectrum there are no changes in secondary structure. We might see an effect in the near u.v. CD spectrum. Our evidence is basically from NMR and from the change in the pK of the aspartic acid, which is reflected in the pH dependence of k_{cat}. We wanted to know whether the cause of these unexpected effects was the absence of a threonine side chain or the presence of a glutamine side chain, so we made the threonine-to-alanine substitution; this has small local effects and no long-range effects. So, it's the presence of the glutamine side chain that causes the perturbations in the substrate binding site.

Edmundson: How do you visualize the fluctuations in the protein that allow your substrate to move?

Roberts: The best picture that I know of was produced some years ago by Martin Karplus and Bruce Gelin (Gelin & Karplus 1975; see also Brooks et al 1988). The kind of observations I have described for the benzoyl ring of methotrexate are analogous to what one sees in this and other proteins for the rings of phenylalanine and tyrosine residues. With only a very small number of exceptions, the evidence is that phenylalanine and tyrosine residues flip over by 180° rapidly. These rates are a few hundred to a few thousand flips per second and so are obviously too slow for molecular dynamics simulations. With pancreatic trypsin inhibitor, Karplus and his colleagues rotated the aromatic ring of each tyrosine residue progressively about the $C_\beta - C_\gamma$ bond and minimized the structure at each step; you could then see how the protein needed to move to accommodate this rotation. This was one of the pieces of evidence that suggested that the movement was a 180° jump rather than a continuous rotation. It showed that these motions could be accommodated by local fluctuations in the protein structure. In other words, it's only the first 'coordination sphere', as it were, of amino acid residues around the ring that needs to move; there needn't be knock-on effects spreading out through the protein. These local fluctuations in protein structure happen very rapidly, all the time. If the fluctuations of the residues above and below the aromatic ring accidentally happen at the same time, the ring can turn over. The rate of a few thousand per second that we measure with fluoromethotrexate is not, then, the rate of the fluctuations—it essentially reflects the probability that the fluctuation of all the necessary residues will happen at the same time, to allow the benzoyl ring to turn over. Karplus's simulation is the nearest we have to a detailed picture of what's going on. There are no experimental results that give us that kind of information at the moment.

Albery: I like the concept of floppiness that you have raised. We speculated some time ago that that might also be extremely important for the kinetics of enzyme reactions; that is to say, that the 10^9–10^{10} vibrations per second may help to provide the right environment in which the actual reaction can take place.

How much information do you have on the free energy profile for the whole of the enzyme-catalysed reaction, with both the natural enzyme and the mutant enzymes? If you could measure the rates precisely enough, the solvent isotope effect might be extremely interesting to investigate.

Roberts: I doubt that we could measure the rates precisely enough, but it would be worth a try.

Albery: D_2O is 18% more viscous than H_2O. The solvent isotope effect could be quite large because rotation is particularly affected by substituting deuterated water molecules.

Roberts: That would be worth investigating. The kinetic mechanism of the enzyme has been established in collaboration with Steve Benkovic (Andrews et al 1989). It's an interesting mechanism in a number of ways; the rate-limiting step is partly hydride ion transfer and partly product release. Product release is in fact promoted by the negative cooperativity with the reduced coenzyme that I mentioned. In most dehydrogenases, after the hydride ion transfer both products dissociate to regenerate free enzyme for another round of catalysis. In this case, the dissociation of the tetrahydrofolate is more than an order of magnitude slower than the turnover number of the enzyme. The oxidized coenzyme dissociates, and then the reduced coenzyme binds in its place; this accelerates the dissociation of tetrahydrofolate. The consequence is that the free enzyme is not on the kinetically preferred pathway, which complicates the understanding of the mechanism of action of mutant enzymes.

Albery: Presumably that's why there is connectivity between binding of the coenzyme and the active site.

Roberts: That's right. We hope that the pathway by which the substitution in the coenzyme binding site in the mutant enzyme T63Q has effects on the substrate binding site will be relevant to this cooperativity, and hence to the catalytic function of the enzyme.

Stoddart: Have you used NMR to find out the barrier to inversion of the two enantiomers, A and B, shown in Fig. 4? You could do the same experiment when they are complexed. In this case, the barrier should be raised and the difference would reflect the rigidity or the floppiness of the cavity.

Roberts: We have done the experiment on the molecules bound to the protein, but not on the free molecules. The barrier to rotation in the complex is greater than the barrier to thermal unfolding of the protein.

Reinhoudt: Could you use T_1-relaxation times to investigate the rate of rotation, or at least the difference in rotation? In small macrocycles you see tremendous changes in T_1 when a molecule rotates around an axis. The T_1 values of the methylene carbon atoms change quite dramatically, by a factor of up to 10, in a series of macrocycles. In your case you could have two methyl substituents, for example, one *meta* and one *para*, and if you studied rotation around the axis you would see a difference in the T_1 relaxation times for both compounds.

Roberts: We have done some ^{13}C T_1 measurements, not on the pyrimethamine series but on the trimethoprim series, which have a benzyl rather than a phenyl substituent. In proteins, ^{13}C T_1 relaxation is determined by rapid fluctuations—that is, any internal motion which is faster than the overall rotational correlation time of the protein, which, for a protein of this size, is of the order of 10 ns. If you look at rings which flip over—the benzyl ring in trimethoprim does that also—you find that flipping of the ring is slow, on the lineshape analysis time-scale of milliseconds. On the relaxation time-scale you see the oscillation of the ring and our measurements suggest that this is ± 25–$30°$ on this sub-nanosecond time-scale. One can imagine that the ring is oscillating backwards and forwards by 25–30° all the time, and every once in a while, when the surrounding residues have moved out of the way, it makes the entire leap over by 180°.

Sutherland: The mononitropyrimethamine complex had almost equal populations of the two diastereoisomers. What happens with 3,5 di-substitution?

Roberts: We haven't looked at that one yet.

Edmundson: Would a 3,5-substituted derivative be sterically too large to fit into the active site?

Roberts: When I first looked at the mononitro derivative, I thought we would see binding of only the isomer with the nitro group pointing out, because I didn't think there would be room for the nitro group to point inwards, but I was wrong. I can't see any reason why the 3,5-dinitro compound shouldn't bind, because we have established that a nitro group can bind both pointing into the site and pointing out towards the solvent, so I don't see why the protein shouldn't be able to accommodate a nitro group in both positions.

Kellogg: Wouldn't it be wise to establish the barrier to racemization between two enantiomers? If this is so low that something is happening during binding, it gives a great deal of information about the reason for a lack of selection of enantiomers.

Roberts: Yes; I am simply saying (and being slightly surprised by the fact) that the enzyme is not selecting between the two enantiomers.

Hamilton: Could the rotation of the difluoro derivative be accommodated simply by rotating the phenylalanine that is directly above it? Rotation of substrate and phenylalanine would then be coupled, resulting in a gearing process. The structure you showed would seem to allow rotation of the difluoro derivative.

Roberts: I can't answer that with certainty, but I think this would not be the case. Part of the reason for the uncertainty is that the electron density for the phenyl ring of that phenylalanine is not particularly well defined in the crystal structure. As far as one can tell, the phenyl ring is not stacked with the benzoyl ring, but nor is it at right angles—the angle between the ring planes

is closer to 45°. Simple rotation would give you some room for manoeuvre, but I suspect not quite enough. In any case, there is a leucine in van der Waals contact on the other side of the benzoyl ring which certainly would have to move.

References

Andrews J, Fierke CA, Birdsall B et al 1989 A kinetic study of wild-type and mutant dihydrofolate reductases from *Lactobacillus casei*. Biochemistry 28:5743–5750

Brooks CL, Karplus M, Pettitt BM 1988 In: Proteins: a theoretical perspective of dynamics, structure, and thermodynamics. Wiley, New York (Adv Chem Phys Ser vol 71) p 95–111

Gao J, Kuczera K, Tidor B, Karplus M 1989 Hidden thermodynamics of mutant proteins: a molecular dynamics analysis. Science (Wash DC) 244:1069–1072

Gelin BR, Karplus M 1975 Sidechain torsional potentials and motion of amino-acids in proteins—bovine pancreatic trypsin inhibitor. Proc Natl Acad Sci USA 72:2002–2006

Protein design: template-assembled synthetic proteins

Robert I. Carey, Karl-Heinz Altmann and Manfred Mutter

Séction du Chimie de Université de Lausanne, Rue de la Barre 2, CH-1005 Lausanne, Switzerland

Abstract. The introduction of the template-assembled synthetic protein (TASP) concept for the *de novo* design of proteins has provided a very powerful new tool for the construction of artificial tertiary structures. TASPs are constructed by the covalent attachment of amphiphilic secondary structure blocks to a topological template, resulting in a non-linear chain architecture of the target molecules. A key feature of this approach is that the template serves to reinforce and direct the folding of the secondary structure elements into predetermined tertiary structures. Having used stepwise solid-phase peptide synthesis (SPPS) in combination with linear oligopeptide templates in the initial stages of our research on TASPs we are now pursuing refined structural and synthetic approaches in the construction of the next generation of TASP molecules. These include the synthesis by SPPS of four-helix bundle TASPs using new cyclic templates (containing either S—S bridges or β-turn mimetics) as well as conventional fragment condensation strategies using protected fragments. The conformational properties of these new TASPs are now under investigation, with special emphasis on the relationship between overall conformation and the nature of the topological template.

1991 Host–guest molecular interactions: from chemistry to biology. Wiley, Chichester (Ciba Foundation Symposium 158) p 187–203

Peptide and protein molecules occur ubiquitously in Nature, where they mediate countless processes of recognition, binding and catalysis. Chemists and biologists alike have long been fascinated by the many functions displayed by globular proteins and, consequently, many innovative attempts to mimic these properties have been made. Previous endeavours have included the preparation of polymeric catalysts designed as enzyme models (Royer 1980) and the preparation and study of carefully designed rigid organic frameworks modelled on the catalytic sites of enzymes (see e.g. Lehn 1989, Dugas 1989, Breslow 1986). In the past decade, however, our ability to study proteins has been greatly increased thanks to the dramatic advances that have been made in the field of molecular biology (see e.g. Davies & Gassen 1983, Botstein & Shortle 1985) and in the methodology that is commercially available for peptide synthesis and purification

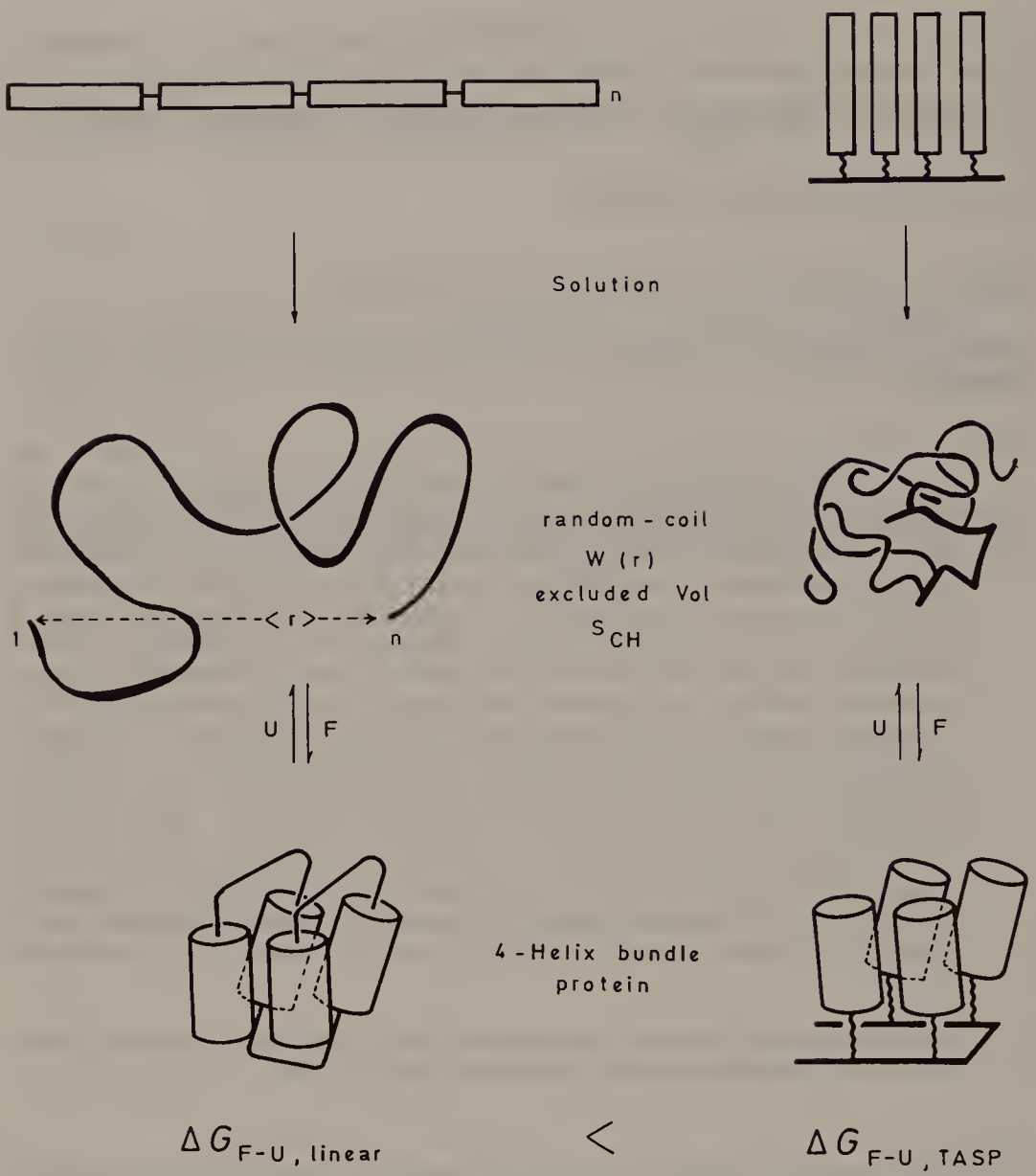


FIG. 1. Schematic representation of thermodynamic aspects of the folding of linear (L) and branched (T) polypeptides. Whereas the unordered (U) conformation of both molecules under poorly solvating conditions (near the so-called theta point) is exclusively determined by short-range interactions, extensive 'excluded volume effects' (Flory 1969) arise as a consequence of long-range interactions when the chain enjoys good solvation. This can be expressed in terms of an increase of the end-to-end vector, r. Because of the higher coil density in the branched molecule, the number of accessible conformations is smaller than in the linear molecule; i.e. the chain entropy, S_{CH}, is decreased. This in turn leads to a destabilization of the unfolded state (T_U) of the branched molecule. If it is assumed that the intramolecular interactions in the folded state (F) are similar in both linear and branched molecules (i.e. ΔH^*, $\Delta S^* = 0$), the branched molecule will be stabilized relative to the linear molecule by a difference in the free energy, $\Delta G^*_{L-T} = \Delta S^*_{CH}$. TASP molecules should therefore be more stable thermodynamically

(see e.g. Marshall 1988, Barany et al 1987). Using recombinant DNA techniques it is now possible to synthesize almost any primary sequence from the twenty naturally occurring amino acids, provided that the gene can be cloned and suitably expressed. Extension of this technique to the incorporation of unnatural amino acids is an area of current investigation (Noren et al 1989). Chemical synthesis as a tool for studying small proteins has also been demonstrated in very impressive fashion (Tam 1988) in the preparation of dozens of analogues of human transforming growth factor α (TGF-α), which is very difficult to isolate from natural sources. In parallel with these recent developments in the preparation of new proteins, the biophysical techniques available for the study of biological macromolecules have advanced rapidly. Of special importance has been the increasingly widespread use of high-field NMR spectroscopy (Wüthrich 1989a,b). Also, our ability to monitor and quantify small differences in protein structure (Pace 1990) and activity (see e.g. Fersht 1985, Schimmel 1990) is increasing as methods are becoming more routine and more powerful. Taken together, these advances have provided a flood of new data on protein molecules and have made possible systematic studies of protein structure and function that could only have been hoped for a short time ago (Knowles 1987).

Protein *de novo* design and the TASP concept

With the knowledge of protein structure that we have been able to accumulate over the past few years and are continuing to accumulate very rapidly, it is now appropriate to ask if it is possible to design from first principles *de novo* and to synthesize completely new protein molecules which contain amino acid sequences non-homologous to those that occur in Nature, but that exhibit all the characteristic properties of natural proteins—the most important of these properties being the ability to fold into a stable, predictable tertiary structure (DeGrado 1989, DeGrado et al 1989, Mutter 1985, 1988, Mutter & Altmann 1986, Mutter & Vuilleumier 1989, Richardson & Richardson 1989).

A descriptive account of the earliest attempts at the *de novo* design of proteins (Gutte et al 1979, Jaenicke et al 1980, Moser et al 1983, Lau et al 1984, Erickson et al 1986, Mutter et al 1986, DeGrado 1989, Hodges et al 1988) was the subject of a recent review (Richardson & Richardson 1989); the design principles on which these molecules were constructed have also been reviewed (Altmann & Mutter 1990, Mutter & Vuilleumier 1989, DeGrado 1989). Although most of these early designs were supported by circular dichroism (CD) and infrared (IR)

---

than linear polypeptides of equivalent packing arrangement (here, a four-helix bundle structure). This aspect of the energetics is important with respect to subsequent modification of the structure designed to introduce, for example, functionality into the molecular structure. Taken from Mutter & Vuilleumier 1989, with permission.

data consistent with the proposed structure, in no case has definite proof of structure been provided. Nevertheless, these pioneering studies established a sound foundation on which to base future studies.

A common feature shared by the early designs was their linear mode of chain connectivity. Although assembling artificial proteins in a linear fashion (analogous to protein-chain assembly in Nature) is a logical synthetic direction to follow, any new protein designed in this way must still overcome the critical hurdle of folding into a well-defined, globular structure. However, despite our rapidly improving understanding of the processes involved in protein folding, we are still unable confidently to predict tertiary structure from an amino acid sequence. The severity of this problem in relation to *de novo* protein design is perhaps best illustrated by the vanishingly small number of primary sequences that are known to adopt a globular folded structure in Nature, compared to the almost infinite number of possibilities which theoretically could exist. Therefore, until we have a better understanding of the complex folding mechanisms through which globular structures are formed from linear polypeptide chains, it is unlikely that even the most sophisticated model systems constructed in a linear manner will consistently contain all the information necessary for the molecule to fold into the desired tertiary structure. Successful designs are more likely to be accidental (single-hit) phenomena than part of a general strategy. This difficult situation, often referred to as the 'protein folding problem', is represented schematically in Fig. 1 and has been discussed in detail (Mutter & Vuilleumier 1989).

These considerations have therefore led us to adopt an alternative approach to protein design (Mutter 1988). We have been synthesizing template-assembled synthetic proteins (TASPs), in which amphiphilic secondary structure blocks are attached to a conformationally directing template molecule. The expected effect of the template is to hold the attached blocks in close proximity to each other, to enhance secondary structure formation, and to facilitate an intramolecular collapse of the amphiphilic peptide blocks to form a globular structure either before or after secondary structure formation has actually begun (for a discussion of current hypotheses for the mechanism of protein folding, see Baldwin 1989, Finkelstein & Ptytsin 1987).

A number of the possible topologies for TASPs are shown in Fig. 2. Artificial proteins constructed with this branched-chain, unnatural architecture should have many advantages over their linear analogues in which the amphiphilic peptide blocks are connected through loop or turn sequences and must reach their final folded state through a more stringently demanding pathway(s). Because of the template effects described above, TASPs should display an increased tendency to adopt a globular folded structure and a decreased tendency to succumb to *inter*molecular aggregation, a process which can fatally disrupt the *intra*molecular association that is required for proper folding. Furthermore, the possibility of synthesizing TASP molecules by chemical means (see next

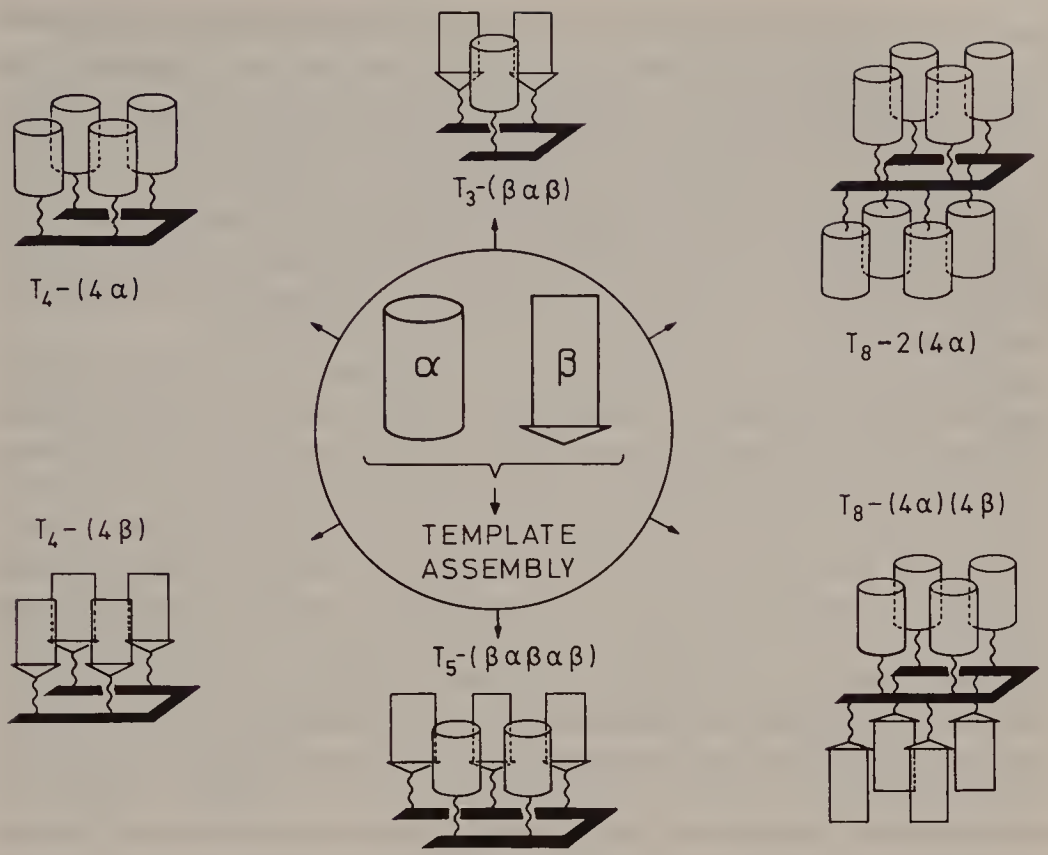


FIG. 2. Schematic representation of different possible packing arrangements (topologies) in TASP molecules (T_n, T, template molecule; n, number of attachment sites on T; α, β, amphiphilic peptide blocks with tendency for helix (α) and β-sheet (β) secondary structure formation). The number, nature and spatial orientation of the attachment sites in the carrier molecule predetermine the topological features of the TASP molecule; thus the carrier acts as a *template* for folding into the desired tertiary structure. Taken from Mutter & Vuilleumier 1989, with permission.

section), gives us the exciting opportunity to incorporate unnatural amino acids with widely varying steric and electronic properties into our target structures—for example, the incorporation of amino acids bearing strongly acidic or basic side-chains into hydrophobic binding pockets.

Construction of TASP molecules: synthetic considerations

To the synthetic chemist, TASP molecules represent a considerable challenge which increases with the size and complexity of the desired TASP structure. The non-linear chain architecture, which is the feature that makes TASP molecules attractive from the viewpoint of protein folding, also unfortunately makes TASP molecules inaccessible to synthesis by recombinant DNA techniques. The approach that has been taken initially by our laboratory has

been that of stepwise solid-phase synthesis using orthogonal protecting schemes (Barany & Merrifield 1980, Barany et al 1987). Starting from templates of the general formula P-Lys-X-Lys-Pro-Gly-Lys-Y-Lys-Z (P, X, Y, Z = Ac-Lys, Lys, Lys, Lys-R or Ac, Lys, Glu, NH_2; R = $NH(CH_2)_5CONHCH_2COOH$; Ac = acetyl) we have synthesized TASPs with βαβ, 4α, and 4α/4β topology by stepwise solid-phase methods and their conformational properties have been evaluated (Mutter et al 1988, 1989, Mutter & Tuchscherer 1988, Mutter & Vuilleumier 1989). A further set of TASPs with 4α topology has been synthesized more recently (Rivier et al 1990) and will be discussed below.

A strategy that involves solution-phase condensation of protected peptide fragments with a given template molecule is also being pursued in our laboratory. This strategy has resulted in the successful preparation of a 4α TASP (Ernest et al 1990a) in which the template molecule was a cyclic peptide containing two β-turn mimetics (Ernest et al 1990b). Although this strategy has been greatly facilitated by developments both in classical solution-phase techniques (Wünsch & Müller 1976) and in the commercial availability of solid-phase resins (Kaiser et al 1989, Rink 1987) that allow for the cleavage of fully protected peptide fragments, one must be careful to recognize the inherent limitations of amide bond-forming processes in which the affinity for attachment involves a weakly basic peptide amine and an electrophilic carbon of an activated acyl derivative (see Kemp 1981 for a thorough discussion of this topic). Certainly, entropy and dilution will set a limit to the size of molecules that can be coupled efficiently in any bimolecular process. The necessity in TASP synthesis to couple usually at least four peptide blocks to a single template further intensifies the problem.

Alternative approaches to the coupling of peptide blocks to template molecules or to the ligating of two or more TASPs to form multidomain TASPs are desirable and will probably be necessary for the efficient, reproducible synthesis of larger (> 100 amino acids) target molecules. Because state of the art enzymic synthesis (Jakubke 1987, Wong et al 1990) does not appear to be sufficiently powerful to meet our present requirements, attempts are under way in our laboratory to utilize the unconventional method of prior thiol capture (Kemp & Galakatos 1986, Kemp et al 1989). This methodology has been successfully employed for the preparation of two peptides (39 and 25 amino acids long, respectively) through the coupling ([13 + 26] and [13 + 12]) of peptide fragments in which all the amino acid side chain protecting groups had been removed, with the exception of those on the thiols of the cysteine residues which were appropriately functionalized for their role in prior thiol capture (D. Kemp & R. Carey, unpublished results). The successful application of this methodology to TASP design will offer the further advantage that the coupling step generates a nascent cysteine thiol to which spectroscopic probes, enzyme cofactors, or various steric and functional groups may be attached.

Properties of TASP molecules

Template-assembled synthetic proteins are macromolecules that exhibit structural features of both synthetic polymers (branched chains, grafted polymers) and globular proteins (peptide building blocks, a hydrophobic core, a hydrophilic surface; Mutter 1988). It is significant to note that all the TASP molecules described so far have been soluble in aqueous buffers; this property had been predicted for TASPs and suggests the formation of a globular structure containing a hydrophobic core and a hydrophilic surface. Conformational analysis of the TASP molecule (Rivier et al 1990) for which the proposed structure is shown in Fig. 3 strongly suggests that this structure is indeed attained. The CD spectrum of this molecule (in which X = H, cf. Fig. 3) in aqueous buffer, pH 7, is shown in Fig 4, together with that of the single helical block (attached to the ϵ-amino group of Ac-Lys-NH_2) for a comparative reference. Both the reference peptide block and the TASP molecule have an unblocked, and therefore charged, amino terminus. Greatly enhanced helicity is observed for the peptide block when it is attached to the template, even though this TASP design contains two features that are destabilizing for a four-helix bundle structure, namely the antiparallel helical arrangement (Chou et al 1988) and the positively charged N-termini of the helices (Shoemaker et al 1987). Interestingly, acetylation of the N-termini of the template-bound helices does

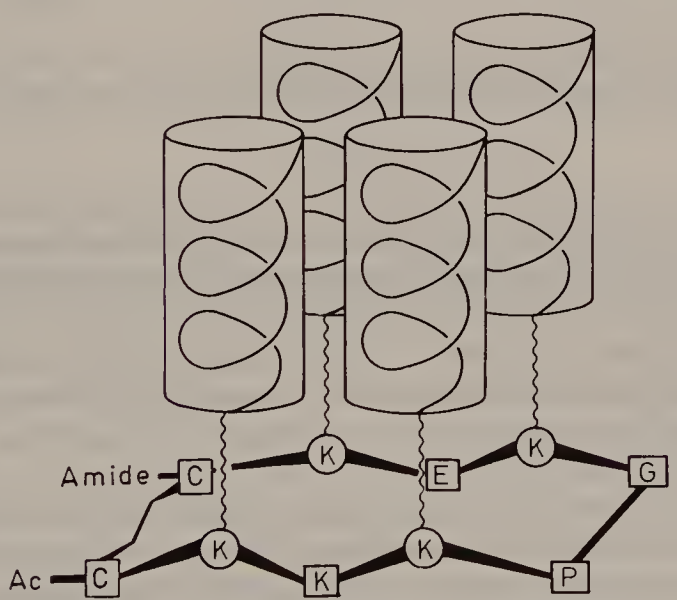


FIG. 3. Schematic representation of the proposed 4-helix bundle structure for a recently synthesized TASP molecule (Rivier et al 1990), containing the template Ac-Cys-Lys-Ala-Lys-Pro-Gly-Lys-Ala-Lys-Cys-NH_2 that has been cyclized via disulphide bond formation between the two cysteine residues. The four helical blocks are identical and have the sequence X-Glu-Ala-Leu-Glu-Lys-Ala-Leu-Lys-Glu-Ala-Leu-Ala-Lys-Leu-Gly (X = H, TASP; X = Ac, TASP′; see text and Fig. 4).

