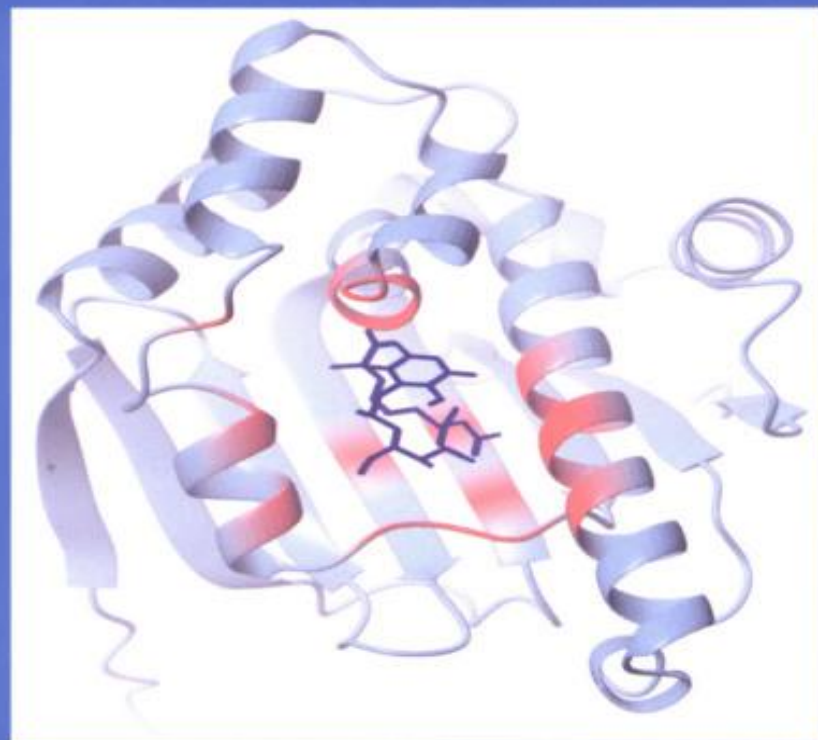




Protein-Ligand Interactions

From Molecular Recognition
to Drug Design

Edited by H.-J. Böhm and G. Schneider



**Methods
and Principles
in Medicinal
Chemistry**

Volume 19

Edited by
R. Mannhold,
H. Kubinyi,
G. Folkers

Protein-Ligand Interactions
From Molecular Recognition to Drug Design

Edited by
H.-J. Böhm and G. Schneider

Methods and Principles in Medicinal Chemistry

Edited by

R. Mannhold

H. Kubinyi

G. Folkers

Editorial Board

H.-D. Höltje, H. Timmerman, J. Vacca,

H. van de Waterbeemd, T. Wieland

Protein-Ligand Interactions From Molecular Recognition to Drug Design

Edited by

H.-J. Böhm and G. Schneider



**WILEY-
VCH**

WILEY-VCH GmbH & Co. KGaA

Series Editors

Prof. Dr. Raimund Mannhold

Biomedical Research Center
Molecular Drug Research Group
Heinrich-Heine-Universität
Universitätsstraße 1
40225 Düsseldorf, Germany
e-mail: raimund.mannhold@uni-duesseldorf.de

Prof. Dr. Hugo Kubinyi

BASF AG Ludwigshafen
c/o Donnersbergstraße 9
67256 Weisenheim am Sand, Germany
e-mail: kubinyi@t-online.de

Prof. Dr. Gerd Folkers

Department of Applied Biosciences
ETH Zürich
Winterthurerstr. 190
8057 Zürich, Switzerland
e-mail: folkers@pharma.anbi.ethz.ch

Volume Editors

Prof. Dr. Hans-Joachim Böhm

F. Hoffmann-La Roche Ltd.
Pharmaceuticals Division
4070 Basel, Switzerland
e-mail: hans-joachim.boehm@roche.com

Prof. Dr. Gisbert Schneider

Institute of Organic Chemistry
and Chemical Biology
Johann Wolfgang Goethe-Universität
Marie-Curie-Straße 11
60439 Frankfurt am Main, Germany
e-mail: g.schneider@chemie.uni-frankfurt.de

Cover illustration

The anti-tumor agent Geldanamycin bound to the N-terminal domain of the chaperone protein HSP90 (Stebbins, C. E., Russo, A. A., Schneider, C., Rosen, N., Hartl, F. U., Pavletich, N. P., *Cell* **89** pp. 239 (1997).
Kindly provided by Doris M. Jacobs, Bettina Elshorst, Thomas Langer, Susanne Grimme, Barbara Pescatore, Krishna Saxena, and Martin Vogtherr; Johann Wolfgang Goethe-Universität Frankfurt am Main, Germany.

This book was carefully produced. Nevertheless, authors, editors and publisher do not warrant the information contained therein to be free of errors. Readers are advised to keep in mind that statements, data, illustrations, procedural details or other items may inadvertently be inaccurate.

Library of Congress Card No. applied for

British Library Cataloguing-in-Publication Data:

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

Bibliographic information published by Die Deutsche Bibliothek

Die Deutsche Bibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data is available in the Internet at <http://dnb.ddb.de>

© 2003 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

All rights reserved (including those of translation in other languages). No part of this book may be reproduced in any form – by photoprinting, microfilm, or any other means – nor transmitted or translated into machine language without written permission from the publishers. Registered names, trademarks, etc. used in this book, even when not specifically marked as such, are not to be considered unprotected by law.

Printed in the Federal Republic of Germany
Printed on acid-free paper

Typesetting K+V Fotosatz GmbH, Beerfelden
Printing Strauss Offsetdruck GmbH, Mörlenbach
Bookbinding J. Schäffer GmbH & Co. KG, Grünstadt

ISBN 3-527-30521-1

Contents

Preface *XI*

A Personal Foreword *XIII*

List of Contributors *XV*

List of Abbreviations *XVII*

Prologue *1*

David Brown

1	Prediction of Non-bonded Interactions in Drug Design	3
	<i>H.-J. Böhm</i>	
1.1	Introduction	3
1.2	Major Contributions to Protein-Ligand Interactions	4
1.3	Description of Scoring Functions for Receptor-Ligand Interactions	8
1.3.1	Force Field-based Methods	9
1.3.2	Empirical Scoring Functions	9
1.3.3	Knowledge-based Methods	11
1.4	Some Limitations of Current Scoring Functions	12
1.4.1	Influence of the Training Data	12
1.4.2	Molecular Size	13
1.4.3	Water Structure and Protonation State	13
1.5	Application of Scoring Functions in Virtual Screening and <i>De Novo</i> Design	14
1.5.1	Successful Identification of Novel Leads Through Virtual Screening	14
1.5.2	<i>De novo</i> Ligand Design with LUDI	15
1.6	Outlook	16
1.7	Acknowledgments	17
1.8	References	17

2	Introduction to Molecular Recognition Models	21
	<i>H.-J. Schneider</i>	
2.1	Introduction and Scope	21
2.2	Additivity of Pairwise Interactions – The Chelate Effect	22
2.3	Geometric Fitting: The Hole-size Concept	26
2.4	Di- and Polytopic Interactions: Change of Binding Mechanism with Different Fit	28
2.5	Deviations from the Lock-and-Key Principle	30
2.5.1	Strain in Host-Guest Complexes	30
2.5.2	Solvent Effects	30
2.5.3	Enthalpy/Entropy Variations	31
2.5.4	Loose Fit in Hydrophobically Driven Complex Formation	32
2.6	Conformational Pre-organization: Flexible vs. Rigid Hosts	32
2.7	Selectivity and Stability in Supramolecular Complexes	34
2.8	Induced Fit, Cooperativity, and Allosteric Effects	36
2.9	Quantification of Non-covalent Forces	38
2.9.1	Ion Pairs and Electrostatic Donor-Acceptor Interactions	38
2.9.2	Hydrogen Bonds	39
2.9.3	Weak Hydrogen Bonds: The Use of Intramolecular „Balances“	42
2.9.4	Polarization Effects	43
2.9.5	Dispersive Interactions	43
2.10	Conclusions	46
2.11	References	46
3	Experimental Approaches to Determine the Thermodynamics of Protein-Ligand Interactions	51
	<i>R. B. Raffa</i>	
3.1	Introduction	51
3.2	Basic Thermodynamics of Protein-Ligand Interactions	51
3.3	Measurement of Thermodynamic Parameters	54
3.3.1	Calorimetric Determination of Thermodynamic Parameters	55
3.3.2	van't Hoff Determination of Thermodynamic Parameters	57
3.3.2.1	Relationship to Equilibrium Constant	57
3.3.2.2	Obtaining the Equilibrium Constant	59
3.4	Applications	60
3.4.1	Calorimetric Determination of Thermodynamic Parameters	60
3.4.2	van't Hoff Determination of Thermodynamic Parameters	63
3.5	Caveats	67
3.6	Summary	68
3.7	References	69
4	The Biophore Concept	73
	<i>S. Pickett</i>	
4.1	Introduction	73
4.2	Methodology for Pharmacophore Detection and Searching	74
4.2.1	Definition of Pharmacophoric Groups	75

4.2.2	Ligand-based Methods for Pharmacophore Perception	78
4.2.3	Protein Structure-based Pharmacophore Perception	84
4.2.4	Methods for Pharmacophore Searching	86
4.3	Pharmacophore Fingerprints	88
4.4	Applications of the Biophore Concept	91
4.4.1	Lead Generation	91
4.4.2	Multi-pharmacophore Descriptors in Diversity Analysis and Library Design	92
4.4.3	Structure-based Design	95
4.5	The Biophore Concept in ADME Prediction	98
4.6	Summary	99
4.7	References	100
5	Receptor-Ligand Interaction	107
	<i>M. M. Höfliger, A. G. Beck-Sickinger</i>	
5.1	Receptors	107
5.1.1	The G-Protein-Coupled Receptors	107
5.2	Ligand-binding Theory	108
5.3	Characterization of the Receptor-Ligand Interaction	111
5.4	Receptor Material	111
5.5	Binding Studies	112
5.6	Binding Kinetics	112
5.7	Binding Assays	115
5.7.1	Separation Assays	115
5.7.2	Radioligand-binding Assay	115
5.8	Fluorometric Assays	116
5.8.1	Fluorescence Labels	116
5.8.2	Fluorescence Correlation Spectroscopy (FCS)	116
5.8.3	Fluorescence Microscopy	117
5.8.4	Fluorescence Resonance Energy Transfer (FRET)	117
5.9	Surface Plasmon Resonance	118
5.10	Molecular Characterization of the Receptor-Ligand Interaction	120
5.10.1	Antibodies	120
5.10.2	Applications of Antibodies	122
5.10.2.1	Receptor and Ligand Detection	122
5.10.2.2	Receptor Characterization	124
5.10.2.3	Functional Characterization of the Receptor-Ligand Interaction	124
5.10.3	Aptamers	125
5.10.4	Receptor Mutation and Ligand Modification	125
5.10.4.1	Receptor Mutagenesis	126
5.10.4.2	Ligand Modification	127
5.10.4.3	Combination of Receptor Mutation and Ligand Modification	129
5.10.5	Cross-linking	130
5.11	Conclusion	132
5.12	References	133

6	Hydrogen Bonds in Protein-Ligand Complexes	137
	<i>M. A. Williams, J. E. Ladbury</i>	
6.1	Introduction	137
6.1.1	The Importance of Hydrogen Bonds	137
6.1.2	Defining the Hydrogen Bond	138
6.2	Physical Character of Hydrogen Bonds	139
6.2.1	Crystallographic Studies of Hydrogen Bonds	139
6.2.2	The Geometry of Hydrogen Bonds	140
6.2.3	Infrared Spectroscopy of Hydrogen Bonds	145
6.2.4	NMR Studies of Hydrogen Bonds	145
6.2.5	Thermodynamics of Hydrogen Bonding	147
6.2.6	Experimental Thermodynamics of Biomolecular Hydrogen Bonds	148
6.3	Interactions with Water	150
6.3.1	Bulk and Surface Water Molecules	150
6.3.2	Buried Water Molecules	151
6.4	Hydrogen Bonds in Drug Design	153
6.4.1	Diverse Effects of Hydrogen Bonding on Drug Properties	153
6.4.2	Optimizing Inhibitor Affinity	154
6.4.3	Computational Tools for Hydrogen Bond Analysis and Design	156
6.5	Conclusion	158
6.6	References	158
7	Principles of Enzyme-Inhibitor Design	163
	<i>D. W. Banner</i>	
7.1	Introduction	163
7.2	The Active Site	165
7.3	The Heuristic Approach	165
7.4	Mechanism-based Covalent Inhibitors	166
7.5	Parallel <i>de novo</i> Design of Inhibitors	168
7.5.1	Evolution of Inhibitors	169
7.6	Inhibitors from Progressive Design	170
7.7	Lessons from Classical Inhibitors	172
7.8	Estimating the Energies of Interactions	176
7.9	Water and Solvent	178
7.9.1	Displacing a Tightly Bound Water	179
7.9.2	Binding of Solvent Molecules	180
7.9.3	Screening	181
7.10	Structure-Activity Relationships (SAR)	181
7.11	Present Clinical Status of Thrombin Inhibitors	182
7.12	Conclusions	183
7.13	Acknowledgments	183
7.14	References	184

8	Tailoring Protein Scaffolds for Ligand Recognition	187
	<i>A. Skerra</i>	
8.1	Introduction	187
8.2	Lipocalins: A Class of Natural Compound Carriers	191
8.3	Anticalins: Lipocalins Reshaped via Combinatorial Biotechnology	194
8.4	Structural Aspects of Ligand Recognition by Engineered Lipocalins	199
8.5	Prospects and Future Applications of Anticalins	205
8.6	References	210
9	Small Molecule Screening of Chemical Microarrays	213
	<i>G. Metz, H. Otteleben, D. Vetter</i>	
9.1	Introduction	213
9.2	Fragment Approaches	214
9.2.1	Conceptual Ideas	214
9.2.2	Choice of Screening Fragments	217
9.2.3	Experimental Approaches	218
9.3	Chemical Microarrays	222
9.3.1	Background	222
9.3.2	On-array Synthesis	223
9.3.3	Off-array Synthesis and Spotting	224
9.4	Screening on Microarrays	229
9.4.1	Detection Technology	229
9.4.2	Protein Affinity Fingerprints	231
9.5	Conclusion	232
9.6	Acknowledgement	234
9.7	References	234

Subject Index	237
----------------------	------------

Preface

The understanding of protein-ligand interactions is the fundamental basis of medicinal chemistry. With only a very few exceptions, drugs interact with macromolecular targets, most often with specific binding sites of membrane-bound or nuclear receptors, enzymes, transporters, or ion channels. Essential for high biological activity are a good geometric fit (the Emil Fischer “lock-and-key” principle) and a high degree of complementarity of hydrophobic and polar parts of both entities, namely, the binding site of the protein and the ligand. However, this short characterization is only part of the story: ligand and binding site flexibility, distortion energies, desolvation effects, entropy, molecular electrostatic field complementarity, and other effects are often equally important.

The chapters of this book, written by leading experts of academia and industry, describe all relevant aspects of intermolecular interactions in great detail. There has been significant progress in the understanding of the forces involved, derived from the inspection of protein-ligand complexes and from systematic investigations of artificial host-guest complexes. Many examples illustrate these effects, as well as the inherent problems of extrapolating from one example to the other. Still, our ability to predict ligand affinities is very limited. Scoring functions for a better estimation of binding affinities (or only their relative differences within congeneric series of compounds) are under active development.

We are sure that this book will be of great value for everybody involved in lead discovery and optimization. It will contribute to further progress in this field and will hopefully pave the way for even better understanding and quantification of the effects governing protein-ligand interactions.

The editors of the book series “Methods and Principles in Medicinal Chemistry” are very grateful to Hans-Joachim Böhm and Gisbert Schneider for their careful selection of authors and their engaging work on this project, to Frank Weinreich for his editorial effort, and to Wiley-VCH for the production of the work.

January 2003

Raimund Mannhold, Düsseldorf
Hugo Kubinyi, Weisenheim am Sand
Gerd Folkers, Zürich

A Personal Foreword

Molecular recognition events are the underlying processes leading to phenomena like “bioactivity”, and understanding molecular recognition is pivotal to successful drug design. This volume gives an overview of current concepts and models addressing the interaction patterns of proteins and their small molecule ligands. The current volume focuses on non-bonding drug-receptor interactions in an aqueous environment as these are most relevant for pharmaceutical drug discovery projects.

Beginning with a general introduction to predictive approaches (Chapter 1) and an overview of molecular recognition models (Chapter 2) providing the conceptual framework on a more theoretical level, important experimental approaches to measuring properties of protein-ligand interactions are treated in Chapter 3. Due to the great importance of pharmacophore modeling in early-phase drug discovery, Chapter 4 is devoted to this topic addressing the many different approaches in this challenging field of research. Structure-based modeling of protein-ligand interactions becomes particularly difficult when a reliable model of the three-dimensional receptor structure is unavailable – a situation the molecular designer is often confronted with when dealing with membrane protein receptors. Chapter 5 shows ways how to address this issue. Since directed polar interactions, in particular hydrogen bonding patterns, are the main determinants of binding specificity, a whole Chapter highlights this central topic (Chapter 6). Chapter 7 describes the practical approach to structure-based drug design taking enzyme-ligand interactions as an example. Finally, Chapter 8 addresses the challenging question how to design the receptor – not the ligand – to obtain desired properties as a host molecule for a small molecular guest; and Chapter 9 extends the treatment of molecular recognition in protein-ligand interactions to the multi-dimensional case, i.e. the field of multiple parallel measurements using modern microarray technology. We are convinced that this compilation of Chapters will provide an entry point to the study of protein-ligand interactions for any interested scientist, in particular medicinal chemists and advanced students of the life sciences.

Editing this book would not have been possible without sustained support from a number of people. We are particularly thankful to Petra Schneider and Martin Stahl, and all our colleagues at F. Hoffmann-La Roche and the MODLAB-Team at Goethe-University for many stimulating discussions and valuable support. Dave

Brown is equally thanked for the Prologue to this volume highlighting the importance of the topic from his long experience in pharmaceutical research. We are very grateful to the series Editors, in particular Hugo Kubinyi, for many helpful comments and encouragement during all phases of the project. Frank Weinreich from Wiley-VCH did an outstanding job putting all the pieces together, and carefully edited this volume. All authors are very much thanked for their great enthusiasm and excellent contributions.

Basel and Frankfurt, December 2002

Hans-Joachim Böhm
Gisbert Schneider

List of Contributors

Dr. DAVID W. BANNER
F. Hoffmann-La Roche Ltd
Pharmaceuticals Division
CH-4070 Basel
Switzerland

Prof. Dr. ANNETTE G. BECK-SICKINGER
Universität Leipzig
Institut für Biochemie
Talstraße 33
D-04103 Leipzig
Germany

Prof. Dr. HANS-JOACHIM BÖHM
F. Hoffmann-La Roche AG
Discovery Chemistry
Pharmaceuticals Division
CH-4070 Basel
Switzerland

Dr. DAVID BROWN
President and CEO
Cellzome AG
Meyerohofstraße 1
D-69117 Heidelberg
Germany

previously
F. Hoffmann-La Roche AG
Pharmaceuticals Division
CH-4070 Basel
Switzerland

Dr. MARTIN M. HÖFLIGER
Universität Leipzig
Institut für Biochemie
Talstraße 33
D-04103 Leipzig
Germany

Dr. JOHN E. LADBURY
Wellcome Trust Senior Research
Fellow
Department of Biochemistry & Molecular Biology
University College London
Gower Street
London, WC1E 6BT
UK

Dr. GÜNTHER METZ
Graffinity Pharmaceutical Design
GmbH
Im Neuenheimer Feld 518–519
D-69120 Heidelberg
Germany

Dr. HOLGER OTTLEBEN
Graffinity Pharmaceutical Design
GmbH
Im Neuenheimer Feld 518–519
D-69120 Heidelberg
Germany

Dr. STEPHEN PICKETT
GlaxoSmithKline Ltd.
Medicines Research Centre
Gunnels Wood Road
Stevenage, Hertfordshire, SG1 2NY
UK

Prof. Dr. ROBERT B. RAFFA
Temple University School of Pharmacy
3307 N. Broad Street
Philadelphia, PA 19140
USA

Prof. Dr. GISBERT SCHNEIDER
Institute of Organic Chemistry
and Chemical Biology
Johann Wolfgang Goethe-Universität
Marie-Curie-Straße 11
60439 Frankfurt am Main
Germany

Prof. Dr. HANS-JÖRG SCHNEIDER
Universität des Saarlandes
FR 8.12 Organische Chemie
D-66041 Saarbrücken
Germany

Prof. Dr. ARNE SKERRA
Technische Universität München
Lehrstuhl für Biologische Chemie
An der Saatzeit 5
D-85350 Freising-Weißenstephan
Germany

Dr. DIRK VETTER
Graffinity Pharmaceutical Design
GmbH
Im Neuenheimer Feld 518–519
D-69120 Heidelberg
Germany

Dr. MARK A. WILLIAMS
University College London
Department of Biochemistry & Molecular Biology
Gower Street
London, WC1E 6BT
UK

List of Abbreviations

2'-CMP	2'-cytidine monophosphate
2-D	Two-dimensional
3-D	Three-dimensional
5-HT	5-Hydroxytryptamine
ACE	Angiotensin converting enzyme
ADME	Absorption, distribution, metabolism, elimination
ADPNP	5'-adenylyl β - γ -imidodiphosphate
Ahx	Aminohexanoic acid
AMP	Adenosine monophosphate
ApoD	Apolipoprotein D
AT	Angiotensin
ATP	Adenosine triphosphate
BBP	Bilin-binding protein
BCUT	Burden chemical abstract service University of Texas
BHK	Baby hamster kidney cells
B_{\max}	Maximal specific binding
Bpa	<i>p</i> -Benzoylphenylalanine
BSA	Bovine serum albumin
C(alpha)	Alpha carbon group of amino acid
cal	Calorie
CATS	Chemically advanced template search
CCD	Charge Coupled Device
CCDC	Cambridge Crystallographic Data Center
CCK	Cholecystokinin
CDK2	Cyclin-dependent kinase 2
CGRP	Calcitonin gene related peptide
CHO	Chinese hamster ovary cells
CMC	Comprehensive Medicinal Chemistry
CoMFA	Comparative molecular field analysis
COS	SV40 transformed African green monkey kidney cells
C_p	Heat capacity (constant pressure)
CYP3A4	Cytochrome P450 3A4
ΔG	Change in free energy

ΔH	Change in enthalpy
ΔS	Change in entropy
ΔX	Change in X
Da	Dalton
deg	Degree
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
Dpm	Decays per minute
DSC	Differential scanning calorimetry
E	Energy
E_a	Energy of association
E_d	Energy of dissociation
EDN	Eosinophil-derived neurotoxin
EDTA	Ethylenediaminetetraacetic acid
ELISA	Enzyme linked immunosorbent assay
ESI-MS	Electron spray ionization mass spectrometry
F_{ab}	Antigen-binding fragment
FCS	Fluorescence correlation spectroscopy
FEB	Free energy perturbation
FKBP	FK506 binding protein
FRET	Fluorescence resonance energy transfer
G	Gibbs free energy
GA	Genetic algorithm
GaP	Gridding and partitioning
GDP	Guanosine diphosphate
GFP	Green fluorescent protein
GH-Score	Goodness-of-hit score
GPCR	G-protein coupled receptor
GRIND	Grid independent descriptors
GTP	Guanosine triphosphate
H	Enthalpy
HDL	High density lipoprotein
HEK	Human embryonic kidney cells
HIV	Human immunodeficiency virus
HIV-RT	HIV reverse transcriptase
hNGAL	Human neutrophil gelatinase-associated lipocalin
HTS	High-throughput screening
IC_{50}	Ligand concentration that causes 50% inhibition
Ig	Immunoglobulin
ITC	Isothermal titration calorimetry
IUPAC	International Union of Pure and Applied Chemistry
J	Joule
K	Association constant
K	Kelvin (measure of absolute temperature; $^{\circ}C + 273.15$)
k_{12}	Association rate (on rate)

k_{21}	Dissociation rate (off rate)
K_d	Dissociation constant
K_{eq}	Equilibrium constant
K_i	Inhibition constant
kJ	Kilojoules
KLH	Keyhole limpet hemocyanin
K_M	Michaelis constant
L	Ligand
L^*	Labeled ligand
LC-MS	Liquid chromatography coupled mass spectrometry
M	mol L^{-1}
MACC	Maximum auto-cross correlation
MALDI-TOF-MS	Matrix assisted laser desorption ionization – time of flight – mass spectrometry
MDDR	MDL Drug Data Report
MDL	Molecular Design Limited
MO	Molecular orbital
MS	Mass spectrometry
MW	Molecular weight
NCI	National Cancer Institute
NK	Neurokinin
NMR	Nuclear magnetic resonance
NOE	Nuclear Overhauser effect
NPY	Neuropeptide Y
OppA	Oligopeptide binding protein A
OSPReY	Orientated substituent pharmacophore PPropErtY space
OWFEG	One window free energy grid
OX	Orexin receptor
P	Pressure
P	Protein
PCA	Principal components analysis
PCR	Polymerase chain reaction
P-gp	P-glycoprotein
pI	Isoelectric Point
PL	Protein-ligand complex
PLS	Partial least squares projection to latent structures
PPACK	D-Phe-Pro-Arg-chloromethylketone
PVDF	Polyvinylidene fluoride
PXR	Pregnane X receptor
pY	Phosphotyrosine
Q	Heat
R	Gas constant ($1.99 \text{ cal mol}^{-1} \text{ deg}^{-1}$; $8.31 \text{ J mol}^{-1} \text{ deg}^{-1}$)
R	Inactive conformation of a G-Protein coupled receptor
R^*	Active conformation of a G-Protein coupled receptor
RBP	Retinol-binding protein

Rh-GAL	Rhodamine-labeled galanin
RI	Ribonuclease inhibitor
RNA	Ribonucleic acid
RNase	Ribonuclease
RSM	Receptor surface model
R_t	Total receptor concentration
RU	Resonance units
S	Entropy
SAM	Self-assembled monolayer
SAR	Structure-activity relationship
SDS	Sodium dodecylsulfate
SDS-PAGE	Sodium dodecylsulfate polyacrylamide gel electrophoresis
SELEX	Systematic evolution of ligands by exponential enrichment
SH2	Src homology 2
SLN	SYBYL line notation
SMILES	Simplified molecular input line entry system
SP	Substance P
SPR	Surface plasmon resonance
T	Temperature
TAR	Transactivation response element
TM	Transmembrane domain
Tmd(Phe)	<i>p</i> -(3-Trifluoromethyl)diazirinophenylalanine
U	Energy
V	Volume
V_H	Variable domain of the heavy chain
V_L	Variable domain of the light chain
W	Watt
W	Work
WDI	World Drug Index
Z	Partition function
z	Charge

Prologue

D. BROWN

Understanding protein-ligand interactions is central to drug design and the discovery of new medicines to benefit human health. It remains true that very few drugs have been designed *de novo*, and this suggests that our level of understanding of protein-ligand interactions remains relatively rudimentary. Why is this? Many protein targets for drugs are embedded in membranes in the form of GPCRs or ion channels, and the difficulty of achieving crystallization of membrane proteins has limited progress in gaining insight into the 3-D structure of these protein targets. And, while we do have 3-D structural data for many soluble protein targets such as enzymes, protein-ligand interaction is always a dynamic process and this has hindered development of a full understanding. In addition, technical barriers have historically limited the rate at which protein-ligand interactions can be studied by methods such as X-ray or NMR spectroscopy.

Recent years have seen a significant change in this situation. During the 1990s, improved methods were devised for protein NMR and X-ray, and, in particular, the number of solved protein X-ray structures increased rapidly. In addition, there were rapid advances in development of 3-D structure prediction methods based on homology modeling of protein folds. We can now expect an even more dramatic rate of progress, particularly in throughput of protein X-ray, because of the implementation of high throughput methods for protein production, crystallization, and structure determination. In the “post-genome” era, focus is turning to the expressed products of the genome, the “proteome.” It is through understanding the function of expressed proteins that drug targets can be selected, and it is through understanding the structures and ligand-binding properties of target proteins that drugs can be designed.

Until quite recently in the drug discovery process, an understanding of protein-ligand interactions was necessary mainly for optimization of leads and, to a more limited extent, for lead identification. Methodologies for molecular recognition are now being used both upstream and downstream in drug discovery. The proteomics revolution is providing the foundation for a new branch of science known as “chemical genomics” (perhaps “chemical proteomics” would be a more appropriate title). The key concept is classification of families of proteins by structure and/or function and correlation with known chemical ligands. This classification can be used predictively to find new ligands for related proteins. Also, key concepts

from molecular recognition studies are driving development of pharmacophore-based descriptors (to move away from a chemistry-biased representation), which provides methods to identify new ligand templates (“scaffold-hopping”). In another key development towards the discovery of new bio-active ligands, virtual screening (*in silico*) has made rapid advances to the extent that screening of virtual libraries of 10^6 – 10^9 molecules will soon be routine in the pharmaceutical and biotechnology industries. In a further development in lead identification, pharmaceutical and biotechnology companies are building compound libraries for “focused” screening based on target class families in an attempt to increase success rates in finding leads by screening. Knowledge of molecular recognition principles is central to this approach, which is a sub-strategy of the chemical genomics approach. Computational approaches to *de novo* ligand design are also now becoming practicable, although current methods generally fail to take chemical accessibility into account. Molecular recognition is also becoming important in activities that have traditionally been “downstream” in the drug discovery process, such as ADME (absorption, distribution, metabolism, excretion). Much of the challenge in the lead optimization process is to attain a molecule with pharmacokinetic properties suitable for use in *in vivo* animal and clinical studies. Drug clearance mechanisms have received much study over the past two decades, and now many of the key determinants of drug clearance are well understood. Cytochrome P450 interactions are central to this process, and the recent availability of 3-D X-ray structures of some key P450s offers the opportunity for a more detailed understanding of the key determinants of ligand interactions with these proteins.