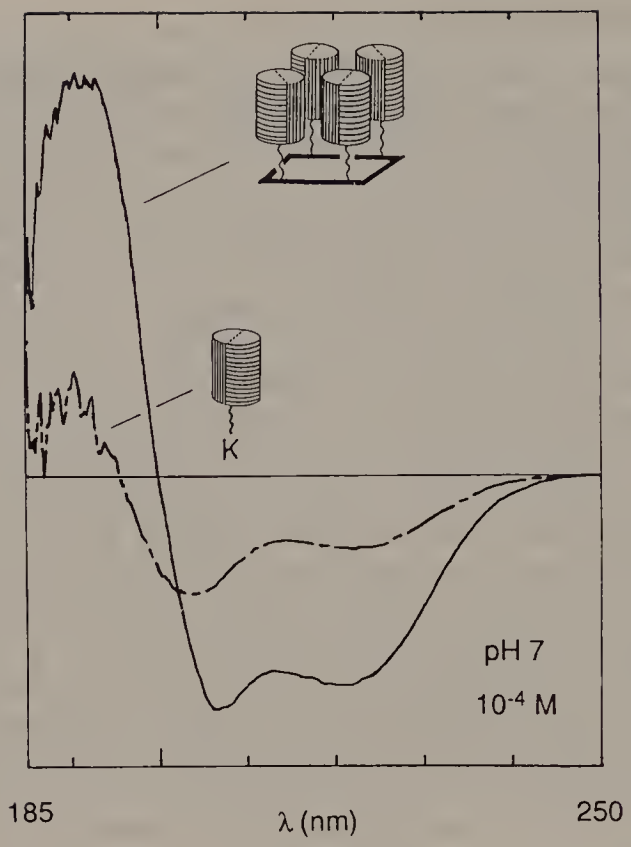


FIG. 4. Circular dichroism (CD) spectra in H_2O at pH 7 of TASP (X = H, cf. Fig. 3) (——) and the corresponding helix block (attached to the ϵ-amino group of Ac-Lys-NH_2) (----).

not result in any significant change of the CD spectrum depicted in Fig. 4, whereas the acetylation of the single block leads to a drastic increase in helicity. Guanidine hydrochloride-induced denaturation of the N-terminally acetylated TASP′ (in which X = Ac, cf. Fig. 3) was monitored as shown in Fig. 5 by the decrease in ellipticity at 222 nm. The helical structure of the TASP′ molecule was found to be stable to concentrations of guanidine hydrochloride of up to 3–4 M, a stability comparable to that of natural proteins (see e.g. Regan & Degrado 1988).

Although we are far from having realized all of the many goals for the *de novo* design of proteins that we established for ourselves at the outset of our research in this area, the syntheses and conformational studies accomplished so far have strongly supported our belief that TASPs represent a very promising approach to the design of novel proteins. It should be mentioned here that our approach has been used in the preparation of a haem-containing artificial protein (Sasaki & Kaiser 1989) in which four amphiphilic peptide blocks with the propensity for helix formation were covalently attached to a haem moiety that served as the template molecule. A further application of the TASP concept

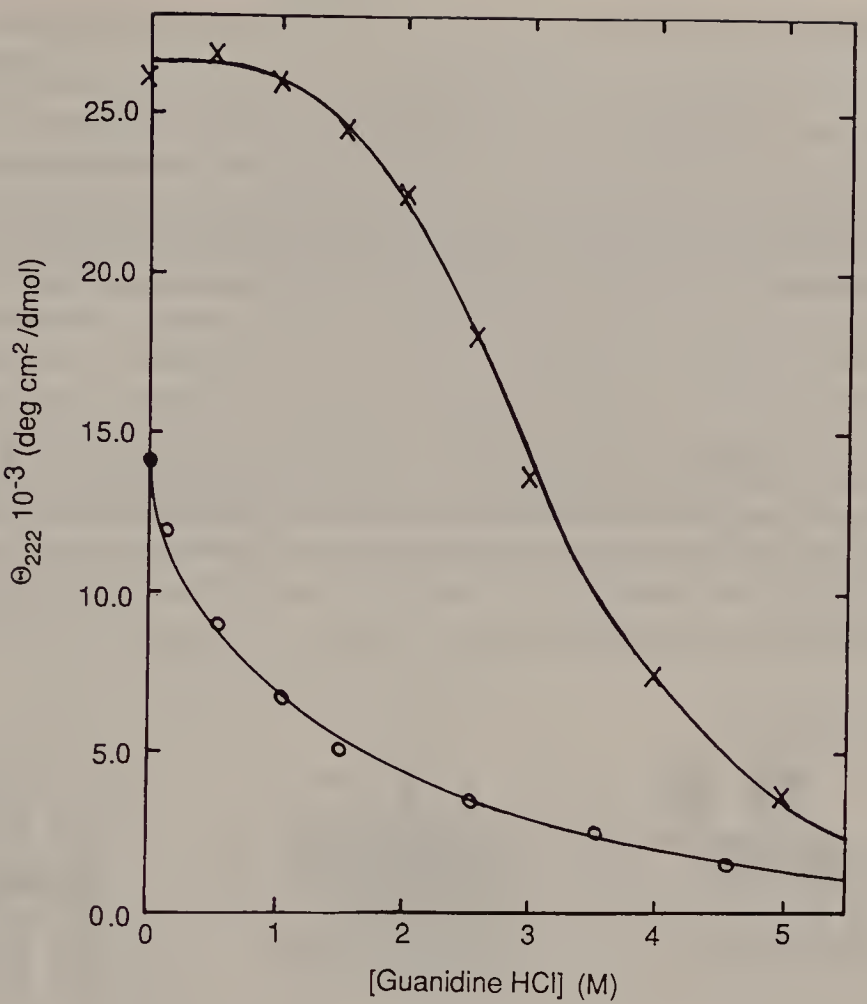


FIG. 5. Denaturation of TASP′ (X = Ac, cf. Fig. 3) (x) by increasing concentrations of guanidine hydrochloride, assessed by monitoring the decrease in ellipticity at 222 nm. The curve for the single-helix block is shown for comparison (o).

which is currently being investigated is the construction of membrane channel models (M. Montal, personal communication). In comparison with site-directed mutagenesis for the mechanistic study of channel proteins (Imoto et al 1986, 1988), the engineering of channel proteins via the TASP framework offers the distinct advantage (Montal 1990) that no ambiguity can arise in the distinction between altered function of the channel protein and problems associated with the cellular processing of the polypeptide or its assembly and folding in the membrane. Further, the channel protein designed by this means can be studied *in vitro* in the absence of any other cellular process or component. In addition, from a structural standpoint there is no ambiguity in the arrangement (parallel or antiparallel) of the channel-forming peptide blocks, because they are covalently attached to a carrier molecule.

Conclusions

'Molecular recognition' has been defined as a process that involves both the binding and the selection of substrate(s) by a receptor molecule, with the inclusion in the definition that the process as a whole should have a function (Lehn 1973). Artificial proteins (like their natural counterparts) possess immense inherent potential as receptor molecules for molecular recognition and have immense potential to fulfil the functionality requirement as well. In this chapter we have presented the template-assembled synthetic protein concept as a new and widely applicable strategy for artificial protein design. Although final structural proof of our proposed structures is still missing, the experimental data on the variety of TASP molecules synthesized so far have encouraged us to proceed with this approach. The challenge immediately ahead of us is to improve on our current synthetic strategies for TASP molecule construction and to achieve unequivocal proof of structure by means of nuclear magnetic resonance or X-ray analysis (if crystallizable products can be obtained). It is our hope that the synthesis and characterization of the possible TASP molecules

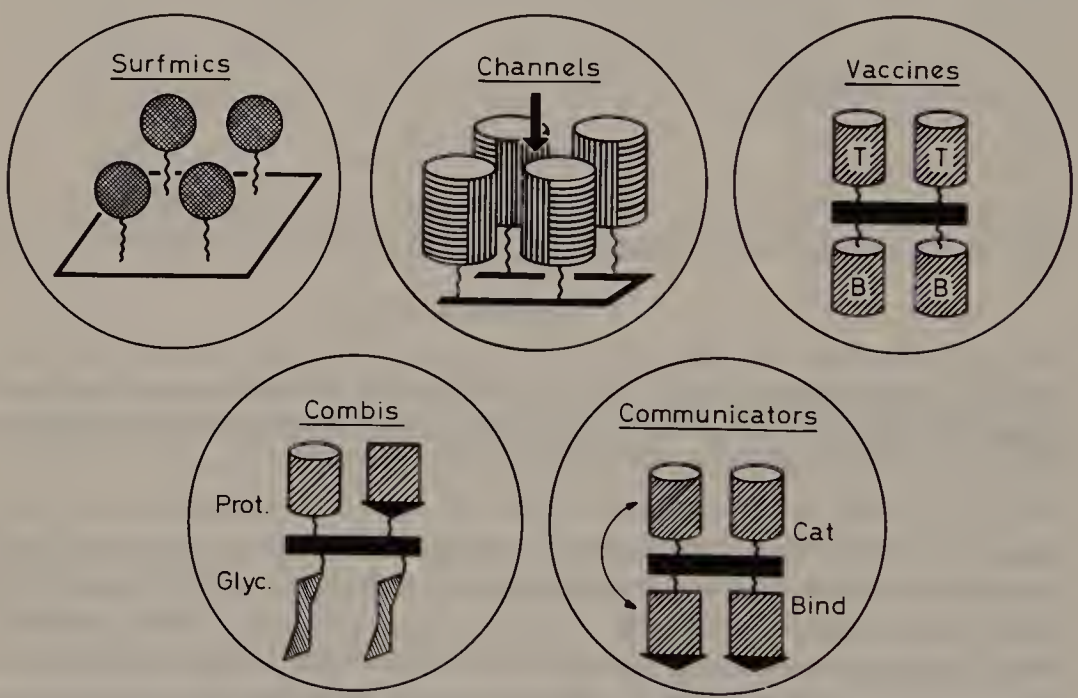


FIG. 6. Schematic representation of potential 'functional' TASP molecules. Surfmics; surface mimetics in which the amino acid side chains of functionally relevant surface residues are assembled on a template matching the spatial requirements of the active sites of proteins. Channels; ion-specific channels could be constructed through the attachment of membrane-spanning helices to an appropriately designed template. Two-domain TASPs could contain both binding and catalytic sites (Communicators), or could be synthetic vaccines, containing both B and T cell epitopes. Combis; a TASP in which peptidic and glycosidic elements are attached to a template molecule.

shown schematically in Fig. 6, as well as other structures like them, will soon be achieved.

Acknowledgements

The authors gratefully acknowledge the financial support of the Swiss National Science Foundation and the Hoffmann–La Roche Research Foundation. R.I.C. gratefully acknowledges the postdoctoral training grant GM 13678-01 from the National Institutes of Health.

References

Altmann K-H, Mutter M 1990 A general strategy for the *de novo* design of proteins—template assembled synthetic proteins. Int J Biochem 22:947–956

Baldwin R 1989 How does protein folding get started? Trends Biochem Sci 14:291–294

Barany G, Merrifield R 1980 Solid-phase peptide synthesis. In: Gross E, Meienhofer J (eds) The peptides—analysis, synthesis, biology. Academic Press, New York, vol 2:1–284

Barany G, Kneib-Cordonier N, Mullen D 1987 Solid-phase peptide synthesis: a silver anniversary. Int J Pept Protein Res 30:705–739

Botstein D, Shortle D 1985 Strategies and applications of in vitro mutagenesis. Science (Wash DC) 229:1193–1201

Breslow R (ed) 1986 International symposium on bioorganic chemistry. Ann NY Acad Sci 471

Chou K-C, Maggoria G, Némethy G, Scheraga H 1988 Energetics of the structure of the four-α-helix bundle in proteins. Proc Natl Acad Sci USA 85:4295–4299

Davies J, Gassen H-G 1983 Synthetic gene fragments in genetic engineering—the renaissance of chemistry in molecular biology. Angew Chem Int Ed Engl 22:13–31

DeGrado W 1989 Design of peptides and proteins. Adv Protein Chem 39:59–124

DeGrado W, Wasserman Z, Lear J 1989 Protein design, a minimalist approach. Science (Wash DC) 243:622–628

Dugas H 1989 Bioorganic chemistry: a chemical approach to enzyme action, 2nd edn. Springer-Verlag, New York

Erickson B, Daniels S, Reddy P et al 1986 Current communications in molecular biology, computer graphics, and molecular modeling. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY p 53–58

Ernest I, Vuilleumier S, Fritz H, Mutter M 1990a Synthesis of a 4-helix bundle-like template assembled synthetic protein (TASP) by condensation of a protected peptide on a conformationally constrained cyclic carrier. Tetrahedron Lett 31:4015–4018

Ernest I, Kalvoda J, Richs G, Mutter M 1990b Three novel mimics for the construction of sterically constrained protein turn models. Tetrahedron Lett 31:4011–4044

Fersht A 1985 Enzyme structure and mechanism, 2nd edn. WH Freeman, San Francisco, CA

Finkelstein A, Ptytsin O 1987 Why do globular proteins fit the limited set of folding patterns? Progr Biophys Mol Biol 50:171–190

Flory P 1969 Statistical mechanics of chain molecules. Interscience, New York

Gutte B, Däumigen M, Wittschieber E 1979 Design, synthesis, and characterization of a 34-residue polypeptide that interacts with nucleic acids. Nature (Lond) 281:650–655

Hodges R, Semchuk P, Taneja A, Kay C, Parker J, Mant C 1988 Protein design using model synthetic peptides. Pept Res 1:19–30

Imoto K, Methfessel C, Sakmann B et al 1986 Location of an α subunit region determining ion transport through the acetylcholine receptor channel. Nature (Lond) 374: 670–674

Imoto K, Busch C, Sakmann B et al 1988 Rings of negatively charged amino acids determine the acetylcholine receptor channel conductance. Nature (Lond) 335:645–648

Jaenicke R, Gutte B, Glatter U, Strassburger W, Wollmer A 1980 Conformation of a synthetic 34-residue polypeptide that interacts with nucleic acids. FEBS (Fed Eur Biochem Soc) Lett 114:161–164

Jakubke H-D 1987 Enzymatic peptide synthesis. In: Udenfriend S, Meienhofer J (eds) The peptides: analysis, synthesis, biology. Academic Press, New York, vol 9:103–165

Kaiser E, Mihara H, Laforet G et al 1989 Peptide and protein synthesis by segment synthesis-condensation. Science (Wash DC) 243:187–192

Kemp D 1981 The amine capture strategy for peptide bond formation—an outline of progress. Biopolymers 20:1793–1804

Kemp D, Galakatos N 1986 Peptide synthesis by prior thiol capture 1. A convenient synthesis of 4-hydroxy-6-mercaptodibenzofuran and novel solid phase synthesis of peptide-derived 4-(acyloxy)-6-mercaptodibenzofurans. J Org Chem 51:1821–1829

Kemp D, Carey R, Dewan J, Galakatos N, Kerkman D, Leung S-L 1989 Rational design of templates for intramolecular O,N-acyl transfer via medium-sized cyclic intermediates derived from L-cysteine. Definition of an experimental maximum in effective molarity through the study of "tunable" templates. J Org Chem 54:1589–1603

Knowles J 1987 Tinkering with enzymes: what are we learning? Science (Wash DC) 236:1252–1258

Lau S, Taneja A, Hodges R 1984 Synthesis of a model protein of defined secondary and quaternary structure. Effect of chain length on the stabilization and formation of two-stranded α-helical coiled-coils. J Biol Chem 259:13253–13261

Lehn J-M 1973 Design of organic complexing agents. Strategies toward properties. Struct Bonding 16:1–71

Lehn J 1989 Supramolecular chemistry—scope and perspectives: molecules, supermolecules, and molecular devices. Angew Chem Int Ed Engl 27:89–112

Marshall G (ed) 1988 Peptides—chemistry and biology. Escom, Leiden (Proc 10th Am Peptide Symp)

Montal M 1990 Channel protein engineering: an approach to the identification of molecular determinants of function in voltage-gated and ligand-regulated channel proteins. In: Narahashi T (ed) Ion channels. Plenum Publishing Corporation, New York vol 2:1–31

Moser R, Thomas R, Gutte B 1983 An artificial crystalline DTT-binding protein. FEBS (Fed Eur Biochem Soc) Lett 157:247–251

Mutter M 1985 The construction of new proteins and enzymes—a prospect for the future. Angew Chem Int Ed Engl 24:639–653

Mutter M 1988 Nature's rules and chemist's tools: a way for creating novel proteins. Trends Biochem Sci 13:261–265

Mutter M, Altmann K-H 1986 The construction of new proteins. Angew Makromol Chemie 145/146:211–229

Mutter M, Tuchscherer G 1988 The construction of new proteins. Template assembled synthetic 4-helix bundle protein. Makromol Chem Rapid Commun 9:437–443

Mutter M, Vuilleumier S 1989 A chemical approach to protein design—template-assembled synthetic proteins. Angew Chem Int Ed Engl 28:535–554

Mutter M, Altmann K-H, Vorherr T 1986 The construction of new proteins. II. Design, synthesis, and conformational properties of folding units with βαβ-topology. Z Naturforsch B 41:1315–1322

Mutter M, Altmann E, Altmann K-H et al 1988 The construction of new proteins. III. Artificial folding units by assembly of amphiphilic secondary structures on template molecules. Helv Chim Acta 71:835–847

Mutter M, Hersperger R, Gubernator K, Müller K 1989 The construction of new proteins. V. A template assembled synthetic protein (TASP) containing both a 4-helix bundle and β-barrel-like structure. Proteins Struct Funct Genet 5:13–21

Noren C, Anthony-Cahill S, Grifith M, Schultz P 1989 A general method for site-specific incorporation of unnatural amino acids into proteins. Science (Wash DC) 244:182–188

Pace N 1990 Measuring and increasing protein stability. Trends Biochem Sci 8:93–97

Regan L, DeGrado W 1988 Characterization of a helical peptide designed from first principles. Science (Wash DC) 241:976–978

Richardson J, Richardson D 1989 The *de novo* design of protein structures. Trends Biochem Sci 14:304–309

Rink H 1987 Solid-phase synthesis of protected peptide fragments using a trialkoxy-diphenyl-methylester resin. Tetrahedron Lett 28:3787–3790

Rivier J, Miller C, Andrews J, Porter J, Tuchscherer G, Mutter M 1990 Total synthesis of several TASP 4α by solid phase methods. In: Epton E (ed) Proc Int Symp Innovation and Perspectives in Solid Phase Peptide Synthesis, (Wolverhampton, UK, 1989) SPPC, Birmingham, p 39–50

Royer G 1980 Enzyme-like synthetic catalysts (Synzymes). Adv Catal 29:197–227

Sasaki T, Kaiser E 1989 Helichrome: synthesis and enzymatic activity of a designed hemeprotein. J Am Chem Soc 111:380–381

Schimmel P 1990 Hazards of deducing enzyme structure-activity relationships on the basis of chemical applications of molecular biology. Acc Chem Res 22:233–234

Shoemaker K, Kim P, York E, Stewart J, Baldwin R 1987 Tests of the helix dipole model for stabilization of α-helices. Nature (Lond) 326:563–567

Tam J 1988 Structure–activity studies of transforming growth factor. In: Marshall G (ed) Peptides—chemistry and biology. Escom, Leiden (Proc 10th Am Peptide Symp) p 561–564

Wong C-H, Chen S-T, Hennen W, et al 1990 Enzymes in organic synthesis: use of subtilisin and a highly stable mutant derived from multiple site specific mutations. J Am Chem Soc 112:945–953

Wünsch E, Müller E (eds) 1976 Houben-Weyl, Methoden der Organischen Chemie. Thieme-Verlag, Stuttgart, vol XV/2

Wüthrich K 1989a Protein structure determination in solution by nuclear magnetic resonance spectroscopy. Science (Wash DC) 243:45–50

Wüthrich K 1989b The development of nuclear magnetic resonance spectroscopy as a technique for protein structure determination. Acc Chem Res 22:36–46

DISCUSSION

Gokel: You mentioned that attempts are being made to construct transmembrane channel proteins using the TASP method. Is the design strategy to have the tubular bundle as an internal transport agent, or for cooperativity among the bundles to produce some kind of pore?

Carey: I think that the idea is to mimic proteins like the nicotinic acetylcholine receptor. The present strategy is to design bundles that will in themselves constitute the pore. However, other means of association and channel formation

should not be ruled out. We are attempting to design molecules that will enable us to test some of the present hypotheses about how membrane channel proteins function. Preliminary experiments have given very promising results (Montal 1990), but we are synthetically limited at present. So far, we have used only stepwise solid-phase synthesis to prepare our membrane channel models, and the highly hydrophobic nature of the products makes their purification and characterization difficult.

Crumpton: There are many channels for which we now know the sequences of the transmembrane regions. Theoretically it would be possible to reproduce the sequences for the multiple membrane-spanning regions of the acetylcholine receptor or, for that matter, of ion channels such as Ca^{2+}-ATPase or Na^+/K^+-ATPase. However, in each of these examples, the multiple membrane-spanning regions have different sequences. I don't understand how, on a template, you could synthesize, in a completely programmed way, separate helices with different sequences. You would have to do them one after another, because you couldn't possibly programme for them to be done simultaneously. Under these circumstances, do you conceive of the linking group within the template being different for the attachment of the different helices? Could you elaborate on how you might achieve three or more independent segments?

Carey: To synthesize a TASP with three or more large hydrophobic sequences attached to the template, the best strategy for us to use would be to construct the TASP by linking the fragments in such a way as to build up the template in the same process. To do this we need a very powerful method for amide ligation of peptide fragments. We are presently investigating the method of prior thiol capture (Kemp 1981, Kemp et al 1989) for this purpose, as it may be the only way that such molecules can be prepared in a highly pure form. The advantage of thiol capture ligations is that they can be performed with deprotected, and thus carefully purified, peptide fragments.

Rebek: Have you considered using non-peptide templates to get greater rigidity and enforce the parallel approach of the several helices?

Carey: That was one of the things which I was hoping to ask people's opinion on. We are using peptide-based templates because they can be prepared rapidly and because our molecular dynamics results indicate that some cyclic decapeptides assume low energy conformations that are ideally suited for the construction of 4α-helix bundle proteins. However, we are interested in investigating rigid organic frameworks which possess convergent functional groups. Through the judicious choice of such molecules we can expect to enhance the template effect of directing and reinforcing the folding of attached peptide blocks.

Rebek: The four methylene groups in the lysine side chain conspire against you when you are trying to fix the conformation to involve a single face.

Carey: The flexibility of the lysine side chain raises the question of whether a more rigid attachment site would be preferable. Also, there is a thermodynamic

price to be paid for having the four methylene groups exposed to solvent. We are now preparing templates containing the amino acids ornithine and diaminobutyric acid, which are structurally analogous to lysine but have only three or two methylene groups respectively in their side chains. We intend to assess quantitatively the effect of side chain length of the template attachment site on the overall stability of a TASP molecule.

Sutherland: The ideal point of attachment would be a smooth surface. Have you thought about packing peptide helices onto surfaces?

Carey: I don't have any ideas about that.

Gutsche: There's a recent report from Stewart of the University of Colorado (Hahn et al 1990) of a synthetic protein with chymotrypsin-like activity consisting of a bundle of four short helical peptide units. Was that made according to your protocol?

Carey: There are many similar reports of functional polymers possessing some enzymic activity. Stewart prepared a branched-chain polymer containing four different peptide chains by stepwise solid-phase synthesis. He observed esterase (not amidase) activity toward a model chymotrypsin substrate. Unfortunately, the molecule was neither purified (by HPLC or ion-exchange chromatography) nor structurally characterized, and many control experiments have yet to be performed. Presently, it is impossible to interpret the results.

Ron Breslow: Stewart has no idea what the products are.

Rebek: It's not peptidase activity.

Carey: This ambiguity illustrates a critical requirement of protein design—molecules must be chemically and structurally well defined, or we shall be unable to learn very much from them. Our goals are first to establish reliable synthetic approaches to TASPs, to purify our products and to obtain convincing proof of structure. Once these foundations have been laid, we shall be able to proceed rationally toward the inclusion of functional properties.

Hamilton: Is it possible to cap the top of the bundle as well as the bottom, particularly with a two-helix bundle?

Carey: I would like to construct a four-helix bundle TASP in which the N-terminal residues are cysteines. There is a good chance that one could successfully 'cap' such a structure with a cyclic peptide containing four appropriately spaced cysteines with thiols activated for unsymmetrical disulphide formation. The product would be a TASP in which the end opposite to the template is attached to a second template through four disulphide bonds. There are other possibilities.

Hamilton: The dipoles of your helices were all opposed. Obviously, for maximum stabilization of the four-helix bundle they should be antiparallel.

Carey: That is a point which is currently being debated. We have compared an *N*-acetylated four-helix bundle which has no positive charge at the N-terminus with the non-acetylated four-helix bundle which has a free, and therefore positively charged, N-terminus. The CD spectra of these molecules (which are both highly pure) are basically identical, indicating that the helix dipole has no

effect. However, if one compares the individual peptide blocks one finds that acetylation of the N-terminus results in a marked increase in helicity. These results support the idea that although the helix dipole plays a major role in the stability of small peptides in solution, it plays only a small role, if any, in the stabilization (or destabilization) of a structure such as a four-helix bundle TASP.

Brändén: In Fig. 2 you showed pictures of different types of structures including βα- and ββ-structures. You must be aware of Kabsch & Sander's (1983) analysis of protein structures in which they show that pentapeptides with identical sequences can form a helix in one structure and a β-strand in another. It is feasible that you could design a 15–18 residue-long stretch that would form a helix, and you showed that this can be achieved. If you have a 6–7 residue piece which you want to fold into a β-structure, how do you ensure that it will fold into that desired structure?

Carey: β-Structure alone is difficult to work with because β-structure-containing peptides tend to interact intermolecularly to form insoluble aggregates. In protein design we rely heavily on the stable formation and folding of supersecondary structure units (βαβ, αβ) because we cannot depend on the long-range interactions that are present in naturally occurring proteins. We now have considerable knowledge concerning supersecondary structure formation of peptides (Mutter et al 1986, 1988) and we hope to be able to apply this knowledge successfully to TASP synthesis.

Dunitz: I have problems with this notion of helix dipoles and its application in calculating interaction energies between neighbouring helices in proteins. Given the charge distribution of a molecule we can always evaluate its various moments. This is a useful exercise if we wish to know the electric field or potential at distances that are large compared with the size of the molecule. As is well known, the potential at large distances is dominated by the first non-vanishing moment. But if we are concerned with the field in the immediate neighbourhood of the distribution, the separation into monopoles, dipoles, quadrupoles etc. is a fruitless exercise. We then need to take the sum over all the point charges or integrate over the entire charge distribution, if it is to be regarded as continuous. I do not think one can calculate the interaction energy between neighbouring helices in terms of the dipole approximation. Of course, one can visualize a lot of little dipoles—plus, minus, plus, minus—and add them up to get an enormous resultant dipole, but I do not think this is a useful exercise. Perhaps Dr Carey's finding that *N*-acetylation has no effect on the CD spectrum of his four-helix bundle is an experimental indication that I am right!

Rebek: What is meant by the helix dipole is the sum of the dipoles of the amides that are parallel in the helix. Stabilization is due to polar groups at the end(s), rather than the creation of the dipole. A positively charged group near one end of the dipole stabilizes the helix by a certain amount.

Dunitz: If all the C—O bonds are pointing one way, then all the C—N bonds point another way and their contributions tend to cancel one another.

Rebek: Amide functions have dipoles; these are all oriented in the same sense in a helix.

Dunitz: But these are not the only contributions that need to be considered. There are all those hydrogen bonds, which are often described in terms of a point-charge model. What about their local dipoles? What is important, I imagine, is the cooperativity. Ten parallel, coupled hydrogen bonds give a greater stabilization than the sum of ten separate hydrogen bonds. I am being distracted from the main point I wished to make, which is that the dipole approximation should not be used for calculating interaction energies at distances that are small in comparison with the dimensions of the helices themselves. Anyway, the notion that one can simply add up all the local amide dipoles would seem to lead to bizarre consequences. If it were correct, when you pressed one end of a macroscopic polar amide crystal a spark would fly out of the other end!

Rebek: There are crystals that do just that!

Dunitz: But they're not made of amides.

Brändén: How do you explain that in protein-engineering studies the introduction of a negatively charged residue at the N-terminus, or a positively charged residue at the C-terminus, leads to a measurable helix-stabilization effect?

Dunitz: I would talk about each case separately.

Brändén: It has now been done in a number of cases (Nicholson et al 1988).

References

Hahn KW, Klis WA, Stewart JM 1990 Design and synthesis of a peptide having chymotrypsin-like esterase activity. Science (Wash DC) 248:1544–1547

Kabsch W, Sander S 1983 Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 22:2577–2637

Kemp D 1981 The amine capture strategy for peptide bond formation—an outline of progess. Biopolymers 20:1793–1804

Kemp D, Carey R, Dewan J, Galaktos N, Kerman D, Leung S-L 1989 Rational design of templates for intramolecular O,N-acyl transfer via medium-sized cyclic intermediates derived from L-cysteine. Definition of an experimental maximum in effective molarity through the study of 'tunable' templates. J Org Chem 54:1589–1603

Montal M 1990 Channel protein engineering: an approach to the identification of molecular determinants of function in voltage-gated and ligand-regulated channel proteins. In: Narahashi T (ed) Ion channels. Plenum Publishing Corporation, New York, vol 2:1–31

Mutter M, Altman K-H, Vorherr T 1986 The construction of new proteins. II. Design, synthesis, and conformational properties of folding units with βαβ-topology. Z Natursforsch B 41:1315–1322

Mutter M, Altman E, Altman K-H et al 1988 The construction of new proteins. II. Artificial folding units by assembly of amphiphilic secondary structures on template molecules. Helv Chim Acta 71:835–847

Nicholson H, Becktel WJ, Matthes BW 1988 Enhanced protein thermostability from designed mutations that interact with α-helix dipoles. Nature (Lond) 336:651–656

General discussion III

Transition state theory and energy requirements for reactions

Ron Breslow: I would like to raise the issue of the need for flexibility or rigidity in enzyme-like reactions. Some years ago we described reactions of esters bound to a cyclodextrin that acylated a hydroxyl group (Czarniecki & Breslow 1978, Breslow et al 1980, 1983). In some ways, that mimics serine protease action. The mechanism of serine proteases involves addition of a hydroxyl group to a carbonyl group to make a tetrahedral intermediate, followed by formation of an ester. The addition reaction requires attack perpendicular to the carbonyl plane, but in the second step, in which the product ester is made, a 90° rotation is required. With rigid substrates that bound into the cyclodextrin cavity and held the group in such a way that perpendicular attack was favoured, we saw accelerations of over 10^6 with good leaving groups. When we put poor leaving groups on the same system we not only lost rate, but we also lost acceleration. Our interpretation was that when the leaving group is good, so that the first step is the rate-limiting step, the restricted geometry is terrific. When the leaving group is poor, the second step becomes rate-limiting and restriction of rotation is a problem.

We then made two new substrates (Breslow & Chung 1990; S. Chung, unpublished work). We put an ester group on a binding unit that was held almost rigid with respect to the hydroxyl, except that the important one degree of motion remained. We found that the first step was now slower; we lost 10 to 100 in acceleration, depending on what we compared it to. The second reaction was now much faster. With poor leaving groups, the extra degree of freedom provided a much better substrate. If the reaction requires motion, you must not freeze that out; you try to have as much rigidity as you can get without preventing the motion which is required for the chemistry.

Some years ago, Dan Koshland postulated a concept which he called orbital steering (Storm & Koshland 1972). His idea was that there might be some overlap reason why transition states are more rigid than molecules. He used this to explain why some internal reactions are very fast (which was not in fact a problem). The difficulty with the idea is that it is not clear why a transition state should be more rigid than the starting material; people pointed out that it was preposterous to suggest that a partial bond in a transition state could be more rigid than a real bond.

I think it's possible that transition states *are* in fact more rigid than molecules, and that the geometric requirement in reactions is a more rigorous requirement than that for molecules themselves. We looked at a series of reactions in which

various hydrogens on steroid molecules were being attacked, using geometric control to hold the attacking group so that it could reach one of the hydrogens. These hydrogens are only a couple of ångströms apart, so rather good geometric control is required. We achieved this, getting high selectivity for one hydrogen over another. When we did molecular mechanics calculations, we duplicated the data well over a range of molecular parameters, provided that we assumed that the bending force constant in the transition state of the hydrogen abstraction was stiff (i.e. high), at least as stiff as that for a real molecule. You could say that this is an artifact of the calculation, but it probably isn't, because the hydrogens are not far apart and if you allow a lot of bending there's no real chance that you could hit one and not another; so there probably is such a restriction in real life.

We wondered what this meant; why should it be difficult to bend a partial bond in a transition state? George Flynn, a chemical physicist, had been looking at reactions in which chlorine atoms in the gas phase react with cyclohexane to produce HCl molecules (unpublished). He did a laser enquiry into the state of the HCl and found that the vibrational temperature was 14 000 K and the rotational temperature was 170 K. The product is cool, rotationally, relative to the actual experimental temperature, and all the excess energy is in vibration. This means that the chlorine had to come end-on into the H—C bond; any angular collision of that kind would have angular momentum which would show up in product rotation. Similarly, if you fire oxygen atoms at hydrocarbons to make a hydroxyl radical, all the energy ends up in vibration—there's essentially no tumbling of the hydroxyl radical. Again, that means that the oxygen came directly in on the C—H bond and that reaction at an angle does not in fact result in product formation. If you fire hydrogen atoms at carbon dioxide and determine what it takes for hydrogen atom translation energy to be converted into vibrational energy in CO_2, you find that an almost end-on collision is required (Flynn 1989); if the collision occurs at an angle, the hydrogen glances off.

Imagine hitting a nail with a hammer. In a collision which is more or less end-on, there is a stiff potential and all the translational energy can be turned into the vibrational energy that is required to first compress the distance—to flatten the nail—and then to excite the vibration. My proposition is that there is an extra geometric requirement for transfer of energy into the correct modes so that reaction results. It's not sufficient to say that somehow or other some collision energy is used to get the reaction over the barrier; the question is *how* do you get over the barrier? It must be done by taking translational energy and somehow turning it into compression of the H—Cl bond, for example, or excitation of the C—H bond, which is required to make the reaction go. If this is true, that there is an extra geometric requirement for a chemical reaction, it explains how a transition state could have greater geometric requirements than a molecule has. The dependence of energy on geometry may be an extra factor associated with the channels through which an energy transfer can be made.

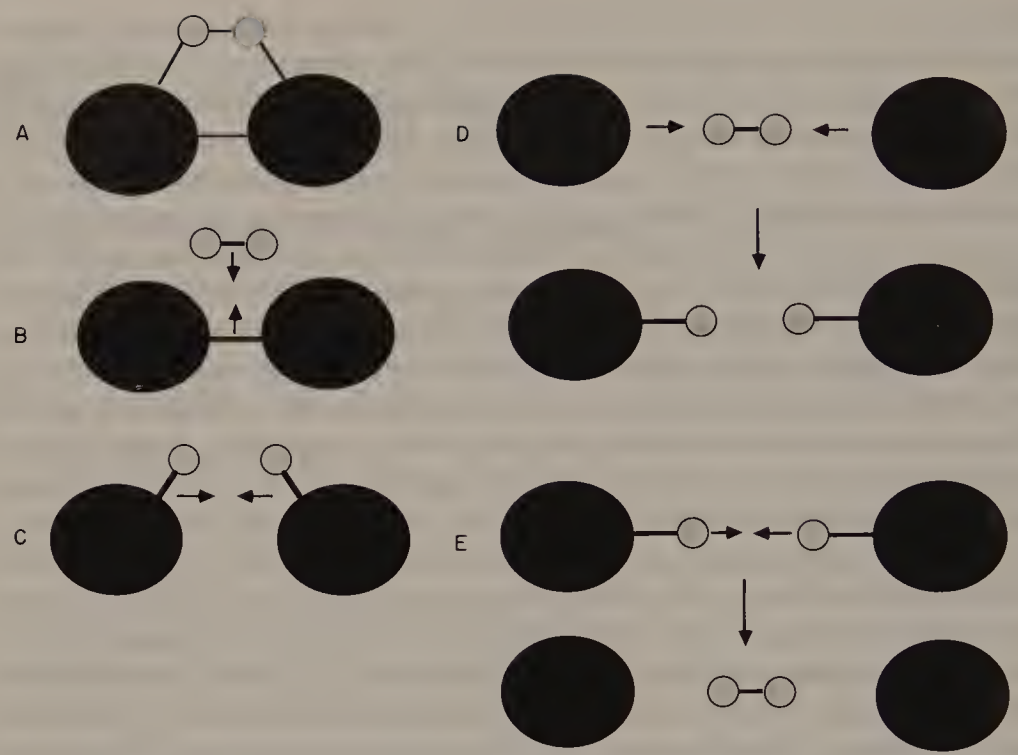


FIG. 1 (*Albery*) The reaction between H_2 and I_2. (A) The classical transition state; (B) and (C) represent ineffective collisions for the forward and backward reactions respectively; (D) the effective termolecular collision for the forward reaction; (E) the effective bimolecular collsion for the backward reaction.

Albery: Consider the hydrogen–iodine reaction, which we learnt at school as *the* example of transition state theory. We were all told that this reaction obeys transition state theory brilliantly, passing through the transition state depicted in Fig. 1A. It has now been shown by Sullivan (1967) that this reaction proceeds via the following mechanism: iodine atoms in I_2 have to come apart and then they crash together in a termolecular collision to break the H—H bond. You can do this reaction at low temperature by making iodine atoms from I_2 photochemically or at high temperatures by making them thermally. The data lie on a common Arrhenius plot, so the two reactions follow the same mechanism. Therefore, the iodine atoms act like a pair of nutcrackers crunching the H_2 molecule (see Fig. 1D). If you try to throw hydrogen against iodine to go through the transition state shown in Fig. 1A, the molecules won't react, because hydrogen is so light that you can't get the necessary energy into the bonds to be broken (Fig. 1B). That's a very clear example of what Ron Breslow was talking about.

Rebek: Is that a termolecular reaction?

Albery: It is a termolecular collision, but the reaction shows second order kinetics because there is a pre-equilibrium between I_2 and 2I.

Rebek: What is the rate-determining step?

Albery: The rate-determining step is the H_2 molecule being crushed between the two iodine atoms (Fig. 1D). In the reverse reaction two HI molecules crash head-on, as shown in Fig. 1E. The momentum from the iodines will go into the H—I bonds and break them to produce hydrogen. If you try to crash them sideways in the conformation shown in Fig. 1C there will be no reaction.

On a more general point, consider the potential energy surface shown in Fig. 2 with a wide valley leading into a narrow gulch. A question which Laidler raised a long time ago is whether the transition state is at the maximum energy or the maximum free energy. The answer that I come up with is that if the reaction follows a straightforward trajectory and it creeps up over the barrier and just gets over the saddle, then that will be the crisis point for the reaction. If, on the other hand, the reaction pathway is off-centre, the molecule may get reflected back, even though there is enough energy to get over the col. When the system has the wobbly trajectory it has to go further into the gulch before it is actually committed to reaction. In terms of transition state theory, this means that the partition function ($q_{\ddagger}$) for the transition state does not have a single geometry. The narrowing of the valley into the gulch means that there's a greater order in the transition state.

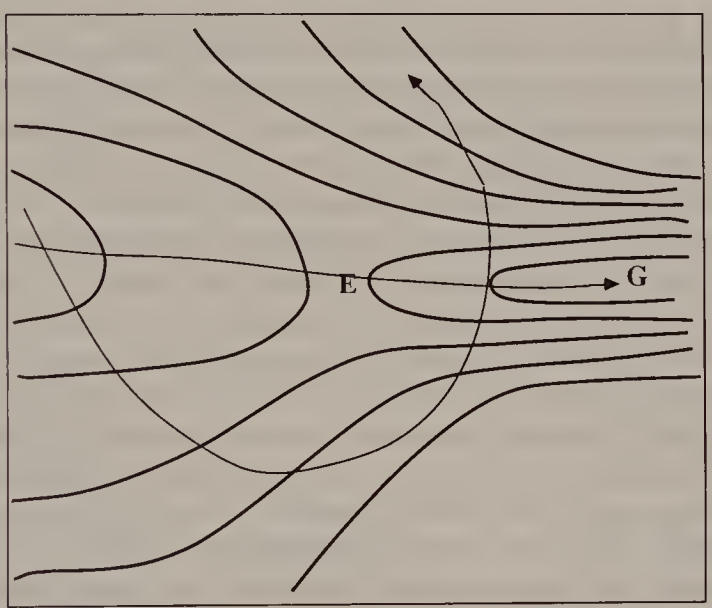


FIG. 2. (*Albery*) A potential energy surface where the maximum in E and that in G are at different positions. The linear trajectory has to go further into the gulch before it is committed, thereby shifting the transition state towards the maximum in G.

Ron Breslow: If the reaction is fairly exothermic, you can afford a glancing shot because there will still be enough energy transferred to do the reaction. If the reaction isn't enormously exothermic, you are always trying to get these reactions to just make it. Our reactions don't have huge excess energy involved and we may require precise collisions to get the energy transfer. There's another argument: if H collides with H_2, the minimum energy path is the linear one, but the path of maximum probability is bent at about 30°. There are a lot of paths that come in at 30°, whereas there's only one path that comes in at 0°. Thus, off-centre attack is predicted to be the best geometry. If that were true in our system, you couldn't possibly select between the accessible hydrogens.

Dunitz: It's news to me that Dan Koshland was thinking about gas-phase reactions when he proposed orbital steering. I thought he was thinking about enzymic reactions and was proposing that precise trajectories are necessary for these, in addition to all the other entropic factors with picturesque names listed by Jencks in his influential review article (1975).

Ron Breslow: Koshland was concerned with enzymic reactions but his discussion was quite general. It did not invoke energy transfer factors.

Dunitz: It's true that no one bothered much about trajectories, about the requirement that the reacting molecule had to come in at a particular angle. In 1973 or 1974, after we had discovered the importance of the approach angle in nucleophilic addition to carbonyl groups (Bürgi et al 1973), I saw Koshland and asked him if that is what he had in mind, and he said it was.

When we talk about 'a reaction' we may be encompassing a whole collection of different reactions with the same stoichiometry. For example, let's talk about the simplest of all chemical reactions, $H + H_2 \rightarrow H_2 + H$. Imagine a hydrogen molecule, A—B, in the gas phase being hit by a hydrogen atom, C: we call this *a* reaction, but there are at least four possible scenarios.

In the first, the incoming atom, C, collides with B, knocks it to infinity and sticks to A to form a new molecule, C—A. In the second, C collides with B and sticks to it to give C—B, while B's former partner, A, shoots off to infinity. In the third, the collision is so violent that all three atoms separate; if one waits long enough, two of them come together to form a molecule. In the fourth, the three atoms stick together to form an excited, triangular H_3 molecule which exists for a few vibrations before one of the three atoms separates. At this level of analysis, there are four distinct reactions, and they all have different 'transition states'. I once saw a vivid film made by Martin Karplus in which all four reaction paths were portrayed. Of course, gas-phase reactions can now be done under carefully tuned conditions such that only one of these reactions will occur. (But there must be reactions in solution, such as acetal hydrolysis, where so many detailed mechanisms have been proposed that it is quite possible that several reaction paths with similar rates may be operating simultaneously. One might call these polydromic or many-pathed reactions.)

Ron Breslow: There's no ambiguity as long as the atoms are not all hydrogen atoms. There's no ambiguity in real life; when an oxygen atom hits a hydrocarbon, a hydroxyl radical is made.

Dunitz: Nevertheless, this is the prototype which is often used.

Albery: Child (1967) showed that in this reaction, depending on where the energy was in the collision, it went into vibrations and rotations in the transition state. Collision theory and transition state theory gave identical results. For example, if the H atom comes from the side, the whole transition state will be set spinning.

Ron Breslow: What has been calculated is the first reaction that Jack Dunitz described. The energy minimum is linear, but there is this business about a set of trajectories that come in off-angle; the most probable angle is 20° or 30°. Experimentally, it doesn't appear that the most probable angle is 30°. In a 30° collision, the products have to be born with rotational excitation, not only vibrational energy. In fact, if a hydrogen atom attacks a CO_2 molecule at 30° there is inefficient transfer of energy into vibration; the hydrogen will bounce off. You can't hit a nail with a hammer at 30° to drive the nail in.

Rebek: I don't know if this is a special feature of intramolecular reactions, but I know of several examples of such reactions for which we concluded that these stereoelectronic effects are very soft indeed and one doesn't need colinear arrangements for S_N2 reactions or proton transfer reactions (Ballester et al 1990).

Ron Breslow: Can you give an example?

Rebek: Yes: in the enolization of the ketone with a xanthine skeleton shown in Fig. 3, you can watch the exchange of the methyl hydrogens, which go quite well (Tadayoni 1990). What's disturbing is that the methylene also exchanges at about the same rate, and no amount of modelling can give a reasonable product-like transition state. You can't get a decent-looking carboxylic acid and you have to distort the amide to get it close enough to react.

Ron Breslow: Internal S_N2 reactions have a strong preference for linear geometry.

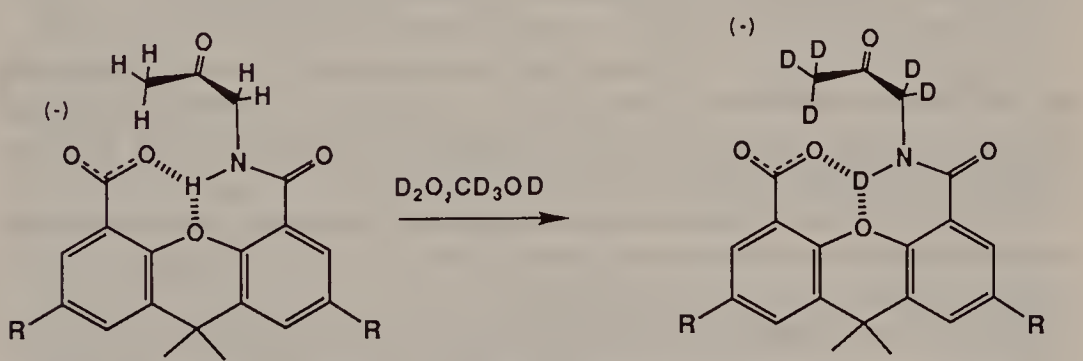


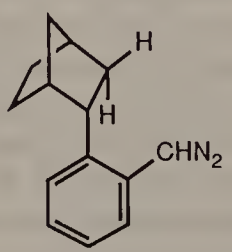
FIG. 3. (*Rebek*) Enolization of a ketone with a xanthine skeleton.

Rebek: Which reactions do you mean?

Ron Breslow: Internal displacements in which you can't do the 120° angles. The question is: how strict is the requirement for an angle of 180° in an S_N2 reaction?

Rebek: The system couldn't have reacted in an intramolecular sense.

Gutsche: The carbene generated from *endo*-2-(*o*-diazomethylphenyl)-bicyclo[2.2.1]heptane (**1**) intramolecularly inserts at the *exo* C—H bond at C-3 (to give a highly strained *trans* ring juncture) rather than the *endo* C—H bond (which would give a less strained *cis* ring juncture) (Baer & Gutsche 1971). We rationalized this on the basis of a perpendicular pathway approach—that is, a flankside approach, rather than a head-on approach.



1

Ron Breslow: In this reaction, attack is on the electrons, not on the nucleus.

Gutsche: This is an intramolecular example of a carbene C—H insertion. There are two competing theories: in the Benson–DeMore mechanism (Benson & DeMore 1964) a 180° approach (along the axis of the C—H bond) is proposed, but the Doering–Skell mechanism (Skell & Woodworth 1956, von Doering & Prinzbach 1959) involves a flankside approach to give a triangular transition state.

Ron Breslow: Essentially, H_3^+ is being made here; surely you want to go for the electrons, not for the end of the molecule.

Gutsche: I am happy that you support that idea. Roald Hoffman comes to the opposite conclusion on the basis of theoretical calculations.

Ron Breslow: Does he think it attacks at the back of the C—H bond?

Gutsche: According to the Benson–DeMore mechanism, which is supported by Hoffman's calculations, the hydrogen is extracted in the manner that I thought you were suggesting—that is, push the hydrogen in, pull it back out to create a biradical and then couple the biradical.

Ron Breslow: Even if you pull a hydrogen off, you surely don't do it from the back. I think it's important to have all these examples, because I will be perfectly happy if my idea is shot down.

Kellogg: The Sorensen experiment (Kirchen et al 1986) is, in my opinion, very informative. In the cyclodecyl carbonium ion a hydride is donated for a ring methylene group to form a 1,5-hydridocyclodecyl carbocation. This is a good

model of the hydride transfer reaction from a saturated carbon to a secondary carbonium system and has been used as a model for reactions involving NADH. This is a very selective process and it stops with the hydride essentially half-transferred under stable ion conditions (low temperature, non-nucleophilic solvent, non-nucleophilic counter ion). There is a choice between the formation of a symmetrical 1,6-hydrido system or an asymmetrical 1,5-hydrido system and the latter clearly wins.

Ron Breslow: Again, that involves an electrophilic attack on the C—H bond. The interesting question is whether attacks on nuclei that can't be linear are still fairly good.

Albery: You are not claiming that this happens in every reaction. There can be differently shaped passes at the transition state. Occasionally, it will be a good thing to come in head-on; equally, there must be other cases in which the pass is shallow and sloppy and therefore the steric requirements are less.

Ron Breslow: Yes, particularly if a complex, not a recoil, is involved. That's a different situation. In a complex there can be time for the energy to become distributed.

Albery: We should distinguish between bimolecular reactions in the gas phase, in which everything has to happen in a single position to get the energy right and the geometry right at the same time, and reactions in solution. In enzyme kinetics or in solution there is a chance for the reactants to get into the same region of space, to allow them to bounce about a bit and then acquire the energy necessary for reaction.

Ron Breslow: Within a collision complex there is a series of collisions. I am proposing that the successful subgroup of those collisions must be those which couple the motion into vibration to break the bonds.

References

Baer TA, Gutsche CD 1971 Chemistry of bivalent carbon intermediates. VI. Photolysis of 2-*n*-butylphenyldiazomethane. J Am Chem Soc 93:5180–5186

Ballester P, Tadayoni BM, Branda N, Rebek J 1990 Stereoelectronic effects in cyclization reactions. J Am Chem Soc 112:3685–3686

Benson SW, DeMore WB 1964 J Am Chem Soc 2:219

Breslow R, Chung S 1990 Additional flexibility solves the leaving group problem in cyclodextrin acylation. Tetrahedron Lett 31:631–634

Breslow R, Czarniecki MF, Emert J, Hamaguchi H 1980 Improved acylation rates within cyclodextrin complexes from flexible capping of the cyclodextrin and from adjustment of the substrate geometry. J Am Chem Soc 102:762–768

Breslow R, Trainor G, Uneo A 1983 Optimization of metallocene substrates for β-cyclodextrin reactions. J Am Chem Soc 105:2739–2746

Bürgi HB, Dunitz JD, Shefter E 1973 Geometrical reaction coordinates. II. Nucleophilic addition to a carbonyl group. J Am Chem Soc 95:5065–5067

Child MS 1967 The H + H_2 reaction in two dimensions. Faraday Discuss Chem Soc 44:68–73

Czarniecki MF, Breslow R 1978 Very fast acylation of beta-cyclodextrin by bound p-nitrophenyl ferrocinnamate. J Am Chem Soc 100:771-779

Flynn G 1989 Chemical cartography: finding the keys to the kinetic labyrinth. Science (Wash DC) 246:1009–1015

Jencks WR 1975 Binding energy, specificity, and enzymic catalysis; the Circe effect. Adv Enzymol 43:219–410

Kirchen RP, Sorensen TS, Wagstaff K, Walker AM 1986 The characterization of and the evolution of molecular hydrogen by the μ-hydrido-bridged cyclodecyl cation. Relevance to the question of hydrocarbon protection by superacids. Tetrahedron 42:1063–1070

Skell PS, Woodworth RC 1956 J Am Chem Soc 78:4496–4497

Storm DR, Koshland DE 1972 Effect of small changes in orientation on reaction rate. J Am Chem Soc 94:5815–5825

Sullivan JH 1967 Mechanism of the 'bimolecular' hydrogen–iodide reaction. J Chem Phys 46:73–78

Tadayoni BM 1990 Stereoelectronic effects at carboxyl oxygen. PhD thesis, University of Pittsburgh, PA, USA

von E Doering W, Prinzbach H 1959 Tetrahedron 6:24–30

Binding of peptides to proteins: an exercise in molecular design

Allen B. Edmundson*, Debra L. Harris†, Gordon Tribbick‡ and H. Mario Geysen‡

**Harrington Cancer Center, Amarillo, Texas 79106, USA, †Department of Biology, University of Utah, Salt Lake City, Utah 84112, USA and ‡Coselco Mimotopes Pty Ltd, Clayton, Victoria 3168, Australia*

Abstract. Peptides coupled to solid supports were systematically tested for binding activity with a polyreactive immunoglobulin light chain dimer by the methods of Geysen and colleagues. Once identified, peptides of progressively increasing affinity for the dimer were synthesized in milligram quantities and diffused into crystals of the protein. The three-dimensional structures of the peptide–protein complexes were determined by X-ray analysis and crystallographic refinement. Criteria for the design of ligands to fill the binding cavity in incremental stages could be formulated from the combined results of peptide scanning and crystallographic analyses. Histidine proved to be an important substituent in the binding series. It was possible to manipulate the properties of this amino acid residue to alter the structures and binding patterns of the ligands. For example, if two β-alanine residues werre added to the carboxyl end of a tetrapeptide ligand, the terminal carboxyl group formed an intramolecular ion pair with the imidazolium group (N-3) of histidine. This interaction was accompanied by cleavage of the intra molecular hydrogen bond between N-1 of histidine and the amide group of a glutamine side chain. The shape of the ligand shifted from a compact to an extended form and the mode of binding changed from a lock-and-key to an induced-fit type. The direction of entry of dipeptides of histidine and proline into the binding cavity (normally amino end first) could be reversed (carboxyl end first) by protonation of the histidine ring.

1991 Host–guest molecular interactions: from chemistry to biology. Wiley, Chichester (Ciba Foundation Symposium 158) p 213–230

Selection of host and guest molecules

Selection of a suitable host protein is crucial in an exercise for which the objective is the rational molecular design of peptide ligands. In the work described here, potential host molecules were screened to meet the following requirements. (1) It was of prime importance for the proteins to be amenable to peptide mapping (Geysen et al 1987) and also to produce crystals for the X-ray analysis of complexes. (2) Knowledge of the three-dimensional structure of the native

(unliganded) protein would facilitate progress. (3) Examinations of a series of related ligand–protein complexes would be further expedited if the crystal lattice of the protein contained large solvent channels through which peptides could be diffused. Attempts to co-crystallize ligands and proteins can be time-consuming and problematical when *all* members of a series need to be available in crystalline form.

Immunoglobulins and their fragments are excellent candidates for host molecules because they collectively offer many types of binding sites for guest compounds (e.g. cavities, pockets, slots, grooves, indentations, even flat surfaces and protrusions; see Marquart et al 1980, Amit et al 1986, Edmundson & Ely 1986a, Colman 1988, Anderson et al 1988, Davies et al 1990).

Among our present collection there are seven human and murine proteins for which the crystal structures have been determined, including an intact IgG immunoglobulin (Rajan et al 1983, L. Guddat & A. B. Edmundson, unpublished work), two antigen-binding fragments (Fabs; Gibson et al 1985, 1988, Herron et al 1989), a variable (V) domain dimer (Fv; Z. C. Fan, L. Shan, L. Guddat & A. B. Edmundson, unpublished work), and one homologous (Schiffer et al 1973, Edmundson et al 1975, Ely et al 1989) and two heterologous light chain dimers (Ely et al 1990, Guddat et al 1990). The parent IgM protein from which the Fv was derived did not bind peptides and the two Fabs were too tightly packed in the crystals to allow diffusion of peptides into the active sites. Supplies of heterologous light chain dimers prepared by hybridization procedures (Peabody et al 1980, Shaw et al 1987) are currently inadequate for a systematic design study. However, a monoclonal IgG immunoglobulin and a light chain dimer (Bence-Jones protein), both from one patient (Mcg) with multiple myeloma and amyloidosis, have been mapped for peptide binding and found to have closely overlapping recognition patterns (Tribbick et al 1989). The light chain dimer was available in much larger quantities than the IgG, produced an abundance of crystals and possessed a large combining site which was accessible to ligands diffusing through the crystal lattice (Edmundson & Ely 1986b, Edmundson et al 1987). The dimer thus met the stipulated requirements for a host molecule and was the logical choice for exploratory trials.