One area where molecular recognition has made a relatively limited impact so far is in toxicology. A significant percentage of potential drugs are lost during either late lead optimization or early in the development phase because of unacceptable toxicity. The observed toxicity is likely to be governed by specific protein-ligand interactions, but our ability to predict potential liabilities remains low.

In summary, we are seeing rapid advances in our understanding of molecular recognition, and, indeed, molecular recognition itself is now recognized as a branch of science. For these reasons, this volume of studies in “Molecular Recognition in Protein Ligand Interactions” is particularly timely. The authors are all world-renowned experts in their area of study, and they offer clear and comprehensive overviews of the state of the art in molecular recognition.

1

Prediction of Non-bonded Interactions in Drug Design

H.-J. BÖHM

1.1

Introduction

The discovery of novel drugs to treat important diseases is still a major challenge in pharmaceutical research. Structure-based design plays an increasingly important role in this endeavor and is now an integral part of medicinal chemistry. It has been shown for a large number of targets that the 3-D structure of the protein can be used to design small molecules binding tightly to the protein. Indeed, several marketed drugs can be attributed to a successful structure-based design [1–4]. Several reviews summarize the recent progress [5–9]. A key to success and further progress in this field is a detailed understanding of the protein-ligand interactions. The purpose of the present contribution is to provide a short introduction into some of the underlying concepts and then to discuss some recent methods that are currently used to predict protein-ligand interactions. Chapter 1.2 will provide a brief introduction to some key features of non-bonded protein-ligand interactions, and Chapter 1.3 summarizes the presently used scoring functions to predict ligand-binding affinity. This is followed by a description of how these scoring functions are currently used in drug discovery. Finally, some applications will highlight that despite their limitations the available methods already prove to be useful.

The vast majority of the currently available drugs act via non-covalent interaction with the target protein. Therefore, non-bonded interactions are of particular interest in drug design. In view of the continuous exponential growth of the number of solved relevant 3-D protein structures, there is an increasing interest in computational methods to predict protein-drug interactions. The goal is to develop a rapid method that could predict the bound conformation of a small molecule and the binding affinity. Having such a robust and reliable method in hand, it is possible to steer synthetic efforts more effectively towards the most promising compounds and then focus the experimental optimization towards other challenging properties such as bioavailability and toxicity.

1.2

Major Contributions to Protein-Ligand Interactions

The selective binding of a low-molecular-weight ligand to a specific protein is determined by the structural and energetic recognition of a ligand and a protein. The binding affinity can be determined from the experimentally measured binding constant K_i (Eq. 1.1):

$$\Delta G = -RT \ln K_i = \Delta H - T\Delta S \quad (\text{Eq. 1.1})$$

The experimentally determined binding constant K_i is typically in the range of 10^{-2} to 10^{-12} M, corresponding to a Gibbs free energy of binding ΔG between -10 and -70 kJ/mol in aqueous solution [6, 9].

There is now a large body of experimental data available on 3-D structures of protein-ligand complexes and binding affinities. These data clearly indicate that there are several features found basically in all complexes of tightly binding ligands:

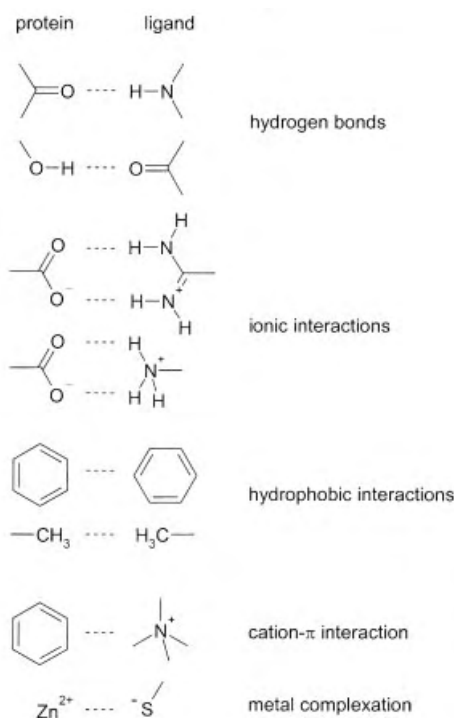
1. There is a high level of steric complementarity between the protein and the ligand. This observation is also described as the lock-and-key paradigm.
2. There is usually high complementarity of the surface properties between the protein and the ligand. Lipophilic parts of the ligands are most frequently found to be in contact with lipophilic parts of the protein. Polar groups are usually paired with suitable polar protein groups to form hydrogen bonds or ionic interactions. The experimentally determined hydrogen bond geometries display a fairly small scatter – in other words, the hydrogen bond geometry is strongly preserved. With very few exceptions, there are no repulsive interactions between the ligand and the protein.
3. The ligand usually binds in an energetically favorable conformation.

Generally speaking, direct interactions between the protein and the ligand are very important for binding. The most important direct interactions are highlighted in Fig. 1.1. Structural data on unfavorable protein-ligand interactions are sparser, partly because structures of weakly binding ligands are more difficult to obtain and are usually considered less interesting by many structural biologists. However, these data are vital for the development of scoring functions. Some conclusions can be drawn from the available data: unpaired buried polar groups at the protein-ligand interface are strongly adverse to binding. Few buried CO and NH groups in folded proteins fail to form hydrogen bonds [10]. Therefore, in the ligand design process one has to ensure that polar functional groups, either of the protein or the ligand, will find suitable counterparts if they become buried upon ligand binding. Another situation that leads to a decreased binding affinity is imperfect steric fit, leading to holes at the lipophilic part of the protein-ligand interface.

The enthalpic and entropic components of the binding affinity can be determined experimentally, e.g., by isothermal titration calorimetry (ITC). Unfortun-

Fig. 1.1 Typical non-bonded interactions found in protein-ligand complexes.

Usually, the lipophilic part of the ligand is in contact with the lipophilic parts of the protein (side chains of the amino acids Ile, Val, Leu, Phe, and Trp, perpendicular contact to amide bonds). In addition, several hydrogen bonds are formed. Some of them can be charge assisted. Cation- π interactions and metal complexation can also play a significant role in individual cases.



nately, these data are still sparse and are difficult to interpret [9]. The available data indicate that there is always a substantial compensation between enthalpic and entropic contributions [11–13]. The data also show that the binding may be enthalpy-driven (e.g., streptavidin-biotin, $\Delta G = -76.5$ kJ/mol, $\Delta H = -134$ kJ/mol) or entropy-driven (e.g., streptavidin-HABA, $\Delta G = -22.0$ kJ/mol, $\Delta H = 7.1$ kJ/mol) [14].

Data from protein mutants yield estimates of 5 ± 2.5 kJ/mol for the contribution from individual hydrogen bonds to the binding affinity [15–17]. Similar values have been obtained for the contribution of an intramolecular hydrogen bond to protein stability [18–20]. The consistency of values derived from different proteins suggests some degree of additivity in the hydrogen bonding interactions.

The biggest challenge in the quantitative treatment of protein-ligand interactions is still an accurate description of the role of water molecules. In particular, the contribution of hydrogen bonds to the binding affinity strongly depends on solvation and desolvation effects (Fig. 1.2). It has been shown by comparing the binding affinities of ligand pairs differing by just one hydrogen bond that the contribution of an individual hydrogen bond to the binding affinity can sometimes be very small or even adverse to binding [21]. Charge-assisted hydrogen bonds are stronger than neutral ones, but this is paid for by higher desolvation penalties. The electrostatic interaction of an exposed salt bridge is worth as much as a neu-

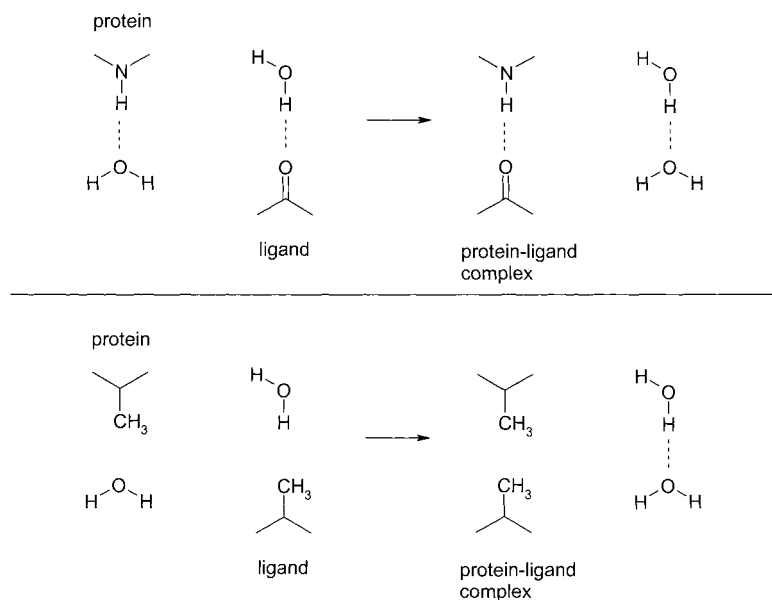


Fig. 1.2 Role of water molecules in hydrogen bonds (upper part) and lipophilic interactions (lower part). In the unbound state (left side), the polar groups of the ligand and the protein form hydrogen bonds to water molecules. These water molecules are replaced upon complex formation. The hydrogen bond inven-

tory (total number of hydrogen bonds) does not change. In contrast, the formation of lipophilic contact increases the total number of hydrogen bonds due to the release of water molecules from the unfavorable lipophilic environment.

tral hydrogen bond (5 ± 1 kJ/mol according to [22]), while the same interaction in the interior of a protein can be significantly larger [23].

Lipophilic interactions are essentially contacts between apolar parts of the protein and the ligand. The generally accepted view is that lipophilic interactions are mainly due to the replacement and release of ordered water molecules and are therefore entropy-driven [24, 25]. The entropy gain results when the water molecules are no longer positionally confined. There are also enthalpic contributions to lipophilic interactions. Water molecules occupying lipophilic binding sites are unable to form hydrogen bonds with the protein. If they are released, they can form strong hydrogen bonds with bulk water. It has been shown in many cases that the contribution to the binding affinity is proportional to the lipophilic surface area buried from solvent with values in the range of $80\text{--}200$ J/(mol \AA^2) [26, 27].

Many protein-ligand complexes are characterized by the presence of both polar and lipophilic interactions. The bound conformation of the ligand is determined by the relative importance of these contributions. An interesting example highlighting several important aspects was recently described by Lange and co-workers using the binding of non-peptidic inhibitors to the SH2 domain of src kinase [28]. The inhibitors are essentially tetrapeptide mimetics with tyrosine-phosphate or a

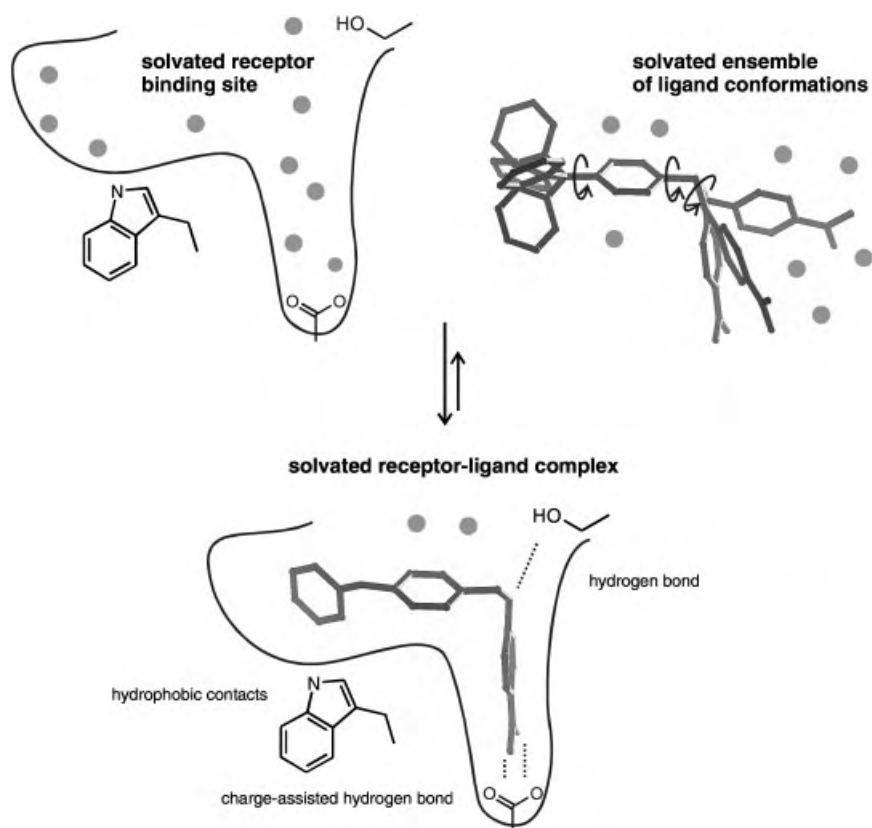


Fig. 1.3 Overview of the receptor-ligand-binding process. All species involved are solvated by water (symbolized by gray spheres). The binding free energy difference between the bound and unbound state is a sum of enthalpic components (breaking and formation of

hydrogen bonds, formation of specific hydrophobic contacts) and entropic components (release of water from hydrophobic surfaces to solvent, loss of conformational mobility of receptor and ligand).

tyrosine-phosphate mimic at one end and a lipophilic group at the other end. As is evident from 11 reported structures of the src SH2 domain with different inhibitors bound, the bound conformation always aims to maximize the interaction between the lipophilic substituent and the lipophilic binding pocket. This is achieved either by an alternative binding mode of the polar end of the inhibitor or by including water molecules that mediate hydrogen bonds between the inhibitor and the protein.

In spite of many inconsistencies and difficulties in interpretation, most of the experimental data suggest that simple additive models for the protein-ligand interactions might be a good starting point for the development of empirical scoring functions. Indeed, the first scoring functions actually built upon experimental work published in 1994 by Böhm [29].

Fig. 1.3 is an attempt to summarize the various interactions that play a role in receptor-ligand binding. It is a complex equilibrium between ensembles of solvated species. In the next section, we will discuss various approaches to capture essential elements of this equilibrium in computationally efficient scoring functions. The discussion focuses on general approaches rather than individual functions.

1.3

Description of Scoring Functions for Receptor-Ligand Interactions

The rigorous theoretical treatment of reversible receptor-ligand binding is difficult and requires full consideration of all species involved in the binding equilibrium. In the unbound state, both the ligand and the receptor are separately solvated and do not interact. In the bound state, both partners are partially desolvated and form interactions with each other. Since it is the free energy of binding one is interested in, the energies of the solvated receptor, the solvated ligand, and the solvated complex have to be calculated as ensemble averages. Their accurate statistical mechanics treatment has been reviewed elsewhere [30] and is not the topic of this review. Large-scale Monte Carlo or Molecular Dynamics simulations are necessary to arrive at reasonably accurate values of binding free energies. These methods are suitable for only small sets of compounds, since they require large computational resources, and even the most advanced techniques are reliable only for calculating binding free energy differences between closely related ligands [31–33]. However, a number of less rigorous but faster scoring schemes have been developed, which should be amenable to larger numbers of ligands. For example, recent experience has shown that continuum solvation models can replace explicit solvent molecules at least in the final energy evaluation of the simulation trajectory [34]. Another less expensive alternative is the use of linear response theory [35, 36] in conjunction with a surface term [37].

Scoring functions that can be evaluated quickly enough to be applied in docking and virtual screening applications can be only very crude measures of the free energy of binding. They usually take into account only one receptor-ligand complex structure and disregard ensemble averaging and properties of the unbound state of the binding partners. Furthermore, all methods have in common that the free energy is decomposed into a sum of terms. In a strict physical sense, this is not allowed, since the free energy of binding is a state function but its components are not [38]. In addition, simple additive models cannot describe subtle cooperativity effects [39]. Nevertheless, it is often useful to interpret receptor-ligand binding in an additive fashion [40–42], and estimates of binding free energy are in this way available at very low computational cost. Fast scoring functions can be categorized into three main classes, i.e., force field-based methods, empirical scoring functions, and knowledge-based methods, and will be discussed here in this order.

1.3.1

Force Field-based Methods

An obvious idea to circumvent parameterization efforts for scoring is to use non-bonded energies of existing, well-established molecular mechanics force fields for the estimation of binding affinity. In doing so, one substitutes estimates of the free energy of binding in solution with an estimate of the gas-phase enthalpy of binding. Even this crude approximation can lead to satisfying results. A good correlation was obtained between non-bonded interaction energies calculated with a modified MM2 force field and IC₅₀ values of 33 HIV-1 protease inhibitors [43]. Similar results were reported in a study of 32 thrombin-inhibitor complexes with the CHARMM force field [44]. In both studies, however, experimental data represented rather narrow activity ranges and little structural variation.

A very recent addition to the list of force field-based scoring methods has been developed by Charifson and Pearlman. This so-called OWFEG (one window free energy grid) method [45] is an approximation to the expensive first-principles method of free energy perturbation (FEP). For the purpose of scoring, a molecular dynamics simulation is carried out with the ligand-free, solvated receptor site. During the simulation, the energetic effects of probe atoms on a regular grid are collected and averaged. Three simulations are run with three different probes: a neutral methyl-like atom, a negatively charged atom, and a positively charged atom. The resulting three grids contain information on the score contributions of neutral, positively charged, and negatively charged ligand atoms located in various positions of the receptor site and can thus be used in a very straightforward manner for scoring. This approach seems to be successful for K_i prediction as well as virtual screening applications [46]. Its conceptual advantage is the implicit consideration of entropic and solvent effects and some protein flexibility.

1.3.2

Empirical Scoring Functions

The underlying idea of empirical scoring functions is that the binding free energy of a non-covalent receptor-ligand complex can be interpreted as a sum of localized, chemically intuitive interactions. Such decompositions can be a useful tool to gain an understanding of binding phenomena even without analyzing 3-D structures of receptor-ligand complexes. Andrews and colleagues calculated average functional group contributions to binding free energy from a set of 200 compounds whose affinity to a receptor was experimentally known [40]. The average functional group contributions can be used to estimate a receptor-independent binding energy for a compound that can be compared to experimental values. If the experimental value is approximately the same as or higher than the calculated value, there is a good fit between receptor and ligand, and essentially all functional groups of the ligand are involved in protein interactions. If it is significantly lower, the compound does not fully utilize its potential to form interactions. Similarly, experimental binding affinities have been analyzed on a per-atom

basis in quest of the maximal binding affinity of non-covalent ligands [47]. It was concluded that in the strongest binding ligands, each non-hydrogen atom on average contributes 1.5 kcal/mol to the binding energy.

With 3-D structures of receptor-ligand complexes at hand, the analysis of binding phenomena can of course be much more detailed. The binding affinity $\Delta G_{\text{binding}}$ can be estimated as a sum of interactions multiplied by weighting coefficients ΔG_i :

$$\Delta G_{\text{binding}} \approx \sum \Delta G_i f_i(r_l, r_p), \quad (\text{Eq. 1.2})$$

where each f_i is a function of the ligand coordinates r_l and the protein coordinates r_p . Scoring schemes that use this concept are called “empirical scoring functions.” Several reviews summarize details of individual parameterizations [48–51]. The individual terms in empirical scoring functions are usually chosen such that they intuitively cover important contributions of the total binding free energy. Most empirical scoring functions are derived by evaluating the functions f_i on a set of protein-ligand complexes and fitting the coefficients ΔG_i to experimental binding affinities of these complexes by multiple linear regression or supervised learning. The relative weight of the individual contributions depends on the training set. Usually, between 50 and 100 complexes are used to derive the weighting factors.

Empirical scoring functions usually contain individual terms for hydrogen bonds, ionic interactions, hydrophobic interactions, and binding entropy. Hydrogen bonds are often scored by simply counting the number of donor-acceptor pairs that fall in a given distance and angle range favorable for hydrogen bonding, weighted by penalty functions for deviations from preset ideal values [29, 52]. The amount of error tolerance in these penalty functions is critical. When large deviations from ideality are tolerated, the scoring function cannot sufficiently discriminate between different orientations of a ligand, whereas small tolerances lead to situations where many structurally similar complex structures obtain very different scores. Attempts have been made to reduce the localized nature of such interaction terms by using continuous modulating functions on an atom-pair basis [53]. Other workers have avoided the use of penalty functions and introduced separate regression coefficients for strong, medium, and weak hydrogen bonds [54]. The Agouron group has used a simple four-parameter potential that is a piecewise linear approximation of a potential well without angular terms (“PLP scoring function”) [55]. Most functions treat all types of hydrogen bond interactions equally. Some attempts have been made to distinguish between different donor-acceptor functional group pairs. Hydrogen bond scoring in the docking program GOLD [56, 57] is based on a list of hydrogen bond energies for all combinations of 12 donor and 6 acceptor atom types derived from *ab initio* calculations of model systems incorporating these atom types.

Hydrophobic interactions are usually estimated by the size of the contact surface at the receptor-ligand interface. Often, a reasonable correlation between experimental binding energies can be achieved with a surface term alone [58–60]. The weighting factor ΔG_i of the hydrophobic term depends strongly on the train-

ing set. It might have been underestimated in most derivations of empirical scoring functions [61] because most training sets contain an overly large proportion of ligands with many donor and acceptor groups (many peptide and carbohydrate fragments).

1.3.3

Knowledge-based Methods

Empirical scoring functions “see” only those interactions that are part of the model. Many less common interactions are usually disregarded, even though they can be strong and specific, e.g., NH- π hydrogen bonds. It would be a difficult task to generate a comprehensive and consistent description of all these interactions in the framework of empirical scoring functions. But there exists a quickly growing body of structural data on receptor-ligand complexes that can be used to detect favorable binding geometries. “Knowledge-based” scoring functions try to capture the knowledge about receptor-ligand binding that is hidden in the protein data bank by means of statistical analysis of structural data alone – without referring to often inconsistent experimentally determined binding affinities [62]. They have their foundation in the inverse formulation of the Boltzmann law:

$$E_{ij} = -kT \ln(p_{ijk}) + kT \ln(Z) , \quad (\text{Eq. 1.3})$$

where the energy function E_{ij} is called a potential of mean force for a state defined by the variables i , j , and k ; p_{ijk} is the corresponding probability density, and Z is the partition function. The second term of the sum is constant at constant temperature T and does not have to be regarded, since $Z=1$ can be chosen by definition of a suitable reference state leading to normalized probability densities p_{ijk} . The inverse Boltzmann technique has been applied to derive potentials for protein folding from databases of protein structures [63]. For the purpose of deriving scoring functions, the variables i , j , and k can be chosen to be protein atom types, ligand atom types, and their inter-atom distance. The frequency of occurrence of individual contacts is a measure of their energetic contribution to binding. When a specific contact occurs more frequently than should be expected from a random or average distribution, this is indicative of an attractive interaction. When it occurs less frequently, one can interpret this as a repulsive interaction between two atom types. The frequencies can thus be converted to sets of atom-pair potentials that are straightforward to evaluate. The PMF function by Muegge and Martin [64] and the DrugScore function by Gohlke et al. [65] belong to this category.

1.4

Some Limitations of Current Scoring Functions

1.4.1

Influence of the Training Data

All fast scoring functions share a number of deficiencies that one should be aware of in any application. First, most scoring functions are in some way fitted to or derived from experimental data. The functions necessarily reflect the accuracy of the data that were used in their derivation. For instance, a general problem with empirical scoring functions is the fact that the experimental binding energies necessarily stem from many different sources and therefore form inconsistent datasets containing systematic experimental errors. Furthermore, scoring functions reflect not only the quality but also the type of experimental data they are based on. Most scoring functions are still derived from data on mostly high-affinity receptor-ligand complexes. Many of these are still peptidic in nature, whereas interesting lead molecules in pharmaceutical research are usually non-peptidic. This is reflected in the relatively high contributions of hydrogen bonds in the total score. The balance between hydrogen bonding and hydrophobic interactions is a very critical issue in scoring, in particular for non-peptidic, drug-like ligands, and its consequences are especially obvious in virtual screening applications, as will be illustrated in Section 1.4.3.

A possible approach to increase the accuracy of scoring functions is to divide up the set of known inhibitors into clusters of structurally related compounds and then derive an individual scoring function for each of the compound sets. Clearly, the application range of such a scoring function is limited to the particular chemotype. However, in practice, industrial pharmaceutical research often focused over a fairly long period on one particular set of compounds, and it may be favorable to work with a scoring function that works only for this particular set of compounds but has a higher accuracy than a general scoring function. A nice example for this approach was recently provided by Rizzo et al. [66] for HIV reverse transcriptase inhibitors using binding data of more than 200 non-nucleoside HIV RT inhibitors representing 8 chemotypes. The average error in the predicted binding energies is 0.50 kcal/mol if an individual scoring function is derived for each of the eight sets. If one single scoring function is fitted to the full dataset, the average error is 0.86 kcal/mol.

Another possibility to increase the accuracy of docking calculation is to take into account information about important characteristics of protein-ligand binding modes as demonstrated recently by Hindle et al. using the docking tool FlexX [67]. For example, when dealing with metalloproteases, the assumption that the ligand must directly interact with the metal ion in the active site improves the accuracy of the docking calculation and also significantly increases the speed.

1.4.2

Molecular Size

The simple additive nature of most fast-scoring functions often leads to large molecules obtaining high scores. While it is true that small molecules with a molecular weight below 200–250 are rarely of very high affinity, there is of course no guarantee that larger compounds are more active. When it comes to comparing scores of two compounds of different size, it therefore makes sense to include a penalty term that diminishes the dependence of the score on molecular size. In some applications, a constant penalty value has been added to the score for each heavy atom or a penalty term proportional to the molecular weight has been used [68]. The scoring function of the docking program FLOG, which contains force field and empirical terms, has been normalized to remove the linear dependence of the crude score from the number of ligand atoms that was found in a docking study of a 7500-compound database [69].

1.4.3

Water Structure and Protonation State

Insecurities about protonation states and water structure at the receptor-ligand interface also make scoring difficult. These effects play a role in the derivation as well as in the application of scoring functions. The entropic and energetic contributions of the reorganization of water molecules upon ligand binding are very difficult to predict (see, e.g., [70]). The only reasonable approach to this problem is to concentrate on conserved water molecules and make them part of the receptor. For example, the docking program FLOG has been applied to the search of inhibitors for a metallo- β -lactamase [71] within the Merck in-house database. Docking was performed with three different configurations of bound water in the active site. The top-scoring compounds showed an enrichment in biphenyl tetrazoles, several of which were found to be active below 20 μM . A crystal structure of one tetrazole ($\text{IC}_{50}=1.9 \mu\text{M}$) not only confirmed the predicted binding mode of one of the inhibitors but also displayed the water configuration that had – retrospectively – been the most predictive one of the three models.

Scoring functions rely on a fixed assignment of a general atom type to each protein and ligand atom. This also implies a fixed assignment of a protonation state to each acidic and basic functional group. Even though these estimates can be reliable enough for conditions in aqueous solution, significant pK_a shifts can be witnessed upon ligand binding [72]. This finding can be attributed to local changes of the dielectric conditions inside the binding pocket.

1.5

Application of Scoring Functions in Virtual Screening and *De Novo* Design

In recent years, virtual screening of large databases has emerged as the central application of scoring functions. In the following sections we will outline the special requirements scoring functions have to fulfill for successful virtual screening and indicate the level of accuracy that can nowadays be expected from virtual screening.

As discussed in the introductory sections, the goal of virtual screening is to use computational tools together with the known 3-D structure of the target to select a subset from chemical libraries for synthesis and biological testing. This subset typically consists of some 100–2000 compounds selected from libraries of 100,000–500,000 compounds. Therefore, it is essential that the computational process, including the scoring function, is fast enough to handle several thousand compounds in a short period of time. As a consequence, only the fastest scoring functions are currently used for this purpose. This is especially true for those scoring functions that are used as objective functions during the docking calculations, since they are evaluated several hundred to several thousand times during the docking process of a single compound.

After a successful virtual screening run, the selected subset of compounds contains a significantly enhanced number of active compounds as compared to a random selection. A key parameter to measure the performance of docking and scoring methods is therefore the so-called enrichment factor. It is simply the ratio of active compounds in the subset selected by docking divided by the number of active compounds in a randomly chosen subset of equal size. In practice, enrichment factors are far from the ideal case where all active compounds would be placed on the top ranks of a prioritized list. Insufficiencies of current scoring functions, as discussed in the previous section, are partly responsible for moderate enrichment rates. Another major cause is the fact that the receptor is still treated as a rigid object. To generate correct binding modes of different molecules, it would be necessary to predict induced fit phenomena. However, predicting protein flexibility is extremely difficult and computationally expensive and therefore is not taken into account in many applications.

1.5.1

Successful Identification of Novel Leads Through Virtual Screening

A respectable number of publications have shown that virtual screening is an efficient way of finding novel leads. The program DOCK, one of the most widely used docking programs, has been applied in many published studies [73–78]. Usually, the DOCK AMBER force field score is applied. The docking program SANDOCK [79] comprises an empirical scoring function that evaluates steric complementarity, hydrophobic contacts, and hydrogen bonding. SANDOCK has been used to find a variety of novel FKBP inhibitors [80].

Docking routines in the program packages DOCK and ICM [81] have been used in two published studies to identify novel nuclear hormone receptor antagonists [82] and for an RNA target, the transactivation response element (TAR) of HIV-1 [83].

A very recent study by Grueneberg et al. resulted in subnanomolar inhibitors of carbonic anhydrase II [84]. The study is a textbook example of virtual screening focusing on successively smaller subsets of the initial database in several steps and employing different methods at each step. Carbonic anhydrase II (CAII) is a metalloenzyme that catalyzes the reversible hydration of CO_2 to HCO_3^- [85]. In the human eye, an isoform of the enzyme is involved in water removal. CAII inhibitors can thus be used to reduce intraocular pressure in the treatment of glaucoma. The top-ranking 13 hits were chosen for experimental testing. Nine of these compounds showed activities below $1\text{ }\mu\text{M}$, while the sulfonamides **9** and **10** (Fig. 1.4) have K_i values below 1 nM .

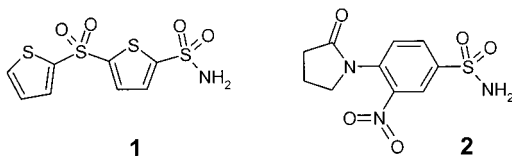
1.5.2

De novo Ligand Design with LUDI

LUDI is a fragment-based *de novo* design computer program developed by Böhm [86, 87]. The software constructs novel protein ligands by joining molecular fragments. In a first step, the program calculates interaction sites, which are discrete positions in the protein-binding site suitable to form hydrogen bonds or to fill a hydrophobic pocket. The interaction sites are derived from a statistical analysis of non-bonded contacts found in crystal packings of small organic molecules. The second step is the fit of molecular fragments onto the interaction sites. The software can fit fragments into the binding site independent of each other, but it also can append new fragments onto an already positioned fragment or lead compound, thus generating novel compounds. The final step is the scoring of the generated protein-ligand complex.

LUDI is commercially available and is the most widely used software for *de novo* design [88]. A large number of prospective applications have been reported where LUDI was used to design or select a compound that was then tested afterward and found to be active. Examples (see Fig. 1.5) are the design of the thrombin inhibitor **3** available from a one-step chemical reaction [89]; design of a novel class, a gyrase inhibitor, exemplified by **4** [90]; and the discovery of the novel inhibitors **5** for tRNA-guanin-transglycosylase [91] and **6** for FKBP-12 [92].

Fig. 1.4 Inhibitors of carbonic anhydrase II. Compounds **1** and **2** are subnanomolar inhibitors identified through virtual screening.



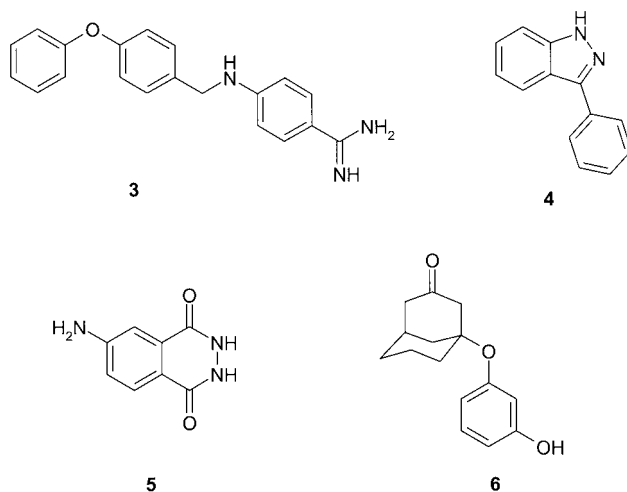


Fig. 1.5 Novel protein ligands discovered using the computer program LUDI.

1.6 Outlook

The first scoring functions were published about 10 years ago. Since then, we have gained much experience in using scoring functions and assessing their accuracy. Significant progress has been made over the last few years, and it appears as if there are now scoring functions available that can be applied to a wide range of different proteins and consistently yield considerable enrichment of active compounds. As a consequence, many small and large pharmaceutical companies are increasingly using virtual screening techniques to identify possible leads. In fact, structure-based design is now seen as a very important approach to drug discovery that nicely complements HTS [93]. HTS has a number of serious disadvantages: it is expensive [94], and it leads to many false positives and few real leads [95, 96]. Not all tests are amenable to HTS techniques. Finally, despite the large size of the chemical libraries available to the pharmaceutical industry, it is far from possible to cover the whole space of drug-like organic molecules. This means that the focused design of novel compounds and compound libraries will gain importance. Given the current aggressive patenting strategies, one may speculate that *de novo* design will become much more important in the near future. Thus, there is every reason to believe that the value of structure-based approaches will continue to grow. The development of improved scoring functions is certainly vital for their success. We would therefore like to inform the reader of what in our eyes are the major challenges in the further development of scoring functions:

1. Polar interactions are still not treated adequately. It is somewhat strange that while the role of hydrogen bonds in biology has been well known for a long time and hydrogen bonds are qualitatively well understood, a quantitative treatment of hydrogen bonds in protein-ligand interactions is still missing.

2. All scoring functions are essentially simple analytical functions fitted to experimental binding data. Presently, there is still a heavy bias in the public domain data towards peptidic ligands. This in turn leads to an overestimation of polar interactions in many scoring functions. The development of better scoring function clearly requires access to more data on non-peptidic, low-molecular-weight, drug-like ligands.
3. Unfavorable interactions and unlikely docking solutions are not penalized strongly enough. Methods for taking account of undesired features of complex structures in the derivation of scoring functions are still lacking.
4. So far, fast scoring functions only cover part of the whole receptor-ligand binding process. A more detailed picture could be obtained by taking into account properties of the unbound ligand, i.e., solvation effects and energetic differences between the low-energy solution conformations and the bound conformation.

1.7

Acknowledgments

The author would like to thank Martin Stahl for his significant contributions to the manuscript.

1.8

References

- 1 J. GREER, J. W. ERICKSON and J. J. BALDWIN, *J. Med. Chem.*, **1994**, 37, 1035.
- 2 M. VON ITZSTEIN, W.-Y. WU, G. B. KOK, M. S. PEGG, J. C. DYASON, B. JIN, T. V. PHAN, M. L. SMYTHE, H. F. WHITE, S. W. OLIVER, P. M. COLMANT, J. N. VARGHESE, D. M. RYAN, J. M. WOODS, R. C. BETHELL, V. J. HOTHAM, J. M. CAMERON and C. R. PENN, *Nature*, **1993**, 363, 418.
- 3 W. LEW, C. X. and U. CHOUNG, *Curr. Med. Chem.*, **2000**, 7, 663.
- 4 S. W. KALDOR, V. J. KALISH, J. F. DAVIES, V. S. BHASKER, J. E. FRITZ, K. APPELT, J. A. BURGESS, K. M. CAMPANALE, N. Y. CHIRGADZE, D. K. CLAWSON, B. A. DRESSMAN, S. D. HATCH, D. A. KHALIL, M. B. KOSA, P. P. LUBBEHUSEN, M. A. MUESING, A. K. PATICK, S. H. REICH, K. S. SU and J. H. TATLOCK, *J. Med. Chem.*, **1997**, 40, 3979.
- 5 R. S. BOHACEK, C. McMARTIN and W. C. GUIDA, *Med. Res. Rev.*, **1996**, 16, 3.
- 6 R. E. BABINE and S. L. BENDER, *Chem. Rev.*, **1997**, 97, 1359.
- 7 R. E. HUBBARD, *Curr. Op. Biotechnology*, **1997**, 8, 696.
- 8 M. A. MURCKO, P. R. CARON and P. S. CHARIFSON, *Ann. Reports in Med. Chem.*, **1999**, 34, 297.
- 9 H. GOHLKE and G. KLEBE, *Angew. Chem. Int. Ed.*, **2002**, 41, 2644.
- 10 I. K. McDONALD and J. M. THORNTON, *J. Mol. Biol.*, **1994**, 238, 777.
- 11 J. D. DUNITZ, *Chem. Biol.*, **1995**, 2, 709.
- 12 P. GILLI, V. FERRETTI, G. GILLI and P. A. BREA, *J. Phys. Chem.*, **1994**, 98, 1515.
- 13 D. H. WILLIAMS, D. P. O'BRIEN and B. BARDSLEY, *J. Am. Chem. Soc.*, **2001**, 123, 737.
- 14 P. C. WEBER, J. J. WENDOLOSKI, M. W. PANTOLIANO and F. R. SALEMME, *J. Am. Chem. Soc.*, **1992**, 114, 3197.
- 15 A. R. FERSHT, J.-P. SHI, J. KNILL-JONES, D. M. LOWE, A. J. WILKINSON, D. M. BLOW, P. BRICK, P. CARTER, M. M. Y. WAYE and G. WINTER, *Nature*, **1985**, 314, 235.

- 16 Y.W. CHEN and A.R. FERSHT, *J. Mol. Biol.*, **1993**, 234, 1158.
- 17 P.R. CONNELLY, R.A. ALDAPE, F.J. BRUZZESE, S.P. CHAMBERS, M.J. FITZGIBBON, M.A. FLEMING, S. ITOH, D.J. LIVINGSTON, M.A. NAVIA, J.A. THOMSON and K.P. WILSON, *Proc. Natl. Acad. Sci. USA*, **1994**, 91, 1964.
- 18 B.P. MORGAN, J.M. SCHOLTZ, M.D. BALINGER, I.D. ZIPKIN and P.A. BARTLETT, *J. Am. Chem. Soc.*, **1991**, 113, 297.
- 19 B.A. SHIRLEY, P. STANSSENS, U. HAHN and C.N. PACE, *Biochemistry*, **1992**, 31, 725.
- 20 U. OBST, D.W. BANNER, L. WEBER and F. DIEDERICH, *Chem. Biol.*, **1997**, 4, 287.
- 21 H. KUBINYI, in *Pharmacokinetic optimization in drug research*, B. TESTA, H. VAN DE WATERBEEMD, G. FOLKERS, R. GUY (Eds.), Wiley-VCH, Weinheim, **2001**, pp. 513.
- 22 H.-J. SCHNEIDER, T. SCHIESTEL and P. ZIMMERMANN, *J. Am. Chem. Soc.*, **1992**, 114, 7698.
- 23 A.C. TISSOT, S. VUILLEUMIER and A.R. FERSHT, *Biochemistry*, **1996**, 35, 6786.
- 24 A. BEN-NAIM, *Hydrophobic Interactions*, Plenum, New York, **1980**.
- 25 C. TANFORD, *The Hydrophobic Effect*, Wiley, New York, **1980**.
- 26 F.M. RICHARDS, *Annu. Rev. Biophys. Bioeng.*, **1977**, 6, 151.
- 27 K.A. SHARP, A. NICHOLLS, R. FRIEDMAN and B. HONIG, *Biochemistry*, **1991**, 30, 9686.
- 28 G. LANGE, D. LESUISSE, P. DEPREZ, B. SCHOOT, P. LOENZE, D. BENARD, J.P. MARQUETTE, P. BROTO, E. SARUBBI, E. MANDINE, *J. Med. Chem.* **2002**, 45, 2915–2922.
- 29 H.-J. BOEHM, *J. Comput.-Aided Mol. Design*, **1994**, 8, 243.
- 30 M.K. GILSON, J.A. GIVEN, B.L. BUSH and J.A. MCCAMMON, *Biophys. J.*, **1997**, 72, 1047.
- 31 P.A. KOLLMAN, *Acc. Chem. Res.*, **1996**, 29, 461.
- 32 M.L. LAMB and W.L. JORGENSEN, *Curr. Op. Chem. Biol.*, **1997**, 1, 449.
- 33 M.R. REDDY, M.D. ERION and A. AGARWAL, in *Reviews in Computational Chemistry*, K.B. LIPKOWITZ, D.B. BOYD (Eds.), Wiley-VCH, New York, **2000**, Vol. 16, pp. 217.
- 34 M.K. GILSON, J.A. GIVEN and M.S. HEAD, *Chem. Biol.*, **1997**, 4, 87.
- 35 J. AQUIST, C. MEDINA and J.-E. SAMUELSSON, *Prot. Eng.*, **1994**, 7, 385.
- 36 T. HANSSON, J. MARELIUS and J. AQUIST, *J. Comput.-Aided Mol. Design*, **1998**, 12, 27.
- 37 R.C. RIZZO, J. TIRADO-RIVES and W.L. JORGENSEN, *J. Med. Chem.*, **2001**, 44, 145.
- 38 A.E. MARK and W.F. VAN GUNSTEREN, *J. Mol. Biol.*, **1994**, 240, 167.
- 39 D. WILLIAMS and B. BARDSLEY, *Persp. Drug Disc. and Design*, **1999**, 17, 43.
- 40 P.R. ANDREWS, D.J. CRAIK and J.L. MARTIN, *J. Med. Chem.*, **1984**, 27, 1648.
- 41 H.-J. SCHNEIDER, *Chem. Soc. Rev.*, **1994**, 227.
- 42 T.J. STOUT, C.R. SAGE and R.M. STROUD, *Structure*, **1998**, 6, 839.
- 43 M.K. HOLLOWAY, J.M. WAI, T.A. HALGREN, P.M.D. FITZGERALD, J.P. VACCA, B.D. DORSEY, R.B. LEVIN, W.J. THOMPSON, L.J. CHEN, S.J. DESOLMS, N. GAFFIN, T.A. LYLE, W.A. SANDERS, T.J. TUCKER, M. WIGGINS, C.M. WISCOUNT, O.W. WOLTERS DORF, S.D. YOUNG, P.L. DARKE and J.A. ZUGAY, *J. Med. Chem.*, **1995**, 38, 305.
- 44 P.D.J. GROOTENHUIS and P.J.M. VAN GALEN, *Acta Cryst.*, **1995**, D51, 560.
- 45 D.A. PEARLMAN and P.A. CHARIFSON, *J. Med. Chem.* **2001**, 44, 502.
- 46 D.A. PEARLMAN, *J. Med. Chem.*, **1999**, 42, 4313.
- 47 I.D. KUNTZ, K. CHEN, K.A. SHARP and P.A. KOLLMAN, *Proc. Natl. Acad. Sci. USA*, **1999**, 96, 9997.
- 48 AJAY and M.A. MURCKO, *J. Med. Chem.* **1995**, 38, 4953.
- 49 J.D. HIRST, *Curr. Op. Drug Disc. Dev.* **1998**, 1, 28.
- 50 J.R.H. TAME, *J. Comput.-Aided Mol. Design*, **1999**, 13, 99.
- 51 H.-J. BOEHM and M. STAHL, *Med. Chem. Res.* **1999**, 9, 445.
- 52 H.-J. BOEHM, *J. Comput.-Aided Mol. Design*, **1998**, 12, 309–323.
- 53 A.N. JAIN, *J. Comput.-Aided Mol. Design*, **1996**, 10, 427.
- 54 R. WANG, L. LIU, L. LAI and Y. TANG, *J. Mol. Model.* **1998**, 4, 379.
- 55 D.K. GEHLHAAR, G.M. VERKHIVKER, P.A. REJTO, C.J. SHERMAN, D.B. FOGEL, L.J.

- FOGEL and S. T. FREER, *Chem. Biol.*, **1995**, *2*, 317.
- 56 G. JONES, P. WILLETT and R. C. GLEN, *J. Mol. Biol.*, **1995**, *245*, 43.
 - 57 G. JONES, P. WILLETT, R. C. GLEN, A. R. LEACH and R. TAYLOR, *J. Mol. Biol.*, **1997**, *267*, 727.
 - 58 H.-J. BOEHM and G. KLEBE, *Angew. Chem. Int. Ed.*, **1996**, *35*, 2588.
 - 59 M. MATSUMARA, W. J. BECKTEL and B. W. MATTHEWS, *Nature*, **1988**, *334*, 406.
 - 60 V. NAUCHATEL, M. C. VILLAVARDE and F. SUSSMAN, *Protein Sci.*, **1995**, *4*, 1356.
 - 61 A. M. DAVIS and S. J. TEAGUE, *Angew. Chem. Int. Ed.*, **1999**, *38*, 736.
 - 62 H. GOHLKE and G. KLEBE, *Curr. Op. Struct. Biol.*, **2001**, *11*, 231.
 - 63 M. J. SIPPL, *J. Comput.-Aided Mol. Design*, **1993**, *7*, 473.
 - 64 I. MUEGGE and Y. C. MARTIN, *J. Med. Chem.*, **1999**, *42*, 791.
 - 65 H. GOHLKE, M. HENDLICH and G. KLEBE, *J. Mol. Biol.*, **2000**, *295*, 337.
 - 66 R. C. RIZZO, M. UDIER-BLAGOVIC, D. P. WANG, E. K. WATKINS, M. B. KROEGER SMITH, R. H. SMITH, J. TIRADO-RIVES, W. L. JORGENSEN, *J. Med. Chem.* **2002**, *45*, 2970.
 - 67 S. A. HINDLE, M. RAREY, C. BUNNING, T. LENGAUER, *J. Comput.-Aided Molec. Des.* **2002**, *16*, 129.
 - 68 Y. SUN, T. J. A. EWING, A. G. SKILLMAN and I. D. KUNTZ, *J. Comput.-Aided Mol. Design*, **1998**, *12*, 579.
 - 69 M. D. MILLER, S. K. KEARSLEY, D. J. UNDERWOOD and R. P. SHERIDAN, *J. Comput.-Aided Mol. Design*, **1994**, *8*, 153.
 - 70 T. G. DAVIES, R. E. HUBBARD and J. R. H. TAME, *Protein Sci.*, **1999**, *8*, 1432.
 - 71 J. H. TONEY, P. M. D. FITZGERALD, N. GROVER-SHARMA, S. H. OLSON, W. J. MAY, J. G. SUNDELOF, D. E. VANDERWALL, K. A. CLEARY, S. K. GRANT, J. K. WU, J. W. KOZARICH, D. L. POMPLIANO and G. G. HAMMOND, *Chem. Biol.*, **1998**, *5*, 185.
 - 72 G. KLEBE, F. DULLWEBER and H.-J. BOEHM, in *Thermodynamics of Protein-Ligand Interactions*, R. RAFFA (Ed), J. Wiley, **2001**, p. 83.
 - 73 D. A. GSCHWEND, W. SIRAWARAPORN, D. V. SANTI and I. D. KUNTZ, *Proteins*, **1997**, *29*, 59.
 - 74 R. L. DESJARLAIS, G. L. SEIBEL, I. D. KUNTZ, P. S. FURTH, J. C. ALVAREZ, P. R. ORTIZ DE MONTELLANO, D. L. DECAMP, L. M. BABÉ and C. S. CRAIK, *Proc. Natl. Acad. Sci. USA*, **1990**, *87*, 6644.
 - 75 B. K. SHOICHET, R. M. STROUD, D. V. SANTI, I. D. KUNTZ and K. M. PERRY, *Science*, **1993**, *259*, 1445.
 - 76 R. L. DESJARLAIS and J. S. DIXON, *J. Comput.-Aided Mol. Design*, **1994**, *8*, 231.
 - 77 I. MASSOVA, P. MARTIN, A. BULYCHEV, R. KOCZ, M. DOYLE, B. F. P. EDWARDS and S. MOBASHERY, *Bioorg. Med. Chem. Lett.*, **1998**, *8*, 2463.
 - 78 D. TONDI, U. SŁOMCZYŃSKA, M. P. COSTI, D. M. WATTERSON, S. GHELLI and B. K. SHOICHET, *Chem. Biol.*, **1999**, *6*, 319.
 - 79 P. BURKHARD, P. TAYLOR and M. D. WALKINSHAW, *J. Mol. Biol.*, **1998**, *277*, 449.
 - 80 P. BURKHARD, U. HOMMEL, M. SANNER and M. D. WALKINSHAW, *J. Mol. Biol.*, **1999**, *287*, 853.
 - 81 R. ABAGYAN, M. TROTOV and D. KUZNETSOV, *J. Comput. Chem.*, **1994**, *15*, 488.
 - 82 M. SCHAPIRA, B. M. RAAKA, H. H. SAMUELS and R. ABAGYAN, *Proc. Natl. Acad. Sci. USA*, **2000**, *97*, 1008.
 - 83 A. V. FILIKOV, V. MONAN, T. A. VICKERS, R. H. GRIFFEY, P. D. COOK, R. A. ABAGYAN and T. L. JAMES, *J. Comput.-Aided Mol. Design*, **2000**, *14*, 593.
 - 84 S. GRUENEBERG, B. WENDT and G. KLEBE, *Angew. Chem. Int. Ed.*, **2001**, *40*, 389.
 - 85 D. W. CHRISTIANSON and C. A. FIERKE, *Acc. Chem. Res.*, **1996**, *29*, 331.
 - 86 H. J. BÖHM, *J. Comput.-Aided Molec. Des.* **1992**, *6*, 61.
 - 87 H. J. BÖHM, *J. Comput.-Aided Molec. Des.* **1992**, *6*, 593.
 - 88 LUDI IS AVAILABLE FROM ACCELRY, SAN DIEGO
 - 89 H.-J. BOEHM, D. W. BANNER and L. WEBER, *J. Comput.-Aided Mol. Design*, **1999**, *13*, 51.
 - 90 H.-J. BOEHM, M. BOEHRINGER, D. BUR, H. GEMUENDER, W. HUBER, W. KLAUS, D. KOSTREWA, H. KUEHNE, T. LUEBBERS, N. MEUNIER-KELLER and F. MUELLER, *J. Med. Chem.*, **2000**, *43*, 2664.
 - 91 G. KLEBE, U. GRÄDLER, S. GRÜNEBERG, O. KRÄMER, H. GOHLKE, in *Virtual Screen-*

- ing for Bioactive Molecules, G. SCHNEIDER, H.-J. BOEHM (Eds.), VCH, Weinheim, 2000, p. 207.
- 92 R. E. BABINE, T. M. BLECKMAN, C. R. KISSINGER, R. SHOWALTER, L. A. PELLETIER, C. LEWIS, K. TUCKER, E. MOOMAW, H. E. PARGE, J. E. VILAFRANCA, *Bioorg. Med. Chem. Lett.* **1995**, 5, 1719.
- 93 D. BAILEY and D. BROWN, *Drug Discovery Today*, **2001**, 6, 57.
- 94 R. M. EGLIN, G. SCHEIDER and H.-J. BOEHM, in *Virtual Screening for Bioactive Molecules*, G. SCHNEIDER, H.-J. BOEHM (Eds.), VCH, Weinheim, 2000, p. 1.
- 95 R. LAHANA, *Drug Discovery Today*, **1999**, 4, 447.
- 96 C. A. LIPINSKI, F. LOMBARDO, B. W. DOMINY and P. J. FEENEY, *Adv. Drug Delivery Rev.*, **1997**, 23, 3.

2

Introduction to Molecular Recognition Models

H.-J. SCHNEIDER

2.1

Introduction and Scope

Molecular recognition is the basis of both biological systems and many chemical technologies. When Emil Fischer in 1894 put forward the first model for molecular recognition in the form of his famous lock-and-key principle, he could not anticipate that chemists would one day produce fully synthetic systems of this kind. It took almost 100 years until completely artificial complexes were developed, in which a host molecule embraces a guest molecule in the way that Fischer believed to be the basis of enzyme function. In 1987 the Nobel Prize award to Cram, Lehn, and Pedersen highlighted how far chemistry had gone in these directions. In recent decades, the field, which Cram named “host-guest chemistry,” and Lehn called “supramolecular chemistry,” has experienced a virtual explosion (see monographs [1–8]). Countless groups over the world are now synthesizing host structures with intricate binding properties for a large array of targets and analyzing supramolecular complexes with rapidly developing physical methods. Coordination chemistry is traditionally directed towards transition metal ion complexation but can provide much additional, and sometimes overlooked, information on principles ruling the spontaneous formation of host-guest complexes.

Empirical analyses of structures and energetics in synthetic supramolecular complexes can provide insight into the non-covalent interaction mechanisms and attribute energy values to each of them. Much of the principles and quantitative information learned from these complexes can be of use for the understanding of biological systems and, e.g., the design of bioactive ligands. Most of the efforts in modern supramolecular chemistry are of course directed towards new technologies in separation, sensors, materials, information storage and processing, energy conversion, artificial enzymes, etc. At the same time, these systems provide many new models for molecular recognition processes and a wealth of information on the underlying interactions. Synthetic chemistry is able to deliver biomimetic as well as unnatural host compounds in which every desired function can be implemented. These functions can be directed towards any given substrate site and can be designed to work in any environment, be it in the ground state or the transition state.

Geometric fitting between host and guest, or lock and key, is a prerequisite that can undergo significant modifications. Shape compliance itself will of course not bring molecules together. Non-covalent interactions provide the driving force for this, and the tightness between lock and key is a function of the underlying interaction mechanisms [9]. The distance dependence for attraction between binding sites varies between r^{-1} for Coulombic interactions and r^{-6} for dispersive interactions; for solvophobic interactions, there is no clear-cut boundary definition at all. Obviously, one needs to consider “soft” and “hard” lock-and-key systems and to analyze the underlying binding mechanisms in order to apply Emil Fischer’s idea in a more rigorous way. In the present chapter, effort is made to illustrate the development and the implications of the lock-and-key model and to highlight conclusions mainly from the study of synthetic recognition models. Particular emphasis will be given to the possibilities to derive information on mechanisms and magnitudes of non-covalent interactions in solution from properly designed host-guest complexes.

2.2

Additivity of Pairwise Interactions – The Chelate Effect

Multi-site interactions can lead to very stable associations [10], also in fully synthetic model complexes. This is illustrated in Fig. 2.1 with, e.g., complexes between ATP and an azacrown ether [11] and between Fe^{3+} ions and an artificial siderophore [12]. Another relatively open host structure in Fig. 2.1 contains three vancomycin moieties around a benzene ring; it binds a cell wall component in the form of a trimeric dipeptide with an association constant of almost $K=10^{17}$ [13], which is powers of magnitude higher than the natural biotin-streptavidin complex [14]. Other high-affinity receptors based on polytopic interactions between separate binding sites were reported, e.g., for cyclodextrin dimers with divalent ligands ($\lg K$ up to 7) [15]. Cyclodextrins equipped with additional stacking units also can selectively bind steroids with $\lg K$ up to 7 [16]. Some more examples will be discussed below. While these complexes form in water, association in aprotic solvents can be made equally strong, as shown recently, e.g., with a semi-cavitand complex with a binding energy of over 42 kJ/mol ($\lg K=16.5$) in chloroform [17]. Highly pre-organized ionophores like **2** in Fig. 2.1, which have all possible binding atoms directed towards the guest cation, can complex, e.g., Cs^+ ions in chloroform with $\Delta G=90$ kJ/mol [18].