Selection of guest molecules appropriate for initial crystallographic experiments was based on ELISA (enzyme-linked immunosorbent assay) results with peptides linked to solid supports (Geysen et al 1987, Tribbick et al 1989). To identify minimal binding units, we tested the Mcg dimer with the 400 possible dipeptides of 20 common L-amino acids. *N*-Acetylhistidylproline and *N*-acetyl-tryptophanylmethionine were found to be equally effective (i.e., they acted as 'mimotopes'). Because of the oxidative side-reactions that are possible with tryptophan and methionine, preference was given to histidylproline as the starting peptide. In the next step the combinations of D- and L-enantiomers were tested to produce an improved ligand (Ac-Hp). The following abbreviations will be

used throughout this chapter: Ac for the *N*-acetyl group; one-letter codes for amino acids in the text and three-letter codes in the figures; capital letters for L-amino acids and lower-case letters for D isomers; NH_2 when an amide group is present at the C-terminus of the peptide and OH when there is a carboxyl group at the C-terminus. β-Alanine (3-aminopropanoic acid) is symbolized as β in the one-letter sequences.

Successive additions to the amino and carboxyl ends of the minimal unit established a binding series with major components of Ac-Hp $\ll$ Ac-fHp $<$ Ac-QfHp (Tribbick et al 1989). Only small increases in binding activity were noted when the lengths of the peptides exceeded four residues. Peptides were tethered via their carboxyl ends and were therefore assumed to enter the active site of the protein with their acetyl groups in the lead.

X-ray analyses of peptide–protein complexes

For crystallographic experiments a single crystal of the Mcg dimer was placed in each of several tubes containing 0.3 ml of 1.9 M ammonium sulphate buffered at pH 6.2 with 0.2 M sodium phosphate (crystallizing medium). The solution over each crystal was saturated with one of the peptides and allowed to stand at 12 °C for one month. After this period, X-ray diffraction data were collected to 2.7 Å resolution with a Nicolet P21 diffractometer (see Edmundson et al 1989). Locations, structures and relative occupancies of the bound peptides were determined by difference Fourier analyses. Protein phases were obtained from the crystallographically refined structure of the native protein (Edmundson et al 1987, Ely et al 1989). After models were fitted to electron density maps with the graphics programs FRODO (Jones 1978, Pflugrath et al 1984) and PMS (R. J. Athay, as quoted in Rajan et al 1983), the structures of the ligand–protein complexes were crystallographically refined with the program PROLSQ (Hendrickson 1985). Alternating cycles of refinement and model building were continued until the reliability factors (R-factors) reached a plateau at 20–22% and the stereochemistry appeared to be satisfactory.

In X-ray analyses of 20 complexes the peptide ligands were all found to be accommodated in the designated active site and not in any other parts of the protein dimer. An example of such a complex is shown in Fig. 1. The protein is represented as an α-carbon skeletal model and the peptide ligand (Ac-QfHpβ-OH) as a Corey-Pauling-Koltun (CPK) space-filling model.

As judged from the relative peak heights of electron density assigned to peptides in difference Fourier maps, the crystal analyses echoed the general findings of the peptide-scanning procedures. For example, if the occupancy of the most favoured ligand Ac-QfHp-OH was set at 1.0 mole per mole of protein, the relative occupancies of Ac-Hp-NH_2 and Ac-fHp-NH_2 were 0.25 and 0.89 under the same soaking conditions. The dipeptide was not represented by continuous electron density and may have been present in two overlapping

FIG. 1. α-Carbon tracing of the Mcg light chain dimer with a Corey-Pauling-Koltun (CPK) solid model of the peptide Ac-QfHpβ-OH in its binding site between the two variable (V) domains. The two constant (C) domains are on the left. Monomer 1 of the light chain dimer is on top and monomer 2 at the bottom of the protein model. Hydrogen atoms, which are not detectable by X-ray analysis at the resolution (2.7 Å) of the present structural study, are added to the CPK model in their probable locations. This ligand occupies most of the space available in the main binding cavity but does not distort the protein constituents (lock-and-key model of binding). D-Proline and β-alanine lie outside the cavity on the far right.

subsites. It was therefore necessary to develop special procedures to obtain unambiguous binding patterns for dipeptides (see later section).

Patterns for larger ligands were more definitive, as illustrated with Ac-fβHp-NH_2 in Fig. 2. In the upper panel the electron density corresponding to this ligand is co-displayed with the protein segments making up the binding site. A skeletal model fitted to the electron density by interactive computer graphics is shown in the lower panel. This peptide is part of the following crystallographic binding series, which will be the major topic of this discussion: Ac-Hp-$NH_2 \ll$ Ac-fHp-NH_2 = Ac-fβHp-NH_2 < Ac-QfHp-OH = Ac-QfHpβ-OH.

Incremental improvements in the design of peptide ligands

The largest jump in occupancy and positional stability occurred when D-phenylalanine was added to the dipeptide (binding pattern not shown). This

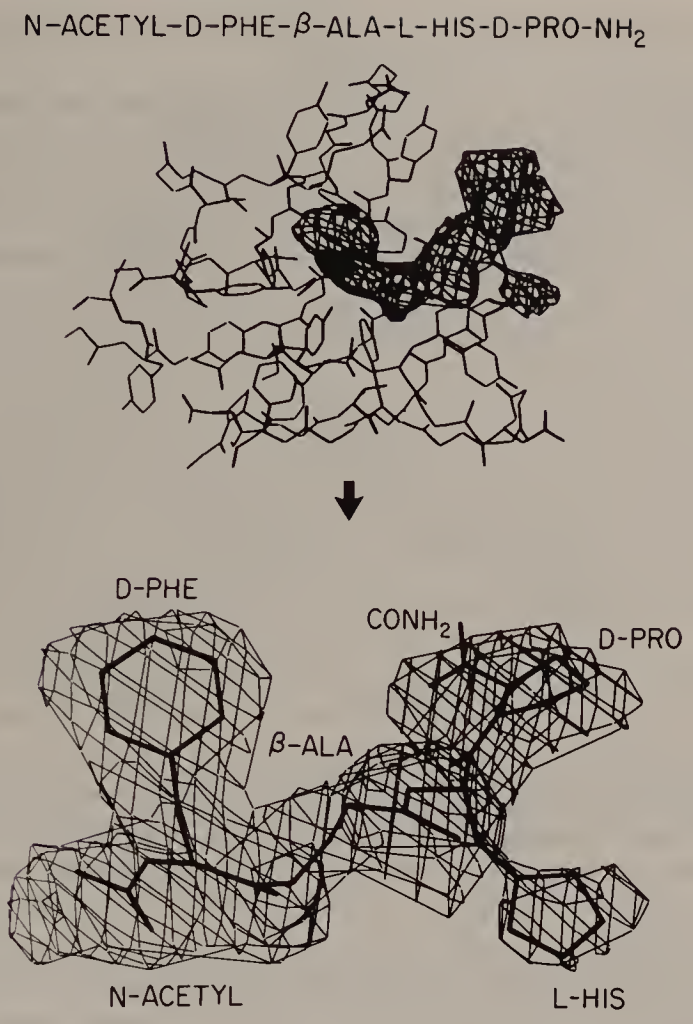


FIG. 2. *Top*: Three-dimensional 'cage' electron density for a tetrapeptide as bound in the binding cavity of the Mcg dimer (skeletal models of polypeptide backbone and amino acid side chains). Monomer 1 constituents line the top of the cavity and monomer 2 the bottom. The cage density was obtained by difference Fourier analysis of X-ray diffraction data collected with crystals of the peptide–protein complex and similar data for the native (unliganded) protein.
Bottom: Skeletal model of the tetrapeptide fitted to the cage density by interactive computer graphics and then subjected to cyrstallographic refinement. D-Phenylalanine (D-Phe) occupies a hydrophobic pocket (biased toward monomer 1) and L-histidine (L-His) lodges near monomer 2 at the entrance of the cavity. These subsites were favoured by the phenylalanyl and histidyl side chains in the tri-, tetra- and pentapeptides in the binding series, although the direction of entry changed as the ligands were elongated from the tripeptide stage.

residue lodged in a hydrophobic pocket (the same as the one seen in Fig. 2) and accounted for 20 of the 45 'contacts' Ac-fHp-NH_2 made with the protein. Two atoms were considered to be in contact if they were separated by $\leqslant 4$ Å (sum of van der Waals radii plus an average tolerance of about 0.4 Å). Strength

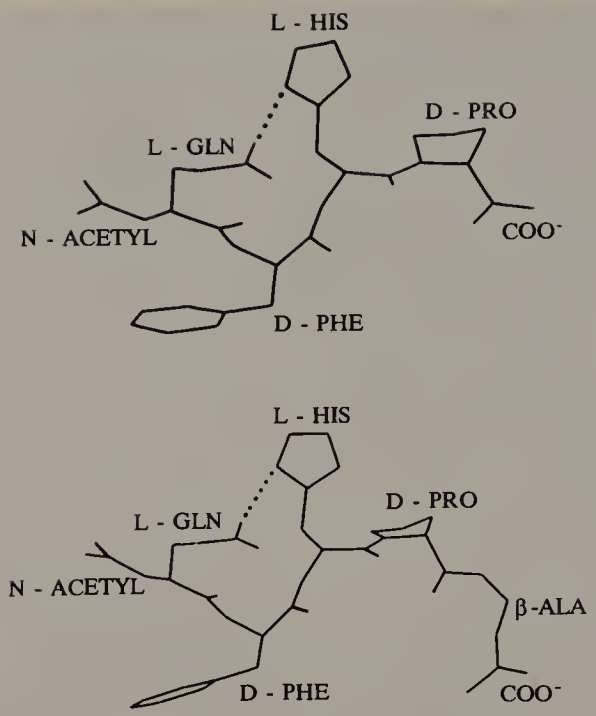


FIG. 3. *Top*: Skeletal model of the optimized tetrapeptide ligand as bound in the cavity of the Mcg dimer. The *N*-acetyl group is located on the floor of the cavity (not shown) and D-proline is near the entrance. This view is roughly 180° from that of Fig. 2. The dotted line indicates an intramolecular hydrogen bond between the carbonyl oxygen atom on the L-glutaminyl side chain and the hydrogen atom on the imidazolium group of L-histidine.
Bottom: Skeletal model of a bound pentapeptide differing from the upper peptide only by the addition of β-alanine at the carboxyl end. Note that there are no major changes in the structure assumed by the ligand inside the cavity (entrance near D-proline).

of binding was roughly correlated with the number of such interactions.

If the peptide was elongated by the insertion of a flexible β-alanine 'spacer' between the phenylalanine and histidine residues, the *N*-acetyl group shifted position, but phenylalanine remained in the same pocket (see Fig. 2). Though displaced to a different starting location, the histidine side chain was directed into the same general subsite that it had occupied as a substituent of Ac-fHp-NH_2. Proline was shunted out of the cavity into bulk solvent.

The crystal structure of the complex with Ac-QfHp-OH revealed why glutamine was strongly favoured for the elongation of Ac-fHp in the peptide-scanning assays (see Fig. 3, top panel). Though emanating from the peptide near the floor of the cavity, the side chain of glutamine extended toward the entrance and participated in interactions which helped to anchor the ligand and to stabilize its internal structure. The methylene groups closely conformed to the hydrophobic surfaces of protein side chains (Tyr-93 and Phe-99 of monomer 1 of the light chain dimer), while the amide group approached to within

hydrogen-bonding distance (2.8 Å) of the imidazolium group of L-histidine (see Fig. 3).

Proline played an important but indirect role in the binding of the ligands. It did not participate in many interactions with the protein, but added rigidity to one end of the peptide and helped maintain histidine in an orientation favourable for binding. If desired, a flexible β-alanine residue could be inserted to counterbalance the rigidity of proline.

The *N*-acetyl group proved to be one of the active substituents in the binding process (see Figs. 2–4). In peptides containing three or more residues, this substituent was located on or near the floor of the cavity. The number of interatomic contacts with the protein ranged from 12 to 22 in the series. The orientation of the acetyl group in each peptide was dictated by the formation of a hydrogen bond between the carbonyl oxygen and a hydroxyl group of tyrosine or serine. Tyr-38, Ser-36 and Ser-91 of monomers 1 and 2 of the light chain dimer are clustered on the floor of the cavity. Singly or in pairs, all have been observed to be within hydrogen-bonding distance of the *N*-acetyl group in at least one member of the binding series.

In both the peptide-scanning procedure and the X-ray series, the *N*-acetylated tetrapeptide Ac-QfHp seemed to represent an excellent design for a space-filling ligand. It was compact, internally stabilized, and closely matched to the topography of the binding site without any major conformational distortions of the protein constituents. The structure of this complex thus conformed to a lock-and-key model of binding.

It is interesting that Ac-qf was the dipeptide most favoured for binding in the IgG immunoglobulin from patient Mcg (Tribbick et al 1989). In this protein the light chain components have the same amino acid sequence as those in the Mcg dimer, although the binding sites are formed from a combination of light and heavy chains instead of two light chains. On the assumption that the light chain could influence binding in the IgG protein, the latter was tested with the same sets of peptides as the Mcg dimer. At the tripeptide level the most reactive ligand for IgG was Ac-fHp and the two most active extensions of Ac-fHp were Ac-qfHp and Ac-QfHp. Thus the two proteins showed marked similarities in binding patterns, with the dominant dipeptide sequences coming together at the tetrapeptide stage.

Effects of altering the design

Borrowing concepts used in traditional drug design, we deliberately modified the seemingly optimized ligand to see if the alterations led to changes in binding activity. The first modification was an extension of Ac-QfHp with one β-alanine residue having a terminal carboxyl group (Ac-QfHpβ-OH). X-ray analysis indicated that the binding pattern inside the cavity was practically indistinguishable from that of Ac-QfHp-OH (see Fig. 3), although the number

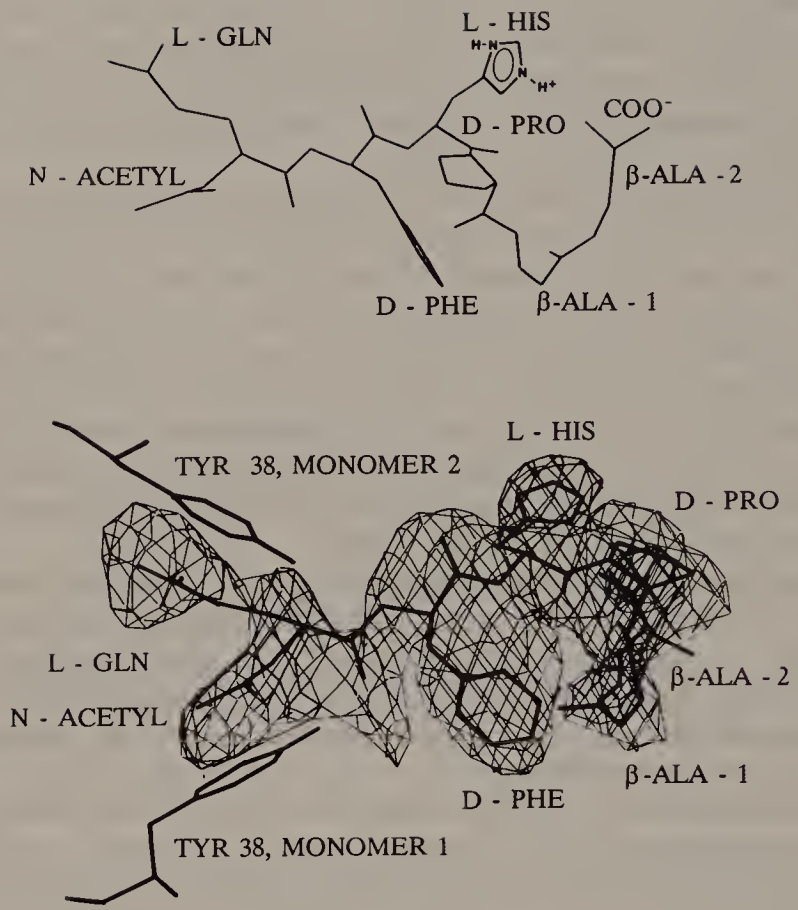


FIG. 4. *Top*: Conformation of a hexapeptide produced by adding two β-alanine residues to the optimized tetrapeptide ligand shown in Fig. 3. An ion pair is formed by the carboxyl group on β-alanine-2 and the protonated nitrogen on L-histidine. The hydrogen bond between L-glutamine and L-histidine in the tetrapeptide is broken and the bound peptide adopts an extended conformation.
Bottom: Cage electron density and superimposed skeletal model of the hexapeptide in a slightly different view from the top panel. Note that the *N*-acetyl group and the L-glutaminyl side-chain form a double wedge forcing an opening in the floor of the cavity. These groups interact with the under-surfaces of the phenolic rings of corresponding tyrosine residues (Tyr-38) of monomers 1 and 2. The binding pattern thus conforms to an induced-fit model of binding.

of interatomic contacts increased marginally from 80 to 87. The extra residue was accommodated outside the cavity and was classified as a neutral adduct.

When two β-alanine residues were added to produce Ac-QfHpββ-OH, the basic binding pattern was thoroughly disrupted (see Fig. 4). Two adjacent β-alanine residues permitted the tail of the peptide to be turned sharply toward

the L-histidine side chain (upper panel of Fig. 4). The driving force for this movement appeared to be the formation of an ion pair between the terminal carboxyl group and the protonated form of histidine (N-3). This electrostatic interaction was obviously favoured over the hydrogen bonding between N-1 of histidine and the carbonyl group on the glutamine side chain, as seen earlier in Fig. 3. We assume that the histidine ring was pulled slightly beyond the hydrogen-bonding range of the glutamine side chain. Irrespective of the mechanism, the glutamine residue swung away from histidine and penetrated through a cluster of four aromatic side chains which are in van der Waals contact in the unliganded (native) structure (i.e., Tyr-38 and Phe-101 from each of the two protein monomers). The planar amide group of glutamine stacked with the ventral surface of the phenolic ring of Tyr-38, monomer 2 (lower panel, Fig. 4). Surprisingly, the carbonyl group of the *N*-acetyl substituent participated in an analogous stacking interaction with the corresponding tyrosine of monomer 1. D-Phenylalanine still resided in the hydrophobic pocket occupied by other members of the series, but its direction of entry was more similar to that of Ac-fHp-NH_2 than Ac-QfHp-OH.

A stacking interaction of glutamine (Gln-112) with tyrosine (Tyr-145) had previously been observed in the crystal structure of monomer 2 of the dimer (Edmundson et al 1975, Ely et al 1989). There are also many examples of electronegative carbonyl oxygen atoms interacting with the hydrogen atoms of unsubstituted phenyl rings, the proposed mechanisms being electrostatic ones (Thomas et al 1982, Rowland et al 1990). Though unexpected with this ligand, such interactions can now be considered as possible stabilizing features in the formation of ligand–protein complexes.

Overall the number of interatomic contacts more than doubled (from 87 to 182) with the addition of a second β-alanine residue. As the equilibria involving histidine shifted from hydrogen bonding to stronger electrostatic interactions, the compact conformation associated with Ac-QfHpβ-OH changed to an extended form which could no longer be confined to the main cavity. We assumed that the weak van der Waals interactions among aromatic side chains on the floor could be transiently disrupted by 'breathing' or thermal motion to allow the *N*-acetyl group and glutamine side-chain to wedge through and enlarge the opening. Clearly, the binding of the maverick ligand can be explained by an 'induced-fit' mechanism.

Predictive value of current results for the molecular design of peptide ligands

In peptide-scanning assays at neutral pH, there was a decided preference for Ac-Hp over other combinations of optical isomers and over any isomeric variations of Ac-PH. The planar and polar, but uncharged, histidine residue was therefore thought to enter the cavity before the puckered five-membered ring of proline. At pH 6.2—the approximate pK of a typical histidine ring and the

nominal pH of the crystallizing medium used for the Mcg dimer—the directionality arguments are different for a dipeptide. As outlined in the preceding section, equilibria involving histidine residues can be manipulated to alter peptide binding patterns.

With histidine being partially charged in Ac-Hp-NH_2, we surmised that proline might enter the cavity first. This supposition seemed plausible, especially in view of the relatively low affinity for the peptide in crystals. Moreover, there were three negatively charged residues on the rim of the cavity, but none in the interior.

Initial tests of the predictions were inconclusive because of the low solubilities and occupancies of Ac-Hp-NH_2 and Ac-hP-NH_2. To increase their solubilities we redesigned the ligands to have a terminal α-carboxyl group instead of an amide (i.e., Ac-Hp-OH and Ac-hP-OH). For crystal-binding studies the soaking periods were extended to more than four months to encourage maximal occupancy and positional stability.

X-ray analysis of the complexes resolved previous ambiguities. As predicted, the direction of entry into the cavity was the reverse of that of the tethered dipeptides binding at pH 7. Protonated histidine was retained near the entrance

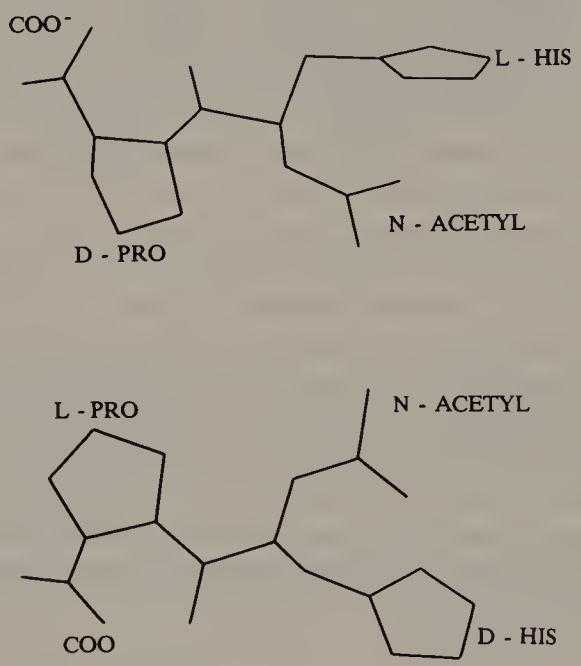


FIG. 5. Conformations of bound dipeptides of enantiomers of histidine and proline in crystals of the Mcg dimer at pH 6.2. Proline is near the floor of the cavity (not shown) and histidine is at the entrance. The mirror plane between the peptides is only approximate because the subsites occupied in the cavity are not quite equal and opposite. In peptide-scanning assays at neutral pH the uncharged histidine residue leads C-terminal proline into the cavity. At pH 6.2 in crystals the direction of entry is reversed, presumably because the histidine ring is partially protonated.

and C-terminal proline was accommodated near the floor of the cavity. Crystallographically refined structures of the two bound peptides are shown in Fig. 5.

The burial of a negatively charged carboxyl group in each peptide was partially compensated for by hydrogen bonding with the phenolic hydroxyl group of a tyrosine residue. Ac-Hp-OH interacted with Tyr-38 of monomer 1 (measured distance of 3.2 Å between the oxygen atoms of the carboxyl and hydroxyl groups), while Ac-hP-OH approached the corresponding tyrosine side chain in monomer 2 (inter-oxygen distance of 2.8 Å).

The view in Fig. 5 was chosen to emphasize the mirror symmetry between the appropriate pairs of enantiomers. The two binding patterns were not strictly equal and opposite because the two protein monomers are not identical and the cavity has only approximately two-fold symmetry.

These results support the assumption of a dominant role for protonated histidine in the binding of the dipeptides at pH 6.2 in these protein crystals. In both cases there was a clear choice: a negatively charged carboxyl group was buried in preference to the positively charged imidazolium ion. With tethered dipeptides at higher pH values the uncharged histidine residue presumably penetrated further into the cavity, but there is not as yet a crystallographic model for this type of binding.

Conclusions

Peptide-scanning techniques (Geysen et al 1987) could be used to select a series of ligands with progressively increasing affinities for a light chain dimer. When synthesized and diffused into crystals of the dimer, these peptides closely conformed to the topography of the binding cavity without major distortion of the protein constituents (lock-and-key model of binding). When criteria had been established for the design of a site-filling ligand, the structure of such an optimized peptide (Ac-QfHp-OH) was modified to shift the intramolecular interactions and the equilibria involving histidine (Ac-QfHpββ-OH). The conformation of the bound ligand changed from a compact, hydrogen-bonded form to an extended molecule which disrupted the protein structure by penetrating the floor of the cavity (induced-fit model of binding). Dipeptides of histidine and proline were designed to have free α-carboxyl groups to increase solubility and to overcome problems of positional instability when they were bound to the dimer. We found that protonation of the imidazole ring of histidine could be used to programme an end-to-end reversal of the direction in which the dipeptide was inserted into the binding cavity.

Acknowledgements

We gratefully acknowledge the advice and support of James Herron and Xiao-min He, particularly at the beginning of this project. We also thank the entire peptide synthesis

group at Coselco Mimotopes Pty Ltd. The group synthesized and characterized all peptides cited in this chapter. Barbara Staker, Brad Nelson, Joanne Pruitt, Brian Schley, Leif Hanson, Luke Guddat, Larry Litke and Melody Robinson offered helpful suggestions and assisted with the illustrations. Suzanne Klause and Twyla Slay prepared the manuscript for publication. The work of Allen B. Edmundson and Debra L. Harris was supported by Grant CA 19616 awarded by The National Cancer Institute, Department of Health and Human Services, and by the Research Institute, Harrington Cancer Center.

This chapter is dedicated to Dr Neville McCarthy, on the occasion of his retirement as Managing Director of the Commonwealth Serum Laboratories, Parkville, Victoria, Australia.