The stability increase of host-guest complexes with the number of individual interactions between the non-covalent binding sites was analyzed in great detail decades ago in coordination chemistry, which provides clues to many more recent observations with purely organic complexes. Under certain conditions, the chelate effect and the resulting total free energy of binding ΔG_{T} can be quantified by Eq. 2.1, in which ΔG_{i} , ΔG_{j} , ΔG_{k} , etc., represent the contributions of interactions between single host and guest sites, and i , j , and k represent the number of each

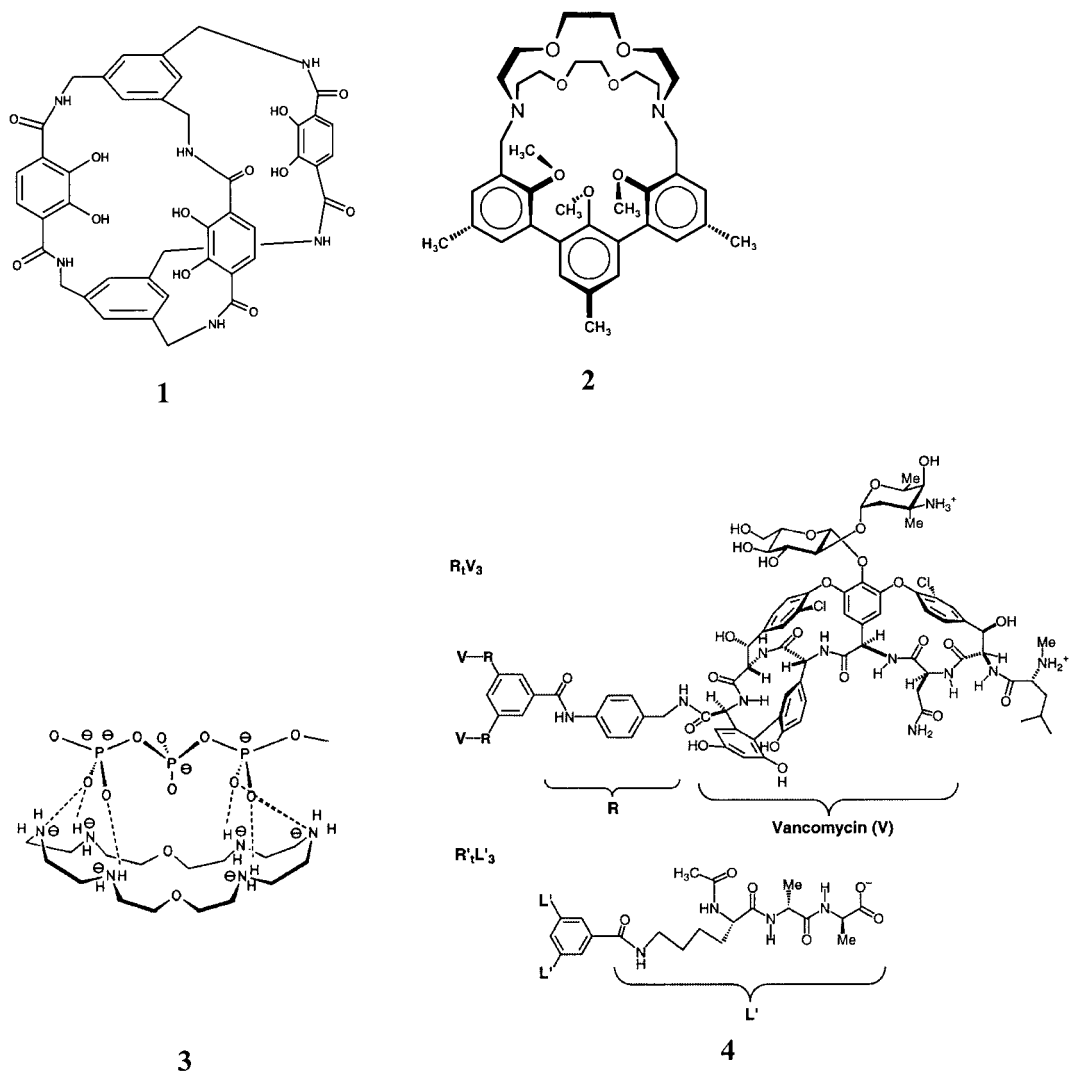


Fig. 2.1 Chelate effects in high affinity artificial complexes: **(1)** an artificial siderophore with $K=10^{59} \text{ M}^{-1}$; **(2)** a ionophore binding Cs^+ ions in chloroform with $\Delta G=90 \text{ kJ/mol}$; **(3)** an azacrown ether and triphosphate residue (as in ATP) as guest, with $K=10^{11} \text{ M}^{-1}$ (only 7 out of the possible 10 to 12 charge-charge

bridges are shown by dashed lines); **(4)** a trivalent vancomycin derivative R_1V_3 [C_6H_3 -1,3,5-(CONHC_6H_4 -4- CH_2NHCOV) $_3$; V =Vancomycin] and trivalent derivative of DADA, $\text{R}'_1\text{L}'_3$ [C_6H_3 -1,3,5-($\text{CON}(\text{N-acetyl})$ -L-Lys-D-Ala-D-Ala) $_3$], with $K=10^{17} \text{ M}^{-1}$.

kind of interaction, which can be salt bridges, hydrogen bonds, van der Waals forces, etc. (Fig. 2.2).

$$\Delta G_t = i\Delta G_i + j\Delta G_j + k\Delta G_k \dots \quad (\text{Eq. 2.1})$$

Additive group contributions have been used for long time for thermodynamic estimations of organic compounds [19] and the description of non-covalent interactions [20, 21]. Limitations due to entropic contributions have been discussed in detail before [4, 22] and will be considered in Section 2.6 also with respect to the enthalpy-entropy compensation that is typical for host-guest complexes. Additivity of pairwise free interaction energies implies that the association constant K_t would be the multiplicative product of the corresponding single constants K_i , etc. It has been shown that the use of the dimensionless association constant K circumvents the problem of dimensions resulting from the multiplication of K units [23] and removes the need to invoke entropic reasons for the chelate effect, also for associations with protein [24]. Calorimetric measurements show that, in fact, the advantage of implementing many binding sites within one ligand, and the so-called macrocyclic effect, which is the affinity increase by placing all interaction sites within a macrocycle, is primarily due to an enthalpy gain. In a number of cases, one even observes an entropic *disadvantage* by complexation with pre-organized macrocyclic ligands [9].

Additivity of pairwise interaction energies is often taken for granted implicitly in force field calculations of supramolecular complexes. In fact, the decomposition of total free energies into single components is surprisingly successful, with both smaller host-guest complexes [25] and protein-ligand interactions [26–28], in spite of noticeable limitations [22, 29]. The additivity of single interactions is limited, for instance, by varying entropic factors in single interactions, by a possible geo-

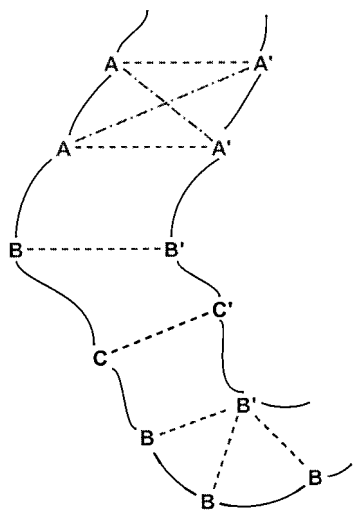


Fig. 2.2 Additive binding interactions between host and guest structures, usually attractive (dashed lines); secondary interactions can also be repulsive (broken lines).

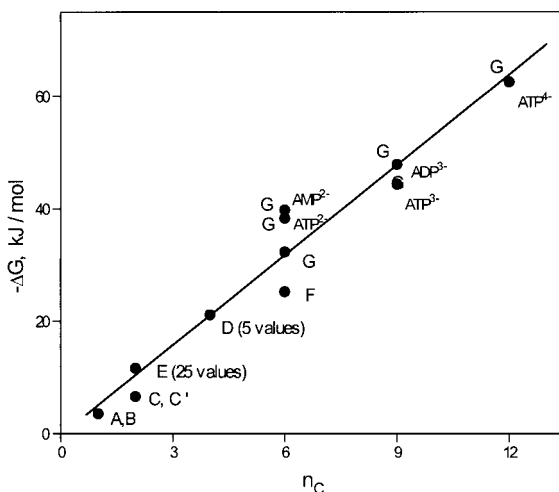


Fig. 2.3 Quantification of the chelate effect: a plot of experimental free complexation energies against the number of pairwise interac-

tions, here with ion pairs; for identification of the complexes see [4], p. 9. Reproduced with permission of the publisher.

metric mismatch between binding sites, by secondary interactions between neighboring groups, or by changes in the microscopic environment, e.g., like dielectrics (see also Sections 2.5 and 2.6) [22]. In order to arrive at safe conclusions for the identification of binding mechanisms and at reliable free energy increments (ΔG_i , etc.), it is necessary to measure complexes in which the numbers i , j , and k of single interactions are systematically varied. For host-guest complexes, in which strain-free matching between donor and acceptor sites is maintained, one indeed observes more often than not a linear increase of the experimental total free energy with the numbers i , j , and k according to Eq. 2.1 (Fig. 2.3) [30]. The slope of the correlation line then gives a statistically meaningful free energy value (ΔG_i , ΔG_j , ΔG_k , etc.) for a specific non-covalent interaction [25]. As with the extraction of reaction or substituent constants from classical linear free energy correlations of the Hammett type, a sufficiently large number of experimental observations is necessary in order to arrive at reliable ΔG values. Synthetic host-guest complexes allow one to construct such a broad experimental basis under planned conditions. In other cases one “mutates” one interaction against others and observes systems in which more than one interaction mechanism is at work. Here one can use either a two-term correlation according to Eq. 2.1 or terms (e.g., ΔG_i) known from independent analyses with only one kind of interaction and then plot the remaining ΔG_j values vs. the number j of the second interactions (Fig. 2.4). This approach is preferable over multi-linear correlations because at least one of the interaction increments $\Delta\Delta G$ is based on independent measurements with sufficiently large numbers of observables. The use of such analyses with respect to the different mechanisms of intermolecular forces will be discussed in Section 2.9, where other examples also illustrate the additivity of non-covalent interactions.

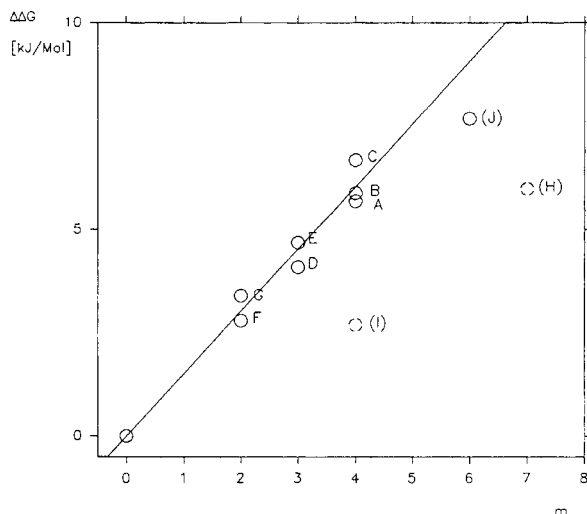


Fig. 2.4 Quantification of the chelate effect in the presence of two interaction mechanisms: plot of experimental free complexation energies on aromatic ion pairs against the number m of pairwise interactions, after deducting the contribution of a primary interac-

tion ΔG_i (salt bridges), see text. Structure of the complexes see [30]. Complexes H and I deviate due to geometric mismatch or too flexible spacers between binding sites. Reproduced with permission of the publisher.

2.3

Geometric Fitting: The Hole-size Concept

For the seemingly simplest case of spherical metal ion complexation, the hole-size fit often, but not necessarily, holds. Fig. 2.5 illustrates the classical case where the cavity diameter of an ionophore determines the selectivity of cation complexation according to its radius [31]. As long as sufficient contact between the metal ion and the donor atom of the ligand is possible, the complexation free energy will be just a linear function of the number of such interactions and their donor quality [32]. If the ionophore size becomes too large, the selectivity vanishes (Fig. 2.6) [33].

Formal consideration of only the number of available binding functions can be misleading. Thus, the 18-crown-6 ether binds K^+ ions by orders of magnitude better than the 18C5 macrocycle, which has the same ring size but five instead of six oxygen atoms. The discrepancy results from the single CH_2 group replacing one of the oxygen atoms, which forces one C-H bond inside the cavity and thus prevents optimal contact of the ion with the oxygen donor atoms (Fig. 2.7) [34]. This example demonstrates how small distortions can greatly influence complexation strength and emphasizes the role of computer-aided molecular modeling to control lock-and-key interactions and to design proper host-guest complexes.

Fig. 2.5 Size selectivity of cryptands: logarithms of the binding constants $\lg K$ vs. ion diameters; (a) – values with $\lg K < 2.0$, (b) in 95% MeOH, (c) in MeOH; see [4]. Reproduced with permission of the publisher.

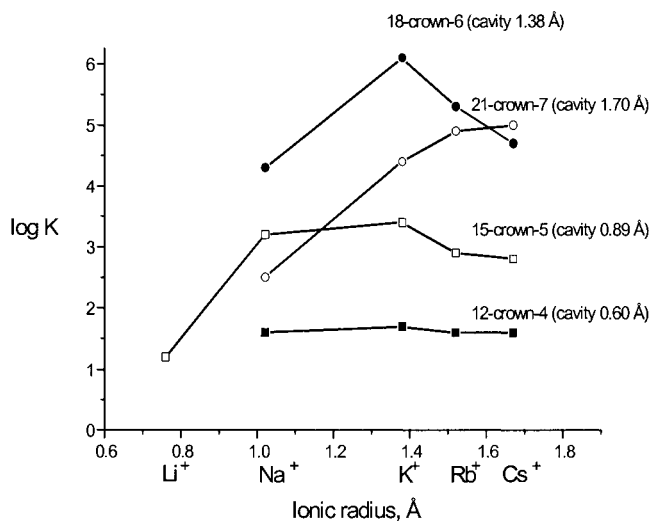
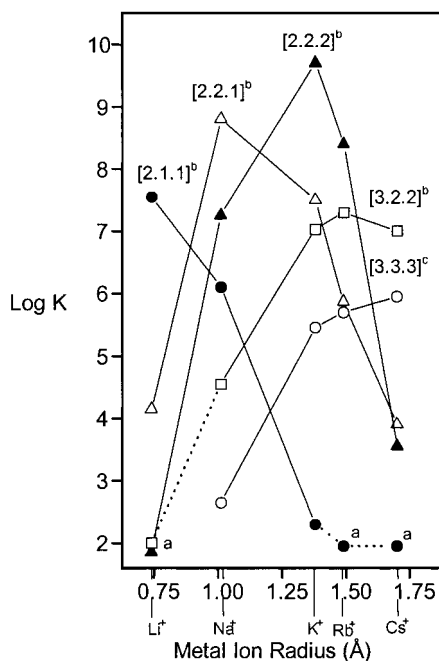


Fig. 2.6 The decrease of selectivity with decreasing fit: Logarithms of the binding constants (average of published results [33], of alkali cations by crown ethers in methanol vs.

ionic radii. In the case of Li^+ with a majority of crown ethers, one observes $\log K < 1$; see [4]. Reproduced with permission of the publisher.

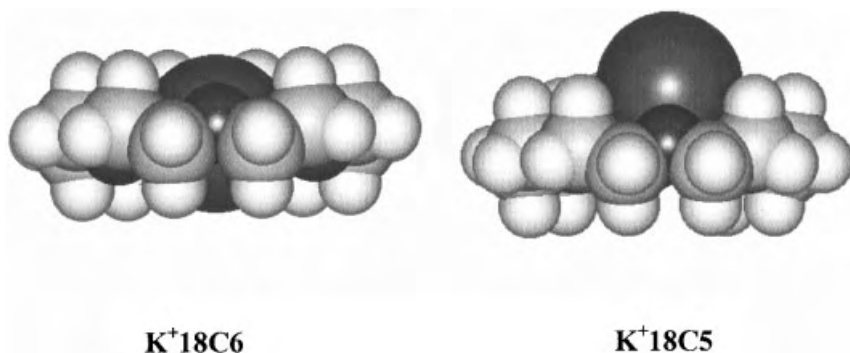


Fig. 2.7 Structures of potassium complexes of 18-crown-6 and 18-crown-5; see [4]. Reproduced with permission of the publisher.

2.4

Di- and Polytopic Interactions: Change of Binding Mechanism with Different Fit

Different binding sites, each equipped with a number of suitable binding functions, can be covalently bound together with a spacer to form polytopic receptors (Fig. 2.8). Such heterotopic host compounds, providing separate binding sites for the anion and the cation, can be highly effective, e.g., for the binding of salts. In this way, considerable enhancements of hydrophilic ion pair transport into and through a lipophilic medium can be achieved [35–37]. The concept of ditopic recognition is also useful for the transport of zwitterionic amino acids through membranes [38]. An interesting extension is to provide binding sites as two separate host compounds, which allows more freedom of host structure choice and at the same time can disrupt very strong associations of guest compounds, such as ion pairs in unpolar media (Fig. 2.9) [39]. In a related approach, multivalent ligands have been used to remove, e.g., strongly bound selectins from cell surfaces [40].

The performance of ditopic receptors will suffer if the spacer is not long and/or flexible enough to allow simultaneous full contact at all binding centers. In some cases one observes only weakening of affinities [41], while in other cases one of the possible intermolecular forces cannot materialize at all. Thus, additive ion pairing as well as dispersive interactions with positively charged polyaromatic host compounds are present, e.g., in complexes of AMP with tetrapyridinium porphyrins.

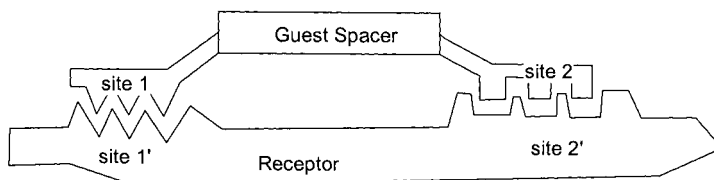


Fig. 2.8 A ditopic receptor with a spacer separating two binding sites.

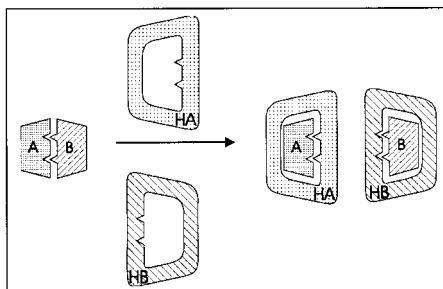
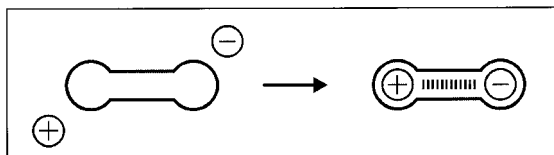


Fig. 2.9 (a) Ditopic host for cooperative binding by additional interaction between cation and anion; (b) Two separate hosts for binding and dissociation of two strongly associated guest molecules.

rin derivatives [42]. With larger contributions of stacking, the geometric matching between the charged sites can be so distorted that one observes, e.g., the same affinities with electroneutral nucleosides as with charged nucleotides [43]. The high affinities of nucleosides compared to nucleotides with the cleft-like receptor shown in Fig. 2.10 [44] illustrate that one interaction mechanism can “overwhelm” another one.

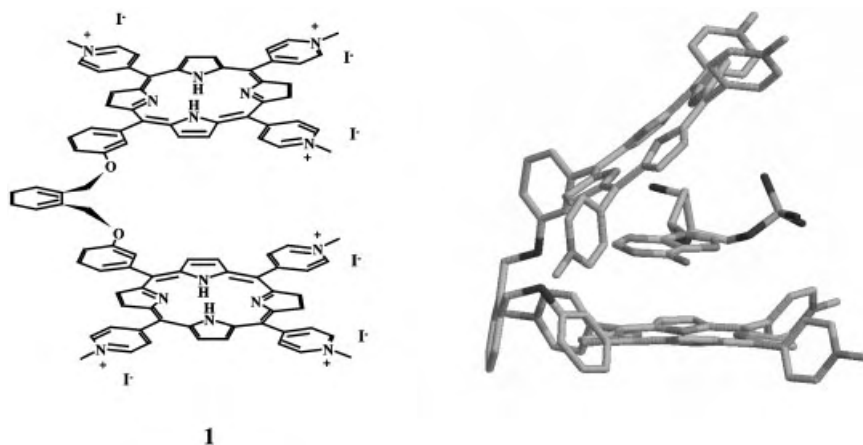


Fig. 2.10 A cleft-like receptor with similar affinity to electroneutral nucleosides and charged nucleotides.

2.5

Deviations from the Lock-and-Key Principle

2.5.1

Strain in Host-Guest Complexes

A classical case in which the building up of strain hampers complexation is illustrated in Fig. 2.11. The open tetramine needs to form several unfavorable gauche interactions for binding copper ions and therefore exhibits an affinity that is 10 powers of magnitude smaller [45]. Exceptions from the lock-and-key analogous “hole-size rule” are also seen if a bidentate ligand interacts with transition metal ions of a different radius. At first sight, unexpectedly, a large cation such as Pb^{2+} prefers the shorter ethylenediamine as ligand, whereas the smaller Ni^{2+} prefers the longer propylenediamine. The reason is that the shorter Ni-N bond length allows formation of an almost strain-free metallo-cyclohexane ring with almost equally long intra-ring bonds, whereas the longer Pb-N bond is better accommodated in a pseudo-cyclopentane ring and would produce more strain in a then heavily distorted metallo-cyclohexane [45, 46]. Possible strain energy changes must also be considered in cases of induced fit and allosteric complexes, where geometry deviations necessarily are accompanied by the building up of less favorable interactions.

2.5.2

Solvent Effects

The influence of solvents can lead to profound deviations from simple geometric fitting rules. Complexation studies in the gas phase with MS techniques have problems in deriving exact associations constants but have given relative values in general agreement with the basic lock-and-key concept [47]. Under most experimental conditions, host and guest molecules, in particular cations, are solvated to a different degree before complex formation, and even in cryptands also within the complex [48]. As a consequence, the selectivity varies with the solvent as a

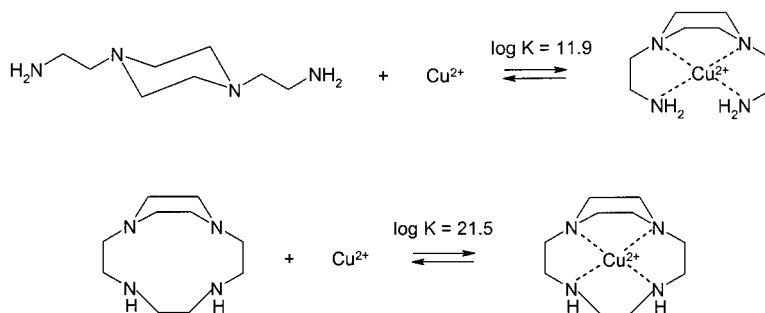


Fig. 2.11 Complex formation with Cu^{2+} requiring different strain in the ligands.

function of solvation and desolvation energies. Thus, the transfer free energy from water to methanol is 10 kJ/mol for K^+ and 8 kJ/mol for Na^+ ; in acetonitrile the sequence is reversed, with 8 kJ/mol for K^+ and 14 for Na^+ [49]. In acetonitrile the Na^+ with the higher charge density is much less stable than the larger K^+ ion; thus, the 15-crown-5 ether in this solvent complexes Na^+ 100 times better than K^+ . In contrast, one observes a small preference for K^+ in methanol [50], where the transfer free energies are less variable. Obviously, solvent effects modify binding properties significantly, in particular with polar substrates.

2.5.3

Enthalpy/Entropy Variations

A fundamental limitation for the application of geometric fitting procedures is that the complexation free energies are the sum of enthalpic and entropic contributions, with the consequence that selectivity can be inversed at different temperatures. Positive cooperativity between different interactions in a complex will usually lead to tighter association at the expense of motional freedom and thus of entropy [51]. The interplay and often observed compensation of enthalpic and entropic contributions have been discussed in several reviews [10, 52, 53], particularly with emphasis on biological systems, and cannot be dealt with in detail here. Unfortunately, many published enthalpy-entropy compensations are blurred by possible artifacts, as the two underlying parameters do not represent independent variables [54].

Intuitively, one may associate pairwise interactions between the lock and key binding sites with an enthalpic gain. In polyvalent complexes such as the trimeric vancomycin discussed above (Fig. 2.1), the total ΔH is indeed about three times larger for the single vancomycin complex; the same about three-fold enhancement applies to the $T\Delta S$ contribution [13]. However, some intermolecular interactions, such as ion pairing in water, are entirely entropy-driven, whereas, e.g., long-range Coulomb interactions or hydrogen bonds are primarily enthalpy driven. The different distance dependence of non-covalent mechanisms necessarily modifies the size-matching requirements. Cyclodextrin complexes with tightly fitting and highly polarizable guest molecules form mainly by enthalpy gain, whereas those with loose fit and aliphatic substrates in their cavity show some hydrophobic entropic contributions. Related changes are observed for aqueous complexes with cyclophanes; they all are closely associated with the interplay of van der Waals and hydrophobic interactions and will therefore be discussed below in Section 2.9. One also has to take into account that the magnitude and even the sign of ΔS and therefore ΔG are a function of the chosen standard state, in contrast to ΔH . The complex between α -cyclodextrin and benzene is characterized by $\Delta H = -19.2$ kJ/mol and a negative $\Delta S = -15$ kJ/mol, e.g., if calculated for the standard state of 1 M, but by a positive $\Delta S = 18$ kJ/mol if calculated for mole fractions [55]. Fortunately, one can often rely on free energy considerations, as non-covalent interactions are often characterized by enthalpy-entropy compensations [52], although many observations of this kind might be experimental artifacts [54].

2.5.4

Loose Fit in Hydrophobically Driven Complex Formation

Particularly in aqueous solutions, there is evidence that loose “fit” can be preferred over tight fit, as exemplified by complexes with the cavitand in Fig. 2.12 [56] or with cyclodextrins [57]. With tetramethylammonium chloride in water, the cycloveratrylene host shows a binding free energy of $\Delta G = 15$ kJ/mol for the smaller cavity, with the number $n = 3$ of methylene groups. For $n = 5$, ΔG is 20 kJ/mol, although the geometric fit is better with the smaller host ($n = 3$). In another experiment, a larger part of a fluorescence dye as guest molecule can fill the cavity of cyclodextrin, or, alternatively, a smaller guest part can move more freely within the cavity [57]. The observed NMR shifts on the guest molecule in line with intermolecular NOEs demonstrate that the preferred binding mode is the latter. Only with the wider cavity of the γ -cyclodextrin does one observe encapsulation of the larger naphthyl ring as well.

2.6

Conformational Pre-organization: Flexible vs. Rigid Hosts

One paradigm in supramolecular chemistry is that a high affinity requires optimal pre-organization of the host or guest structure to each other, so that the binding sites can geometrically match without conformational changes and with a minimal loss of conformational freedom. This calls for the design of host compounds in which all binding functions are largely prefixed to take up the substrate, as exemplified by the complexes in Fig. 2.1, and has been the incentive for demanding synthetic efforts in supramolecular chemistry. The question is, then, which free energy cost is involved by the presence of flexible bonds. Complexation of α -cyclodextrin with cyclohexane shows $\Delta G = -15$ kJ/mol, with the much more flexible *n*-heptane an even higher value of $\Delta G = -22$ kJ/mol, with quite similar ΔH parameters but *less* entropy disadvantage for inclusion of the more flexible *n*-alkane [55]. Such data suggest that conformational freedom in supramolecular complexes, at least in aqueous medium, may be better maintained with more flexible

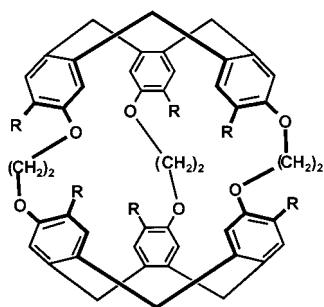


Fig. 2.12 The cavity of the cycloveratrylene host with $n = 3$ ($R = O-CH_2COOH$) fits geometrically better to the tetramethylammonium guest but shows with $\Delta G = 15$ kJ/mol a smaller affinity than the larger cavity with $n = 5$ ($\Delta G = 20$ kJ/mol).

systems, which is in line with the results discussed above with Fig. 2.12. That flexibility usually does not significantly lower stabilities has been stressed also for polyvalent aggregations with biological material [10]. Literature values on the cost of such restrictions vary between $T\Delta S=1.5$ and ~ 5 kJ/mol per single bond [58]. An experimental quantification was obtained with a series of host-guest systems in which the number of single bonds was systematically increased, maintaining the interacting binding elements at the ends of the chains and ensuring that no unfavorable gauche interactions have to build up upon complexation (Fig. 2.13). There is a linear decrease in ΔG with the increasing number n of single bonds in the complexes, but the slope of the correlation indicates only a disadvantage of $\Delta G=1.3$ kJ/mol per single bond [59]. Noticeably, a similar value emerges from studies of artificial peptide β -sheets in which a variable number of single bonds must be offset also by hydrogen bonds in chloroform [60]. Somewhat larger values around 2.0 kJ/mol for freezing rotations around C–C single bonds were re-

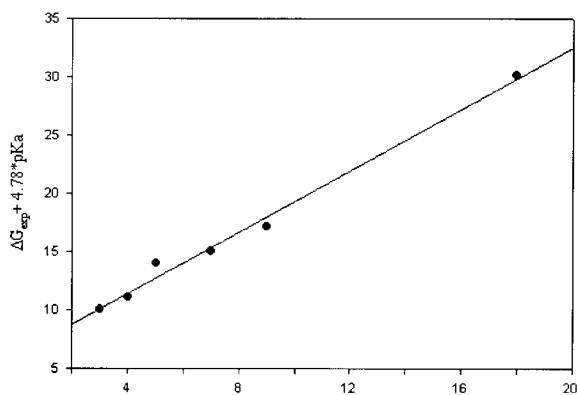
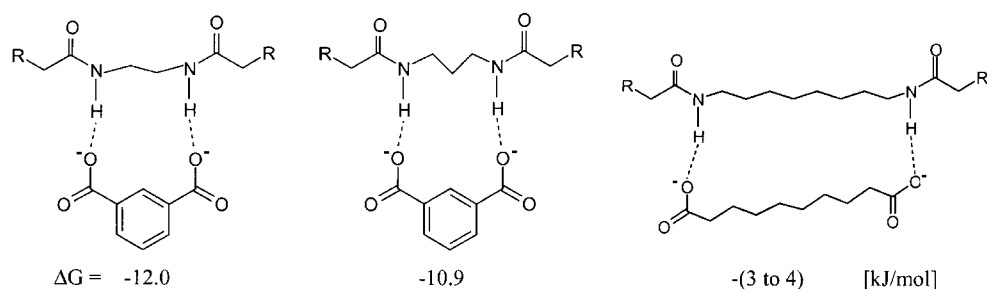


Fig. 2.13 Decrease of complexation free energy ΔG with increasing number of single bonds in complexes like those illustrated

above (from measurements in CDCl_3 , ΔG corrected for pK changes, see [56]). Reproduced with permission of the publisher.



Fig. 2.14 Complexation free energies with ligands of variable flexibility and transition metal ions (see [63]).

ported from other investigations [61]. A related study with ion pairs in water, also showing a linear correlation of ΔG with n , yielded an even smaller increment of only 0.5 kJ/mol [62]. These data together with the observations discussed above indicate that the effect of pre-organization has been overestimated and that the conformational freedom in such complexes is less restricted than expected. One might assume that the small effect of increasing flexibility could be due to the rather weak associations in the complexes discussed above. However, some very strong transition metal ion complexes [63] also give no evidence of a particular advantage of pre-organized bonds (Fig. 2.14). These rather low free energy losses due to the presence of flexible bonds agree well with $G_{\text{rot}} = 1.3$ kJ/mol per rotatable but fixed bond, which was derived for the analysis of protein complexes [64]. It should be remembered that quite efficient non-covalent interactions can develop in structures containing many single bonds, both in natural systems like peptides or proteins and in ionophores like monensin, as well as in synthetic podand or lariat host compounds where many interaction sites are positioned on flexible chains [65].

2.7

Selectivity and Stability in Supramolecular Complexes

High selectivity in molecular recognition coupled with high sensitivity is the goal in synthetic supramolecular as well as in medicinal chemistry. Unfortunately, both aims cannot always be met at the same time. An interesting strategy to overcome the problem of often small selectivity with a single host-guest complex consists in the parallel arrangement of several receptor units [66]. If all binding functions in a host molecule are pre-oriented for optimal contact with the guest molecule, one should in fact expect a maximum of selectivity and affinity. This is indeed observed with some synthetic ionophores, which can discriminate, e.g., Na^+ and K^+ ions with a selectivity surpassing even that of the natural antibiotic valinomycin [67].

Receptors for guest molecules larger than simple ions make use of interactions at different sites and are necessarily more limited with respect to a simultaneous optimization of selectivity and sensitivity. In favorable cases, the selection site will provide additional binding forces, as illustrated by the model peptide receptor in

Fig. 2.15. Here, the primary binding force is produced at the N-terminus by a crown ether unit and at the C-terminus by an ammonium ion; sequence selectivity is brought about by a stacking unit – which can be a fluorophore helping also optical detection – which at the same time leads to an affinity increase depending on the selected amino acid side chain [68].

In other cases, one of the interactions can be so strong that optimal contact with other binding sites cannot materialize. Examples for this have been discussed above, e.g., with the porphyrin-based host, which cannot differentiate between nucleotides and nucleosides due to the dominating stacking effects. Even adverse, anti-cooperative effects between selectivity and affinity sites can be tolerated, in particular if the aim is stereoselectivity. In the chiral crown ether (Fig. 2.16), which is the basis of Cram's "chiral resolution machine" [69], stereoselection is due to interactions between amino acid side groups and the crown ether

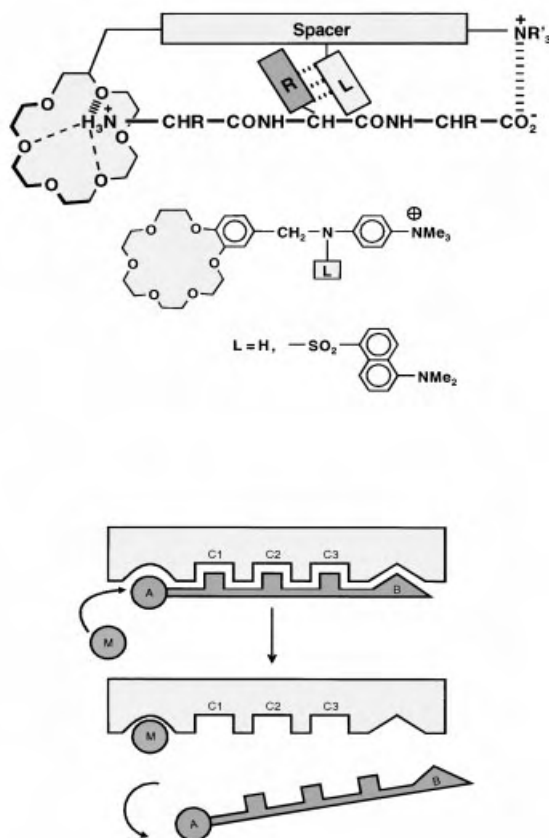


Fig. 2.15 A sequence selective peptide receptor with cooperativity between all possible interaction sites; the peptide can be released upon complexation with a K^+ cation (see [65]).

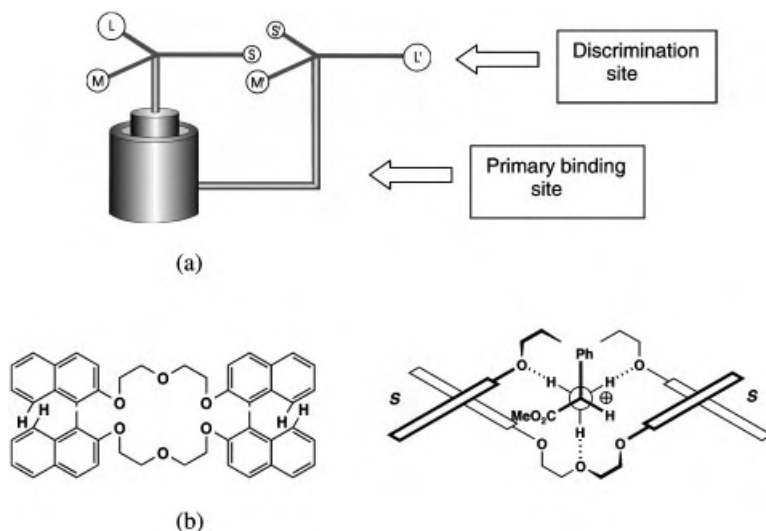


Fig. 2.16 Discrimination with non-binding selection sites. (a) Selection principle; (b) Chiral selection of protected amino acids by a crown ether with remote binaphthyl selection site.

naphthyl moieties, which may be rather repulsive. The principle of spatially separated binding and discrimination sites is illustrated by Fig. 2.16 and is used in many applications. Thus, the low affinity brought about by Hoogsteen base pairing in antisense strategies with nucleic acids can be increased by covalent coupling of oligonucleotides to rather unselective intercalators. Obviously, the primary binding site securing a high affinity should have low selectivity in order not to mitigate the selectivity at the second site.

2.8

Induced Fit, Cooperativity, and Allosteric Effects

Cooperativity in proteins seems to be a very general phenomenon, not restricted to allosteric systems [70–72]. It plays an essential role not only in cooperative control of different substrates as in the classical case of hemoglobin [73], but also, e.g., in protein folding [74, 75]. For biological macromolecules, it is difficult to assess the individual contributions of separate binding sites, although considerable progress has been made, particularly by the use of site-specific protein mutants [70]. Artificial allosteric systems not only open the way to interesting new analytical technologies [76, 77] but also allow a very direct control of positive or negative cooperativity between different binding sites [78]. Conformational coupling in synthetic allosteric models is based on much more rigid elements than are available in proteins; consequently, the strength of allosteric effects in those simple complexes can easily be higher than usually observed in proteins. Thus, binding of Zn^{2+} ions in a structure such as 1 in Fig. 2.17 leads to a strong complexation of

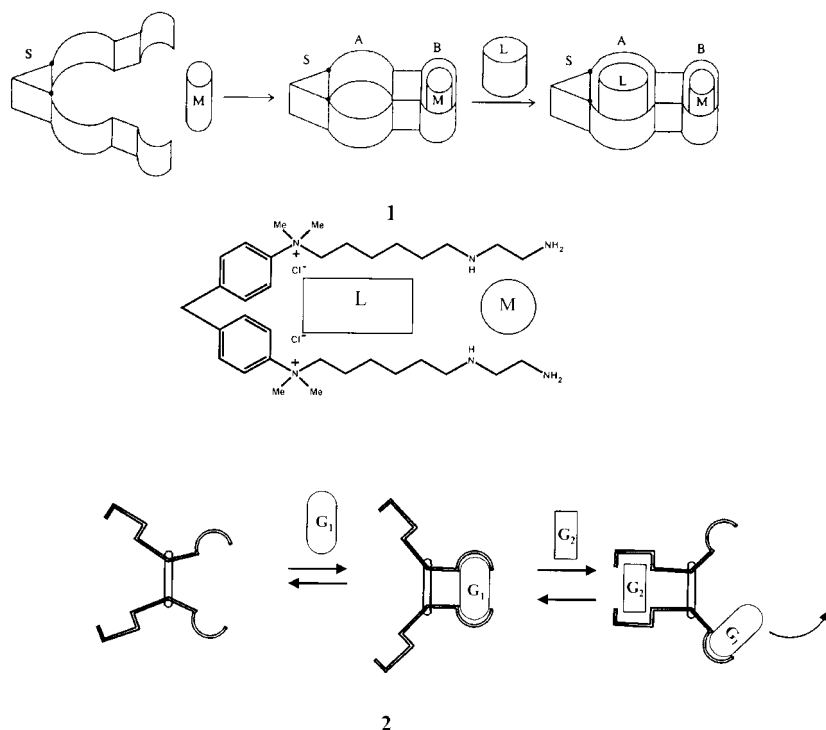


Fig. 2.17 Positive and negative cooperativity in synthetic allosteric models.

lipophilic fluorophores, the association of which is below the detection limit in the absence of the metal inducing the conformational switching [79]. As a consequence, the ratio of association constants with occupation (K_o) and without (K_u) occupation of the metal binding site is $K_o/K_u > 1000$ and is much higher than for the strongest cooperativity found in proteins [80]. Negative cooperativity can also be realized if occupation of one binding site leads to release of a substrate at the second binding site, which does not fit anymore after the induced conformational change (2, Fig. 2.17) [81].

Positive cooperativity without conformational coupling between binding sites is possible if the two guest molecules attract each other after being brought together within a host providing binding sites for both molecules. A corresponding example has been discussed already for salt binding and transport in Section 2.4 and Fig. 2.9 [35–37]. The same principle is at work if two host molecules interact which each other upon complexation with either two guest molecules or a single ditopic guest. Such a positive cooperativity has been realized with synthetic models and plays an important role, e.g., in gene regulation by oligomeric transcription factors RXR, which reach high affinities towards DNA only as pentamers [10, 82].

2.9

Quantification of Non-covalent Forces

The quantification of intermolecular forces is of fundamental importance for the understanding of structure and functions of artificial and biological complexes and for the design of new materials, devices, and drugs. Detailed analyses of accurately determined structures, mostly in crystals, have provided a wealth of information about the occurrence and geometric conditions of non-covalent interactions but cannot attribute energy values. The intriguing strategy to obtain energy contributions from the large body of structure and affinity data for protein-ligand complexes available mostly from medicinal chemistry will be dealt with in other chapters of the present monograph. On this basis, scoring functions have been developed based on reference sets with about 80 [26] or up to 170 [83] protein ligand structures, with affinities ranging from 40 mM to 10 fM. Energy contributions were derived, e.g., for the non-distorted hydrogen bond between 1.7 and 4.4 kJ/mol, for ion pairing 3.0 to 7.9 kJ/mol, and 0.1 kJ (mol Å²) for hydrophobic interactions. Such scoring values reproduce observed affinities in protein complexes with a standard deviation of around 8 kJ/mol or a factor of about 100 in terms of equilibrium constants. Smaller synthetic models, which will be discussed in this section, not only can provide more accurate predictions but also can be made to deliver information about single interaction mechanisms under better-defined conditions. Such complexes are often designed to derive information and free energy values on binding mechanisms that are difficult to identify and quantify in large biomolecules, such as cation- π or C-H- π interactions, or dispersive forces and to discriminate those from hydrophobic interactions.

2.9.1

Ion Pairs and Electrostatic Donor-Acceptor Interactions

The evaluation of binding free increments for salt bridges with simple host-guest complexes was demonstrated in Section 2.2. The linear correlation shown in Fig. 2.3, reaching from, e.g., zinc sulfate to the azacrown ether triphosphate complex shown in Fig. 2.1, yields – on the basis of measurements with more than 80 mostly organic ion pairs in water – an average value of $\Delta\Delta G = 5 \pm 1$ kJ/mol for a bridge between single charges, if the ionic strength corresponds to typical buffer concentrations [30]. With respect to the salt effect, the $\lg K$ values correlate surprisingly well with the Debye-Hückel equation, i.e., they are not only linear but also exhibit a sensitivity (slope) near to the theoretical value, even for very anisotropic organic ions [62]. With a number of such ion pairs, the correlations show an increment of $\Delta\Delta G = 8$ kJ/mol per charge-charge interaction, if extrapolated to pure water at zero ionic strength. Several of the analyzed systems rely on complexation with protonated amines. It has been demonstrated that they can also be described by ion pairing without noticeable hydrogen bond contributions, by observation of the affinities after methylation to peralkylammonium salts. This is also seen in the example of polyamine binding to nucleic acids [84]. The empirically derived

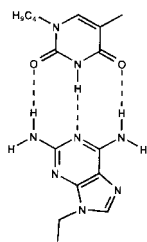
increments are approximately in line with those predicted from the Bjerrum theory, as evident also from related correlations of the $\lg K$ values as a function of the charge products $z_A z_B$ [4]. Noticeably, with both correlations the same $\Delta\Delta G$ values are obtained for hard and small as for large and polarizable organic ions, or those with significant charge delocalization, such as phenolates [25].

Supramolecular complexes with aromatic components are often stabilized by Coulombic attraction of electron-poor and electron-rich π -systems and are usually called donor-acceptor complexes in analogy to Lewis-type complexes between acids and bases. The binding mechanisms are sometimes difficult to distinguish from dispersive van der Waals charge-transfer and sometimes from hydrophobic interactions. However, the strength of complexes, measured in aprotic solvents, with, e.g., electron-rich molecular clips containing naphthalene side walls, strongly increases with electron-withdrawing substituents on the enclosed phenyl guest compound, in line with calculated electron densities of the π -systems [85]. Similar characteristics hold for other synthetic complexes, including many rotaxanes and catenanes, which form the basis of intriguing models of supramolecular machines and of devices for energy transfer, for information transmission and storage and other possible applications [86–88].

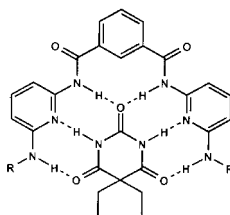
2.9.2

Hydrogen Bonds

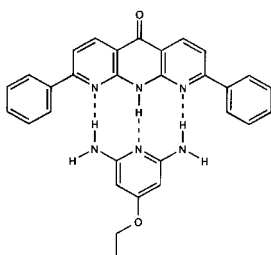
Hydrogen bonds involving amide or amide-type functions as donor, D, and acceptor, A, form the basis of many synthetic and biological complexes. Synthetic host-guest complexes of the type shown in Fig. 2.18 allow the use of well-defined conformations and measuring conditions [89]. Their analysis has given consistent values for the free energy contributions $\Delta\Delta G$ of each participating hydrogen bond. Stability constants of many simple complexes with barbiturates and other model compounds with amide functions yielded an average value $\Delta\Delta G = 5$ kJ/mol per bond in chloroform [90]. However, the examples I–V in Fig. 2.18 already indicate that the total binding energies ΔG are only approximately a function of the number n of hydrogen bonds in each complex [91]. Thus, the complexes I, II, and III, all with three bonds, exhibit all a much lower stability than V, with only two bonds. As first shown by calculations of the partial charges involved in Watson-Crick base pairs [92], the reason for weaker bonds is due to often repulsive secondary interactions: if a positively charged donor D is flanked by a negatively charged acceptor A there must be an unfavorable repulsion between opposing DA functions (broken lines in Fig. 2.2). Complexes I, II, and III all represent ADA-DAD-combinations, while V stands for a DD-AA case and gains from the additional secondary interactions. The nucleobase G-C base pair (DDA-AAD) has only one of these repulsive secondary interactions, and, therefore, a more than 50% greater $\Delta\Delta G$ value is observed in comparison to the A-T base pair (an AD-DA combination). Synthetic combinations bearing more A groups at one side and D groups at the other side show a correspondingly higher affinity, which is increased by the secondary interactions. Surprisingly, one can describe the total



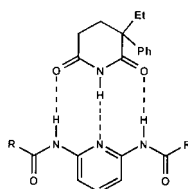
I Rich 1967: **12**
calc. **12.1**



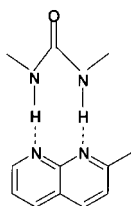
II Hamilton 1988: **24.5**
calc. **24.2**



III Zimmermann 1993: **10.5**
calc. **12.1**



IV Schneider 1989: **13.0**
calc. **12.1**



V Schneider 1996: **21.1**
calc. **21.6**

Fig. 2.18 Examples of recognition models by hydrogen bonds with experimental association free energy $-\Delta G$ (in CDCl_3 , [kJ/mol]) and $-\Delta G$ values (*in italics*) calculated with $\Delta\Delta G$ incre-

ments of 7.9 kJ/mol for primary and 2.9 kJ/mol for secondary interactions (see text and refs. [4], [91]).

binding energies quite well with additive ΔG contributions, which in chloroform are for the primary interaction 7.9 kJ/mol, and for the secondary interaction 2.9 kJ/mol, regardless of whether the latter is attractive (as in AAA-DDD combinations) or repulsive (as in the more frequently occurring non-homogenous combinations). These increments have been derived from the analysis of 58 complexes

in chloroform and reproduce the experimental data with deviations of ± 1.8 kJ/mol or less, as illustrated with a few complexes in Fig. 2.18. The ΔG values are almost twice as large in tetrachloromethane compared to the weak donor medium chloroform and become almost zero upon addition of only 10% methanol [90]. This, and the strong effect of neighboring functions on the donor/acceptor strength of hydrogen bond functions [93], sheds light on the problem of applying related scoring factors to biopolymers.

The acceptor and donor strength of many functions besides those of the amide type have been characterized by the analysis of associations between simple molecules, such as, e.g., phenols and anilines, for which thousands of experimental data exist, mostly measured in chloroform or in carbon tetrachloride [94, 95]. Although these data are hampered by less well-defined structures compared to supramolecular complexes, they not only give a fairly consistent basis for the prediction of hydrogen-bonded associations but also can be used, e.g., for crown ether and cryptand complexes with alkali or ammonium ligands [32].

Hydrogen bonds also play an important role in anion binding, both in proteins [96] and in recently developed artificial receptors [97]. Systematic association measurements with model amides (Fig. 2.19) in chloroform show binding increments (Tab. 2.1) between a single amide group and different anions, which are approximately additive [98]. The ΔG values for chloride complexation increase from monodentate to bidentate to tridentate hosts (Fig. 2.19, 1–3), i.e., from 6 to 12 to 18 kJ/mol, respectively. Noticeable deviation from additivity is observed if an an-

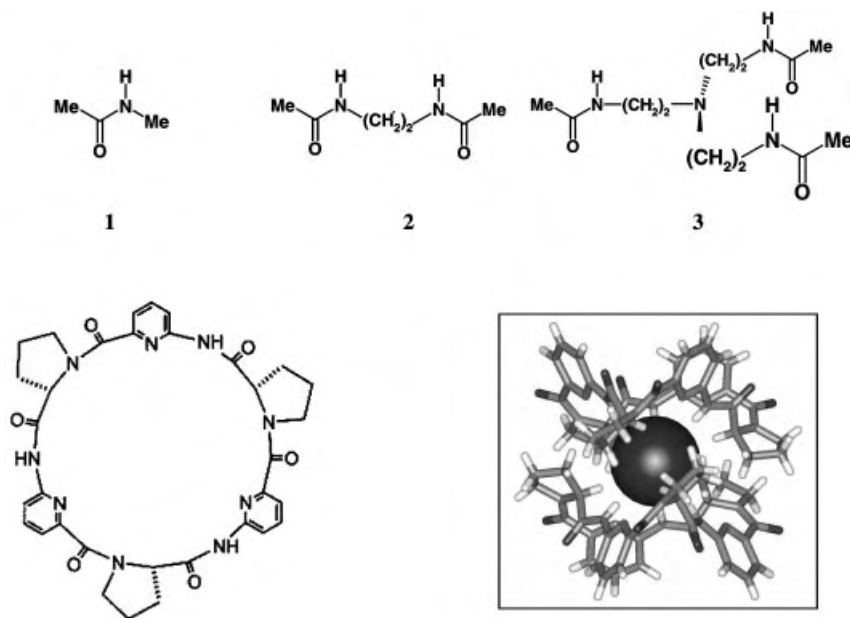


Fig. 2.19 Amides as receptors for anions.

Tab. 2.1 Complexation free energies ($-\Delta G$, kJ/mol)^{a)}. (a) For anions with model amides **1–3**; (b) Of these anions with carbohydrate models **C** and **G**

	Cl^-	Br^-	$C_6H_5CO_2^-$	$R_2PO_4^-$
1	5.7	4.6	6.6	8.4
2	11.6	7.2	14.4	–
3	18.2	12.6	14.0	–
C	2.1	1.4	2.1	1.7
G	3.75	2.9	4.1	3.5

a) In $CDCl_3$ with tetraalkylammonium salts; measurements for the mono-, di-, and tridentate amides **1,2,3** (Fig. 2.19) with $H_2PO_4^-$; for the carbohydrate models **C**: *trans*-1,2-cyclohexanediol and **G**: *n*-dodecyl β -D-galactopyranoside or β -D-glucopyranoside with $(C_6H_5O)_2PO_2^-$.

ion-like carboxylate can take advantage of only two hydrogen bonds, for which reason there is no ΔG increase with receptor **3** in this case. Urea or thiourea components in anion receptors can build up twice as many hydrogen bonds per unit and show considerable affinities in the more competitive solvent DMSO. With macrocyclic oligopeptides in which all N-H groups are pre-organized to point to an anion in the center, considerable affinities can be achieved even in aqueous medium [99]. The cyclopeptide in Fig. 2.19 yields with iodide a stable 1:2 complex in solution as well as in the solid state, with six hydrogen bonds directed towards the iodide anion.

The affinities observed for complexes between amides and anions are remarkably parallel to those found for the interactions between such anions and carbohydrate models. The data in Tab. 2.1 show the same affinity increase in the sequence $I^- < Br^- < Cl^- < RCOO^-$ [100]; the carboxylate is again a particularly strong acceptor as the result of two geometrically matching hydrogen bonds with vicinal diols. The formation of two almost linear and parallel hydrogen bonds is also responsible for the efficiency of the guanidinium residue for carboxylate complexation in artificial receptors [101] as well as in proteins (cf. Chapter 6) [102].

2.9.3

Weak Hydrogen Bonds: The Use of Intramolecular “Balances”

In this section we will briefly discuss non-covalent interactions that are usually too weak to be measured directly in host-guest equilibria and instead have been studied with the help of “balances,” in which the influence of non-covalent interactions on sensitive conformational equilibria is studied. Vibrational spectroscopy with haloalkanes and 4-nitrophenol in carbon tetrachloride has revealed single hydrogen bond energies reaching for iodine to fluorine as acceptors, e.g., from about 4 to 7 kJ/mol [103]. Weak hydrogen bonds with, e.g., C-H bonds as donor have been identified largely on the basis of extensive analyses of solid-state structures [104, 105] but also with computational methods, e.g., in nucleic acids [106]

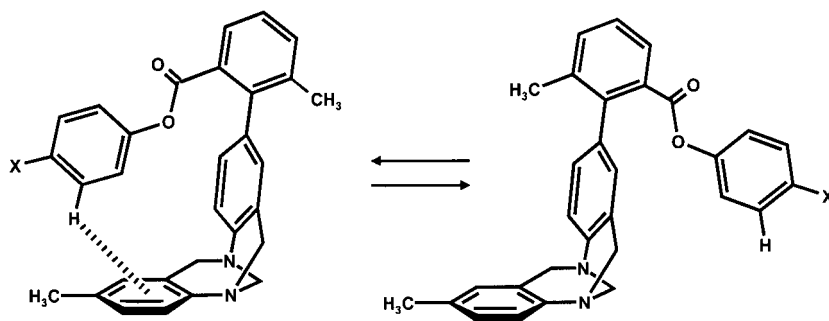


Fig. 2.20 An intramolecular “balance” for measuring CH- π interactions (see [105]).

and in proteins [107]. Edge-to-face interactions of arene systems, a frequent motif also in proteins [108], are prototypes of hydrogen bonds between weakly acidic aromatic C-H bonds and π -moieties [105]. They have been quantified in solution with the help of conformational balances [109, 110], as illustrated in Fig. 2.20. There, the energy advantage of the folded conformer is shifted from $\Delta G=1.0$ kJ/mol for the substituent $X=H$ to $\Delta G=2.9$ kJ/mol for $X=NO_2$ [110].

2.9.4

Polarization Effects

Cation- π interactions were first identified in alkali metal ion complexes of a Coulombic nature [111]. The high-order effects in organic and biologically important systems between onium ions and π -moieties have been discovered by the crucial role of ammonium ions opposing aryl groups in water-soluble cyclophane complexes, where a larger hydrophobicity of electroneutral components other than expected lead to a smaller binding strength [112]. Polarization induced on the π -system plays an essential role in complexes with onium ions; this is evident from the observation that anions, which also can lead to polarization, also show complexation with aromatic clefts [113]. The binding free energy increment for an ammonium-benzene interaction was estimated to amount to about 2 kJ/mol from an analysis of associations between aromatic ion pairs (Fig. 2.4) [30].

2.9.5

Dispersive Interactions

Van der Waals interactions are the most difficult ones with respect to both theoretical and experimental evaluations. Computational descriptions need to include polarization functions and solvent effects [114, 115]. Experimentally determined stability constants also may be due to electrostatic effects, in particular with stacking between aromatic units, and to solvents effects. The latter may dominate in water, which at the same time is the most suitable medium for dispersive interactions due to its low molar polarizability. The problems are most evident with recent in-

interpretations of nucleobase stacking, which probably is the biologically most important manifestation of intermolecular interactions with aromatic moieties. While hydrophobic effects were proposed to dominate, mostly on the basis of solvent effects [116, 117], other results speak for the predominance of polarization effects [118].

Again, most of these weak interactions have been identified in the solid state [104], including, e.g., those with halogen atoms [119]. The underlying energy components are experimentally accessible in principle by measurements with conformational balances such as those shown in Fig. 2.21 [120]. For the substituent $R=H$, one observes the same small preference $\Delta G_{E/Z}$ for the E conformer in water (D_2O) as in chloroform. This is different with substituents R in the para-position of the phenyl ring, which can reach the naphthyl moiety in the E conformer. In water there is an increase of $\Delta G_{E/Z}$, which is *larger* for $R=\text{phenyl}$ than for $R=\text{cyclohexyl}$, although the higher hydrophobicity of cyclohexane compared to benzene would in contrast speak for a larger hydrophobic effect. Noticeably, in chloroform the $\Delta G_{E/Z}$ values are independent from the different substituents R . The same evidence for dispersive instead of hydrophobic interactions between aromatic systems is seen if the aryl groups in the balance bear nitrogen atoms, e.g., with $R=\text{pyridyl}$, pyrimidyl, or quinolyl residues [121]. The results emphasize the propensity of heteroaromatic systems for stacking like, e.g., in nucleic acids. However, the observed variations are quite small, with, e.g., $-\Delta G_{E/Z}=1.8$ vs. 3.3 kJ/mol for $R=\text{phenyl}$ or $R=\text{pyrimidyl}$, respectively.

Analysis of porphyrin complexes with a large range of substrates has allowed for the first time the quantification of intermolecular dispersive interactions in model complexes in aqueous solution and their differentiation from hydrophobic effects. The examples shown in Fig. 2.22 [122] demonstrate that, in line with the results discussed above with the balance (Fig. 2.20) [120, 121], hydrophobic contributions of aliphatic groups are small in comparison to substituents bearing electron lone pairs or multiple bonds. In accordance with this observation, benzoate has a much larger affinity to positively charged, water-soluble porphyrins than does cyclohexane-carboxylate, even though the latter has the same surface size and is more hydrophobic [42c]. Unsaturated substrates exhibit association energies that are a linear function of the number of double bonds; cyclopropanes behave more as olefinic than as aliphatic substituents. The binding free energy contributions $\Delta\Delta G$ for the different compounds are independent of the substituent location on aliphatic or aromatic frameworks. Furthermore, they are additive within the error unless the substituents are in a vicinal position. Such a deviation is nor-

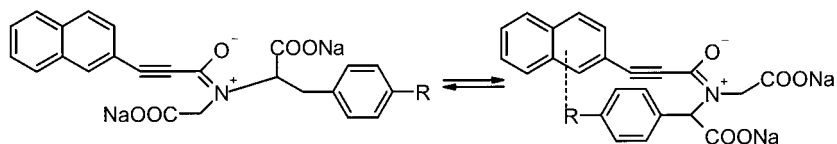


Fig. 2.21 An intramolecular “balance” for measuring stacking interactions (see [114, 115]).