References

Amit AG, Mariuzza RA, Phillips SEV, Poljack RJ 1986 Three-dimensional structure of an antigen–antibody complex at 2.8 Å resolution. Science (Wash DC) 233:747–753

Anderson WF, Cygler M, Braun RP, Lee JS 1988 Antibodies to DNA. Bioessays 8:69–74

Colman PM 1988 Structure of antibody–antigen complexes: implications for immune recognition. Adv Immunol 43:99–132

Davies DR, Padlan EA, Sheriff S 1990 Antibody–antigen complexes. Annu Rev Biochem 59:439–473

Edmundson AB, Ely KR 1986a Determination of the three-dimensional structures of immunoglobulins. In: Weir DM, Herzenberg LA, Blackwell C, Herzenberg LA (eds) Handbook of experimental immunology, 4th edn. Blackwell Scientific Publications, Oxford, vol 1:15.1–15.23

Edmundson AB, Ely KR, 1986b Three-dimensional analyses of the binding of synthetic chemotactic and opioid peptides in the Mcg light chain dimer. In: Synthetic peptides as antigens. Wiley, Chichester (Ciba Found Symp 119) p 107–129

Edmundson AB, Ely KR, Abola EE, Schiffer M, Panagiotopoulos N 1975 Rotational allomerism and divergent evolution of domains in immunoglobulin light chains. Biochemistry 14:3953–3961

Edmundson AB, Ely KR, Herron JN, Cheson BD 1987 The binding of opioid peptides to the Mcg light chain dimer: flexible keys and adjustable locks. Mol Immunol 24:915–935

Edmundson AB, Ely KR, He XM, Herron JN 1989 Cocrystallization of an immunoglobin light chain dimer with bis(dinitrophenyl) lysine: tandem binding of two ligands, one with and one without accompanying conformational changes in the protein. Mol Immunol 26:207–220

Ely KR, Herron JN, Harker M, Edmundson AB 1989 Three-dimensional structure of a light chain dimer crystallized in water; conformational flexibility of a molecule in two crystal forms. J Mol Biol 210:601–615

Ely KR, Herron JN, Edmundson AB 1990 Three-dimensional structure of a hybrid light chain dimer: protein engineering of a binding cavity. Mol Immunol 27:101–114

Geysen HM, Rodda SJ, Mason TJ, Tribbick G, Schoofs PG 1987 Strategies for epitope analysis using peptide synthesis. J Immunol Methods 102:259–274

Gibson AL, Herron JN, Ballard DW et al 1985 Crystallographic characterization of the Fab fragment of a monoclonal anti-ss-DNA antibody. Mol Immunol 22:499–502

Gibson AL, Herron JN, He XM et al 1988 Differences in crystal properties and ligand affinities of an antifluorescyl Fab (4-4-20) in two solvent systems. Proteins Struct Funct Genet 3:155–160

Guddat L, Fan Z, Shan L, Edmundson AB 1990 Structure of a hybrid dimer of two immunoglobin light chains. Abstr Am Crystallogr Assoc p 88

Hendrickson WA 1985 Stereochemically restrained refinement of macromolecular structures. Methods Enzymol 115:252–270
Herron JN, He XM, Mason ML, Voss EW Jr, Edmundson AB 1989 Three-dimensional structure of a fluorescein–Fab complex crystallized in 2-methyl-2,4-pentanediol. Proteins Struct Funct Genet 5:271–280
Jones TA 1978 A graphics model building and refinement system for macromolecules. J Appl Crystallogr 11:268–272
Marquart M, Deisenhofer J, Huber R, Palm W 1980 Crystallographic refinement and atomic models of the intact immunoglobulin molecule Kol and its antigen-binding fragment at 3.0 Å and 1.9 Å resolution. J Mol Biol 141:369–392
Peabody DS, Ely KR, Edmundson AB 1980 Obligatory hybridization of heterologous immunoglobulin light chains into covalently linked dimers. Biochemistry 19:2827–2834
Pflugrath JW, Saper MA, Quiocho FA 1984 New generation graphics system for molecular modeling. In: Hall S, Ashiaka T (eds) Methods and applications in crystallographic computing. Clarendon Press, Oxford, p 404–407
Rajan SS, Ely KR, Abola EE et al 1983 Three-dimensional structure of the Mcg IgG1 immunoglobulin. Mol Immunol 20:787–799
Rowland RS, Carson WM, Bugg CE 1990 Applications of crystallographic databases in molecular design. In: Laver WG, Air GM (eds) Use of X-ray crystallography in the design of antiviral agents. Academic Press, San Diego, p 261–281
Schiffer M, Girling RL, Ely KR, Edmundson AB 1973 Structure of a λ-type protein at 3.5-Å resolution. Biochemistry 12:4620–4631
Shaw DC, Harris DL, He XM et al 1987 A mild method for the preparation of disulfide-linked hybrids of immunoglobulin light chains. Mol Immunol 3:267–274
Thomas KA, Smith GM, Thomas TB, Feldman RJ 1982 Electronic distributions within protein phenylalanine aromatic rings are reflected by the three-dimensional oxygen atom environments. Proc Natl Acad Sci USA 79:4843–4847
Tribbick G, Edmundson AB, Mason TJ, Geysen HM 1989 Similar binding properties of peptide ligands for a human immunoglobulin and its light chain dimer. Mol Immunol 26:625–635

DISCUSSION

Crumpton: How much do you think your data are influenced by the technique that you use? You start with a crystal, which has restricted flexibility compared with the protein in solution. Although you showed elegantly that there is an opportunity for some change in structure of the crystalline protein, surely you would have to admit that that opportunity is limited. A crystal often shatters when you diffuse ligands into it. We explain this by saying that there is a change in conformation on ligand binding and that this change cannot be accommodated within the relatively rigid framework (restricted flexibilty) of the crystal. Would you agree that within this context, you are superimposing a limitation on the analysis that you are able to achieve, in that you can accommodate only a restricted amount of structural change—that which preserves the crystal—whereas those situations in which binding of a ligand results in crystal disruption are obviously not available for analysis. The alternative would be to start with a protein and a peptide in solution, and co-crystallize the two. It would be interesting to make the comparison between these two approaches.

Edmundson: I always like Michael's questions because I agree with everything he says! Working with a crystal is restrictive. We don't consider the formation of complexes that destroy the crystals, because we don't do those experiments. We certainly can cite examples of such complexes, but not from the present series, which are preselected peptide ligands that have already been demonstrated to bind to the light chain dimer in solution. In the past we have diffused fluorescent compounds like rhodamine B into the dimer and found that the crystal shatters within five hours. We have chosen to work within a crystallographic framework and feel that we are lucky to detect conformational changes—most crystallographic groups do not report any changes in crystals. Provided that the ligands are preselected, they can often be diffused into crystals without significant changes in the protein. We can also engineer changes in the protein by deliberately altering the ligand, but, as you said, we cannot force major changes or we will destroy the crystal.

Esther Breslow: I was impressed by how high the binding constants are relative to the flexibility there is inside the cavity. We have been studying a protein that is specific for the binding of a particular peptide or related peptides (Breslow & Burman 1990). The contacts seem much tighter than yours and the cavity is intolerant of substitutions. The binding constants are lower, partly because of the conformational change that the protein undergoes on going from an unliganded to a liganded state, but this conformational change results in a tight fit around the peptide. The peptide cannot move around much and substitution of D residues for most L residues leads to loss of binding. In the antibodies you described the cavity is remarkably tolerant and at the same time it provides the opportunity for tight interactions.

Edmundson: In all likelihood the main cavity of the Mcg dimer is an example of a very large binding site. The binding site for fluorescein in the 4-4-20 monoclonal antibody is much smaller. 4-4-20 is a high affinity molecule and there is not much room for a ligand to move in the active site (Herron et al 1989). We can therefore specify the types of interaction that keep fluorescein firmly in the site with a binding constant of $10^{10}\,M^{-1}$. There are only six contact residues to account for this high affinity constant. In the Mcg dimer there are 21 potential contact residues in the main cavity alone. Almost every amino acid in the interface between the two variable domains can be utilized over a span of more than 20 Å, depending on the nature of the ligand to be bound. There is an example in which one ligand molecule, bis(dinitrophenyl)lysine, is bound in the main cavity and another ligand is bound in the deep pocket. The bis(dinitrophenyl)lysine ligand fits into the main cavity like a hand in a glove, and binding of the second is associated with major conformational changes in the lining of the deep pocket. In the main cavity the polypeptide backbone is usually not involved in binding because it is shielded by so many side chains. The cavity is sufficiently voluminous for these side chains to rotate and translate into new orientations.

In the case of your protein, the site may not be large enough to allow the side chains to move around as much as they do in our system. The side chains in the Mcg dimer can move around a ligand and hold it in a particular position. We seldom see changes in the backbone; when we do, the shifts are of the order of 1–2 Å.

The hypervariable loops of antibodies tend to be flexible and have a number of aromatic residues which act as constituents of combining sites. In the 4-4-20 antibody against fluorescein, which is a dianion, there has to be compensation for two charges, but there is also a large aromatic core to be immobilized. The two negative charges are balanced by formation of an ion pair with arginine and hydrogen bonds with serine and tyrosine. The aromatic rings of the ligand interact with two tryptophans, a tyrosine and a histidine residue.

The Mcg dimer is a model for a primitive antibody of unknown specificity. These properties were considered advantageous when we were looking for a host molecule to use for *de novo* design of ligands. What shocked us was the finding that the light chain dimer and the IgG immunoglobulin from Mcg were virtually indistinguishable in their binding properties when compared by the peptide-scanning techniques of Geysen et al. That is a surprising result from a structural viewpoint, since the IgG molecule is composed of a heavy chain and a light chain, whereas the Mcg dimer is made up of two light chains.

Sutherland: Can you say what type of protein will be flexible? I realize this protein has a large cavity, which must increase flexibility, but can you relate flexibility to protein structure itself?

Edmundson: The hypervariable regions contributing to the binding sites are all loops. They are not tied down in organized secondary structures at the entrance to the binding region. In the Mcg dimer the parts of the polypeptide chains that move the most are unrestricted by interactions with other chains. The side chains lining the cavity generally have space to move freely. At the floor of the cavity there are four large side chains (two tyrosines and two phenylalanines) in van der Waals contact. These side chains form a barrier between the main cavity and the deep pocket. For a ligand to breach this barrier, it would probably be necessary for several fluctuations of the protein molecule to occur at the same time (e.g. 180° flips on a millisecond time scale). That is possible, as Gordon Roberts showed with dihydrofolate reductase (p 173). The active site of this enzyme contains leucine and phenylalanine, which have bulky side chains well suited for general binding of the substrate. However, a small side chain (Asp-26) exerts the fine control of the mode of substrate binding, in which the orientation of the pteridine ring can differ by 180° in two conformational states.

Ed Voss and his co-workers made a series of anti-fluorescyl monoclonal antibodies, and the high affinity 4-4-20 antibody is the only one that has an arginine residue in a position appropriate to neutralize one of the negative charges on the ligand (Bedzyk et al 1990). The other antibodies all

have histidine in this position, and have 2–3-fold lower binding affinity for the ligand.

In the large hydrophobic cavity of the Mcg dimer, most of the binding of peptides can be accounted for by general interactions with bulky side chains. The specificity is provided by fine adjustments in the amino acid sequence of the ligand and alteration of its conformation when bound.

Esther Breslow: What's in the cavity when the peptide isn't there?

Edmundson: Water, but it's not present in an organized shell. We have seen no ions anywhere near the structure of the binding cavity. Only 12 ordered water molecules were found in the main cavity although there's room for about 25. In crevices not occupied by peptide ligands in the complexes there were occasionally small blips of electron density that might correspond to trapped water molecules.

Crumpton: In immunology, specificity is the hallmark of antibodies. There are many examples of exquisite specificity. It seems to me that your data are challenging the basic premise of antibody specificity. If the effect was seen solely with the Bence-Jones protein you might be able to get away with it, because the Bence-Jones protein isn't actually an antibody, being simply two light chain variable regions making a cavity; we know that evolution has rejected this in favour of making the binding site from a heavy chain and a light chain. But the whole immunoglobulin, assayed using peptides on pins, shows the same spectrum of specificity as the light chain dimer. So, I imagine you consider the light chain dimer not to be unusual and that variable specificity is a characteristic feature of the whole antibody as well as the light chain dimer. This raises the question of whether there is something odd about the immunoglobulin that you have chosen to study. Is it unusual, or is it really representative of antibodies in general?

Edmundson: I hope that we have challenged the field with these simple concepts. I don't think that specificity is incompatible with polyreactivity coupled with different binding affinities. It seems likely that specificity can be overlaid on general patterns of binding. The example of arginine substitution in the anti-fluorescyl antibodies is a good one. These antibodies all bind fluorescein, but 4-4-20 binds the ligand with substantially higher affinity. I would bet that these antibodies would also bind peptides with moderate affinities.

There is no reason to believe that the Mcg light chain is an unusual molecule. Shark light chains that have been cloned have a sequence that can fit into the Mcg structure with almost no changes (Schluter et al 1989). To cover ourselves, we have always considered the Mcg dimer as a model for a primitive antibody—that helps us deal with immunologists who yell and scream!

Ron Breslow: I wonder whether to have high specificity, high affinity is also needed. That might be the real correlation. Your antibodies are of relatively low affinity, and only some of the available contacts are being used. The ligands seem to fill the whole cavity in the high affinity antibodies. Perhaps your

antibodies and ligands are unusual in the sense that the whole of the binding site is not used.

Edmundson: The optimized ligand Ac-QfHp-OH filled the main cavity and used most of the available contacts. When this ligand was extended to Ac-QfHpββ-OH, the interaction patterns were drastically changed because the ligand was spread between two binding compartments. In this case the whole of the binding site was not utilized. I agree that we haven't looked at enough high affinity sites. In the only one we have considered (4-4-20), the binding pattern was consistent with your suggestion—fluorescein occupied almost all of the available space. It is interesting that the ligand did not enter the active site from the top, but from the side in this case. Because it was presented to the immune system conjugated to a lysine side chain on a large carrier molecule (keyhole limpet haemocyanin), the ligand could gain entry to the active site of the antibody only from the side where there is a docking surface for a carrier molecule. In the tight slot where binding occurs in the antibody there is an opening of the right size and shape for a lysine side chain to extend to a macromolecule outside the site.

Brändén: There are antibodies in which the binding site is not a deep pocket, but a flat surface. Your structure is representative of hapten-binding antibodies, not protein-binding ones. With this deep binding pocket you have a system that could be used to test the tools that are available for predicting how a ligand will bind.

Edmundson: In a way, we do that every day now that we know the basic rules about what will bind to the Mcg dimer. However, I have only about 50% success in guessing which new molecules will bind and we have investigated more than 45 complexes. The one assurance that we can give is that all ligands that bind in Mario Geysen's system also bind in our system and, so far, binding has always occurred in the designated active site. However, these ligands are preselected and we have now reached the point at which we should guess what will bind. Many years ago, I predicted that opioid and chemotactic peptides would bind to the Mcg dimer and they did (Edmundson & Ely 1986). We sometimes regret such an era predated the more sensible systematic approach of today.

References

Bedzyk WD, Herron JN, Edmundson AB, Voss EW Jr 1990 Active site structure and antigen binding properties of idiotypically cross-reactive anti-fluorescein monoclonal antibodies. J Biol Chem 265:133–138

Breslow E, Burman S 1990 Molecular, thermodynamic and biological aspects of recognition and function in neurophysin–hormone complexes: a model system for the analysis of protein–peptide interactions. Adv Enzymol 63:1–67

Edmundson AB, Ely KR 1986 Three-dimensional analyses of the binding of synthetic chemotactic and opioid peptides in the Mcg light chain dimer. In: Synthetic peptides as antigens. Wiley, Chichester (Ciba Found Symp 119) p107–129

Herron JN, He XM, Mason ML, Voss EW Jr, Edmundson AB 1989 Three-dimensional structure of a fluorescein–Fab complex crystallized in 2-methyl-2,4-pentanediol. Proteins Struct Funct Genet 5:271–280

Schluter SF, Hohman VS, Edmundson AB, Marchalonis JJ 1989 Evolution of immunoglobulin light chains: cDNA clones specifying sandbar shark constant regions. Proc Natl Acad Sci USA 86:9961–9965

Involvement of water in host-guest interactions

R. U. Lemieux*, L. T. J. Delbaere†, H. Beierbeck* and U. Spohr‡

Department of Chemistry, University of Alberta, Edmonton, Alberta, Canada T6G 2G2, †Department of Biochemistry, University of Saskatchewan, Saskatoon, Saskatchewan, Canada S7N 0W0 and ‡Chembiomed Ltd, Edmonton Research & Development Park, 2011-94 Street, Edmonton, Alberta, Canada T6H 4N9

Abstract. As predicted by inhibition studies the X-ray crystal structure of the complex formed between the tetrasaccharide α-L-Fuc(1→2)-β-D-Gal(1→3)[α-L-Fuc-(1→4)]-β-D-GlcNAc-OMe (Le^b-OMe) and the lectin IV of *Griffonia simplicifolia* (GS-IV) shows three hydroxyl groups (referred to as the polar key) hydrogen bonded within the combining site and flanked by hydrophobic surfaces. Apart from OH-6 of the β-D-GlcNAc unit, the six other hydroxyl groups reside at or near the periphery of the combining site. Linear enthalpy–entropy compensation is observed for complex formation with monodeoxy and other derivatives of Le^b-OMe involving one of these six hydroxyl groups. Decreases in both the thermodynamic parameters ($-\Delta H°$ and $-\Delta S°$) are largest when a hydroxyl group is in contact with water at the periphery of the combining site. The experimental evidence indicates that the binding reactions involve very similar if not identical changes in the conformations of both the lectin and the ligands; it is therefore proposed that the enthalpy–entropy compensations arise because water molecules hydrogen bonded to the amphiphilic surfaces of the unbound oligosaccharide and the protein are more mobile (higher entropy content) and less strongly hydrogen bonded than are water molecules in bulk solution. Monte Carlo simulations of the hydration of Le^b-OMe appear to support this idea. In accordance with this proposal the association of complementary amphiphilic molecular surfaces from aqueous solution is driven by the release of water molecules from both non-polar and polar regions of the amphiphiles to form stronger hydrogen bonds in bulk water. In the case of highly amphiphilic molecules such as the oligosaccharide Le^b-OMe the negative contributions to entropy change dominate positive contributions that may arise from hydrophobic effects. The GS-IV(Le^b-OMe)$_2$ complex is stabilized by the hydrogen-bonding networks involving an asparate, an asparagine and a serine residue within the combining site and the above-mentioned key hydroxyl groups. Improved packing of the molecules may also be involved.

1991 Host–guest molecular interactions: from chemistry to biology. Wiley, Chichester (Ciba Foundation Symposium 158) p 231–248

At a recent Ciba Foundation symposium on carbohydrate recognition in cellular function a presentation on protein–carbohydrate interactions by Carver et al

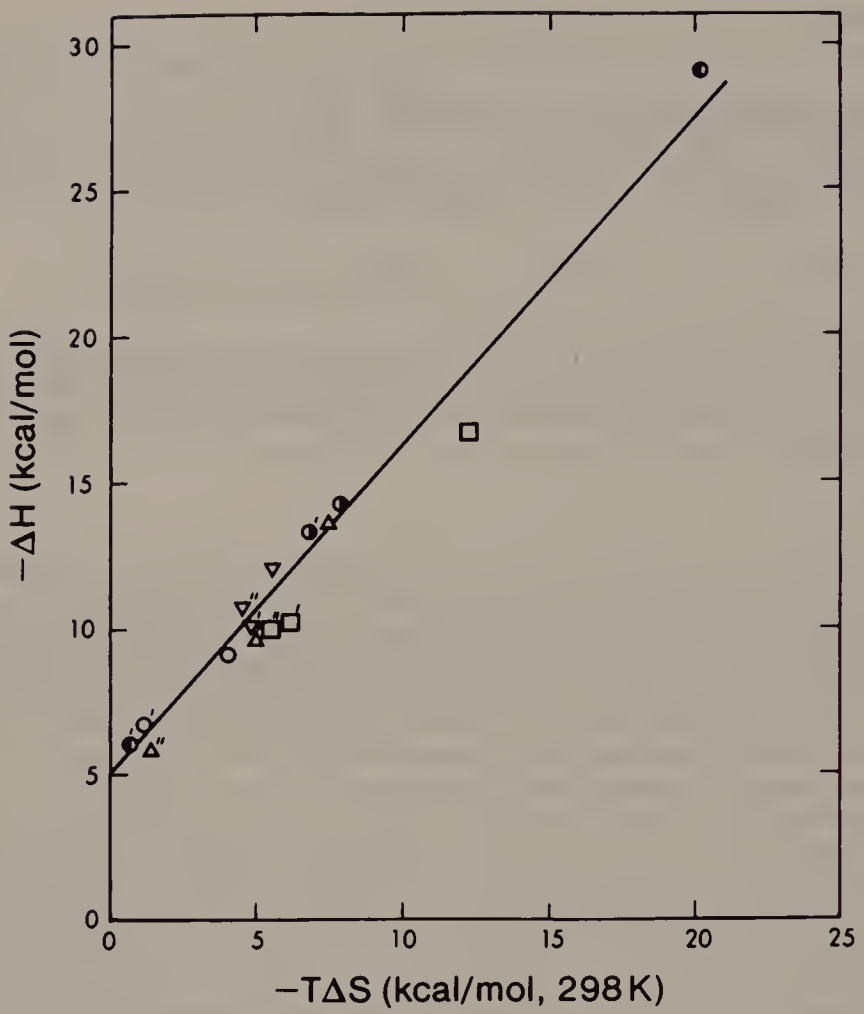


FIG. 1. A plot of selected literature values for the changes in enthalpy and entropy at 25 °C (298 K) that occur on the binding of a variety of carbohydrates by a variety of plant lectins (Hindsgaul et al 1985). Lectin I of *Ulex europaeus*: ◐ α-L-Fuc(1→2)-β-D-Gal(1→4)-β-D-GlcNAc-OMe and ◐′ α-L-Fuc-OMe. Lectin IV of *Griffonia simplicifolia*: ◑ α-L-Fuc(1→2)-β-D-Gal(1→3)[α-L-Fuc(1→4)]-β-D-GlcNAc-OMe (**1**) and ◑′ Methoxymethyl(1→2)-β-D-Gal(1→3)[α-L-Fuc(1→4)]-β-D-GlcNAc-OMe (**2**). Peanut agglutinin: □ β-D-Gal(1→4)-β-D-Glc-OMe, □′ β-D-Gal-OMe and □″ α-D-Gal-OMe. *Erythrina cristagalli* agglutinin: Δ β-D-Gal(1→4)-β-D-GlcNAc(1→2)-D-Man, Δ′ β-D-Gal(1→4)-D-Glc, Δ″ α,β-D-GalNAc. Soybean agglutinin: ▽ β-D-GalNAc-OPh-*p*-NO_2, ▽′ α-D-Gal-OMe and ▽″ α-D-GalNAc-OMe. Concanavalin A: ○ α-D-Man-OMe and ○′ α-D-Man-OPh-*p*-NO_2.

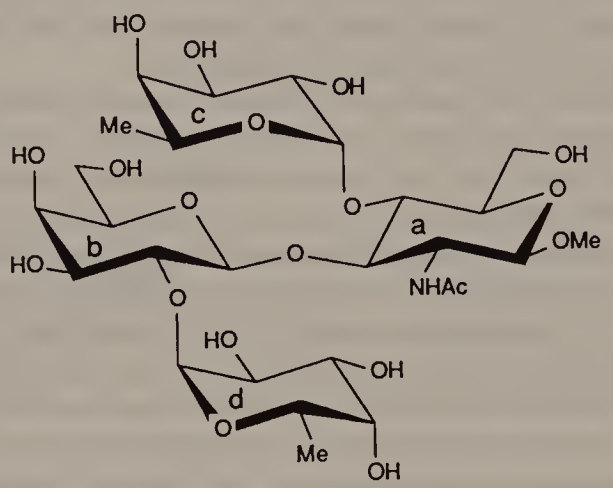
(1989) provoked discussion about the observation, reproduced in Fig. 1, that a remarkable near linear relationship exists between the decreases in enthalpy and the compensating decreases in entropy for the binding of a wide variety of oligosaccharides by a variety of lectins (Hindsgaul et al 1985). Similar compensations have long been recognized and it was tentatively concluded in an extensive review (Lumry & Rajender 1970) that the enthalpy–entropy

compensation pattern is 'real, very common and a consequence of the properties of liquid water as a solvent regardless of the solutes and the solute processes studied.' It was pointed out that the available data suggested that the compensation behaviour for processes occurring in aqueous solution is due to a single unique characteristic of water. Furthermore, it was suggested that one might expect to find such a basis for compensation in reactions involving proteins because during evolution proteins have adapted to the characteristics of water. It was also considered that compensation phenomena may provide a particularly easy pathway toward an understanding of the role of solvent and surface water in protein reactions. The authors noted that there was no shortage of imaginative proposals to explain this remarkable phenomenon.

Belleau (1967) first proposed the direct participation of liquid water in protein reactions on the basis of a linear-compensation pattern. More recently, it has been considered (Eftink et al 1983) that enthalpy–entropy compensation in ligand binding by a protein results from a conformational change in either the protein or the ligand. Hindsgaul et al (1985) accepted this model to rationalize the data plotted in Fig. 1, reasoning that it could be expected that the larger the combining site (the greater the decrease in enthalpy), the greater the effort may be for organization of protein structure (as reflected by a decrease in entropy). Carver et al suggested (1989) that although this model may be correct in many cases the opposite is true for oligosaccharide–protein interactions; that is to say, it is the loss of conformational entropy in *the ligand* on binding that accounts for the entropy barrier to binding and this loss of entropy increases as the extent of the oligosaccharide interaction with the protein surface increases, because more glycosidic linkages are restricted in their conformation. We shall present evidence that challenges these points of view. Specifically, we have experimental evidence that seems to favour the original theory that all enthalpy–entropy compensations involve manifestation of a unique property of liquid water. Our contention is that the formation of a complex between water molecules and a polar group that is adjacent to a non-polar (hydrocarbon) group (i.e. interaction at an amphiphilic surface) results in hydrogen bonds that are on average weaker than those in bulk water and in water molecules that are more mobile (less ordered) than in bulk water. The compensation arises because the less strongly the water molecules are held in the complex, the higher the entropy becomes. Thus, the release of the water molecules to bulk solution will inevitably contribute to a decrease in enthalpy at the cost of a decrease in entropy. It follows from this that the compensation should be linear, as is found. The high correlation at physiological temperatures must be relevant to the general subject of biological associations.

It has long been established (Belleau 1967, Lumry & Rajender 1970) that under fixed experimental conditions, as the chemical structure of the ligand is changed, the intrinsic thermodynamic parameters ($\Delta H°$ and $\Delta S°$) for binding by the protein may also change and, if they do, enthalpy–entropy compensation is

observed. Accordingly, we decided to examine the thermodynamics of the binding of derivatives of a tetrasaccharide known as the Lewis b human blood group determinant by the lectin IV of *Griffonia simplicifolia* (GS-IV) (Shibata et al 1982) using the methyl glycoside derivative of the tetrasaccharide (**1**) as the reference ligand. Structure **1** consists of a methyl *N*-acetyl-β-D-glucosaminide (a unit) substituted at its 3 position by a β-D-galactosyl group (b unit) and at its 4 position by an α-L-fucosyl group (c unit). The β-D-galactosyl unit is in turn substituted at its 2 position by an α-L-fucosyl group (d unit).



1 (Le^b-OMe)

It was attractive to study the binding of Le^b-OMe and its derivatives (congeners) by GS-IV for a number of reasons.

(a) Methods exist for the synthesis of the reference ligand Le^b-OMe and a wide range of structural variations (a congener series consisting mainly of deoxy, deoxyhalogeno and nor derivatives) (Spohr & Lemieux 1988, Lemieux et al 1988).

(b) Studies based on ^1H NMR spectroscopy showed that Le^b-OMe and its congeners possess similar conformations when dissolved in D_2O (U. Spohr & R. U. Lemieux, unpublished). These conclusions were drawn in part from the data reported in Table 1. For reasons previously discussed in detail (Lemieux et al 1980, 1988), remarkable specific inter-unit deshielding effects are observed in the ^1H NMR spectrum of Le^b-OMe. Of particular significance are the chemical shifts of H-5c and H-5d, which are near 0.80 and 0.30 p.p.m., to lower field than H-5 in methyl α-L-fucopyranoside. This observation alone fixes the preferred conformation for this tetrasaccharide (Lemieux & Driguez 1975, Thøgersen et al 1982). The fact that very similar inter-unit deshieldings were found for the various deoxy-Le^b-OMe compounds (see Table 1) demonstrates unequivocably that these compounds all maintain the same conformational preference. Therefore, any conformational changes required for the binding of these ligands would be the same and the energetics of conformational change are likely to be similar.

TABLE 1 Evidence that the deoxygenation of the fucose units of Le^b-OMe does not cause a significant change in the conformational preference

		Differences in ^1H NMR chemical shift[a] for deoxy derivatives of Le^b-OMe					
	Le^b-OMe (1)	*6b (12)*	*2c (11)*	*3c (16)*	*2d (5)*	*3d (13)*	*4d (15)*
H-1c	0.24	0.23	0.25	0.26	0.24	0.25	0.24
H-5c	0.80	0.85	0.73	0.83	0.80	0.84	0.80
H-1d	0.37	0.37	0.39	0.37	0.30	0.40	0.38
H-5d	0.30	0.30	0.28	0.31	0.28	0.37	0.34

[a]Differential chemical shifts obtained by subtraction of the chemical shifts obtained for methyl-α-L-fucopyranoside from those found in fucose units and of the corresponding methyl-deoxy-α-L-fucopyranoside from those found in the deoxy-Le^b-OMe. Bold numbers in brackets refer to derivatives of Le^b-OMe listed in the legend to Fig. 2; c and d refer to units of Le^b-OMe, as shown in **1**.

(c) Not only did the protein GS-IV and the complex GS-IV(Le^b-OMe)$_2$ provide fine crystals for X-ray structural analysis, but a wide variety of the GS-IV(Le^b-OMe congener)$_2$ complexes also crystallized and all these crystals are strictly isomorphous (Vandonselaar et al 1987). Thus, it was apparent that although the lectin undergoes some conformational change to form the complex with Le^b-OMe, the conformational change involved in the formation of the complexes with the various congeners must be very nearly the same.

(d) The relative potencies of the various Le^b-OMe congeners as inhibitors of binding could be determined using a solid-phase competitive radioimmunoassay (Spohr et al 1985). The 50% inhibition data suggested that although the replacement of certain hydroxyl groups by hydrogen caused little change in free energy, the thermodynamic parameters may be substantially affected. Some of the data were in fact best explained by assuming that certain hydroxyl groups underwent intramolecular hydrogen bonding for acceptance into a lipophilic region of the combining site.

(e) A convenient method was available for determining the thermodynamic parameters; the method is based on ultraviolet difference spectroscopy and has been described in detail (Spohr et al 1985). The basis of the method is that the binding causes a significant change in the value of A_{290} minus A_{315}, and analysis of the saturation plot allows the calculation of association equilibrium constants. Measurements of the association constants in the temperature range 283–313 °K yield the van't Hoff plots from which the changes in enthalpy and entropy are calculated. Once the structure of the complex became known (Delbaere et al 1990) it was evident that the source of the change in ultraviolet absorption ($A_{290} - A_{315}$) was the two tryptophan residues present in the combining site.