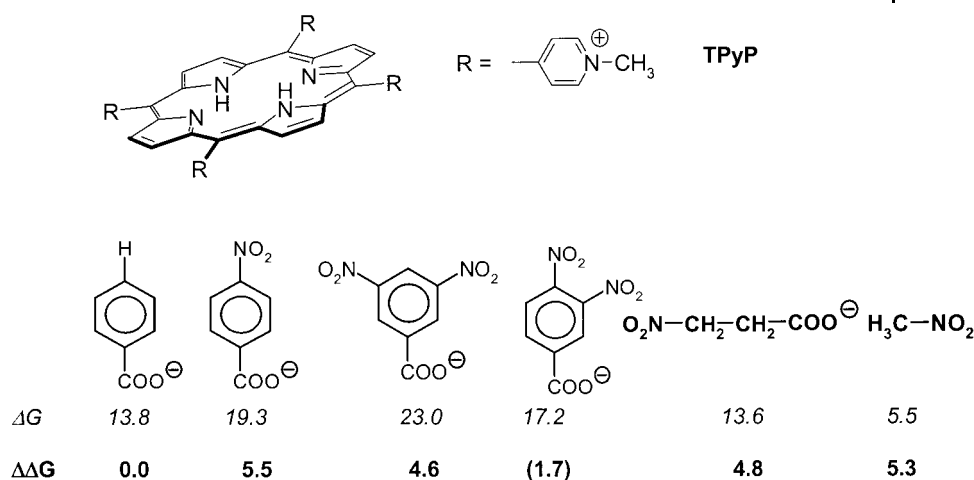


Fig. 2.22 Dispersive interactions measured with tetrapyrroline porphyrin **TPyP** in water; binding energies ΔG [kJ/mol, *italics*], with in-

crements $\Delta\Delta G$ obtained by subtracting ΔG values measured with the substituted ligand from that of the unsubstituted (see [122]).

mal in free energy correlations and, in the case of, e.g., nitro substituents, is due to steric interference between the groups. Measurements of over 50 complexes with positively or negatively charged porphyrins can be used to extract substituent increments $\Delta\Delta G$, which quite accurately describe the observed free energies. Because both electron-withdrawing and electron-pushing substituents on benzene rings increase the affinity in the same way as positively or negatively charged porphyrins, and because the affinity increases with the polarizability of the groups, other mechanisms besides dispersive interactions can be excluded. Of particular importance for protein interactions are the relatively large $\Delta\Delta G$ values found for sulfur and for amide groups; in line with this, one observes a regular affinity increase with the number of amino acids in oligoglycines. Organic solvents such as methanol lead to a strong decrease of binding energies, which is linear in the volume percent of organic solvent in a binary mixture with water [122]. The group contribution $\Delta\Delta G$ values have to be taken as relative numbers, very much like substituent constants in, e.g., the Hammett equation. Their absolute magnitude will depend on the size of the acceptor molecule, which in the case of porphyrins is several times larger than that of a single benzene unit, and will change with the reaction medium, including the salt concentration. Dispersive interactions, which can be scaled with the help of model compound studies, are believed to play an important role in protein folding [123], beyond the usually considered hydrophobic forces [124].

2.10

Conclusions

The mechanisms ruling interactions between molecules are most clearly visible in synthetic complexes of moderate size. The analysis of large biomolecular aggregations is hampered by many simultaneously operating forces that are not independent of each other and by problems in determining the accompanying structural details in solution. Measurements of associations between simple molecules with only one binding site have yielded an often overlooked wealth of valuable data [94, 95] but suffer from less well-defined structural organization and from restrictions with respect to the number of possible interactions. Highly pre-organized synthetic host-guest models can be designed to provide detailed insight into all kind of possible mechanisms responsible for molecular recognition. They reveal the additivity of binding free energy increments, as well as its limitations, and provide numbers that can be applied to more complex large systems. Medium effects such as the influence of ionic strength or solvent changes can be analyzed in detail with synthetic complexes. Secondary interactions have been identified and scaled in complexes where interacting groups are in close proximity, such as in nucleobase associations where, in contrast to peptides, different hydrogen bond donor and acceptor sites come close. The energy factorizations available from the analysis of synthetic model systems can be used to test computational methods, to further develop and parameterize force fields, and to evaluate interactions in and with large biopolymers.

2.11

References

- 1 J.-M. LEHN, *Supramolecular Chemistry: Concepts and Perspectives*, Wiley-VCH Weinheim etc, **1995**.
- 2 F. VÖGTLE, *Supramolecular Chemistry: An Introduction*, Wiley New York, **1993**.
- 3 P. D. BEER, P. A. GALE, D. K. SMITH, *Supramolecular Chemistry* (Oxford Chemistry Primers, 74) Oxford University Press, Oxford, **1999**.
- 4 H.-J. SCHNEIDER, A. YATSIMIRSKI, *Principles and Methods in Supramolecular Chemistry*, Wiley, Chichester etc, **2002**.
- 5 J. W. STEED, J. L. ATWOOD, *Supramolecular Chemistry*, Wiley, Chichester etc, **2000**.
- 6 J. M. LEHN, J. L. ATWOOD, J. W. D. DAVIES, D. D. MACNICOL, F. VÖGTLE (Eds.) *Comprehensive Supramolecular Chemistry*, Vol. 1–11, Pergamon/Elsevier Oxford etc, **1996**.
- 7 A. D. HAMILTON (Ed.) *Supramolecular Control of Structure and Reactivity (Perspectives in Supramolecular Chemistry)*, Wiley, New York etc. **1996**.
- 8 For historical perspectives of the lock-and-key principle see: F. CRAMER, *Pharm. Acta Helv.* **1995**, 69, 193; R. U. LEMIEUX, U. SPOHR, *Adv. Carbohydr. Chem. Biochem.* **1994**, 50, 1; F. W. LICHTENTHALER, *Angew. Chem., Int. Ed. Engl.* **1994**, 33, 2364; A. ESCHENMOSER, *ibid.* **1994**, 33, 2363.
- 9 For a recent summary on non-covalent interactions in host-guest complexes see ref. 4.
- 10 For a recent review with biologically relevant examples see e.g., M. MAMMEN, S.-K. CHOI, G. M. WHITESIDES *Angew. Chem. Int. Ed. Engl.* **1998**, 37, 2749.

- 11 M.W. HOSSEINI, J.-M. LEHN, M.P. MERTES, *Helv. Chim. Acta* **1983**, 66, 2454.
- 12 Y. TOR, J. LIBMAN, A. SHANZER, C.E. FELDER, S. LIFSON, *J. Am. Chem. Soc.*, **1992**, 114, 6661; see also Z. HOU, C.J. SUNDERLAND, T. NISHIO, K.N. RAYMOND, *J. Am. Chem. Soc.*, **1996**, 118, 5148; P. STUTTE, W. KIGGEN, F. VÖGTLE, *Tetrahedron*, **1987**, 43, 2065.
- 13 J.H. RAO, J. LAHIRI, L. ISAACS, R.M. WEIS, G.M. WHITESIDES, *Science* **1998**, 280, 708.
- 14 V. HEGDE, C.-Y. HUNG, P. MADHUKAR, R. CUNNINGHAM, T. HÖPFNER, R.P. THUMMEL, *J. Am. Chem. Soc.*, **1993**, 115, 872.
- 15 R. BRESLOW, B.L. ZHANG, *J. Am. Chem. Soc.* **1996**, 118, 8495.
- 16 HAMASAKI, K., USUI, S., IKEDA, H., IKEDA, T., UENO, A. *Supramol. Chem.*, **1997**, 8, 125S.
- 17 D. STARNES, D.M. RUDKEVICH, J. REBEK, JR. *J. Am. Chem. Soc.* **2001** 123, 4659.
- 18 D.J. CRAM, *Angew. Chem., Int. Ed. Engl.*, **1986**, 25, 1039.
- 19 R.M. ROSENBERG, *Chemical Thermodynamics*, 5th edition, Wiley, New York, **1994**, chapter 22; N. COHEN, S.W. BENSON, *Chem. Rev.*, **1993**, 93, 2419; P. GIANINI, L. LEPROT, *J. Solut. Chem.*, **1996**, 25, 1.
- 20 V. GUTMANN, *The Donor-Acceptor Approach to Molecular Interactions*, Plenum Press, New York, **1978**; R.S. DRAGO, *Structure and Bonding*, Springer Heidelberg, **1973**, 73.
- 21 See also W.P. JENCKS, *Proc. Nat. Acad. Sci. USA* **1981**, 78, 4046.
- 22 (a) W.P. JENCKS, *Catalysis in Chemistry and Enzymology*, McGraw Hill, New York, **1989**; (b) W.P. JENCKS, *Adv. Enzymol.*, **1975**, 43, 219; M.I. PAGE, *Chem. Soc. Rev.*, **1973**, 2, 295; (c) W.P. JENCKS, *Proc. Natl. Acad. Sci. USA*, **1978**, 78, 4046
- 23 A.W. ADAMSON, *J. Am. Chem. Soc.*, **1954**, 76, 1578; new reviews see e.g., R.D. HANCOCK, A.E. MARTELL, *Chem. Rev.* **1989**, 89, 1875
- 24 W. KAUZMANN, *Adv. Protein Chem.*, **1959**, 14, 1.
- 25 H.-J. SCHNEIDER, *Angew. Chem., Int. Ed. Engl.*, **1991**, 30, 1417; H.-J. SCHNEIDER, *Chem. Soc. Rev.*, **1994**, 22, 227.
- 26 H.-J. BÖHM, M. STAHL *Med. Chem. Res.*, **1999**, 9, 445, and references cited therein.
- 27 R.P. ANDREWS, D.J. CRAIK, J.-L. MARTIN, *J. Med. Chem* **1984**, 27, 1648.
- 28 See also T.J. SAGE, C.R. STOUT, R.M. STROUD, *Structure* **1998**, 6, 839.
- 29 G. KLEBE, H. J. BÖHM, *J. Recept. Signal Transduct. Res.* **1997**, 17, 459; F. BRÜGE, S.L. FORNILI, G.G. MALENKOV, M.B. PALMAVITTORELLI, M.U. PALMA, *Chem. Phys. Lett.* **1996**, 254, 283.
- 30 H.-J. SCHNEIDER, T. SCHIESTEL, P. ZIMMERMANN, *J. Am. Chem. Soc.*, **1992**, 114, 7698.
- 31 J.-M. LEHN, J.P. SAUVAGE, *J. Am. Chem. Soc.*, **1975**, 97, 6700.
- 32 H.-J. SCHNEIDER, V. RÜDIGER, O.A. RAEVSKY, *J. Org. Chem.*, **1993**, 58, 3648
- 33 After data from Y. INOUE, G.W. GOKEL, (Eds.) *Cation Binding by Macrocycles*, Marcel Dekker, New York, **1990**; R.M. IZATT, K. PAWLAK, J.S. BRADSHAW, R.L. BRUENING, *Chem. Rev.*, **1991**, 91, 1721, and references cited therein.
- 34 O.A. RAEVSKY, V.P. SOLOV'EV, A.F. SOLOTOVNOV, H.-J. SCHNEIDER, V. RÜDIGER, *J. Org. Chem.*, **1996**, 61, 8113.
- 35 G.J. KIRKOVITS, J.A. SHRIVER, P.A. GALE, J. L. SESSLER, *J. Incl. Phenom. Macrocycl. Chem.* **2001**, 41, 69.
- 36 See N. PELIZI, A. CASNATI, A. FRIGGERI, R. UNGARO, *J. Chem. Soc.-Perkin Trans. 2* **1998**, 1307; L.A.J. CHRISSTOFFELS, F. DE JONG, D.N. REINHOUDT, S. SIVELLI, L. GAZZOLA, A. CASNATI, R. UNGARO, *J. Am. Chem. Soc.* **1999**, 1212, 10142, and references cited therein.
- 37 J.M. MAHONEY, A.M. BEATTY, B.D. SMITH, *J. Am. Chem. Soc.* **2001**, 123, 5847
- 38 M.T. REETZ, J. HUFF, J. RUDOLPH, K. TÖLLNER, A. DEEGE, R. GODDARD, *J. Am. Chem. Soc.* **1994**, 116, 11588.
- 39 K. KAVALLIERATOS, B.A. MOYER *J. Chem. Soc., Chem. Commun.* **2001**, 1620, and references cited therein.
- 40 C.R. BERTOZZI, L.L. KIESSLING, *Science* **2001**, 291, 2357–2364; L.L. KIESSLING, J.E. GESTWICKI, L.E. STRONG *Curr. Opin. Struct. Biol.* **2000**, 4, 696.
- 41 H.-J. SCHNEIDER, I. THEIS, *Angew. Chem. Int. Ed. Engl.* **1989**, 28, 753.
- 42 (a) R.F. PASTERNAK, E.J. GIBBS, A. GAUDEMER, A. ANTEBI, S. BASSNER, DE L. POY, D.H. TURNER, A. WILLIAMS, F. LAPLACE, M.H. LANSARD, C. MERIENNE,

- PERREE- M. FAUVET, *J. Am. Chem. Soc.* **1985**, 107, 8179; (b) V. KRAL, A. ANDRIEVSKY, J. L. SESSLER, *Chem. Com.* **1995**, 2349; (c) H.-J. SCHNEIDER, M. WANG, *J. Org. Chem.*, **1994**, 59, 7464.
- 43 P. ČUDIĆ, M. ŽINIĆ, V. TOMIŠIĆ, V. SIMÉON, J.-P. VIGNERON, J.-M. LEHN, *Chem. Commun.* **1995**, 1073; I. PIANTANIDA, V. TOMIŠIĆ, M. ŽINIĆ, *J. Chem. Soc., Perkin Trans. 2*, **2000**, 375.
- 44 M. SIRISH, H.-J. SCHNEIDER, *J. Am. Chem. Soc.* **2000**, 112, 5881.
- 45 A. E. MARTELL, R. D. HANCOCK, R. J. MOTEKAITIS, *Coord. Chem. Rev.*, **1994**, 133, 39.
- 46 R. D. HANCOCK, *Progr. Inorg. Chem.*, **1989**, 37, 187.
- 47 J. S. BRODBELT, D. V. DEARDEN, in *Comprehensive Supramolecular Chemistry*, Vol. 8 (J. E. D. DAVIES, J. A. RIPLESTER, Eds.), Pergamon/Elsevier Oxford etc, **1996**, 567, ff; M. PRZYBYLSKI, M. O. GLOKER, *Angew. Chem., Int. Ed. Engl.*, **1996**, 35, 806; R. D. SMITH, J. E. BRUCE, Q. WU, Q. P. LEI, *Chem. Soc. Rev.*, **1997**, 26, 191.
- 48 L. TROXLER, G. WIPFF, *J. Am. Chem. Soc.* **1994**, 116, 1468.
- 49 Value after G. GRITZNER, *Pure Appl. Chem.* **1988**, 60, 1743.
- 50 Data see Inoue, Y. GOKEL, G. W. (Eds.) *Cation Binding by Macrocycles*, Marcel Dekker, New York, **1990**; R. M. IZATT, K. PAWLAK, J. S. BRADSHAW, R. L. BRUENING, *Chem. Rev.*, **1991**, 91, 1721, and references cited therein.
- 51 C. T. CALDERONE, D. H. WILLIAMS, *J. Am. Chem. Soc.* **2001**, 123, 6262.
- 52 J. D. DUNITZ, *Chem. Biol.* **1995**, 2, 709.
- 53 D. H. WILLIAMS, M. S. WESTWELL, *Chem. Soc. Rev.* **1998**, 27, 57.
- 54 See (a) O. EXNER, *Progr. Phys. Org. Chem.*, **1973**, 10, 411; (b) R. LUMRY, S. RAJENDER, *Biopolymers*, **1970**, 9, 1125; (c) W. LINERT, *Chem. Soc. Rev.*, **1994**, 23, 429; (d) K. SHARP, *Protein Science* **2001**, 10, 661.
- 55 See data in E. E. TUCKER, S. D. CHRISTIAN, *J. Am. Chem. Soc.*, **1984**, 106, 1942, and in ref 4, p. 106.
- 56 L. GAREL, B. LOZACH, J.-P. DUTASTA, A. COLLET, *J. Am. Chem. Soc.*, **1993**, 115, 11652.
- 57 H.-J. SCHNEIDER, T. BLATTER, S. SIMOVA, *J. Am. Chem. Soc.*, **1991**, 113, 1996.
- 58 a) D. H. WILLIAMS, M. S. SEARLE, M. S. WESTWELL, V. MACKAY, P. GROVES, D. A. BEAUREGARD, *CHEMTRACTS- Organic Chemistry* **1994**, 7, 133 b) M. S. SEARLE, D. H. WILLIAMS, U. GERHARD, *J. Am. Chem. Soc.* **1992**, 114, 10697; c) P. GROVES, M. S. SEARLE, M. S. WESTWELL, D. H. WILLIAMS, *Chem. Commun.*, **1994**, 1519, and references cited therein.
- 59 F. EBLINGER, H.-J. SCHNEIDER, *Angew. Chem. Int. Ed. Engl.*, **1998**, 37, 826.
- 60 J. S. NOWICK, J. M. CARY, J. H. TSAI, *J. Am. Chem. Soc.* **2001**, 123, 5176.
- 61 M. MAMMEN, E. I. SHAKHNOVICH, J. M. DEUTCH, G. E. WHITESIDES *J. Org. Chem.* **1998**, 63, 3168.
- 62 A. MD. HOSSAIN, H.-J. SCHNEIDER, *Chem. Eur. J.*, **1999**, 5, 1284.
- 63 L. D. PETTIT, J. M. L. SWASH, *J. Chem. Soc., Dalton Trans.* **1977** 697; b) P. TILUS, *Finn. Chem. Lett.* **1979**, 76.
- 64 H.-J. BÖHM, *J. Comp.-Aided Mol. Design*, **1994**, 8, 243–256, 623.
- 65 G. W. GOKEL, *Chem. Soc. Rev.*, **1992**, 20, 39.
- 66 J. J. LAVIGNE, E. V. ANSLYN, *Angew. Chem. Int. Ed. Engl.* **2001**, 40, 3119.
- 67 A. CASNATI, A. POCHINI, R. UNGARO, C. BOCCHI, F. UGOZZOLI, R. J. M. EGBERINK, H. STRUIJK, R. LUGTENBERG, F. DE JONG, D. N. REINHOUDT, *Chem. Eur. J.* **1996**, 2, 436 (Published in: *Angew. Chem. Int. Ed. Engl.* **1996**, 35).
- 68 A. MD. HOSSAIN, H.-J. SCHNEIDER, *J. Am. Chem. Soc.*, **1998**, 120, 11208.
- 69 (a) D. CRAM, J. M. CRAM, *Acc. Chem. Res.*, **1978**, 11, 9; for related approaches see (b) V. PRELOG, *Angew. Chem., Int. Ed. Engl.*, **1989**, 28, 114; (c) T. M. GEORGIADIS, M. M. GEORGIADIS, F. DIEDERICH, *J. Org. Chem.*, **1991**, 56, 3362.
- 70 A. LEVITZKI, *Quantitative Aspects of Allosteric Mechanism*, Springer, New York, **1978**.
- 71 D. E. KOSHLAND JR., *Angew. Chem., Int. Ed. Engl.*, **1995**, 33, 2475.
- 72 E. DI CERA, *Adv. Protein Chem.* **1985**, 51, 59; E. DI CERA, *Chem. Rev.* **1998**, 98, 1563.
- 73 M. F. PERUTZ, *Quart. Rev. Biophysics* **1989**, 22, 139.
- 74 T. E. CREIGHTON, *Protein Folding*, Freeman, New York etc **1992**.

- 75 A.R. DINNER, A. SALI, L.J. SMITH, C.M. DOBSON, M. KARPLUS, *Trends Biochem. Sci.* **2000**, 25, 331; C.M. DOBSON, A. SALI, M. KARPLUS, *Angew. Chem.-Int. Edit.* **1998**, 37, 868; S. VAJDA, M. SIPPL, J. NOVOTNY, *Curr. Opin. Struct. Biol.* **1997**, 7, 222.
- 76 See e.g., J.W. CANARAY, B.C. GIBB, *Progr. Inorg. Chem.* **1997**, 45, 1.
- 77 M. TAKEUCHI, M. IKEDA, A. SUGASAKI, S. SHINKAI, *Accounts Chem. Res.* **2001**, 34, 865.
- 78 S. SHINKAI, *Crown Ethers and Analogous Compounds*, M. HIRAOKA, Ed., **1992**, 45, 335, Elsevier; J. REBEK, *Acc. Chem. Res.*, **1984**, 17, 258.
- 79 H.-J. SCHNEIDER, D. RUF, *Angew. Chem., Int. Ed. Engl.*, **1990**, 29, 1159.
- 80 G.H. CZERLINSKI, *Biophys. Chem.*, **1989**, 34 169.
- 81 Examples for negative cooperativity see: J.C. RODRIGUEZ-UBIZ, O. JUANEZ, E. BRUNET, *Tetrahedron Lett.*, **1994**, 35, 8461; H.-J. SCHNEIDER, F. WERNER, *J. Chem. Soc., Chem. Commun.*, **1992**, 490.
- 82 H. CHEN, M.L. PRIVALSKY, *Proc. Nat. Acad. Sci. USA* **1995**, 92,422.
- 83 R.X. WANG, L. LIU, L.H. LAI, Y.Q. TANG, *J. Mol. Model.* **1998**, 4, 379.
- 84 H.-J. SCHNEIDER, T. BLATTER, *Angew. Chem., Int. Ed. Engl.*, **1992**, 31, 1207.
- 85 F.-G. KLÄRNER, U. BURKERT M. KAMIETH R. BOESE *J. Phys. Org. Chem.*, **2000**, 13, 604; F.-G. KLÄRNER, J. PANITZKY, D. BLÄSER, R. BOESE *Tetrahedron* **2001**, 57, 3673, and references cited therein.
- 86 P.R. ASHTON, V. BALDONI, V. BALZANI, A. CREDI, H.D.A. HOFFMANN, M.V. MARTINEZ-DIAZ, F.M. RAYMO, J.F. STODDART, M. VENTURI, *Chem.-Eur. J.* **2001**, 7, 3482., and references cited therein.
- 87 V. BALZANI, A. CREDI, M. VENTURI, *Proc. Natl. Acad. Sci. U.S.A.* **2002**, 99, 4814.
- 88 M.B. NIELSEN, C. LOMHOIT, J. BECHER, *Chem. Soc. Rev.* **2000**, 29, 153, and references cited therein.
- 89 (a) S.C. ZIMMERMAN, *Top. Curr. Chem.*, **1993**, 165, 71; (b) A.D. HAMILTON, *Adv. Supramol. Chem.*, **1991**, 1, 1; (c) J. REBEK, JR., *Acc. Chem. Res.*, **1990**, 23, 399.
- 90 H.-J. SCHNEIDER, R.K. JUNEJA, S. SIMOVA, *Chem. Ber.*, **1989**, 112, 1211.
- 91 J. SARTORIUS, H.-J. SCHNEIDER, *Chemistry-Eur. J.*, **1996**, 2, 1446.
- 92 J. PRANATA, S.G. WIERSCHKE, W.L. JORGENSEN, *J. Am. Chem. Soc.*, **1991**, 113, 2810.
- 93 S. SHAN, D. HERSCHLAG, *J. Am. Chem. Soc.*, **1996**, 118, 5515
- 94 M.H. ABRAHAM, *Chem. Soc. Rev.*, **1993**, 22, 73; O.A. RAEVSKY, *J. Phys. Org. Chem.*, **1997**, 10, 405, and references cited therein.
- 95 M.H. ABRAHAM, J.A. PLATTS, *J. Org. Chem.* **2001**, 66, 3484.
- 96 H. LUECKE, F.A. QUIOCHO, *Nature* **1990**, 347, 402; J.J. HE, F.A. QUIOCHO, *Science* **1991**, 251, 1497.
- 97 (a) A. BIANCHI, BOWMAN-K. JAMES, GARCIA-E. ESPANA (Eds.) *Supramolecular Chemistry of Anions* **1997**, Wiley-VCH, New York etc.; (b) P.D. BEER, P.A. GALE, *Angew. Chem., Int. Ed. Engl.* **2001**, 40, 487; (c) F.P. SCHMIDTCHEN, M. BERGER, *Chem. Rev.* **1997**, 97 1609; (d) M.M. ANTONISSE, D.N. REINHOUDT, *Chem. Commun.* **1998**, 443.
- 98 F. WERNER H.-J. SCHNEIDER, *Helv. Chim. Acta* **2000**, 83, 465.
- 99 S. KUBIK, R. GODDARD, R. KIRCHNER, D. NOLTING, J. SEIDEL, *Angew. Chem. Int. Ed. Engl.* **2001** 40, 2648.
- 100 J.M.F. COTERÓN, HACKET, H.-J. SCHNEIDER, *J. Org. Chem.*, **1996**, 61, 1429.
- 101 M. BERGER, F.P. SCHMIDTCHEN, *J. Am. Chem. Soc.* **1999**, 121, 9986–9993; T. HAAACK, M.W. PECZUH, X. SALVATELLA, J. SÁNCHEZ-QUESADA, J. DE MENDOZA, A.D. HAMILTON, E. GIRALT, *J. Am. Chem. Soc.* **1999**, 121, 11813, and references cited therein.
- 102 W.D. MORGAN, B. BIRDSALL, P.M. NIETO, A.R. GARGARO, J. FEENEY, *Biochemistry* **1999**, 38, 2127; L.S. CHEN, Z.P. ZHANG, A. SCAFONAS, R.C. CAVALLI, J.L. GABRIEL, K.J. SOPRANO, D.R. SOPRANO, *J. Biol. Chem.* **1995**, 270, 4518, and references cited therein.
- 103 C. OUVARD, M. BERTHELOT, C. LAURENCE, *J. Phys. Org. Chem.* **2001**, 14, 804, C. LAURENCE, P. NICOLET, M.T. DALATI, J.L. M. ABBOUD, R. NOTARIO, *J. Phys. Chem.* **1994**, 98, 5807.
- 104 G.R. DESIRAJU, T. STEINER, *The Weak Hydrogen Bond in Structural Chemistry and Biology*, Oxford Univ. Press, Oxford, **1999**.

- 105 M. NISHIO, M. HIROTA, Y. UMEZAWA, *The CH-Interaction*, Wiley, New York etc **1998**.
- 106 J. SÜHNEL, *Biopolymers* **2001**, 61, 32–51; M. BRANDL, M. MEYER, J. SÜHNEL, *J. Biomol. Struct. Dyn.* **2001**, 18, 545.
- 107 M. BRANDL, M. S. WEISS, A. JABS, J. SÜHNEL, R. HILGENFELD, *J. Mol. Biol.* **2001**, 307, 357.
- 108 S. K. BURLEY, G. A. PETSKO, *Adv. Protein Chem.*, **1988**, 39, 125.
- 109 M. KARATSU, H. SUEZAWA, K. ABE, M. HIROTA, M. NISHIO, E. OSAWA, *Tetrahedron*, **1983**, 39, 3091.
- 110 S. PALIWAL, S. GEIB, C. S. WILCOX, *J. Am. Chem. Soc.*, **1994**, 116, 4497.
- 111 (a) N. G. ADAMS, L. M. BABCOCK, Eds., *Advances in Gas Phase Ion Chemistry*, JAI Press, C. T. Greenwich, Vol.1, **1992** Vol. 2, **1996**; (b) M. T. BOWERS, Ed., *Gas Phase Ion Chemistry*, Academic Press, New York, Vol., 1+2, **1979**, Vol 3, **1984**.
- 112 J. C. MA, J. C. D. A. DOUGHERTY, *Chem. Rev.* **1997**, 97, 1303; T. J. SHEPPOD, M. A. PETTI, D. A. DOUGHERTY *J. Am. Chem. Soc.* **1988**, 110, 1983; F. DIEDERICH, *Angew. Chem., Int. Ed. Engl.* **1988**, 27, 362; H.-J. SCHNEIDER, T. BLATTER, *Angew. Chem., Int. Ed. Engl.* **1988**, 27, 1163.
- 113 H.-J. SCHNEIDER, F. WERNER, T. BLATTER, *J. Phys. Org. Chem.* **1993**, 6, 590
- 114 For recent reviews see K. MÜLLER-DETHLEFS, P. HOBZA, *Chem. Rev.*, **2000**, 100, 143; D. FELLER, E. R. DAVIDSON: *Basis Sets for Ab Initio Molecular Orbital Calculations and Intermolecular Interactions*, in *Computational Chemistry*, K. B. LIPKOWITZ, D. B. BOYD, Eds., VCH Publishers, New York, vol 17 pp. 1–43, **2001**.
- 115 See e.g. A. VARNEK, S. HELISSEN, G. WIPFF, A. COLLET, *J. Comput. Chem.* **1998**, 19, 820.
- 116 K. M. GUCKIAN, B. A. SCHWEITZER, R. X. F. REN, C. J. SHEILS, D. C. TAHMASSEBI, E. T. KOOL *J. Am. Chem. Soc.* **2000**, 122, 2213.
- 117 Y.-P. PANG, J. L. MILLER, P. A. KOLLMAN, *J. Am. Chem. Soc.* **1999**, 121, 1717.
- 118 P. HOBZA, J. SPONER *Chem. Rev.* **1999**, 99, 3247.
- 119 P. METRANGOLO, G. RESNATI, *Chem.-Eur. J.* **2001**, 7, 2511.
- 120 R. R. GARDNER, L. S. CHRISTIANSON, S. H. GELLMAN, *J. Am. Chem. Soc.* **1997**, 119, 5041.
- 121 S. L. MCKAY, B. HAPTONSTALL, S. H. GELLMAN, *J. Am. Chem. Soc.* **2001**, 123, 1244.
- 122 H.-J. SCHNEIDER, T. LIU, M. SIRISH, V. MALINOVSKI, *Tetrahedron* **2002**, 58, 779; H.-J. SCHNEIDER, T. LIU, *Angew. Chem., Int. Ed. Engl.* **2002**, 41, 1368.
- 123 A. R. FERSHT, *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding*, W. H. Freeman & Co., New York, **1999**.
- 124 See e.g., W. E. STITES, *Chem. Rev.* **1997**, 97, 1233–1250; A. D. ROBERTSON, K. P. MURPHY, *Chem. Rev.* **1997**, 97, 1251.

3

Experimental Approaches to Determine the Thermodynamics of Protein-Ligand Interactions

R. B. RAFFA

3.1

Introduction

Used appropriately and judiciously, thermodynamic parameters can offer insight into the energetics of protein-ligand interactions that is not readily attainable by other means. The utility or application of thermodynamic analysis has traditionally been considered more the domain of (bio)chemistry than biology. However, the modern recognition of an interface in the case of protein-ligand interactions, particularly when the protein is an enzyme or a drug receptor, has kindled an integration with pragmatic benefit to basic understanding and to drug-discovery efforts [1].

Because the nature of most protein-ligand interactions involves relatively weak forces resulting from electrostatic attractions such as ion–ion, ion–dipole, dipole–dipole (hydrogen bonds), induced transient fluctuating dipoles (van der Waals), or hydrophobic effects, they are typically readily reversible and thus amenable to standard equilibrium thermodynamic analysis. Also convenient is that most protein-ligand interactions occur as closed systems, namely, they contain a fixed amount of matter, and the exchange of work is confined to expansion ($\int PdV$). Because other types of energy exchange, such as radiation, or other types of exchange of work, such as electrical, surface, or photophysical, are negligible (or are approximated to be), the thermodynamic analysis of protein-ligand interactions is simplified.

This chapter provides a broad overview of the purpose and experimental approaches for determining thermodynamic parameters of protein-ligand interactions.

3.2

Basic Thermodynamics of Protein-Ligand Interactions

Thermodynamics, originally the study of the more limited phenomena of heat and heat transfer, evolved to become the more broad study of energy and energy transfer with the recognition – through the cumulative work of Count Rumford

(Benjamin Thompson), Robert Mayer, Sadi Carnot, James Joule, and others (see [2–4] for historical accounts) – that heat is a form of energy. A vast amount of experience and experimentation can be generalized in the following way (e.g., [5]): in any defined “system,” although the work done on the system (W) or the heat absorbed by the system (Q) in going from one “state” of the system to another varies with the path taken, the sum of W and Q is a constant and depends only on the initial and final “states” of the “system” under consideration. This generalization is formalized as follows:

$$\Delta U = Q + W, \quad (\text{Eq. 3.1})$$

where ΔU represents the change in the energy¹⁾ of the circumscribed “system.” This equation defines energy in terms of the measurable entities of heat and work and ΔU as dependent only on the state of the system (i.e., independent of the path by which the system moves from one state to another). ΔU around a closed path is zero, and only *changes* in energy can be measured (in terms of heat and work), not absolute values.

The First Law of thermodynamics (colloquially, the law of “conservation of energy”; Mayer, Helmholtz) does not explain why or guarantee that a defined system change will occur spontaneously or, if it does, in which direction the change will occur. This shortcoming is addressed by the Second Law of thermodynamics. Again, a vast amount of experience and experimentation can be generalized by (Carnot, Kelvin, Clausius),

$$\sum (Q/T) \geq 0 \quad (\text{Eq. 3.2})$$

or

$$\int d(Q_{\text{reversible}}/T) \geq 0 \quad (\text{Eq. 3.3})$$

where T is temperature in Kelvin. By defining change in “entropy” as $\Delta S \equiv Q/T$,

$$\sum \Delta S_{\text{system}} + \sum \Delta S_{\text{surroundings}} \geq 0, \quad (\text{Eq. 3.4})$$

or

1) U (or E) was previously termed the “internal” energy (no longer used). For a “closed” system (defined as one in which there is no exchange of mass with the ‘surroundings’) at rest, $\Delta U = Q + W$ if there is no other mechanism of exchange of energy. By convention, Q is the heat absorbed *by* the system (hence,

positive if heat flows into the system and negative if heat flows out of the system) and W is the work done *on* the system (hence, positive if the surroundings do work on the system and negative if the system does work on the surroundings).

$$\int dS \geq 0. \quad (\text{Eq. 3.5})$$

Spontaneous change or equilibrium is described when the RHS of Eq. 3.4 or 3.5 is $>$ or $=0$, respectively. To restrict the evaluation to measurable properties of the system rather than of the surroundings, free energy functions have been derived (Gibbs, Helmholtz). Most protein-ligand interactions occur at constant temperature and pressure, so that the only work is $-P\Delta V$. The second law then is represented by

$$\Delta S_{\text{system}} - \frac{(\Delta U + P\Delta V)_{\text{system}}}{T} \geq 0. \quad (\text{Eq. 3.6})$$

Since $\Delta U + P\Delta V$ is the change in “enthalpy²⁾” for these conditions,

$$\Delta S - \frac{(\Delta H)}{T} \geq 0, \quad (\text{Eq. 3.7})$$

which upon rearrangement becomes

$$T\Delta S - \Delta H \geq 0. \quad (\text{Eq. 3.8}).$$

With the definition of (J. Willard Gibbs) “free energy” as

$$\Delta G = \Delta H - T\Delta S, \quad (\text{Eq. 3.9}),$$

where $\Delta G < 0$ describes spontaneous change and $\Delta G = 0$ describes equilibrium.³⁾ These and other fundamentals of thermodynamics are reviewed in several excellent texts [6–25]. In terms of protein-ligand interactions, energy changes occur in the dissociation of the ligand molecules from the molecules of the solvent and the association of ligand molecules with the protein molecules. Ligand with protein is associated with changes in ΔH and ΔS . In addition, because the solvent environment is structured due to hydrogen bonds, London forces, or van der Waals interactions, particularly near membrane surfaces, the leaving of ligand molecules is associated with a reversal of the solvation process, which generally involves a decrease in entropy and an increase in energy level. Thus, the change in free energy upon protein-ligand interaction is the net result of dual rearrangement processes: first of the protein molecule (usually involving a change in degrees of freedom or

2) Change in enthalpy is defined as $\Delta H = \Delta U + P\Delta V$, where P and V are the pressure and volume, respectively, of the system. $P\Delta V$ is negligibly small in most protein-ligand interactions, so $\Delta H \approx \Delta U$, and the change in the enthalpy is used as an indication of the molecular forces involved in the interaction.

3) This is the fundamental criterion for a spontaneous transformation in a system, typical of most protein-ligand interactions, of constant temperature and pressure. The interaction proceeds spontaneously in the direction in which $\Delta G < 0$. It is important to note that the rate of the interaction is not determined by the sign or magnitude ΔG .

exposure to water molecules) and then of the solvent molecules (usually involving a decrease in structural constraint and hence an increase in entropy).

3.3

Measurement of Thermodynamic Parameters

For an interaction between a protein (P) and a ligand (L) that forms a protein-ligand complex (PL) according to a simple, reversible, bimolecular step represented as



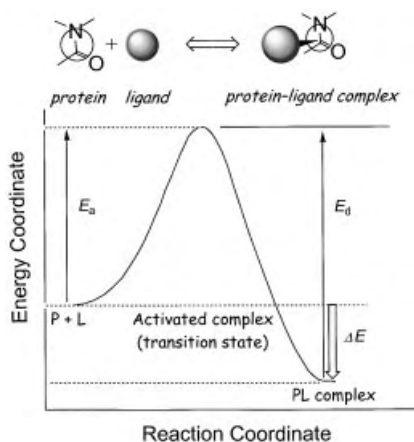
the reaction can be characterized, with appropriate caveats, by the equilibrium constant ($K_{\text{eq}} = [PL]/[P][L]$).⁴⁾ In practice, the reciprocal of the equilibrium constant is commonly used and is termed the Michaelis constant (K_M) when the protein is an enzyme and the ligand is a substrate and is termed the dissociation constant (K_d or K_i) when the protein is a receptor and the ligand is a neurotransmitter, hormone, or drug.

The interaction can be visualized as a reaction-energy diagram as shown in Fig. 3.1. Changes in the energy coordinate (the ordinate) are plotted as a function of the position of the interaction as it proceeds in either direction along the reaction coordinate (the abscissa). This highly schematized representation indicates the overall change in energy (ΔE) for the protein-ligand interaction and the activation energies for the association (ΔE_a) and dissociation (ΔE_d) steps. The diagram applies to the elementary step of the interaction. Associated processes, such as migration to the interaction site, catalytic activity (enzymes), activation of second-messenger transduction processes (receptors), etc., are not included.

For the interaction represented by Eq. 3.10, the relationship between the change in free energy (ΔG), change in enthalpy (ΔH), and change in entropy (ΔS) is given by Eq. 3.9. There are two major ways of obtaining the thermodynamic parameters. One way is by direct measurement of the heat of reaction, which for no ΔPV work is the same as ΔH . The recent development of highly sensitive calorimeters allows such measurement for a relatively wide variety of protein-ligand interactions and is described in more detail below. An alternative procedure employs a more indirect measure, which utilizes a simplified relationship (the van't Hoff equation) between the thermodynamic parameters and the temperature dependence of the equilibrium constant of Eq. 3.10.

4) The relationship between these constants and the forward and reverse rate constants of the interaction is not automatically known except for an elementary reaction step.

Fig. 3.1 Reaction-energy diagram for the reversible interaction between a protein and a ligand that forms a protein-ligand complex. ΔE is the overall change in energy for the interaction. ΔE_a and ΔE_d are the activation energies for the association and dissociation processes, respectively. Intermediate between the dissociated and associated components is a transition state comprised of an activated complex.



3.3.1

Calorimetric Determination of Thermodynamic Parameters

The use of calorimetry to measure the heat of a reaction is a time-honored technique. Presently, two modernized high-accuracy automated types of equipment are available with accompanying convenient software. One is known as “differential scanning calorimetry” (DSC), and the other is known as “isothermal titration calorimetry” (ITC). DSC measures the heat capacity (which at constant pressure is the temperature derivative of enthalpy) of the protein-ligand interaction under investigation by incrementally varying the temperature of the system over a specified range (the “scan”). Ultrasensitive isothermal titration microcalorimetry (the use of instruments for which the sensitivity is better than $1 \mu\text{W}$) [26] measures the heat change that is associated with reactions in solution at a constant temperature and, by the sequential addition of ligand to the solution, also yields thermodynamic parameters. It is a well-characterized and widely accepted technique because the interaction is carried out at a constant pressure, $V\Delta P=0$. Therefore, the energy change associated with the interaction is ΔH the change in enthalpy ($\Delta U = \Delta H + \Delta PV$). An advantage of ITC over other methods is that it measures the enthalpy change directly. Other techniques, also described below, determine the enthalpy change indirectly. For this reason, DSC or ITC is the preferred method of obtaining interaction parameters, provided that the experimental conditions allow the use of these techniques. Because of the greater use of ITC for protein-ligand interactions to date, the details of this technique are provided below.

In the standard ITC apparatus, the protein-ligand interaction proceeds in a sample cell of relatively small volume (usually 1–3 mL). One component (e.g., protein) of the reaction is placed in the reaction cell, and the other component (e.g., ligand) is added in stepwise fashion by an automated injection system in preset measured amounts for preset measured times. A built-in stirrer ensures that the reaction is continuously and well mixed. The reaction cell is composed of material

that has high thermal conductivity such that energy changes (heat of reaction) that occur within the reaction cell are transmitted with minimal loss as changes in temperature. In modern ITC equipment, the change in temperature is measured as the amount of differential current (power) that is required to maintain the reaction cell at the same preset temperature as that of a reference cell filled with distilled water or the same buffer solution as the reaction cell. As a consequence of this design, the measurements are extremely precise because the dependent variable is power and essentially the only limitation is the electronic thermal motion.

If the protein-ligand interaction is endothermic, more power ($\mu\text{cal s}^{-1}$) is required relative to the reference cell. The power that is required, over baseline, comprises the raw data output of the ITC equipment. If the reaction is exothermic, less power is required, which is recorded as a downward deflection in output (Fig. 3.2). The overall interaction between a protein (enzyme or receptor) and a ligand (substrate, inhibitor, neurotransmitter, hormone, or drug) is carried out in a sequence of automated titrations. At each injection step, the power is recorded as a function of time. Each subsequent injection in the series is made after the power function returns to baseline. The output, therefore, forms an S-shaped curve, mirroring the progression of binding of the interacting species from initial

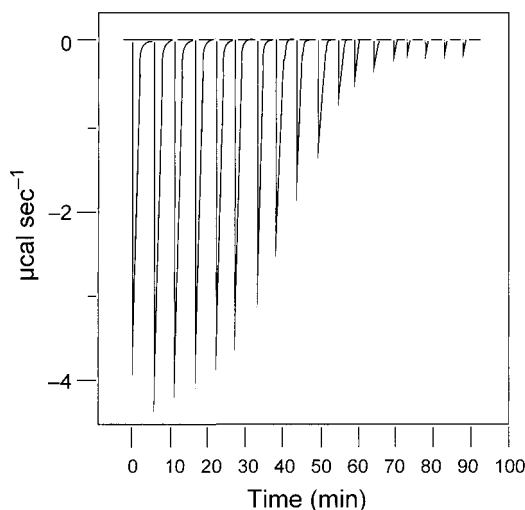
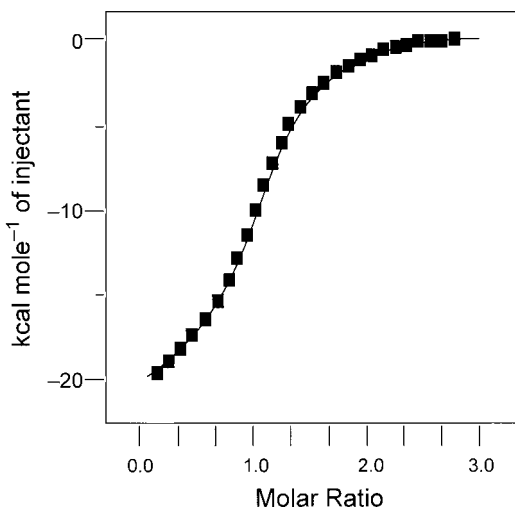


Fig. 3.2 Diagrammatic representation of typical results obtained in an ITC study of a protein-ligand interaction. The raw data output (peak) accompanying each injection of ligand is the power ($\mu\text{cal s}^{-1}$) that is required to maintain the sample cell at the same temperature as a reference cell. A downward deflection indicates an exothermic reaction; an upward deflection indicates an endothermic

reaction. Multiple ligand injections are made at preset intervals. The progressively smaller heat outputs correspond to progressively greater protein-ligand binding until saturation is achieved. The residual deflections at the end of the run yield the heat of dilution, which is subtracted from the other deflections prior to further analysis.

Fig. 3.3 The raw data output of ITC is transformed to show the heat exchange at each injection (kcal mol^{-1} of injectant), obtained by integration of the area of each “spike” in the raw data output, as a function of the molar ratio of the protein-ligand binding interaction. The curve is then computer-generated as the best fit to either a one-site or multi-site binding model.



injection to full saturation (Fig. 3.3). At the end of each run, all of the binding sites are occupied and no further heat of reaction is detected. Any residual power differential is a measure of the heat of solvation of the injected species. In modern ITC equipment, this heat is usually automatically subtracted from the heat of reaction. The raw data obtained for each injection (peak) are then integrated with respect to time, and the integrated heats that are derived from the raw data are plotted against the molar ratio of the interacting species. A best fit of the data is obtained using a non-linear algorithm. From this fit, the stoichiometry, K_d , and ΔH of the interaction are obtained. From the K_d and ΔH , the other thermodynamic parameters, ΔG and ΔS , are easily calculated from standard relationships. Additional details of the design and application of ITC are available in several excellent reviews [27–29].

3.3.2

van't Hoff Determination of Thermodynamic Parameters

3.3.2.1 Relationship to Equilibrium Constant

In the simplest case, the protein-ligand interaction can be represented as, or modeled as, a reversible bimolecular reaction such as depicted by $P + L \rightleftharpoons PL$. The change in Gibbs free energy (ΔG) for the interaction in the direction indicated is related to the standard free energy change (ΔG°) by the following equation:

$$\Delta G = \Delta G^\circ + RT \ln \left(\frac{[PL]}{[P][L]} \right), \quad (\text{Eq. 3.11})$$

where the brackets indicate concentration, $R = 1.99 \text{ cal/mol} \cdot \text{K}$ ($= 8.31 \text{ J/mol} \cdot \text{K}$), and T is the absolute temperature in Kelvin ($^\circ\text{C} + 273.15$). Most protein-ligand in-

teractions are examined at steady state, at which $\Delta G=0$ (the process is not capable of producing work), so that Eq. 3.11 becomes

$$\Delta G^\circ = -RT \ln \left(\frac{[\text{PL}]}{[\text{P}][\text{L}]} \right), \quad (\text{Eq. 3.12})$$

The ratio of complex concentration to the reactant concentrations can be represented by the equilibrium constant K_{eq} , the reciprocal of the equilibrium constant (e.g., K_M , K_d , or K_i), or by some alternative designation in other types of studies. For the example of K_d , substitution into Eq. 3.12 yields

$$\Delta G^\circ = -RT \ln(K_{\text{eq}}) = -RT \ln(1/K_d) = RT \ln(K_d). \quad (\text{Eq. 3.13})$$

Hence, for the conditions under which most protein-ligand interactions are studied, Eq. 3.13 describes the relationship between the thermodynamic parameter ΔG° and a reaction characteristic (the equilibrium constant) that can be measured experimentally. Because the change in Gibbs free energy is related to the change in enthalpy and entropy by $\Delta G^\circ = \Delta H^\circ - T\Delta S^\circ$, Eq. 3.13 can be rearranged to

$$\ln(K_d) = \left(\frac{\Delta H^\circ}{R} \right) \left(\frac{1}{T} \right) - \frac{\Delta S^\circ}{R}. \quad (\text{Eq. 3.14})$$

Eq. 3.14 is an integrated form of the van't Hoff equation

$$\frac{d(\ln K_{\text{eq}})}{dT} = \frac{\Delta H^\circ}{RT^2}, \quad (\text{Eq. 3.15})$$

and is an approximation valid when ΔH° and ΔS° are not temperature dependent. Noting that Eq. 3.14 represents a linear relationship between $\ln(K_d)$ and $1/T$ with the y-intercept $= -\Delta S^\circ/R$ and the slope $= \Delta H^\circ/R$, it is a common practice in thermodynamic analysis of protein-ligand interactions to determine K_d at several different temperatures and then construct a “van't Hoff plot” from which ΔH° and ΔS° are determined from the slope and y-intercept of the resultant data plotted as $\ln(K_d)$ against $1/T$ (which is a straight line if the heat capacity is independent of temperature). A smaller error in ΔH° can be obtained if ΔS° is determined first from the van't Hoff plot and then ΔH° from $\Delta H^\circ = \Delta G^\circ + T\Delta S^\circ$.

Not all such plots turn out to be linear, indicating that in those cases the heat capacity change (ΔC_p) is not independent of temperature for the interaction under study. It has also been suggested that ΔH° values determined using the van't Hoff plot method can differ from the same values determined using direct calorimetric measurement [30]. However, it has subsequently been reported that discrepancies are relatively minor [31].

3.3.2.2 Obtaining the Equilibrium Constant

In order to apply the van't Hoff method of obtaining thermodynamic parameters, some means of measuring the association or dissociation constant of the protein-ligand interaction must be used. The basic principles and many of the experimental methodologies available for obtaining these constants have recently been summarized [32, 33] and are the subject of more extensive coverage in recent reviews (e.g., [34]) and monographs (e.g., [12, 35]). The methods include (extracted from [32] and [33]):

- Equilibrium dialysis – Two compartments of a dialysis cell are divided by a semi-permeable membrane. The protein-ligand complex is allowed to associate or dissociate across the membrane until equilibrium is attained. By measuring the constituents of the interaction, the binding constant can be obtained from standard formulas.
- Steady-state dialysis – The equilibrium dialysis technique is accelerated by having buffer flow at a constant rate on one side of the semi-permeable membrane and by stirring both sides in order to minimize the concentration gradients [36].
- Diafiltration – A type of dialysis equilibrium in which pressure is used to force the ligand-containing solution from one chamber into the protein-containing chamber [37].
- Ultrafiltration – Pressure or centrifugation is used to force a mixture of known total concentrations of protein and ligand through a semi-permeable membrane [38].
- Partition equilibrium – Separation occurs between two phases rather than across a semi-permeable membrane. Examples include partition between aqueous and lipid phases or partition between a liquid and a solid phase (e.g., where the binding sites are embedded on a solid matrix).
- Gel (exclusion) chromatography – Counterpart to equilibrium dialysis when there is sufficient difference in size between protein and ligand and when the protein and protein-ligand complexes co-migrate.
- Spectroscopy – Binding-induced changes in either a chromophore or fluorophore absorbance or emission are used to measure the ratio of free to bound ligand concentration. Examples include circular dichroism (differential absorption of left- and right-handed circularly polarized light), fluorescence emission (energy loss as radiation as a fluorophore returns to ground state from photon-excited state) methods, including fluorescence anisotropy (binding of ligand changes the relative depolarization of the emission spectrum compared with that of a polarized exciting light).
- Electrophoresis – The components are separated on the basis of differential rates of migration toward an anode or cathode.
- Sedimentation equilibrium – An analytical ultracentrifuge is operated at a relatively slow speed that leads to a measurable equilibrium distribution of the constituents of a protein-ligand interaction.
- Radioligand binding – The most commonly used technique for the determination of binding to receptors is commonly called radioligand binding because of

the use of a radioactive-labeled ligand for the quantification of the amount of bound material. As typically used, a radiolabeled ligand is incubated with the receptor preparation for a time sufficient for equilibrium to be attained. Bound and unbound ligands are then separated using any of a variety of techniques such as dialysis, centrifugation, or vacuum filtration (the most widely used method) (see [33] and [39] for details).

- Others – Affinity chromatography, biosensor techniques, and radioimmunoassay are among some of the other available techniques. In addition, perhaps a special mention should be made of the technique of estimating dissociation constants in pharmacological studies using irreversible antagonists (for the K_d of an agonist) or a reversible antagonist (for the K_d of the antagonist). These estimates, although not as intimate to the receptor-ligand interaction as some of the others, nevertheless have been used to some distinct advantage.

3.4

Applications

3.4.1

Calorimetric Determination of Thermodynamic Parameters

There are now well over 200 publications in which microcalorimetry has specifically been used to study protein-ligand interactions of a variety of types. A list of these studies is readily available by a MEDLINE search or from ITC equipment suppliers. Since the studies are too numerous to review here, perhaps a recent one might serve as a representative example of the technique and of its application. In this example [40] we determined the thermodynamic parameters associated with the binding of the reversible inhibitor 2'-CMP (2'-cytidine monophosphate) to RNase-A (ribonuclease A). We were specifically interested in the binding under conditions that were relatively "physiological," i.e., at body temperature and in a buffer that contained multiple ions at roughly cellular concentrations.

RNases are exo- and endonucleases (EC 3.1.27.5), present in vertebrates and also in several bacteria [41–43], mold [44], and plant species [45, 46], that participate in a variety of RNA-processing pathways. Several members of the RNase superfamily, commonly referred to as the "non-secretory" type, function in predominantly intracellular roles, whereas others, termed the "secretory" type, have evolved [47] roles that are predominantly extracellular, presumably contributing to digestive and cytoprotective functions. (There are actually several systems of nomenclature for RNases. This came about through historical factors, such as different names for the same RNase being studied in different species and subsequently recognized as the same RNase, identification of RNase activity after naming the enzyme for other reasons, etc.). For the cytoprotective function of RNases, cytotoxicity against external threats is a desirable and self-protective characteristic that is manifested under normal physiologic conditions. Usually, an intracellular ribonuclease inhibitor (RI) with exceptionally high affinity for RNase protects the

cell from any secretory RNase that does not leave the cell. However, under two circumstances the secretory RNases can be cytotoxic: failure of RI activity or unchecked RNase activity. The first circumstance is a consequence of genetic defects that result in deficiencies in RNase production or function. The second is a consequence of excess activity or inappropriate activity in pathological states. Perhaps the best-known example of the latter is the enhanced tumor growth that is attributed to angiogenesis stimulated by the blood-borne RNase angiogenin. However, there are other RNases, specifically those designated as the ribonuclease 2 type, that are implicated in pathophysiological conditions where eosinophils appear in increased numbers, as in asthma and other inflammatory disorders in which tissue damage occurs as part of an allergic response [48–50].

Members of the human RNase-A superfamily include

- (“secretory”) pancreatic type (ribonucleases 1);
- (“non-secretory” or “neurotoxin” type) liver, spleen, and urine (Us) RNases (ribonucleases 2), also known as eosinophil-derived neurotoxin (EDN);
- plasma RNase (HT-29) (ribonucleases 4);
- and angiogenins [47].

They constitute a group of homologous enzymes that display a preference for pyrimidine bases of RNA. Although some of the details are yet to be delineated, the catalytic mechanism of RNA cleavage by RNases is hypothesized to occur as depicted in Fig. 3.4. The overall reaction is thought to occur in two steps [51]. In the first step, a 2',3'-cyclic phosphodiester is formed by a “transphosphorylation” reaction from the 5' carbon (starting from the base) to the 2' carbon of the next nucleotide in the RNA chain (Fig. 3.5). The catalytic reaction domain is formed by specific amino acid residues of the RNase (Fig. 3.6), the details of which have been investigated by several strategies such as chemical modification and site-di-

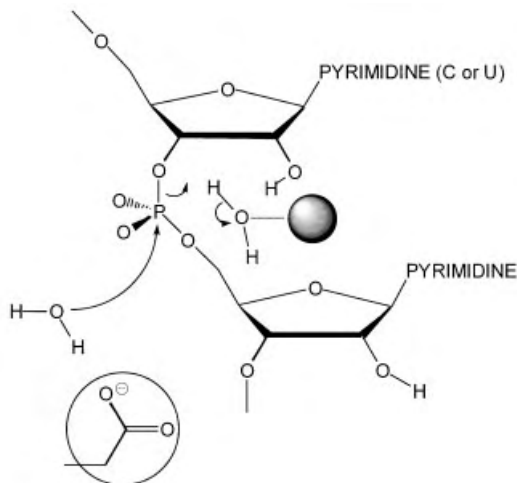


Fig. 3.4 The proposed mechanism for the catalytic cleavage of RNA by RNase. The spheres represent amino acid residues of RNase or metal ions (e.g., Mg^{2+}). Modified from [41].

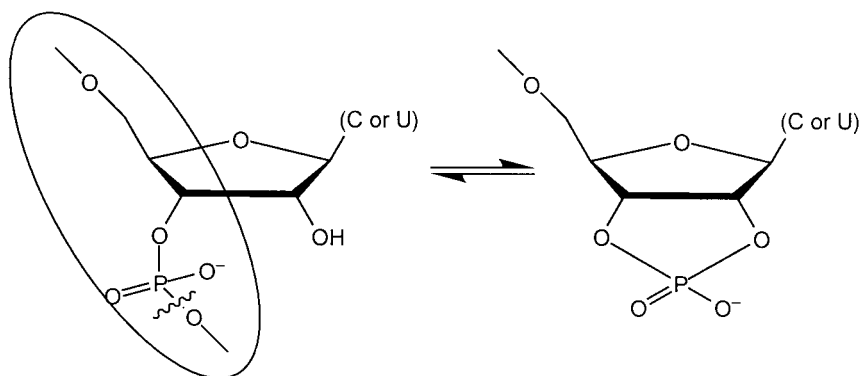


Fig. 3.5 The proposed depolymerization reaction catalyzed by RNase-A. The RNA backbone is indicated by the ellipsoid. Modified from [41].