Although the $\Delta H°$ and $\Delta S°$ values obtained by ultraviolet difference spectroscopy were highly reproducible and appeared to provide values for the change in enthalpy to within $\pm$ 0.5 kcal mol^{-1}, this tribute to technical skill does not guarantee significance in terms of the actual values of the

thermodynamic parameters being measured. However, we became confident that the values are sufficiently precise for our present purposes for the following reasons (Spohr et al 1986).

(a) Experiments to probe the structure of the binding site based on the radioimmunoassay had indicated that the β-D-GlcNAc a unit of Le^b-OMe is not involved in its binding to the lectin (Spohr et al 1985) and this finding was confirmed by the crystal structures of the complexes GS-IV(6a-deoxy-6a-iodo-Le^b-OMe)$_2$ (Vandonselaar et al 1987) and GS-IV(Le^b-OMe)$_2$ (Delbaere et al 1990), which display the 6a groups in a channel of water molecules well removed from the protein. It could therefore be expected that changes at the 6a position would not affect the thermodynamic parameters. Indeed, this proved to be the case and ΔH° and ΔS° values found for the binding of Le^b-OMe were (well within experimental error) the same as those determined for the binding of 6a-deoxy-Le^b-OMe and 6a-*O*-mesyl-Le^b-OMe.

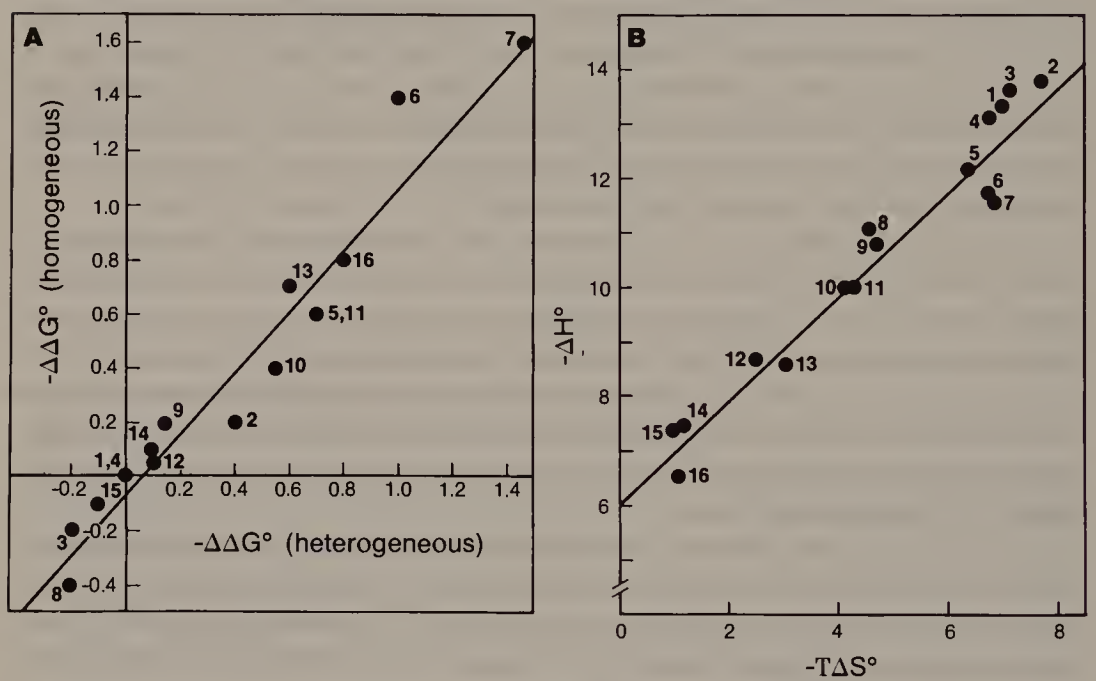


FIG. 2. Thermodynamic parameters in kcal mol^{-1} at 298 K for the binding to a lectin from *Griffonia simplicifolia* (GS-IV) of the Lewis b human blood group determinant (**1**) and the following derivatives: 4d-chloro-4d-deoxy (**3**), 6a-deoxy (**4**), 2d-deoxy (**5**), 6c-nor (**6**), 6c,6d-di-nor (**7**), 4d-deoxy-4d-fluoro (**8**), 6d-nor (**9**), 2c-deoxy (**11**), 6b-deoxy (**12**), 3d-deoxy (**13**), 3-deoxy-3-fluoro (**14**), 4d-deoxy (**15**), 3c-deoxy (**16**) and the structurally related 2-*O*-methoxymethyl-β-D-Gal(1→3)[α-L-Fuc(1→4)]-β-D-GlcNAc-OMe (**2**) and α-L-Fuc(1→2)-β-D-Gal(1→4)[α-L-Fuc(1→3)]-β-D-GlcNAc-OMe (**10**, the Y human blood group determinant) (Spohr et al 1986). A. The plot of the differential changes in free energy determined by the solid-phase radioimmunoassay (heterogeneous) against those determined by ultraviolet difference spectroscopy (homogeneous). B. Plot to display the linear enthalpy–entropy compensation ($T_c = 298$ K).

(b) It is well established that 50% inhibition data provide the differential changes in free energy ($\Delta\Delta G°$) for the various binding reactions. It was comforting to find (Fig. 2A) that the plot of the $\Delta\Delta G°$ values, obtained in the radioimmunoassay (heterogeneous conditions), using Le^b-OMe (**1**) as the reference, against those obtained by ultraviolet difference spectroscopy (homogeneous conditions) showed good correspondence. As shown in Fig. 2A, the correspondence is within ± 0.2 kcal mol^{-1}, with the exception of compound **6**. The spread in $\Delta\Delta G°$ from compound **7** to compound **8** is small, about 2 kcal mol^{-1}, but appears to be real in view of the good correspondence between the $\Delta\Delta G°$ values determined by two different methods. It will be seen below that the changes in free energy resulting from the structural changes can be rationalized on the basis of the structure of the complex. The fact remains that these changes in $\Delta G°$ are much smaller than the changes in $\Delta H°$ which, as seen in Fig. 2B, have a spread of over 7 kcal mol^{-1}.

(c) Enthalpy–entropy compensation was to be expected and, in fact, as shown in Fig. 2B, a plot of $-\Delta H°$ against $-T\Delta S°$ for binding of a variety of Le^b-OMe derivatives yields a near linear relationship. A line with a slope of unity (compensation temperature of 298 K) and an intercept at $\Delta H° = -6$ kcal mol^{-1} provides a good 'best fit'. As shown in Table 2, this decrease in enthalpy, when $\Delta S° = 0$, corresponds closely to the decreases in free energy for the various binding reactions.

In view of the above, we submit that our thermodynamic data deserve serious attention. This belief is, as indicated below, further supported by the data presented in Table 2. As already mentioned, changes at the remote 6a position had negligible effect on the thermodynamic parameters. As previously reported (Spohr et al 1985), changes at the 3b, 4b and 4c positions essentially abolish binding activity and, consequently, these hydroxyl groups are considered to be

TABLE 2 Effects of deoxygenation on the thermodynamic parameters for binding by the lectin GS-IV

Position deoxygenated (compound)	*kcal/mol*		*Contact lost*[b]	
	$-\Delta G°$	$-\Delta\Delta H°$	*bond*	*non-bonded*
None (**1**)	6.3	0[a]	—	—
6a (**4**)	6.3	0.2	—	H_2O only
6b (**12**)	6.2	4.6	—	Trp-133
2c (**11**)	5.7	3.3	—	Arg-48
3c (**16**)	5.5	6.7	Arg-48	—
2d (**5**)	5.7	1.2	Asn-135	Trp-138
3d (**13**)	5.6	4.7	His-114	Trp-138
4d (**15**)	6.4	5.9	—	Phe-108, His 114

[a]$\Delta H° = -13.3$ kcal/mol at 298 K.
[b]At or near van der Waals contact. The key polar interaction involves OH-3b (Asn-135, Asp-89), OH-4b (Asp-89) and OH-4c (Ser-49).

involved in a key polar interaction within the combining site (Lemieux 1985). For reasons of brevity and clarity of presentation, the discussion of these results will be restricted to the deoxy derivatives of Le^b-OMe listed in Table 1.

As shown in Table 2, the deoxygenations resulted in decreases in enthalpy, the parameter that provides the driving force for complex formation. It is evident from the crystal structure of the GS-IV(Le^b-OMe)$_2$ complex (Fig. 3) that the hydroxyl groups referred to in Table 2 all, with the exception of H-6a, reside at, or very near, the periphery of the combining site. Undoubtedly, it is for this reason that their replacement by hydrogen caused only slight changes in the stability of the complex ($\Delta G°$) (Fig. 2A). If these hydroxyl groups are not directly involved in the binding reaction, then why does their replacement by

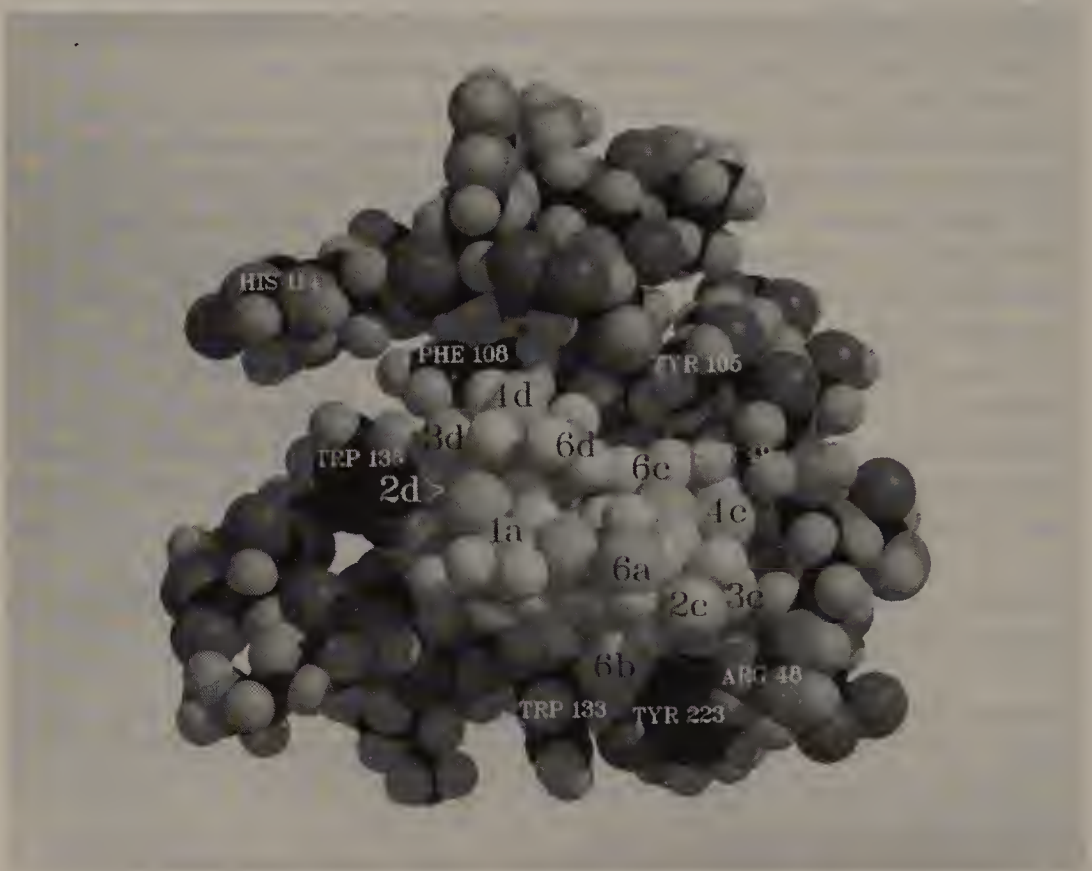


FIG. 3. View of the complex formed between Le^b-OMe (**1**) and the lectin GS-IV, looking down the mean plane of the pyranose ring of the β-D-GlcNAc unit toward 44 amino acids that provide the combining site and its near environment. For Le^b-OMe, only the positions of the hydroxyl groups and the two *C*-methyl groups of the fucose units are labelled. Note that eight of the ten hydroxyl groups in Le^b-OMe are accessible to water. OH-4c and OH-2d are largely imbedded for hydrogen bonding with the protein. OH-6a is far removed from the combining site and not involved. OH-3c is closer to Arg-48 than is OH-2c. The other three hydroxyl groups, at positions 6b, 3d and 4d, are at the periphery of the combining site.

hydrogen so significantly affect the thermodynamic parameters ($\Delta H°$ and $\Delta S°$, see Fig. 2B)? We suggest that it is because, on binding, changes occur in the hydration shell that reflect the substitution of the hydroxyl group by a non-polar group. We believe that because these hydroxyl groups are at amphiphilic surfaces, their hydration places water molecules in a state of greater mobility (i.e. greater rotational and translational degrees of freedom) than they possess in bulk water. Furthermore, the hydration involves formation of hydrogen bonds weaker than those that exist between water molecules in bulk solution. We decided to present these ideas in this gathering of leading researchers on host–guest interactions and I look forward to their reaction with keen anticipation.

To begin with, it seems most useful to examine in detail the structure of the complex formed at one of the GS-IV combining sites, as revealed by X-ray crystallographic analysis at 2.8 Å resolution. The resolution has now been improved to 2.5 Å but, unfortunately, time did not allow inclusion of this new data.

As shown in Fig. 3, which displays a Corey-Pauling-Koltun (CPK) model for Le^b-OMe (**1**) complexed within the shallow combining site of GS-IV, of the ten hydroxyl groups of **1**, only two of the key hydroxyl groups (OH-3b and OH-4b) interact with GS-IV completely within the combining site. In accord with the experiments probing the binding site (Spohr et al 1985), OH-6a is remote from the protein. The remaining seven hydroxyl groups (which include the key OH-4c group) are situated at the periphery of the combining site where interaction is possible with both the protein and water. Of these seven hydroxyl groups, OH-2d is imbedded most deeply within the combining site. It is in close proximity to the aromatic group of Trp-138 and can serve as proton acceptor for hydrogen bonding with one of the hydrogens of the amide group of Asn-135. The other hydrogen of this amide group can hydrogen bond to OH-3b (Delbaere et al 1990). Deoxygenation at position 2d had the least effect on the thermodynamic parameters and appears to lower the stability of the complex by about 0.6 kcal mol^{-1} (see Table 2). The greatest effect came from deoxygenation at position 4c and the hydroxyl group at this position donates a proton for hydrogen bonding to Ser-49. OH-3c was found to reside in close proximity to the highly polar guanidine group of Arg-48. It seems likely (Table 2) that OH-3c and perhaps the neighbouring OH-2c become involved in hydrogen bonding networks as acceptors of protons from this guanidinium group either directly, or by way of intermediate water molecules. Hydrogen bond formation is also indicated between His-114 and OH-3d. Like OH-2d, OH-3d resides in close proximity to a non-polar region of Trp-138. The deoxygenation of either OH-6b or OH-4d has a major effect on the thermodynamic parameters (see Table 2). These hydroxyl groups appear to be involved only in non-polar interactions with the protein, with Trp-133 in the case of OH-6b and with both Phe-108 and His-114 in the case of OH-4d. Notice that the substitution of these hydroxyl groups by hydrogen produced the least effect on the stability ($\Delta G°$) of the complex.

The interactions involving the hydroxyl groups at the periphery of the combining site are of course also likely to be affected by interaction with water molecules in a dynamic equilibrium under the thermal agitation at 298 K. In any attempt to interpret these data in detail we shall therefore have to deal with structures that are time averaged.

As suggested above, the hydration of amphiphilic regions of molecules causes water molecules to have greater mobility (higher entropy) and to be less strongly hydrogen bonded (higher enthalpy) than water molecules in bulk solution; this must be taken into consideration when interpreting the results in Table 2. This idea is derived from the fact that hydrogen bonds are highly directional and bonding is strongest when the nuclei of the oxygen atoms and the nucleus of the bonded hydrogen atom are co-linear. The hydrogen bonding of a hydroxyl group which neighbours a hydrophobic molecular feature by a water molecule may therefore place the water molecule, on the average, at some distance away from the neighbouring non-polar group. Consequently, void space (lack of van der Waals contact) may effectively exist around the non-polar region (Lemieux 1985, 1989). Either such cavities will exist, or the strengths of the hydrogen bonds involving the water molecules will be sacrificed by bending or breaking in order to fill the cavities. This situation is similar to that for water in contact with air, except that a greater mobility is expected for the water molecules at the surface of an amphiphile. This expectation seems reasonable since the gradient of the force field for the movement of the water molecules of a hydrated amphiphilic surface should be much less than that when the movement is against air. Furthermore, it seems reasonable to expect that, because of steric hindrances to ordering provided by the neighbouring non-polar groups, the water molecules may not be able to form the well-ordered arrangements that are expected when the contact is with air (or extended hydrophobic surfaces).

In order to test these ideas, Monte Carlo simulations of the hydration of Le^b-OMe at 300 K were performed. The procedure has been reported in detail (Beierbeck & Lemieux 1990). It involved surrounding the molecule with a fairly uniform layer of three to four water molecules formed from a total of 250 water molecules. Using rather arbitrarily chosen orientations for the hydroxyl groups of Le^b-OMe and the HSEA conformer (that provided by hard sphere exo-anomeric calculation) (Lemieux et al 1980, Thøgersen et al 1982), which appears to be present in the GS-IV $(Le^b\text{-OME})_2$ complex (Delbaere et al 1990), we obtained the results illustrated in Fig. 4. The organizations of the water molecules about the Le^b-OMe molecule after 1000, 2000 and 3000 Monte Carlo sampling cycles (moves per water molecule) are shown, except that the water molecules more than 3 Å away from the plane defined by H-1b, H-3b and H-5b of the β-D-Gal unit are removed. Thus, it is demonstrated that the space above this non-polar region is occupied at times by fewer water molecules than at other times. In fact, the calculations indicate that, on average, the surfaces of C-H groups are more than 1 Å removed from the surface of the nearest water molecule. Thus,

FIG. 4. Views of the hydration cluster of Le^b-OMe (**1**) formed by 250 water molecules after 1000 (a), 2000 (b) and 3000 (c) Monte Carlo sampling cycles cut 3 Å in front of the plane defined by H-1b, H-3b and H-5b of the β-D-Gal unit, to display the hydration over this hydrophobic region.

the simulations support the notion that the hydration of amphiphilic surfaces creates cavities along the non-polar regions.

The apparent virtue of this theory is that it provides an explanation for the linear enthalpy–entropy phenomenon for all molecular associations in water; namely, the release of water molecules from an amphiphilic surface always leads to stronger hydrogen bonding and increased ordering—a unique property of water. The entropy compensation prevents the force for association from becoming too strong, thereby providing a built-in control on the extent of biological associations. Also, as a consequence of the significance of the entropy term, these associations have high temperature coefficients—a biological necessity. It follows that host–guest associations in aqueous solution occur whenever complementary surfaces have the polar groups correctly aligned for attractive interaction. Figure 5 demonstrates the high level of amphiphilicity of the complementary regions of the Le^b-OMe and GS-IV molecules that interact to form the complex. It should be stressed that virtually all organic molecules are amphiphilic and, in contrast to, say, fatty acids or phospholipids,

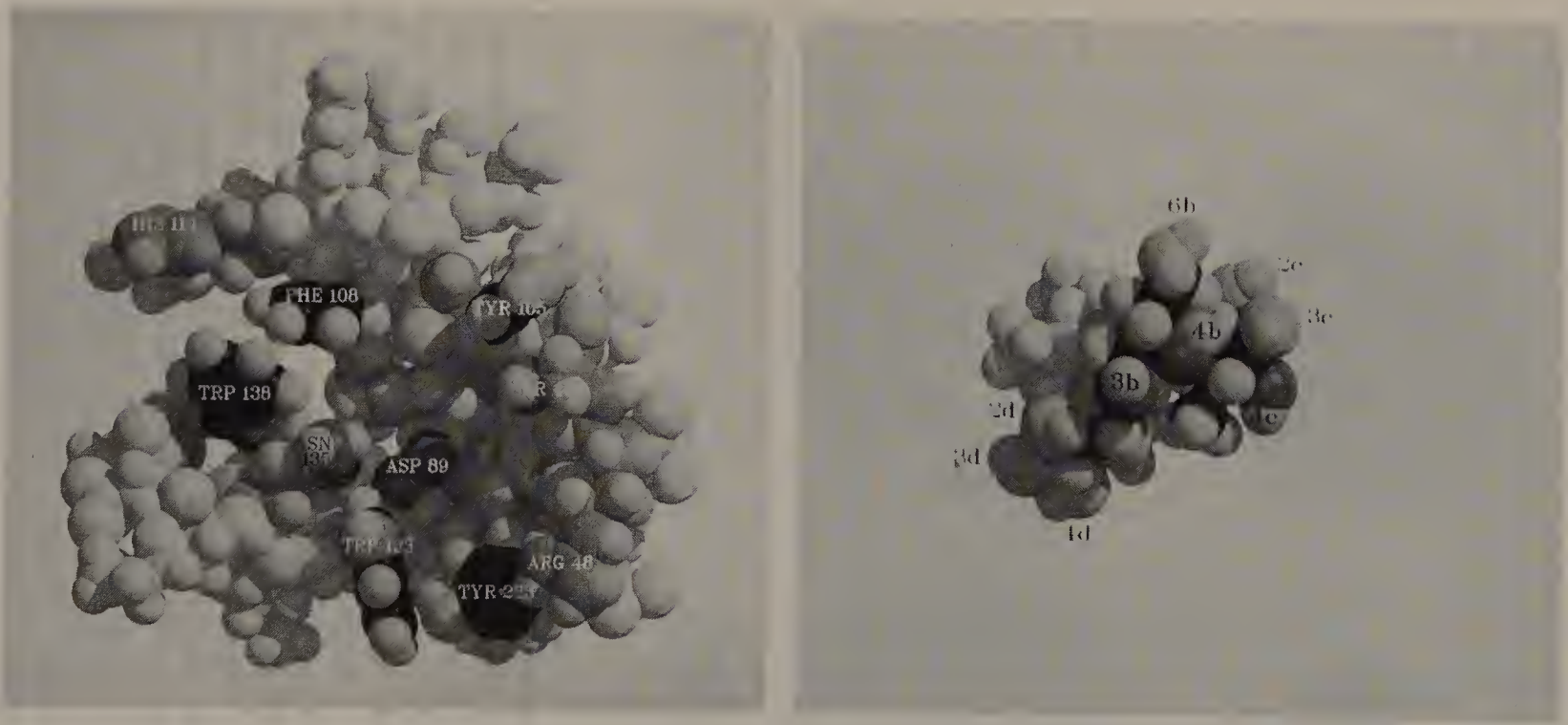


FIG. 5. Views of the complementary surfaces of Le^b-OMe and GS-IV (darker atoms) that interact to form the complex in order to display their highly dispersed amphiphilicities. On complexation, both OH-3b and OH-4b form hydrogen bonds with the carboxylate group of Asp-89 (Delbaere et al 1990). The recently determined X-ray crystal structure at 2.5 Å resolution established that the NH of His-114 is well oriented for hydrogen bonding with OH-3d.

which have localized hydrophobic and hydrophilic areas, most molecules present highly dispersed regions of amphiphilicity for their interaction with water.

The experimentally determined change in enthalpy for the binding of Le^b-OMe by GS-IV is near $-13\,kcal\,mol^{-1}$ (Fig. 2B). However, about $-7\,kcal\,mol^{-1}$ of this change appears to be related to changes in the states of water that are compensated by decreases in entropy. On this basis, the residual $-6\,kcal\,mol^{-1}$ change reflects the forces of adhesion between Le^b-OMe and GS-IV within the complex. This force must mainly be the difference between the strengths of the hydrogen bonds for hydration of both Le^b-OMe and GS-IV prior to the association and the strengths of the hydrogen bonds within and with water about the complex after its formation. The occurrence of additional contributions is suggested by the data in Table 2. As mentioned above, although they are small, the differential changes in free energy for the various binding reactions listed appear to be real (Fig. 2A). This conclusion is supported by the fact that, aside from the deoxygenation at the remote 6a position, the deoxygenations at positions 6b and 4d had the least effect on the change in free energy. Indeed, in the case of position 4d, the deoxygenation stabilized the complex. Certainly, it is not likely to be a coincidence that the hydroxyl groups at these two positions are the only ones that are not involved in the intermolecular hydrogen bonding. Furthermore, both are in or nearly in van der Waals contact with hydrophobic groups of the protein. It is probably significant (Lemieux et al 1988) that both OH-6b and OH-4d are well disposed for intramolecular hydrogen bonding. Consequently, it appears that the complexes formed with the 6b-deoxy (**12**) and 4d-deoxy (**15**) congeners are stabilized to some extent by the release of water molecules from the hydrophobic regions, a release that is more effective for the deoxy derivatives **12** and **15** than for **1** (Tanford 1980). Only OH-2d and OH-3d appear to be directly involved in intermolecular hydrogen bonds as proton acceptors from weak proton donors of the lectin; that is, for OH-3d the proton donor is the NH group of the imidazole ring of His-114 and for OH-2d the donation is from the NH_2 group of Asn-135. This group is also involved in proton donation to OH-3b (Delbaere et al 1990). These hydroxyl groups are near the periphery of the combining site and are therefore positioned to interact with water, so it is not surprising that the loss of these hydrogen bonds had only a minor (less than one kcal per mole) effect on the stability of the complex. The large differential change in enthalpy ($\Delta\Delta H^\circ = -4.7\,kcal\,mol^{-1}$) resulting from deoxygenation of OH-3d may be related to the introduction of a hydrophobic interaction with Trp-138. This appears to be less likely to be true for the deoxygenation at OH-2d, because of the much smaller $\Delta\Delta H^\circ$ value ($-1.2\,kcal\,mol^{-1}$). Certainly, this hydroxyl group appears to be more firmly embedded into the combining site and less accessible to water. The largest change in enthalpy occurred on deoxygenation of OH-3c. This effect and the large compensating change in entropy must be related to the proximity of this hydroxyl group to the positively charged

guanidinium group of Arg-48. Deoxygenation of OH-3c undoubtedly has an important effect on the state of water molecules in this region of the complex. OH-2c is farther removed from both the combining site and Arg-48. Thus, the effect appears attenuated for the 2c-deoxy congener and the longer range effect indicates an involvement of water molecules in these phenomena.

In conclusion, this research appears to establish a dominant role for water in host–guest interactions in an aqueous medium. The evidence for the assignment of the enthalpy–entropy compensations to the release of water from amphiphilic surfaces though not definitive is persuasive, and the suggestion is surely highly compelling and is worthy of serious attention. We expect that greater certainty will come with the knowledge of the X-ray crystal structures for the complexes referred to in Table 2.

Acknowledgements

We wish to thank Professor L. G. Hepler for valuable advice. The research was supported by the Natural Sciences and Engineering Research Council of Canada (grant A-172 to R.U.L.).

References

Beierbeck H, Lemieux RU 1990 Computer simulation studies of the solvation of oligosaccharides. Can J Chem 68:820–827

Belleau B 1967 Water as the determinant of thermodynamic transitions in the interaction of aliphatic chains with acetylcholinesterase and cholinergic receptors. Ann N Y Acad Sci 144:705–719

Carver JP, Michnick SW, Imberty A, Cumming DA 1989 Oligosaccharide–protein interactions: a three dimensional view. In: Carbohydrate recognition in cellular function. Wiley, Chichester (Ciba Found Symp 145) p 6–26

Delbaere LTJ, Vandonselaar M, Prasad L et al 1990 Molecular recognition of a human blood group determinant by a plant lectin. Can J Chem 68:1116–1121

Eftink MR, Anusiem AC, Biltonen RL 1983 Enthalpy–entropy compensation and heat capacity changes for protein–ligand interactions: general thermodynamic models and data for the binding of nucleotides to ribonuclease A. Biochemistry 261:3884–3896

Hindsgaul O, Khare DP, Bach M, Lemieux RU 1985 Molecular recognition. III. The binding of the H-type 2 human blood group determinant by the lectin I of *Ulex europaeus*. Can J Chem 63:2653–2658

Lemieux RU 1985 The hydrated polar-group 'gate' effect on the specificity and strength of the binding of oligosaccharides by protein receptor sites. Proceedings of the VIIIth International Symposium on Medicinal Chemistry, 1984, Uppsala, Sweden. Swedish Pharmaceutical Press, Stockholm, 1:329–351

Lemieux RU 1989 The origin of the specificity in the recognition of oligosaccharides by proteins. Chem Soc Rev 18:347–374

Lemieux RU, Driguez H 1975 The chemical synthesis of 2-acetamido-2-deoxy-4-*O*-(α-L-fucopyranosyl)-3-*O*-(β-D-galactopyranosyl)-D-glucose. The Lewis-a blood-group antigenic determinant. J Am Chem Soc 97:4063–4069

Lemieux RU, Bock K, Delbaere LTJ, Koto S, Rao VS 1980 The conformations of oligosaccharides related to the ABH and Lewis human blood group determinants. Can J Chem 58:631–653

Lemieux RU, Cromer R, Spohr U 1988 Molecular recognition. VIII. The binding of the β-D-galactopyranosyl residue of the Lewis b human blood group determinant by the lectin IV of *Griffonia simplicifolia* and by a monoclonal anti-Lewis antibody. Evidence for intramolecular hydrogen binding. Can J Chem 66:3083–3098

Lumry R, Rajender S 1970 Enthalpy–entropy compensation phenomena in water solutions of proteins and small molecules: a ubiquitous property of water. Biopolymers 9:1125–1227

Shibata S, Goldstein IJ, Baker DA 1982 Isolation and characterization of a Lewis b-active lectin from *Griffonia simplicifolia* seeds. J Biol Chem 257:9324–9329

Spohr U, Lemieux RU 1988 Synthetic, conformational, and immunochemical studies of modified Lewis b and Y human blood-group determinants to serve as probes for the combining site of the lectin IV of *Griffonia simplicifolia*. Carbohydr Res 174:211–237

Spohr U, Hindsgaul O, Lemieux RU 1985 Molecular recognition. II. The binding of the Lewis b and Y determinants by the lectin IV of *Griffonia simplicifolia*. Can J Chem 63:2644–2652

Spohr U, Bach M, Lemieux RU 1986 Thermodynamic parameters for the binding of Lewis B tetrasaccharide by the lectin IV of *Griffonia simplicifolia*. In: Abstracts of the XIIIth International Carbohydrate Symposium, Ithaca, New York, p 353

Tanford C 1980 The hydrophobic effect: formation of micelles and biological membranes. Wiley, New York

Thøgersen H, Lemieux RU, Bock K, Meyer B 1982 Further justification for the *exo*-anomeric effect. Conformational analysis based in nuclear magnetic resonance spectroscopy of the B human blood group determinant. Can J Chem 60:44–57

Vandonselaar M, Delbaere LTJ, Spohr U, Lemieux RU 1987 Crystallization of the lectin IV of *Griffonia simplicifolia* and its complexes with the Lewis b and Y human blood group determinants. J Biol Chem 262:10848–10849

DISCUSSION

Reinhoudt: The entropy–enthalpy compensation that you mentioned is quite a general phenomenon in both water and organic solvents. I have never really understood the reason for it.

Lemieux: As I mentioned, Lumry & Rajender (1970) published an excellent review of this subject. They tentatively concluded that the entropy–enthalpy compensation pattern is real and very common and that for aqueous solutions it is due to a single unique characteristic of water. I agree with this for the following reason: the binding of water molecules to an amphiphilic surface can be expected, because of the directionality of the hydrogen bonds, to lead to the formation of cavities and, therefore, a high degree of disorder. Bending the hydrogen bonds to fill the cavities would weaken the hydrogen bonds. The compromise between cavity formation and the average strength of the hydrogen bonds will depend on the prevailing temperature. The idea appears to be supported by Monte Carlo simulations. Consequently, I think it plausible to

expect that the release of these water molecules to bulk will lead to increased ordering of the water molecules (a decrease in entropy) and a compensating decrease in enthalpy as the result of the formation, on average, of stronger hydrogen bonds.

Reinhoudt: This effect is not restricted to water.

Lemieux: I do not know if that is true—I haven't examined any other solvent.

Stoddart: There is a good analogy between the type of binding that you described and that observed in the cyclodextrin field. Many people have discussed the release of high energy water from the cavities of cyclodextrins on entry of an organic substrate. It seems to me that you described an effect which is opposite to that; in the binding you discussed, the guest is the carbohydrate, in which the hydroxylated region is surrounded by the cavity of the protein, which is largely aromatic in nature.

Lemieux: I agree. I think this may be a very general phenomenon.

Ron Breslow: You are dealing with more than simply the classical idea that hydrocarbon surfaces tend to come together. The interesting question you addressed is why polar molecules bind to proteins coming out of water. Your argument is that even though polar groups can be solvated, the environment around that solvation is not as favourable as the environment around bulk water—real water. Even water that solvates such polar groups isn't in as good a situation as it is in an environment with neighbouring water molecules only; that's very interesting. Joe Collins (personal communication) was interested in this general question of fitting organic structures into the ordered water lattice. He found that most of the natural stereochemistries of polyhydroxy compounds are those which fit into the structure well; that is, the oxygens of the naturally occurring isomers fit into lattice positions in the lattice structure.

As you said, water is a big structure, and with even local solvation there can be problems. This phenomenon shouldn't be referred to as a 'hydrophobic effect', because it doesn't involve just putting two hydrocarbon surfaces together, which is what is normally meant by 'hydrophobic effect'; your neighbours are really different. We could call it the Lemieux effect!

Lemieux: No, at least not yet—further evidence is required. What I find interesting is that it is a very simple idea that took about 10 years to achieve, and then only by way of an attempt to rationalize experimental results. The combination of the thermodynamic data and the crystal structure of the complex was required.

Esther Breslow: You discussed the interaction between the 4-hydroxyl group of the α-L-fucosyl d-unit that is at the 2-position of the β-D-galactosyl residue and aromatic residues in the lectin. Removal of the 4-hydroxyl caused a decrease in negative enthalpy, so the binding of that hydroxyl to those aromatic rings was associated with a negative ΔH and an entropic compensation, such that there was little free energy change.

Lemieux: I cannot be sure whether that hydroxyl maintains interactions only with water or becomes intramolecularly hydrogen bonded, at least part of the time, for association with the protein. When there is a hydrogen in place of the hydroxyl group it appears that a hydrophobic interaction with the neighbouring phenyl group of Phe-108 is created, because both the 4d-deoxy and 4d-chloro-4d-deoxy compounds are bound more strongly than the reference tetrasaccharide (Le^b-OMe).

Esther Breslow: In the crystal, isn't the hydroxyl up against an aromatic ring?

Lemieux: No, but it must be remembered that the 'pictures' were taken at 14 °C—there's a lot of thermal agitation and what we look at is a time-averaged situation. If the phenyl group of Phe-108 were to sit over that hydroxyl group some of the time, then I imagine that the hydroxyl group would become intramolecularly hydrogen bonded to assist the exclusion of water. I say this because we have seen enhancements of binding for deoxy compounds only when the hydroxyl group that was replaced was in a good position to make an intramolecular hydrogen bond with a neighbouring proton acceptor. We have seen this in many cases with both lectins and monoclonal antibodies.

Esther Breslow: Table 2 suggested that the aromatic ring was the hydroxyl's only contact. I wondered whether, to explain the change in ΔH, you also assumed that the water on the surface of that aromatic group was particularly disordered before binding, then became more hydrogen bonded when released. The transfer of aromatic groups from water to less polar solvents is often associated with a negative ΔH.

Lemieux: I don't really know anything about the hydration of the amino acid residues. What we have discovered is that certain hydroxyl groups of the complexed tetrasaccharide reside near the periphery of the combining site but do not appear to be in van der Waals contact with the protein, yet when you replace them by hydrogen you get important changes in the thermodynamic parameters of binding. This is not the case when the hydroxyl group is well removed from the periphery.

Ron Breslow: The essence of your proposition is that the hydration of the hydroxyl in the substrate is not as good and the energy is not as low as it would be if the hydrating water molecules were able to hydrate themselves. The change in ΔH on release of water molecules from hydroxyl groups occurs because the bonding in organized water is stronger than when the hydroxyl is hydrated. So, when you remove the hydroxyl group, regardless of what you do to the interaction with the protein, you have lost that original class of higher energy water molecules.

Lemieux: Yes, that is my point. We seem to be getting a clue as to what might be going on when water molecules are removed from those regions that become embedded within the combining site where the intermolecular hydrogen bonds are established. Inevitably, water molecules will move from these regions to bulk. Since these complementary regions of the substrate and the protein do

not become rehydrated, the mechanism for compensation is removed. Consequently, the changes in energetics that result from the change in the states of water molecules will be more profound and may contribute significantly to the driving force for complexation.

Dunitz: Is the hydration of simple sugars in water really understood? I have been teaching students that glucose is the most stable of the pyranoses, because all the hydroxyls are in equatorial positions, but this explanation, if valid at all, can apply only to the gas phase. Is the hydration in water of glucose versus galactose, for example, understood?

Lemieux: Not to my knowledge. There have been a lot of measurements made, but it is not clear what they mean.

Dunitz: The relative solubilities must be known, and the thermodynamics might be a little simpler than in your rather complicated system.

Reference

Lumry R, Rajender S 1970 Enthalpy–entropy compensation phenomena in water solutions of proteins and small molecules: a ubiquitous property of water. Biopolymers 9:1125–1227

Molecular modelling approaches to host-guest complexes

J. G. Vinter* and M. R. Saunders

SmithKline Beecham, Welwyn, Herts AL6 9AR, UK

Abstract. Host–guest interactions can be modelled as a non-bonding recognition process using long-range electrostatic forces. By using molecular isopotential maps the differences between the methotrexate–dihydrofolate reductase and folate–dihydrofolate reductase complexes can be predicted. By extending the technique to molecule–molecule docking the interaction of formamide with the crown ether 18-crown-6 can be simulated with reasonable accuracy. The closely related problem of predicting the separation of enantiomers of chiral molecules by chromatography has been attempted with encouraging results. A preliminary report is presented on the progress being made towards a better model for simulating stacking arrangements of π systems by charge distribution.

1991 Host–guest molecular interactions: from chemistry to biology. Wiley, Chichester (Ciba Foundation Symposium 158) p 249–265

The value of the molecular electrostatic potential (MEP) as a means of characterizing electrical properties has been well established since its introduction by Scrocco & Tomasi (1973) but extensions into drug research were slower to develop (Weinstein et al 1981). In essence, the 'electrical volume' and specific polar interaction sites can be identified by probing a model molecule, equipped with charge distribution information, with a point charge. By plotting the interaction of this point in three-dimensional space around the molecule we can build up a picture of the electrical environment.

The docking of two molecules is only an extension of the MEP technique, in that the point charge is replaced by a complete structure. Extra care must be taken during the process to allow different molecular orientations to be adopted by the 'bullet' molecule as it docks with the 'target'. Our technique is to drive the target towards the bullet from 204 points evenly distributed around a sphere centred on the target and allow it to find its local minimum energy

**Present addresses*: Molecular Recognition Unit, University Chemical Laboratories, University of Cambridge, Lensfield Road, Cambridge CB2 1EW and Department of Chemistry, University College London, 20 Gordon Street, London WC1H 0AJ, UK.

interaction with the bullet. The method involves an electrostatic minimization process with added van der Waals constraints at short range to stop molecular coalescence.

Probably the most dramatic difference between the chemistry of living systems and that observed more routinely in the chemical laboratory is the sensitivity of natural systems to stereochemistry (Bergmann & Fruton 1937). Reactions carried out *in vitro* in general give a mixture of isomers (except in certain circumstances, such as the use of chiral starting materials or catalysts). In contrast, almost all of the chemistry observed in living systems shows an exquisite sensitivity to the stereochemistry of the participating molecules. This can have profound consequences at all points in the search for molecules with biological activity—for example, the evidence (Ockenfels et al 1977) that the teratogenic and embryotoxic activities of thalidomide are both entirely enantiospecific, and reside in a single isomer; the entirely enantiospecific inactivation of dopamine β-hydroxylase by β-ethynyltyramine (DeWolf et al 1989); and numerous other well-known examples.

If a computerized docking technique is to be useful, it must account for chiral properties. We have undertaken a study of the chiral separation of enantiomers by chromatography on a chiral stationary phase as a test case for chiral modelling. As we shall see in the relatively simple case of the interaction of the crown ether 18-crown-6 with formamide, chemical intuition, exhaustive search on the computer and experiment give us results which are so close as to be effectively indistinguishable. However, most real systems offer considerably more flexibility than does the problem of docking a rigid formamide molecule onto a rigid (by assumption) crown ether. The highly complex energy surfaces of such systems are much more difficult to interpret intuitively. The problem represents a smaller scale version of the interaction between a drug and its receptor. The problem of chiral recognition on a chromatography column is computationally much more tractable than the modelling of drug–receptor interactions. While retaining many of the important features which govern drug activity, it is simple enough for us to be able to make the assumptions implicit in a computational docking approach without introducing such large errors that the certainty of any conclusions we draw is in severe doubt, and small enough for us to be able to investigate different algorithms without the vast investment of computer time that would be necessary for a set of free energy perturbation calculations on a drug–receptor system.

Finally, it is important to remember that a modelling exercise which uses empirical calculations to derive its quantitative data can only ever yield results as good as the data it draws on. It is well established that of all the terms in use within the molecular mechanics formalism, the handling of electrostatic interactions is the least satisfactory. This is in part due to the fact that partial charge distributions are themselves derived from calculation and cannot yet be experimentally determined routinely. Add to this their ubiquitous long-range

effects and the inevitable interference of surrounding solvent, proton transfer and molecular movement, and it is clear that any full description of the electrostatic interactions is at present impossible.

Hunter & Sanders (1990) have recently taken the obvious (yet hitherto untried) step of separating charge on all π systems (i.e., carbon atoms bearing π electrons) by defining more precisely the distribution of charge in space. They have been able to predict the stacking pattern of two porphyrin rings remarkably accurately and their elegant and simple solution opens up new prospects for more successful modelling while, at the same time, casting doubt on the validity of what has gone before. It is worth mentioning that all molecular dynamics techniques, which are used extensively in biological modelling and consume large quantities of human and computer resources, depend on molecular mechanics for their basic energy determination!

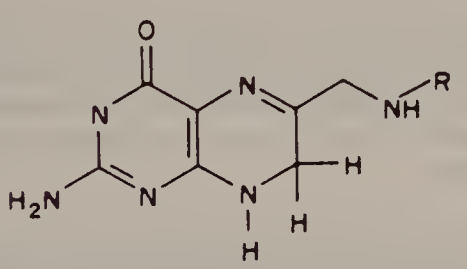
Methods and results

All calculations were performed using the COSMIC suite of molecular modelling programmes (Vinter et al 1987).

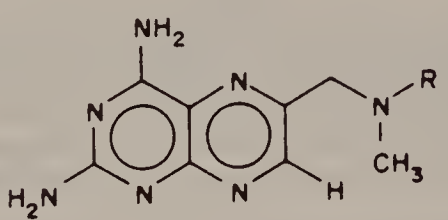
Molecular electrostatic potentials

The technique consists of calculating the energy of interaction of a 'target' molecule with a point charge, placed at defined grid-points around the target. The energy is derived by summing the electrostatic contributions of the point charge with each atom of the target. The important data needed for this calculation are the charge distributions across the atoms of the target. These cannot be derived experimentally and are usually derived from a wave mechanical analysis. The calculated charge densities are placed at the centre of each target atom; the validity of this assumption will be discussed later.

In 1978 we plotted MEPs for the natural substrate of dihydrofolate reductase and an inhibitor of this enzyme, methotrexate. The major difference between



A
Folate



B
Methotrexate

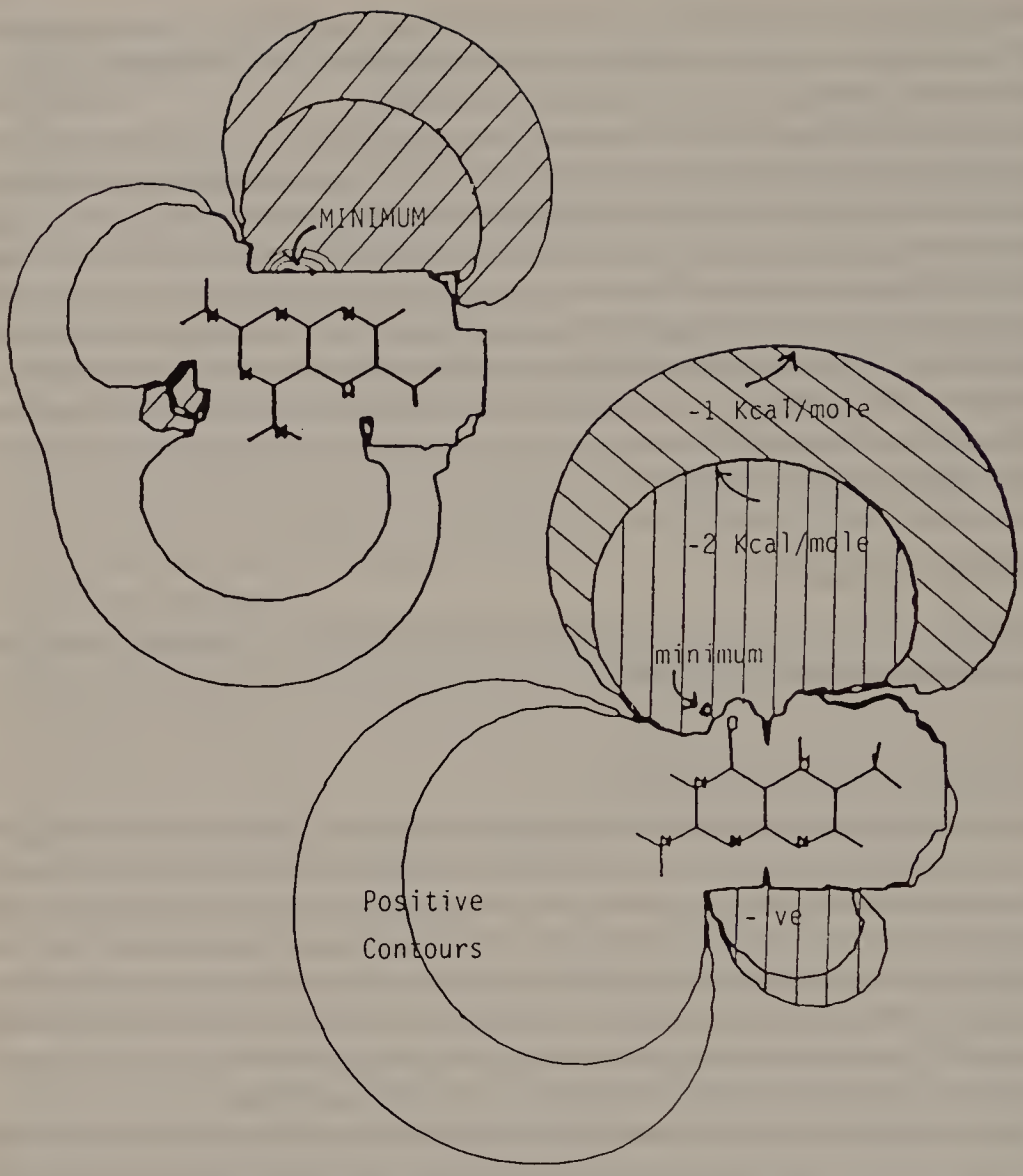


FIG. 1. The two-dimensional molecular electrostatic potential (MEP) maps of the dihydrofolate reductase inhibitor methotrexate (above) and the natural substrate dihydrofolate (below). The MEPs are plotted across the aromatic planes and overlie one another only if the structural correspondence is flipped through 180° for one species. The overlay is with minimum upon minimum and negative (cross-hatched) and positive (clear) regions approximately corresponding.

these two structures is that the pteridine carbonyl of the folate (**A**) is replaced by a trigonal amino group in methotrexate (**B**). Two-dimensional plots were sufficient to examine the ring systems. The results are presented in Fig. 1. The deep electrostatic negative energy well of folate is associated with the carbonyl group, whereas it is the N-1 that attracts the 'proton' most avidly on methotrexate.

These MEPs strongly suggested that, if primary binding is driven by long-range electrostatic forces, the inhibitor may well bind 'upside-down' compared with the natural substrate, despite their structural similarities. The important implication of this idea is that structure is often no guide to behaviour. It was not until a year or so later (Charlton et al 1979, Fontecilla-Camps et al 1979) that experiments proved that the 'upside-down' theory was indeed correct.

Non-bonded docking

In essence, the replacement of the point charge by a complete molecule should lead to a means of simulating docking behaviour. This is a 'rigid body' docking and uses multiple starting points approximately 1 Å apart on a 10 Å sphere. The 'bullet' molecule is rotated in 60° steps about the Cartesian x, y and z axes, and the lowest energy orientation is selected as input to a simplex rigid body geometry optimization. The optimization is allowed to continue until the root mean square change of the energy with respect to the six degrees of freedom is less than 0.0001 kcal mol^{-1} $Å^{-1}$ or 0.0001 kcal mol^{-1} deg^{-1}, or for a maximum of 500 iterations.

This technique was applied to the formation of a complex between 18-crown-6 and formamide. Over 200 distinct docking positions were identified over a

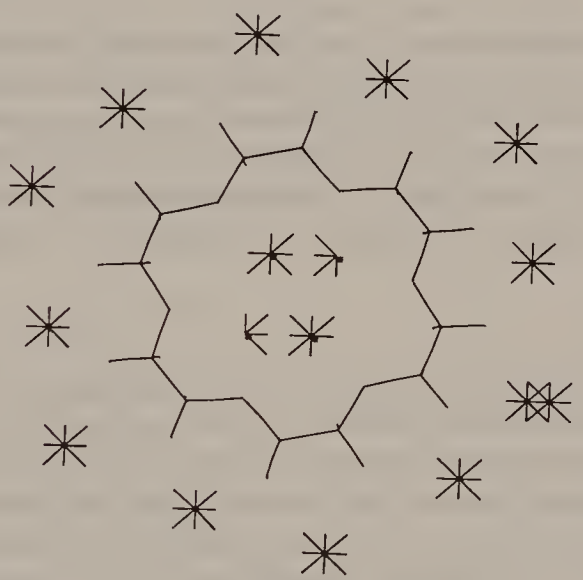


FIG. 2. The electrostatic potential energy minima around 18-crown-6. The points at the centre of the crown ether are deep negative wells derived from the interaction of a point positive charge. Those around the outside are shallow positive wells derived from the interaction of a negative charge with the dimensions of a hydroxide ion. Values were calculated for all points of a $41 \times 41 \times 41$ grid of resolution 0.5 Å centred on the crown ether (Vinter & Harris 1989).

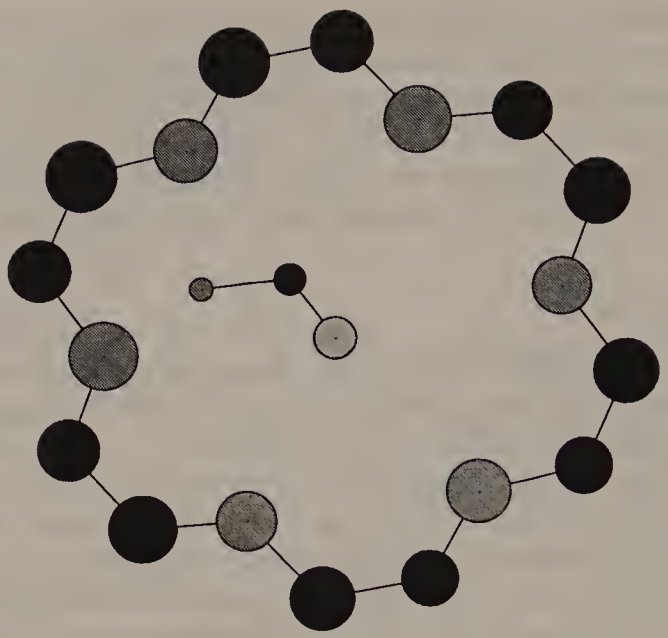


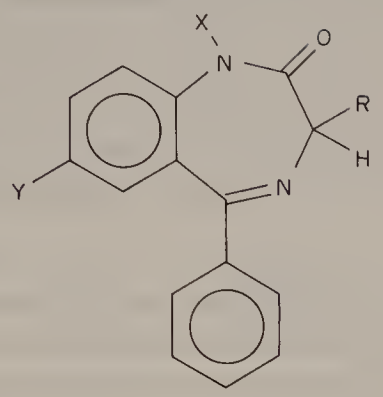
FIG. 3. 18-Crown-6 clathrate structure with formamide (hydrogen atoms removed). Data are from the X-ray crystallographic study of Watson et al (1984).

20 kcal mol^{-1} difference range, suggesting the complicated route that a simple docking process might take as it overcomes multiple minima to reach a final energy minimum position. The electrostatic potential energy minima around 18-crown-6 are shown in Fig. 2. The lowest binary complex was extracted from the overall picture (Vinter & Harris 1989) and compared with the published X-ray data (Watson et al 1984; Fig. 3). The correspondence was within 0.1 Å and the relative orientations of calculation and experiment compared very well.

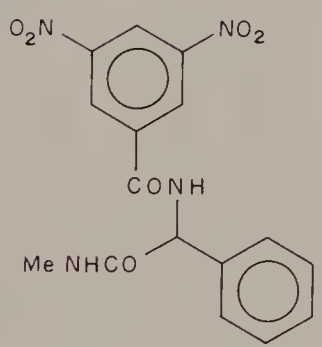
Chiral docking

We have also become interested in the resolution of enantiomers by chromatography on columns packed with one enantiomer of a chiral substance. Clearly, from the introductory discussion above, this represents a real problem to the pharmaceutical industry, and computational chemistry approaches may have a useful role to play in the design or choice of column stationary phases for the resolution of mixtures of enantiomers in order that the properties of pure isomers can be investigated. We have attempted to model the separation of enantiomers of a number of chiral benzodiazepinones (**C**) on two chirally distinct stationary phases, (*R*)-*N*-3,5-dinitrobenzoyl phenylglycine (**D**) and (*S*)-*N*-3,5-dinitrobenzoyl leucine (**E**). Experimental data are from Pirkle & Tsipouras (1984); calculations were performed on 13 of the reported compounds, 12 of which were successfully separated and one of which was not.

C Chirol benzodiozepinones

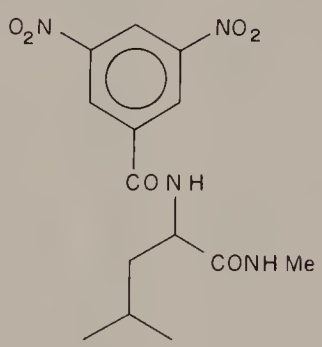


Compound	R	X	Y
1	CH_3	H	Cl
2	CH_2CH_3	H	Cl
3	$CH_2CH_2CH_3$	H	Cl
4	$CH(CH_3)_2$	H	Cl
5	$(CH_2)_3CH_3$	H	Cl
6	$CH_2CH(CH_3)_2$	H	Cl
7	$CH_2CH_2SCH_3$	H	Cl
8	$CH_2C_6H_5$	H	Cl
9	$CH_2p\text{-}C_6H_4OH$	H	Cl
10	CH_3	CH_3	Cl
11	$CH(CH_3)_2$	CH_3	Cl
12	OH	H	Cl
13	cyclohexyl	H	H



D

(R)-N-3,5-Dinitrobenzoylphenyl glycine



E

(S)-N-3,5-Dinitrobenzoyl leucine

The two chiral stationary phases, and both the *R* and *S* isomers of each of the benzodiazepinones, were built; geometries were optimized using molecular mechanics, charges were added using the method of Abraham & Hudson (1985), and the compounds were put through our conformation space sampler MIN01 to find a reasonable low energy conformation. The lowest energy conformation found (which is not necessarily the global minimum) was then taken and reminimized using molecular mechanics to give the starting point for subsequent calculations.

The first stage of the computer-simulated separation of enantiomers was a rigid body docking, as described above. The lowest intermolecular energy (no contributions from intramolecular strain) found for each isomer of each mobile phase on the two chiral column materials is recorded in Table 1. Each of these low energy end points was then used as the starting point for a Metropolis Monte Carlo hunt of conformational space. This sampling was carried out in torsional and intermolecular space with random steps of up to 30° in rotatable bonds, 0.1 Å in all three Cartesian coordinates *x*,*y*,*z* and 30° about these coordinates for intermolecular orientation. Simulations were run using a temperature of 300 K, and were allowed to continue until no successful steps were made in $5x$ (where x = number of atoms) attempts. Lowest energies obtained are recorded in Table 2. These low energy conformations were then further subjected to full molecular mechanics minimizations (Table 3).

For a comparison of the binding of two isomers to a 'receptor' as simple

TABLE 1 Use of rigid body docking to predict the order of elution of enantiomers of chiral compounds (benzodiazepinones) from two chirally distinct stationary phases

	PheGly column				*Leu column*			
	Binding enthalpy		*First eluted*		*Binding enthalpy*		*First eluted*	
	R	*S*	*Predicted*	*Observed*	*R*	*S*	*Predicted*	*Observed*
1[a]	−19.5	−18.1	—	R	−17.0	−17.0	—	S
2	−16.5	−21.0	R	R	−19.7	−17.4	S	S
3	−19.9	−22.5	R	R	−20.9	−18.1	S	S
4	−18.9	−20.7	R	R	−20.7	−17.1	S	S
5	−19.9	−21.8	R	R	−18.7	−18.7	—	S
6	−20.5	−22.4	R	R	−19.5	−18.0	—	S
7	−19.7	−20.8	R	R	−17.9	−19.4	—	S
8	−18.7	−20.8	R	R	−19.6	−20.3	—	S
9	−19.7	−21.2	R	R	−21.2	−18.9	S	S
10	−18.0	−20.3	R	R	−18.2	−16.8	S	S
11	−18.6	−17.3	—	R	−17.9	−17.5	—	S
12	−18.1	−21.8	R	R	−20.5	−18.1	S	S
13	−16.9	−17.8	—	*	−17.9	−16.3	—	*

[a]For structures see **C**.
*, Not separated experimentally.

TABLE 2 Variations recorded after a Metropolis Monte Carlo space search in attempts to predict the order of elution of enantiomers of chiral compounds from two chirally distinct stationary phases

	PheGly column				*Leu column*			
	Binding enthalpy		*First eluted*		*Binding enthalpy*		*First eluted*	
	R	*S*	*Predicted*	*Observed*	*R*	*S*	*Predicted*	*Observed*
1[a]	−49.0	−46.8	S	R	−46.8	−48.2	R	S
2	−47.0	−48.5	R	R	−47.5	−45.1	S	S
3	−48.6	−51.3	R	R	−50.2	−47.4	S	S
4	−47.4	−49.0	R	R	−49.3	−45.5	S	S
5	−42.7	−46.2	R	R	−44.9	−42.1	S	S
6	−45.8	−48.2	R	R	−45.5	−43.8	S	S
7	−44.4	−46.2	R	R	−44.0	−44.5	—	S
8	−45.9	−48.0	R	R	−46.7	−46.8	—	S
9	−51.5	−52.6	R	R	−52.9	−50.3	S	S
10	−31.8	−33.8	R	R	−32.9	−30.7	S	S
11	−27.3	−28.2	—	R	−29.4	−29.1	—	S
12	−28.2	−32.1	R	R	−30.8	−28.4	S	S
13	−44.0	−45.0	—	*	−45.5	−45.0	—	*

[a]For structures see **C**.
*, Not separated experimentally.

TABLE 3 Variations after final Newton–Raphson minimization in attempts to predict the order of elution of two enantiomers of chiral compounds from two chirally distinct stationary phases

	PheGly column				*Leu column*			
	Binding enthalpy		*First eluted*		*Binding enthalpy*		*First eluted*	
	R	*S*	*Predicted*	*Observed*	*R*	*S*	*Predicted*	*Observed*
1[a]	−51.0	−48.7	S	R	−50.0	−51.6	—	S
2	−48.4	−50.6	R	R	−50.0	−46.8	S	S
3	−50.3	−52.9	R	R	−51.4	−48.1	S	S
4	−50.7	−51.3	R	R	−51.4	−48.1	S	S
5	−50.4	−54.5	R	R	−54.7	−51.2	S	S
6	−53.8	−56.5	R	R	−54.0	−51.8	S	S
7	−51.8	−54.4	R	R	−52.8	−53.5	—	S
8	−53.7	−55.3	R	R	−53.5	−54.3	—	S
9	−60.6	−60.0	—	R	−60.1	−58.0	S	S
10	−38.7	−41.1	R	R	−42.8	−38.4	S	S
11	−36.0	−35.7	—	R	−36.4	−37.6	—	S
12	−36.4	−39.8	R	R	−37.4	−35.4	—	S
13	−47.4	−48.3	—	*	−49.1	−46.9	—	*

[a]For structures see **C**.
*, Not separated experimentally.

as a chiral stationary phase, the assumption that entropy changes are approximately the same for both isomers is likely to be much closer to a true one than when binding is to a much more complex biological receptor. If we make that assumption, then we can approximate free energy change by enthalpy change, and use the calculated molecular mechanics energy as a reflection of free energy change. Using these assumptions, the predictions for the isomer eluting first from each pair on each column are extremely encouraging.

Using the energy differences obtained for the pair of enantiomers which were not experimentally separated by the columns (compound **13** in Tables 1–3) as an estimate of the uncertainties in the calculations, for the simple rigid body docking we obtain 10 correct predictions out of 12 for the phenylglycine column, with two undecided; for the more flexible leucine column, we obtain six correct predictions with six undecided.

Allowing more flexibility into the system by using the Metropolis Monte Carlo conformation space search, our correct predictions remain at 10 correct, for the phenylglycine column, but one of the undecided predictions has become wrong. For the leucine column we have improved to eight correct predictions, one wrong, with three undecided.

When the systems are allowed to relax fully, our success rate does not change significantly for the phenylglycine column. However, for the leucine column, the difference in energies obtained for compound **13**, our negative control, is so large at 2.2 kcal mol^{-1} as to make prediction of the first-eluting isomer rather dubious. We still obtain six correct predictions, with six undecided on the grounds of our now much larger error bars. There are no incorrect predictions (outside the error bars at least) but, even making the assumption that the error is entirely in the calculated energies for compound **13** (i.e. giving zero error in measurement for the rest, and predicting on that basis), we still only predict correctly for eight of out 12.

There are a number of possible sources of error. First, our stepwise procedure of docking followed by the Metropolis Monte Carlo search, followed by minimization, may be a choice of wrong starting point. In principle, the Metropolis Monte Carlo search should enable the system to 'visit' all parts of the available conformational space, but in practice, to run a simulation for this length of time would be prohibitively expensive. Examination of the output from our runs shows good sampling of the conformational space near the starting point, but rather few excursions away from it. It may be possible to improve sampling by running the simulation at a higher temperature, or by using a sampling technique with a more realistic physical basis than Monte Carlo—for example, molecular dynamics (though the tendency of non-bonded systems to fall apart completely in such simulations is well known), or a combination of sampling techniques.

An alternative to long sampling runs using a single starting point is to use multiple starting points and shorter simulations. The initial rigid body docking run gives very many potential starting points for the conformational space

sampling, and we have somewhat arbitrarily chosen the lowest of these. Comparison of Tables 2 and 3 (where the reported energies are the same quantity, and hence directly comparable) shows that relatively few cycles of full geometry optimization can have dramatic consequences for the calculated energy of the system, so very slight changes in initial geometry could have led to the selection of an alternative starting point. We have started applying hierarchical cluster analysis methods to extract different families of conformations from the output of programs such as our docking routine, and shall be investigating the possibility of using representatives of such families as starting points for the last two stages of the process. It seems likely to us that this will be more computationally efficient than very long or very high temperature simulations, and more likely to find the global minimum energy conformation in affordable time scales.

Our assumption that the entropy change is approximately constant between enantiomers may be in error. While changes in translational and rotational entropy should be the same for both enantiomers, the entropy changes due to the restriction of internal rotations on complex formation may differ, and we are interested in ways of estimating this from our calculations. Entropy differences due to differential solvation of the different complexes will probably be harder to estimate, but again we are interested in methods of estimating this quantity.

Finally, even if our sampling is perfect (or at least sufficient), and our assumptions about entropy are justified, it is still possible to get meaningless results if our energy calculation is not sufficiently good. For example, the systems described here have two, three or four aromatic rings, and molecular mechanics has a quite poor record of reproducing stacking phenomena in such systems. An error of a few tenths of a kcal mol^{-1} in the calculation of the interaction energy for both the *R* and *S* isomers could dramatically change the calculated energy difference, and hence our prediction of which isomer will be eluted first.

Improvements in charge representation: porphyrin stacking phenomena

Given that molecular mechanics uses the point charge (Mulliken) model for representing electron distribution in a molecule, its use in investigating aromatic and π-system stacking patterns always results in parallel co-centred binary complexes. Experiment suggests that for a simple case like benzene, the preferred binary is an end-on T-shaped pair, although many other arrangements can occur with comparable energy (Cox et al 1958). However, there seems to be no evidence for a parallel, co-centred binary. Under the same conditions of calculation, two porphyrin rings come together as a co-centred species able to spin on the central axis. Experiment has them parallel but staggered by about 4 Å down the N/N′ axis without rotation. Following the more rigorous wave-mechanical Distributed Multipole Analysis of Price & Stone (1987), Hunter & Sanders (1990) have made the simple assumption that the charge must be distributed (by 0.47 Å) in the

π plane. They set 0.5 electrons on each π-point and a +1 charge at the nucleus, giving each π carbon atom an overall neutrality. When a simple docking experiment was done using a conventional van der Waals potential to stop molecular coalescence, the correct geometry was predicted.

We have incorporated the technique of Hunter & Sanders (1990) into a larger molecular modelling environment and extended its application to lone pairs. Partial charges have been incorporated to account for electron distribution and polarizability and their experiments have been successfully repeated. We present these limited results in Figs. 2 and 3 as a preliminary confirmation of a potentially important contribution. In particular, Fig. 4 shows the distribution of porphyrin binaries over a set energy range. Hunter's techniques could not be used to investigate so many interactions and our results suggest that, although the general result obtained by Hunter & Sanders is reproduced (Fig. 5), a large variation of related low energy arrangements is likely to exist in solution. Further work will be done to consolidate the value of these modifications in the broader areas of molecular mechanics, molecular dynamics and biological modelling.

FIG. 4. The first 30 structures (over a 3 kcal mol^{-1} range) from a rigid body docking of porphyrin with itself using the Hunter & Sanders modification (looking along the stacking axis).

FIG. 5. The most stable stacked complex of porphyrin after a rigid body docking simulation (further refinement from Fig. 4).

Conclusion

Computer simulation of the docking of two species, from the simple interaction with a proton to the complex interaction of two chiral structures, can give insight into the factors that are at play. It is reasonable to conclude that the main forces are electrostatic but are complicated by solvent interaction, molecular flexibility and kinetic effects. However, the computational methods are continually being improved and updated and, occasionally, a leap forward (however simple) is made which may improve them to the point of casting doubt on previous work.

Acknowledgements

We are grateful to Jeremy Sanders at Cambridge for helpful discussion. We should also like to thank Andrew Davis at SKB for writing some of the software for this project.

References

Abraham RJ, Hudson B 1985 Charge calculations in molecular mechanics. III: Amino acids and peptides. J Comput Chem 6:173–181

Bergmann M, Fruton JS, 1937 Proteolytic Enzymes III. Specificity of aminopeptidase and carboxypeptidase. A new type of enzyme in the intestinal tract. J Biol Chem 117:189–202

Charlton PA, Young DW, Birdsall B, Feeney J, Roberts GCK 1979 Stereochemistry of reduction of folic acid using dihydrofolate reductase. J Chem Soc Chem Commun 922–924

Cox EG, Cruikshank DWJ, Smith JAS 1958 The crystal structure of benzene at −3°C. Proc R Soc Lond A Math Phys Sci 247:1–21

DeWolf WE, Chambers PA, Southan CD, Saunders D, Kruse LI 1989 Inactivation of dopamine β-hydroxylase by β-ethynyltyramine: kinetic characterisation and covalent modification of an active site peptide. Biochemistry 28:3833–3842

Fontecilla-Camps JC, Bugg CE, Temple C, Rose JD, Montgomery JA, Kisluik RL 1979 Absolute configuration of tetrahydrofolates. A crystallographic determination. J Am Chem Soc 101:6114–6115

Hunter CA, Sanders JKM 1990 The Nature of π-π Interactions. J Am Chem Soc 112:5525–5535

Ockenfels H, Kohler F, Meise W 1977 Teratogenic effect of N-phthaloyl-L-aspartic acid on the mouse. Arzneim Forsch 27:126–128

Pirkle WH, Tsipouras A 1984 Direct liquid chromatographic separation of benzodiazepinone enantiomers. J Chromatog 291:291–298

Price SL, Stone AJ 1987 The electrostatic interactions in van der Waals complexes involving aromatic molecules. J Chem Phys 86:2859–2868

Scrocco E, Tomasi J 1973 The electrostatic molecular potential as a useful tool for the interpretation of some molecular properties. Top Curr Chem 42:95–150

Vinter JG, Harris M 1989 The graphical display of chemical structure. Chem Br 25:1111–1116

Vinter JG, Davis A, Saunders MR 1987 Strategic approaches to drug design. 1. An integrated software framework for molecular modelling. J Comput Aided Mol Des 1:31–51

Watson HW, Galloy J, Grossie AD, Vogtle F, Muller WM 1984 Host–guest complex chemistry. Structures of 18-crown-6 and diaza-18-crown-6 with neutral molecules. J Org Chem 49:347–353

Weinstein H, Osman R, Topiol S, Green JP 1981 Quantum chemical studies on molecular determinants for drug design. Ann N Y Acad Sci 367:434–451

DISCUSSION

Dunitz: You would not claim, would you, that the D_{3d} structure for complexed 18-crown-6 (Fig. 3) is the energy minimum for uncomplexed 18-crown-6?

Vinter: It is the energy minimum in the complexed structure. I tried to reproduce the formamide complex of 18-crown-6 by examining only the paths of approach of one structure towards another.

Dunitz: The D_{3d} structure is observed in crystal structures only where 18-crown-6 is complexed with other molecules or ions.

Vinter: Originally, I was interested in the complexation of diquat with a 30-crown-10 derivative, because when these two species combine the crown ether changes from a twisted, rather messy, mobile structure, to wrap around the diquat like a horseshoe. We were not able to simulate this movement computationally at that time and the 18-crown-6 complex was chosen to investigate only approach using the complex crystal coordinates. We are still working on the problem of mutual conformational movement and are near to a computational solution.

Dunitz: The electric potential is relatively insensitive to where you put the charges, because you integrate over the charge distribution to get the potential. In the D_{3d} structure, all the lone pairs on the oxygens are pointing inside the ring; unless you put something in the centre, such as a potassium ion, it cannot be an energy minimum.

Vinter: The effect of the focusing of the oxygen lone pairs is to produce a deep electrostatic energy minimum at the centre of the hexagonal structure. If you look at phthalocyanin-crown ether ionophores, which are immobile, a similar deep coulombic well exists exactly where an ion would be found if complexed. You are right that we cheated by starting with the empty 18-crown-6 in a conformation already set up for complexation, but we were not in a position to investigate two variables, i.e. approach and movement, at that time.

Dunitz: When you start with a structure which must be close to an energy maximum along some internal coordinates at least, why doesn't it relax into the right structure?

Vinter: The structure of the 18-crown-6 that we used was one of the high energy conformations compared with the overall conformational profile, but when the energy was minimized and the structure relaxed, it retained its open

form and occupied a local minimum position even without its complexed partner. If we had manually taken the ring apart or squashed it well away from the open hexagonal shape, it would have relaxed to a quite different architecture.

Dunitz: It would be nice if you could look at the right structure.

Vinter: As I said, I would like to study the 30-crown-10-series and be able to watch the complementary movements occurring on complex formation. We are now in a position to do this.

Sutherland: 18-Crown-6 would be relatively rigid if one started with the structure found in a cation complex, but in real life, free 18-crown-6 has an enormous number of low energy conformations.

Rebek: In the docking of formamide with 18-crown-6 the hydrocarbon fragments were positively charged (Fig. 2). In docking experiments of the sort that lead to hydrophobic effects between two alkyl groups, will these not be tremendously repulsive?

Vinter: No; those electrostatic minima associated with hydrocarbon areas were very small compared with the deep negative wells at the centre of the crown ether. To plot the electrostatic map we use a point positive charge which can penetrate below the combined van der Waals surfaces of two interacting non-protonic hydrogens. They are therefore of little consequence compared with the van der Waals potential at the same point. Those positive minima are much less significant than they appear.

Sutherland: In Figs. 4 and 5 I assume you have superimposed a charge distribution on the atoms of the porphyrin system.

Vinter: I have, but the original authors (Hunter & Sanders 1990) didn't. When they did this work they ignored the charge distribution in porphyrin and each atom was a neutral system. In fact, Hunter & Sanders separated the charge only on each atom having a π system and each complete atom system was overall neutral.

Sutherland: Do you differentiate between heteroatoms and carbons?

Vinter: Hunter and Sanders didn't; we have done it since.

Sutherland: What happens when you do that?

Vinter: You are then varying the energy of interaction, but the final geometry is very hard to change; once you have put in this idea of charge separation, the geometry remains extremely stable, even though you do not take into account nitrogen lone pairs.

Sutherland: In the model, if you look at the electrostatics without lone pairs it would make sense.

Vinter: Yes, because the stacking problems are solely a function of π-bond interactions. I showed the results of my calculations but it's Chris Hunter's calculations that I want to emphasize. We are working on new molecular mechanics to include π-charge separation to deal with the stacking phenomenon, which is extremely important in biological systems.

Gokel: If you deal with a cation as a point charge, then potassium is equivalent to sodium, but the metal–oxygen bond distances required are obviously somewhat different. One knows from solid-state structure data, not only for 18-crown-6 but also for a number of macrocycles, that 18-crown-6 in particular does not adopt a hexagonal conformation when it binds a sodium ion. In fact, one of the macroring oxygens puckers up and there is also a water molecule in the solvation shell. This puckering effectively reduces the size of the ring so that the shorter Na^+–O distance (compared with the K^+–O distance) can be accommodated. A similar, but perhaps opposite, phenomenon is observed for the complex between 12-crown-4 and sodium. In this case, two small rings sandwich the too-large cation and octa-coordinate sodium. Here, there is no counter ion in the solvation sphere. How do you deal with the fact that bond distances may be different, even though the charges are the same? Another complication is that when there are too few donors, the anion may either be in the solvation sphere or it may be hydrogen bonded to a water molecule in the solvation sphere.

Vinter: Of course, in full-blown docking experiments similar to the one between 18-crown-6 and formamide, each interactor is allocated a normal van der Waals surface which defines its shape and therefore its interaction energy. For electrostatic maps we use point positive and negative charges which are allocated only small bulk. These maps are therefore not size sensitive. We have done experiments (J. G. Vinter, M. R. Saunders & A. Davis, unpublished work 1986) in which specific ions have been used as probes and, as such, were given van der Waals volumes commensurate with their ionic radii. The results were not validated by experiment in these cases, but the results looked reasonable. It is worth mentioning that the experimental verification of partial charge distribution and hence of electrostatic maps is not directly possible. When we were using these mapping methods for structure–activity work in industry, one very useful check was to pick similar structures from Dr Kennard's Cambridge Structural Database and compare our electrostatic energy minima with counter ion and solvent positions in the crystal. It is reasonable to assume that these floating entities would find their way to electrostatic wells, and that proved to be so in the majority of cases.

Dunitz: If one makes statistics, not with Monte Carlo simulations but with observed patterns in crystal structures, one finds that ether oxygen has different structural preferences for the various alkali cations (Chakrabarti & Dunitz 1982). It's not only the oxygen–metal distances that differ, but also the orientation to the R—O—R group is different for lithium, sodium and potassium. The cations invariably lie close to the bisecting plane through the oxygen atoms, but whereas the larger cations show an apparent preference for approaching the ether group along a tetrahedral lone pair direction, Li^+ cations tend to be found along the R—O—R bisector—that is, along the trigonal lone pair direction. It should be clear that an electrostatic model based on the potential round the ether group is totally inadequate to describe these results.

Vinter: That's exactly what I am saying. Everybody uses this overall generalization of point charge distribution. What's fascinating to me is that we are beginning to realize that we must stretch and re-distribute charge about atoms and bonds. Certainly, the polarization problem is grossly underestimated and we can't tackle it properly; we can apply crude approximations, but we don't get a feeling for the dynamic variation. The idea of pulling out orbitals in a quantum mechanical-type of way and putting charge in various places on those pulled out π points and lone pair points may well begin to answer your problems. I am trying to take you back and show you the terrible inadequacies of what we have. The simplicity of this idea of distributing charge sensibly around the atom, which should have been thought of long ago in molecular mechanics, may well give us a new impetus to start to look at factors like polarization of charge, how we can move it in space in response to another incoming entity, and how we can vary it such that we get the patterns that we see in the crystal. It's lunatic, but until now the charge has regularly been put in one place and left there, regardless of dynamics, and treated as a sort of sacred cow, not to be touched—that is nonsense.

References

Chakrabarti P, Dunitz JD 1982 Directional preferences of ether O-atoms towards alkali and alkaline earth cations. Helv Chim Acta 65:1482–1488

Hunter CA, Sanders JKM 1990 The nature of π–π interactions. J Am Chem Soc 112:5525–5535

Summing-up

Ian Sutherland

Department of Chemistry, University of Liverpool, PO Box 147, Liverpool L69 3BX, UK

It's quite worrying for me to have to reflect on a meeting of this type because it covers enormous areas of ignorance as well as the areas about which I know a little. Perhaps it would be appropriate to look into a crystal ball and make predictions, rather than talk about the past. Talking about the future gives one an element of freedom—people may remember what you say, of course, but you can afford to get something wrong because everyone who makes predictions also makes mistakes.

The scope of the meeting was certainly wide; biologists and chemists talked to one another, usually speaking the same language. The complexity of intermolecular interactions obviously differs between biology and chemistry; that was underlined in the contributions on protein complexes. But it is reassuring for organic chemists, who have to make simple molecules, to discover that relatively few binding sites can complex carbohydrates and extract them from aqueous solution. Hydroxycalixarenes can apparently do this, which is encouraging for chemists because we can hope to simulate some of the features of proteins using much simpler molecules.

At the beginning of the meeting, when molecular assembly was the topic for discussion, Fraser Stoddart led us in the direction which molecular assembly has to take. If you cast your minds back to the work that Schill did many years ago, you will remember that molecular assembly was then based upon making covalent bonds and subsequently breaking them. That's a start, but it's obviously synthetically difficult. The next idea came from Jean-Pierre Sauvage, who found that using metals to make similar assemblies meant that the unwanted bonds were easier to break. Dr Stoddart has taken us a step further, in the direction that I'm sure is the one for the future, towards the assembly of molecules using non-covalent interactions.

Of course, he talked about relatively small molecular assemblies; larger three-dimensional molecular assemblies will be a formidable problem until we can predict crystal structures. We are currently quite unable to do this, although I still believe that occasionally we can be successful using analogies. I think synthetic molecular assemblies will remain a problem for quite some time.

Two-dimensional assemblies have been mentioned and we have talked about monolayers and bilayers. Those of us who want well-described, predictable

assemblies should perhaps move away from the very mobile assembly of the type that George Gokel has been using, to more rigid assemblies in two dimensions, before we take the step into three dimensions.

The applications of molecular assemblies could be very wide indeed, but we have yet to learn how to organize them and to make them of an appropriate size. We know about small molecular assemblies and large molecular assemblies in two dimensions, but the sort of assemblies that are required to construct optical or electrical devices seem to be the ones that are very hard to make—those that are of intermediate size. We shall certainly have to think about that in the future and I believe that there will be considerable progress in that direction.

At the moment, the most promising application is in the area of molecular sensors, with the use of either biomolecules to recognize substrates or synthetic molecules to recognize simpler substrates such as cations. This sounds like an advertisement, but molecular sensors will have enormous economic implications. If the tedious, time-consuming methods that are used every day in the health service and in environmental analysis could be replaced by simple, continuously recording, long-lived sensors, a substantial amount of money could be saved. I am sure there are people somewhere in this city who have already taken note of that possibility.

We seemed a little pessimistic about moving from enzyme- and immunoassay-based sensors to sensors using larger molecules and synthetic chemistry, but, having listened to Julius Rebek and Ron Breslow, I believe that we are making a lot of progress towards making small synthetic molecules that can complex small organic molecules. There will be advances in the area of molecular recognition by synthetic systems and I know that there are a few people here who hope that I am right in that prediction. Eventually, this may also produce useful catalysts. We have already seen signs that some useful catalytic systems will emerge from this work, although I would not like to guess when these systems will reach large-scale use. It will be difficult to compete with enzymes that are fairly freely available. Progress is also likely in inserting some of these active and selective molecules into assemblies to achieve selective performance from the various devices that we hope to make.

Let me move on now to natural host molecules. The effectiveness of vancomycin in forming complexes gives us hope for the future for the design of small molecules. It seems that we need relatively few specifically arranged interactions to carry out complexation in rather unfavourable aqueous solvents. It's clear that DNA is a receptor molecule *par excellence*. I doubt we shall make molecules of such sophistication in the near future. The work on DNA will throw light on DNA structure. In addition, using molecules which either simulate a part of the DNA molecule or bind strongly to DNA, it will be possible to inhibit the normal functioning of DNA or to inhibit binding of its regulatory enzymes; this is a promising chemotherapeutic approach.

To move from DNA to the most complex molecules of all, the proteins, is another step up the molecular scale. We are starting to understand much more about not only the structures of proteins but also their dynamics; this comes from research into proteins using NMR spectroscopy and the examination of substrate specificity. The difficult problem of protein dynamics is starting to yield to such investigations. We are beginning to understand, and certainly will understand in the future, the differences between selective, rigid proteins and the more flexible, less selective, but very useful proteins which can bind a wide range of molecules. One hopes, as a chemist, that this will lead ultimately to the design of proteins, not by close analogy with structures of known function, but by *ab initio* design based on knowledge of protein structure. Protein structure will eventually become predictable. The lead in to this is the nice idea of taking fragments of proteins and organizing them together on a template, to obtain some aspects of organized protein structure. Ultimately, I feel we shall be able to translate prediction from one dimension to three dimensions before doing the experiments.

That is quite an optimistic note on which to end, and I must remind you of the final contribution by Andy Vinter. It is clear that our modelling must improve, but it is also clear that modelling plays a useful role in our thinking, as it always has done. The good news for molecular modelling is that the cost is decreasing while the memory and the speed are increasing. In the future, we shall be able to look much more easily at the sort of molecules that we like to think about and have discussed. The role that molecular modelling plays will increase, but I doubt if it will overtake experimental and creative skills in the foreseeable future. That's a rash prediction, but I think the hands and brains of chemists and biologists will still be needed—computers are not going to take over. Predicting that we shall retain our jobs is a nice note on which to end my forecasting!

To return to the present, I should like to thank the splendid speakers who have stimulated discussion and who really are experts in their fields. The discussion at this meeting has been the best that I can recall for many years. It has been the sort of discussion that one normally has over a cup of tea with friends and it's very pleasing to have been able to extend this over a wider circle with new friends from other fields. I think we have all benefited from this discussion and now understand a little more about how the other side thinks; the chemists have learned from the biologists and maybe the biologists have even learned something from the chemists.

Index of contributors

Non-participating co-authors are indicated by asterisks. Entries in bold type indicate papers; other entries refer to discussion contributions.

Indexes compiled by Liza Weinkove

Subject index

Ciba Foundation Symposium 158

HOST-GUEST MOLECULAR INTERACTIONS: FROM CHEMISTRY TO BIOLOGY

Chairman: **I.O. Sutherland**
1991

Host–guest molecular interactions lead to the formation of molecular assemblies containing two or more species. It has long been recognized that such 'supramolecular' association is of fundamental importance in biological processes, but it is only recently that chemists have begun to design and synthesize novel host molecules.

The field is relatively new, but exciting applications are already emerging, for example in the development of molecular sensors, of new chromogenic reagents and of reagents for selective metal ion capture or for the capture of molecules such as urea. In addition, it is expected that improved understanding of the interactions in relatively simple complexes of synthetic host molecules will lead to a better understanding of the much more complex biological systems in which the host molecules are either proteins or nucleic acids.

In this book the contributions from experts in the chemical and biological sciences range over intermolecular interactions in both synthetic and natural (biological) systems. Topics covered include biomimetic chemistry, the design of synthetic host molecules which mimic biological processes such as ion transport across membranes and catalysis, and the concepts of preorganization and template-directed synthesis. Molecular recognition processes involved in binding of antibiotics to peptides and to DNA and in binding of sequence-specific gene inhibitors to DNA are discussed. The potential applications of oligonucleotides with modified phosphate diester linkages as probes of DNA–protein interactions and for anti-HIV treatment are described. Interactions between proteins and other molecules and the computational modelling of interactions are considered.

Related Ciba Foundation Symposia:

No 119 **Synthetic peptides as antigens**
Chairman: G.L. Ada
1986 ISBN 0 471 99838 9

No 145 **Carbohydrate recognition in cellular function**
Chairman: E. Ruoslahti
1989 ISBN 0 471 92307 9

No 159 **Catalytic antibodies**
Chairman: W.P. Jencks
1991 ISBN 0 471 92962 X

JOHN WILEY & SONS
Chichester · New York · Brisbane · Toronto · Singapore
A Wiley–Interscience Publication