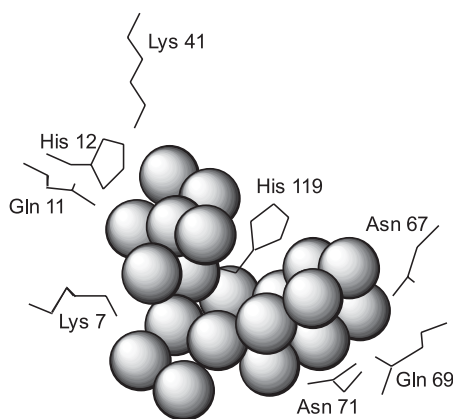


Fig. 3.6 The catalytic cleft of bovine RNase-A (indicated by the stippled region). A segment of RNA is oriented and held in the pocket formed by the amino acids indicated. Modified from [52].

rected mutagenesis studies (e.g., [52–55]). The reaction products of the first step are not enzyme bound and therefore migrate into the solvent [56]. In the second step, which is believed to occur within the solvent, the product of the first step (2',3'-cyclic phosphodiester) is hydrolyzed to a 3' nucleotide [57, 58]. These reactions can then be represented as follows [51]:

Step 1: $\text{RNA} \rightleftharpoons \text{2',3'-cyclic phosphodiesters} + \text{R-OH}$

Step 2: $\text{2',3'-cyclic phosphodiesters} \rightarrow \text{3'-phosphomonoesters.}$

Step 1 is the primary one that is catalyzed by RNases. It is a fairly straightforward reaction and therefore is amenable to analysis by standard procedures [59]. RNase is also susceptible to inhibition by substances such as 2'-CMP. In our study [40], we used ITC to determine the binding affinity and thermodynamic parameters associated with the reversible inhibition of RNase-A by 2'-CMP at body temperature (37°C) and in a more “physiologically relevant” (i.e., multi-ion)

buffer.⁵⁾ These data ultimately might be helpful in drug-design efforts. Consistent isotherms with stable baselines were obtained. Maximal output to the injections of 2'-CMP was about -1.5 to -2.5 $\mu\text{cal/s}$, the negative deflection indicative of an exothermic reaction. As conventional for studies of this sort, the transposed data were plotted as the integrated heats (kcal/mol of 2'-CMP) for each injection against the 2'-CMP/RNase-A molar ratio, and fitting parameters for the single-site nonlinear regression computer-fit of the raw data points yielded values for S (stoichiometry of the interaction), K_{eq} , and ΔH° for each run. The calculated stoichiometry was very close to 1:1, consistent with previous measures by others of a 1 to 1 interaction between 2' CMP and RNase-A (e.g., [59]). The other estimated parameters, means (\pm S.D.) of triplicate runs, were $K_d = 13.9$ (± 3.9) μM ; $\Delta G^\circ = -6.90$ (± 0.16) kcal/mol ; ΔH° (kcal/mol) $= -15.7$ (± 2.0) kcal/mol ; and $\Delta S^\circ = -0.028$ (± 0.006) $\text{kcal/mol} \cdot \text{K}$. The observed negative entropy change is consistent with the location of the ribonucleolytic reaction active site within a cleft that binds and cleaves RNA [60]. The interaction proceeds because of a larger decrease in enthalpy. These results, which were determined in multi-ion buffer, were notably different from those determined in single-ion buffer [61] (Tab. 3.1). This single example, hopefully, serves as an example of the methodology of ITC and also a sense of its versatility.

3.4.2

van't Hoff Determination of Thermodynamic Parameters

The van't Hoff method has been the most commonly applied technique to determine thermodynamic parameters. A MEDLINE search of "van't Hoff" reveals over 500 publications between 1966 and 2002. The application to enzyme reaction is well known. More recently, this method has been applied to ligand-receptor inter-

5) Bovine pancreatic RNase-A, 2'-CMP free acid (98% purity), Na^+ , K^+ , Ca^{2+} , Mg^{2+} acetate, and glacial acetic acid (ACS or molecular biology grade) were purchased from commercial sources. The RNase was dissolved in deionized water and was dialyzed twice for 4 h (in 20 mL solution) in a stirred 1-L beaker maintained at 1.5°C by immersion in an ice-bath. RNase and salt stock solutions (in deionized water) were mixed such that the final concentrations were KCl (3 mM), CaCl_2 (0.1 mM), NaAcetate (10 mM), K_2PO_4 (3 mM), MgSO_4 (0.4 mM), and KAcetate (50 mM) adjusted to pH 5.5 by dropwise addition of 50 mM HAcetate . The concentration of RNase (0.04–0.05 mM), selected to be not much higher than the K_d of interaction with 2'-CMP, was determined by quantitative UV spectrophotometry (277.5 nm; extinction coefficient

$\epsilon = 9800 \text{ M}^{-1} \text{ m}^{-1}$). The concentration of 2' CMP (1.2 mM), selected so that the c value (equal to the product of the binding constant and the total molar concentration of RNase) would be between 1 and 500, was prepared in the same buffers as the RNase-A and verified spectrophotometrically (260 nm, $\epsilon = 7400 \text{ M}^{-1} \text{ cm}^{-1}$). Solutions were degassed at 36.5°C under vacuum (about 686 mmHg). The reference cell of the calorimeter contained degassed deionized water. The reaction cell contents were stirred at 400 rpm at 37°C throughout the experiment (the frictional heat of stirring is incorporated into the baseline). 2'-CMP was introduced into the reaction cell in a series of 35 $4\text{-}\mu\text{L}$ injections, each delivered over 16 s at 3-min intervals. The equipment automatically adjusts for the change in volume. The data were evaluated (sampling rate 2 s^{-1}) with computer software.

Tab. 3.1 Comparison of the dissociation constant and thermodynamic parameters obtained for the 2'-CMP/RNase-A interaction in multi-ion buffer and in a 50 mM potassium acetate buffer [61]

	<i>Multi-ion</i>	<i>Single-ion</i>
ΔG° (kcal/mole)	$-6.90 \pm 0.16^*$	-7.46 ± 0.10
ΔH° (kcal/mole)	$-15.7 \pm 2.0^*$	-21.9 ± 0.9
ΔS° (kcal/mole \cdot K)	$-0.028 \pm 0.006^*$	-0.047 ± 0.003
K_d (μ M)	$13.9 \pm 3.9^*$	5.6 ± 1.0

* Significant difference ($P < 0.05$).

actions [1]. Because these applications are less well known, a short summary is presented. There are also more caveats associated with such applications, a topic considered subsequently.

The basic principles of thermodynamics of course apply to any chemical system, and in this sense the extension of the application of thermodynamic analysis to ligand-receptor interactions is straightforward. Ligand-receptor interactions involve a ligand molecule that has “affinity” for a receptor molecule in biological tissue. There is a requisite complementary 3-D shape for the ligand to able to “fit” the receptor and form chemical bonds – usually weak, reversible ones – with the receptor molecule. A subset of ligands, termed “agonists,” is also capable of inducing a biological effect by binding to receptors. Such molecules are said to have “intrinsic activity,” “efficacy,” or some similar term. Agonists can be “full” or “partial,” depending on their efficacy. Ligands that possess affinity but lack efficacy are “antagonists.” Such ligands do not activate measurable biological effects but block the agonist’s access to the receptor sites. Because it is not always possible to control all variables precisely, the application of thermodynamic analysis to drug-receptor (pharmacological) interactions involves some care in both methodology and interpretation. Nevertheless, such an endeavor is often worthwhile if there is the opportunity to learn more about such systems than can be learned using other measures. The receptor concept was originated during the latter part of the 1800s and early 1900s, but it was the development of methodological techniques during the 1970s, in particular, radioligand binding techniques (e.g., [33]), that allowed the accurate determination of the number of drug-receptor binding complexes. With the wide commercial availability of relatively stable, radioactively labeled ligands, the technique is now almost routine (e.g., [35, 39]).

The study published by Weiland et al. in *Nature* in 1979 [62] was perhaps the first to truly catch the attention of many biologists and remains probably the best-known thermodynamic study of drug-receptor interactions to many pharmacologists. In this study the authors measured the temperature dependency of the binding of 20 agonists and antagonists to the β -adrenoceptor located on turkey erythrocyte membranes. They reported that agonist binding affinity was greater at the lower of the two temperatures they examined. The calculated thermodynamic parameter values

Tab. 3.2 Examples of thermodynamic studies of ligand interaction with opioid receptors (from [1])

<i>Preparation</i>		$\Delta G^{\circ'}$	$\Delta H^{\circ'}$	$\Delta S^{\circ'}$	<i>Reference</i>
<i>Agonists</i>					
a. Radioligand binding					
β -endorphin	Rat brain	<0	>0	>0	64
DAMGO (μ)	Guinea pig brain	<0	>0	>0	65
DAMGO	Bovine adrenal	<0	>0	>0	66
DAMGO	Rat brain	<0	>0	>0	67
DAMGO	r-MOR/(CHO)	<0	<0	>0	68
DADLE (δ)	Guinea pig brain	<0	>0	>0	65
DADLE	Bovine adrenal	<0	>0	>0	66
Deltorphin	Rat brain	<0	>0	>0	67
Dihydromorphine	Rat brain	<0	>0	>0	69
Dihydromorphine	Rat brain	<0	>0	>0	67
DPDPE (δ)	m-DOR-1	<0	<0	>0	70
EKC (κ)	Guinea pig brain	<0	>0	>0	65
EKC	Bovine adrenal	<0	>0	>0	66
EKC (<i>has</i>)	Frog brain	<0	<0	<i>T-dep</i>	71
EKC (<i>las</i>)	Frog brain	<0	<0	<i>T-dep</i>	71
Etorphine	Rat brain	<0	>0	>0	72
Etorphine	Bovine adrenal	<0	>0	>0	66
Methadone	r-MOR/(CHO)	<0	<0	>0	68
Morphine	r-MOR/(CHO)	<0	<0	>0	68
Ohmefentanyl	r-MOR/(CHO)	<0	>0	>0	68
Pentazocine	r-MOR/(CHO)	<0	<0	>0	68
PL017	r-MOR/(CHO)	<0	<0	>0	68
SNC-80 (δ)	m-DOR-1	<0	<0	>0	70
SNC-80 (<i>has</i>)	h-DOR/(CHO)	<0	<0	>0	73
SNC-80 (<i>las</i>)	h-DOR/(CHO)	<0	<0	>0	73
Sufentanil	r-MOR/(CHO)	<0	>0	>0	68
b. Isolated Tissue					
DPDPE	MVD	<0	<0	>0	74
<i>Antagonists</i>					
a. Radioligand Binding					
CTAP	r-MOR/(CHO)	<0	>0	>0	68
Diprenorphine	Rat brain	<0	<0	>0	72
Diprenorphine	r-MOR/(CHO)	<0	>0	>0	68
Naloxone	Rat brain	<0	<0	>0	69
Naloxone	Rat brain	<0	<0	>0	67
Naloxone	h-DOR/(CHO)	<0	>0	>0	73
Naltrexone	r-MOR/(CHO)	<0	<0	>0	68

Tab. 3.2 (cont.)

	<i>Preparation</i>	ΔG°	ΔH°	ΔS°	<i>Reference</i>
Naltriben (δ)	h-DOR/(CHO)	<0	<0	>0	73
Naltrindole (δ)	Mouse brain	<0	>0	>0	75
Naltrindole	Mouse spinal cord	<0	<0	>0	75
Naltrindole	NG 108-15 cells	<0	>0	>0	75
Naltrindole	m-DOR-1	<0	>0	>0	70
Naltrindole	h-DOR/(CHO)	<0	>0	>0	73
TIPP(ψ) (δ)	m-DOR-1	<0	<0	>0	70
b. Isolated Tissue					
Naloxone	MVD	<0	<0	<0	76
<i>Mixed</i>					
Radioligand binding					
Bremazocine	h-DOR/(CHO)	<0	>0	>0	73

CHO=Chinese-hamster ovary cells; CTAP=D-Phe-Cys-Tyr-D-Trp-Arg-Thr-penicillamine-Thr-NH₂; DAMGO=[D-Ala²,NMePhe⁴,Gly-ol⁵]enkephalin; DADLE=[D-Ala²,D-Leu⁵]enkephalin; DPDPE=[D-Pen^{2,5}]enkephalin; EKC=ethylketocyclazocine; *has*=high-affinity binding site; *las*=low-affinity binding site; m-DOR-1=cloned δ receptor from mouse brain; MVD=mouse vas deferens; NG 108-15=mouse neuroblastoma-rat glioma hybrid; PL017=Tyr-Pro-NmePhe-D-Pro-NH₂; SNC80=(+)-4-[(*aR*)-*a*-((2*S*,5*R*)-4-allyl-2,5-dimethyl-1-piperazinyl)-3-methoxybenzyl]-*N,N*-diethylbenzamide; *T-dep*=temperature-dependent; TIPP(ψ)=Tyr-Tic(ψ)[CH₂NH]Phe-Phe-OH.

were reasonable for chemical reactions ($\Delta G^{\circ}=-6.19$ to -12.51 kcal mol⁻¹, $\Delta H^{\circ}=-12.75$ to -18.86 kcal mol⁻¹). In contrast, it was found that antagonist binding to the β -adrenoceptor was largely “entropy driven” (the major contribution to the negative ΔG° was due to a positive ΔS° of 0.013 to 0.042 kcal mol⁻¹ K⁻¹). Thus, there was a clean distinction between agonist and antagonist binding that coincided quite nicely with prevailing views of the actions of agonists and antagonists at receptor sites – that agonists, but not antagonists, induce conformational changes in receptors and that this could account for the induction of the biological response by agonists but not antagonists. We now know, of course, that this does not hold for all receptor binding, but the early publications by Weiland et al. [62, 63] stand out as seminal in the field. Subsequent work has provided insight into a number of drug-receptor interactions. An example of the results of thermodynamic studies on one particular receptor type, the opioid receptor, is given in Tab. 3.2 [64–76]. Similar summaries of thermodynamic studies of other receptors can be found in Raffa [1].

Another application of the method, one that was brought to bear on a puzzling ligand-receptor question, was that reported by Wild et al. [75]. Although previous pharmacological studies had suggested the existence of more than one subtype of opioid δ receptor, only one had been cloned. Wild et al. [75] reasoned that a distinction could be demonstrated if two preparations, each containing a population of opioid δ receptors, had different temperature dependency of the dissociation constant. They measured the temperature dependence of the dissociation constant (using radioligand binding techniques) of the selective opioid δ receptor ligand

[³H]naltrindole in mouse brain tissue and mouse spinal cord tissue. Comparison of the two revealed that the van't Hoff plots for mouse brain and mouse spinal cord had different slopes: one was positive and the other was negative. It was concluded that there are multiple subtypes of opioid δ receptors (at least functionally).

3.5

Caveats

The measurement of thermodynamic parameters for protein-ligand interactions can provide valuable insight into aspects of the interaction that are not easily obtainable by other techniques. However, as with all techniques, there are certain limitations in the approach – some related to the methodology and some related to the complexities of the systems under investigation. It is necessary to remember, for example, that the parameters determined apply to the overall reaction being measured. For most protein-ligand interactions, more than one process may be involved. For receptor-ligand interactions, this is almost certainly the case. For example, as a drug molecule interacts with a receptor and makes the transition from free to bound state, energy changes occur as the result of the alteration of the arrangement of receptor molecules as well as of the solvent molecular matrix from which the ligand leaves. Ion displacement, proton transfer, and other processes can be involved. The thermodynamic parameters that are measured for the interaction include these processes.

Some of the more likely encountered possible limitations in thermodynamic analysis, particularly for ligand-receptor interactions, include the following:

- Most receptors are membrane bound. Thus, the interaction of the receptor with the membrane must be considered (constraints on degrees of freedom, changes in the degrees of freedom upon ligand binding, etc.)
- The thermodynamic analyses most often used, particularly the van't Hoff method, require that measurements be made at steady-state conditions. In the case of radioligand binding determination of equilibrium constants, the time required for the protein-ligand interaction to reach steady state depends on the incubation temperature, and, therefore, the equilibrium constant must be determined for each temperature studied. For the most accurate results, the determination needs to be made at more than two temperatures in order to detect non-linearity. The integrated form of the van't Hoff equation takes the simple form that is commonly used only if ΔH° and ΔS° for the interaction are not temperature dependent; otherwise, non-linearity in the van't Hoff plot can arise. Meaningful information can still be obtained in such cases, but more complex analysis is required.
- The relevant affinity state is not always obvious. If the binding reaction is complicated by other processes, such as degradation of the ligand or internalization of the ligand, receptor, or both, then the data cannot be analyzed by simple ther-

modynamics methods – unless the system is defined in a way to incorporate these additional phenomena.

- Although thermodynamic parameters can be obtained for interaction mechanisms that are complex, interpretation of the results is greatly simplified when the interaction mechanism is simple. For example, tissues in which multiple receptor types are expressed will yield results different from tissues expressing only one type, unless a type-selective ligand is used.
- In radioligand binding studies, non-linear Scatchard plots or competition curves that have abnormally steep slopes imply complex binding phenomena, possibly involving multiple receptor types or affinity states. In such cases, the thermodynamic parameters should be separately determined for each receptor type or affinity state.
- The equilibrium (dissociation) constant (binding affinity) that is measured might depend upon the receptor affinity state, G protein coupling, allosteric influences, or other factors distal to the actual binding site. According to most present models, this is more likely for agonists than antagonists.

3.6

Summary

The determination of thermodynamic parameters of chemical reactions is extremely useful for the characterization and understanding of chemical reaction processes. The recent extension of this strategy to protein-ligand interactions has yielded equally significant insight into the more intimate details of these complicated and intransigent systems. In addition, the pragmatic application of the information obtained from thermodynamic data of protein-ligand interactions to novel drug-discovery efforts offers exciting new opportunities for creative and valuable work.

Prior to the introduction of modern, automated, high-sensitivity calorimetry equipment, the van't Hoff technique (which is based on the temperature dependence of the equilibrium constant of the reaction) was the primary experimental approach available to determine the thermodynamics of protein-ligand interactions. It remains a mainstay of such determinations. The technique requires the measurement of the equilibrium constant (or of its reciprocal, the dissociation constant), and a large variety of methods have been developed to accomplish this. Radioligand binding is presently the most commonly used method for measuring the reaction constants of ligand-receptor interactions. In its most simplified form, the van't Hoff equation assumes a temperature independence of enthalpy, and this requirement is unfortunately not always verified by experimentalists who use it. However, this problem can be easily avoided or overcome by appropriate experimental design or data analysis.

The introduction of highly sensitive and automated calorimetric equipment has added new options for the measurement of thermodynamic parameters of protein-ligand interactions. By varying either the temperature – as in differential

scanning calorimetry (DSC) – or the ligand concentration – as in isothermal titration calorimetry (ITC) – thermodynamic parameters are obtained directly. As with the van't Hoff technique, there are limitations of practice and of interpretation. Perhaps the major pragmatic limitation at the present time is the relatively large amount of sample required for high-affinity interactions. As new strategies are devised to overcome these drawbacks, the application of calorimetric approaches will expand even further.

The ever-increasing interest in the folding and interaction of large biomolecules with endogenous or designed ligands will provide the impetus for continued improvement and implementation of experimental approaches to determine the thermodynamics of protein-ligand interactions. The formalization of this interest in the new field of “proteinomics” will provide a framework for its development, and its application to drug-discovery efforts will demonstrate, as thermodynamics has always done, its utility.

3.7

References

- 1 R. B. RAFFA, *Drug-Receptor Thermodynamics: Introduction and Applications*, John Wiley & Sons, UK, **2001**.
- 2 A. P. LIGHTMAN, *Great Ideas in Physics*, McGraw-Hill, USA, **2000**.
- 3 M. GUILLEN, *Five Equations that Changed the World*, Hyperion, USA, **1995**.
- 4 H. C. VON BAEYER, *Warmth Disperses and Time Passes*, Random House, USA, **1999**.
- 5 I. M. KLOTZ in R. B. RAFFA, ed., *Drug-Receptor Thermodynamics: Introduction and Applications*, John Wiley & Sons, UK, **2001**.
- 6 H. C. VAN NESS, *Understanding Thermodynamics*, Dover Publications, USA, **1969**.
- 7 B. H. MAHAN, *Elementary Chemical Thermodynamics*, W. A. Benjamin, USA, **1964**.
- 8 YA. A. SMORODINSKY (translated by V. I. KISIN), *Temperature*, Mir, USSR, **1984**.
- 9 L. D. LANDAU, A. I. KITAIGORODSKY (translated by M. GREENDLINGER), *Molecules*, Mir, USSR, **1980**.
- 10 J. S. DUGDALE, *Entropy and its Physical Meaning*, Taylor & Francis, UK, **1996**.
- 11 M. PLANCK, *Treatise on Thermodynamics*, Dover, USA, **1945**.
- 12 I. M. KLOTZ, *Ligand-Receptor Energetics: a Guide for the Perplexed*, John Wiley & Sons, USA, **1997**.
- 13 E. GRUNWALD, *Thermodynamics of Molecular Species*, John Wiley & Sons, USA, **1997**.
- 14 G. G. HAMMES, *Thermodynamics and Kinetics for the Biological Sciences*, John Wiley & Sons, USA, **2000**.
- 15 M. GRAETZEL, P. INFELTA, *The Bases of Chemical Thermodynamics*, Universal, USA, **2000**.
- 16 P. PERROT, *A to Z of Thermodynamics*, Oxford University Press, UK, **1998**.
- 17 L. K. NASH, *Elements of Chemical Thermodynamics*, 2nd ed., Addison-Wesley, USA, **1970**.
- 18 A. MACZEK, *Statistical Thermodynamics*, Oxford University Press, UK, **1998**.
- 19 M. GOLDSTEIN, I. F. GOLDSTEIN, *The Refrigerator and the Universe: Understanding the Laws of Energy*, Harvard University Press, USA, **1993**.
- 20 H. J. MOROWITZ, *Foundations of Bioenergetics*, Academic Press, USA, **1978**.
- 21 Y. A. ÇENGEL, M. A. BOLES, *Thermodynamics: an Engineering Approach*, 2nd ed., McGraw-Hill, USA, **1994**.
- 22 S. I. SANDLER, *Chemical and Engineering Thermodynamics*, 2nd ed., John Wiley & Sons, USA, **1989**.
- 23 G. N. LEWIS, M. RANDALL, *Thermodynamics* (revised by K. S. PITZER, L. BREWER), 2nd ed., McGraw-Hill, USA, **1961**.
- 24 E. B. SMITH, *Basic Chemical Thermodynamics*, 4th ed., Clarendon, UK, **1990**.
- 25 G. PRICE, *Thermodynamics of Chemical Processes*, Oxford University Press, **1998**.

- 26 I. WADSÖ, *Thermochimica Acta* **1995**, 267, 45–59.
- 27 E. FREIRE, O. L. MAYORGA, M. STRAUME, *Analytical Chem.* **1990**, 62, 950A–959A.
- 28 M. L. DOYLE, *Curr. Opinion Biotech.* **1997**, 8, 31–35.
- 29 R. R. J. O'BRIEN, I. HAQ, J. E. LADBURY in R. B. RAFFA, ed., *Drug-Receptor Thermodynamics: Introduction and Applications*, John Wiley & Sons, UK, **2001**.
- 30 H. NAGHIBI, A. TAMURA, J. M. STURTEVANT, *Proc. Natl. Acad. Sci. (USA)* **1995**, 92, 5597–5599.
- 31 J. R. HORN, D. RUSSELL, E. A. LEWIS, K. P. MURPHY, *Biochem.* **2001**, 40, 1774–1778.
- 32 D. J. WINZOR, W. H. SAWYER in R. B. RAFFA, ed., *Drug-Receptor Thermodynamics: Introduction and Applications*, John Wiley & Sons, UK, **2001**.
- 33 R. D. O'BRIEN, ed., *The Receptors, a Comprehensive Treatise: vol. 1 General Principles and Procedures*, Plenum, USA, **1979**.
- 34 W. H. SAWYER, D. J. WINZOR, *Curr. Protocols Protein Sci. A* **1999**, 5A, 1–40.
- 35 D. J. WINZOR, W. H. SAWYER, *Quantitative Characterization of Ligand Binding*, John Wiley & Sons, USA, **1995**.
- 36 S. P. COLOWICK, F. C. WOMACK, *J. Biol. Chem.* **1969**, 244, 774–776.
- 37 W. F. BLATT, S. M. ROBINSON, H. J. BIXLER, *Analytical Biochem.* **1968**, 26, 151–173.
- 38 H. PAULUS, *Analytical Biochem.* **1969**, 32, 91–100.
- 39 L. E. LIMBIRD, *Cell Surface Receptors: a Short Course on Theory and Methods*, 2nd ed., Kluwer, USA, **1996**.
- 40 R. B. RAFFA, S. D. SPENCER, R. J. SCHÜLINGKAMP, *Biochem. Pharmacol.*, in press.
- 41 A. W. NICHOLSON in G. D'ALESSIO, J. F. RIORDAN, eds., *Ribonucleases: Structures and Functions*, Academic Press, USA, **1997**.
- 42 R. W. HARTLEY in G. D'ALESSIO, J. F. RIORDAN, eds., *Ribonucleases: Structures and Functions*, Academic Press, USA, **1997**.
- 43 M. IRIE in G. D'ALESSIO, J. F. RIORDAN, eds., *Ribonucleases: Structures and Functions*, Academic Press, USA, **1997**.
- 44 I. G. WOOL in G. D'ALESSIO, J. F. RIORDAN, eds., *Ribonucleases: Structures and Functions*, Academic Press, USA, **1997**.
- 45 P. A. BARIOLA, P. J. GREEN in G. D'ALESSIO, J. F. RIORDAN, eds., *Ribonucleases: Structures and Functions*, Academic Press, USA, **1997**.
- 46 S. K. PARRY, Y.-H. LIU, A. E. CLARKE, E. NEWBIGIN in G. D'ALESSIO, J. F. RIORDAN, eds., *Ribonucleases: Structures and Functions*, Academic Press, USA, **1997**.
- 47 J. J. BEINTEMA, H. J. BREUKELMAN, A. CARSANA, A. FURIA in G. D'ALESSIO, J. F. RIORDAN, eds., *Ribonucleases: Structures and Functions*, Academic Press, USA, **1997**.
- 48 G. J. GLEICH, C. R. ADOLPHSON, *Adv. Immunol.* **1986**, 39, 177–253.
- 49 G. J. GLEICH, H. KITA, C. R. ADOLPHSON in M. M. FRANK, K. F. AUSTEN, H. N. CLAMAN, E. R. UNANUE, eds., *Samter's Immunologic Diseases*, Little, Brown, USA, **1995**.
- 50 M. R. SNYDER, G. J. GLEICH in G. D'ALESSIO, J. F. RIORDAN, eds., *Ribonucleases: Structures and Functions*, Academic Press, USA, **1997**.
- 51 C. M. CUCHILLO, M. VILANOVA, M. V. NOGUÉS in G. D'ALESSIO, J. F. RIORDAN, eds., *Ribonucleases: Structures and Functions*, Academic Press, USA, **1997**.
- 52 N. RUSSO, R. SHAPIRO, Potent inhibition of mammalian ribonucleases by 3',5'-pyrophosphate linked nucleotides. *J. Biol. Chem.* **1999**, 274, 14902–14908.
- 53 M. S. STERN, M. S. DOSCHER, *FEBS Letts.* **1984**, 171, 253–255.
- 54 K. HAYDOCK, C. LIM, A. T. BRUNGER, M. KARPLUS, *J. Amer. Chem. Soc.* **1990**, 112, 3826–3831.
- 55 G. L. GILLILAND in G. D'ALESSIO, J. F. RIORDAN, eds., *Ribonucleases: Structures and Functions*, Academic Press, USA, **1997**.
- 56 C. M. CUCHILLO, X. PARÉS, A. GUASCH, T. BARMAN, F. TRAVERS, M. V. NOGUÉS, *FEBS Letts.* **1993**, 333, 207–210.
- 57 F. M. RICHARDS, H. W. WYCKOFF HW in P. D. BOYER, ed., *The Enzymes* 3rd ed., vol. 4, Academic Press, USA, **1971**.
- 58 M. R. EFTINK, R. L. BILTONEN RL in A. NEUBERGER, K. BROCKLEHURST, ed., *Hydrolytic Enzymes*, Elsevier, The Netherlands, **1987**.
- 59 T. S. WISEMAN, S. WILLISTON, J. F. BRANDTS, L.-N. LIN, (1989) *Analytical Biochem.* **1989**, 179, 131–137.
- 60 J. J. BEINTEMA, J. HOFSTEENGE, M. IWAMA, T. MORITA, K. OHGI, M. IRIE, R. H.

- SUGIYAMA, G. L. SCHIEVEN, C. A. DEKKER, D. G. GLITZ DG, *Biochem.* **1988**, 27, 4530–4538.
- 61 S. D. SPENCER, O. ABDUL, R. J. SCHU-LINGKAMP, R. B. RAFFA, *J. Pharmacol. Exp. Ther.* **2002**, *in press*.
- 62 G. A. WEILAND, K. P. MINNEMANN, P. B. MOLINOFF, *Nature (London)* **1979**, 281, 114–117.
- 63 G. A. WEILAND, K. P. MINNEMANN, P. B. MOLINOFF, *Molec. Pharmacol.* **1980**, 18, 341–347.
- 64 P. NICOLAS, R. G. HAMMONDS JR, S. GOMEZ, C. H. LI, *Arch. Biochem. Biophys.* **1982**, 217, 80–86.
- 65 P. A. BOREA, G. M. BERTELLI, G. GILLI, *Eur. J. Pharmacol.* **1988**, 146, 247–252.
- 66 N. BOURHIM, P. CANTAU, P. GIRAUD, E. CASTANAS, *Compar. Biochem. Physiol. C.: Compar. Pharmacol. Toxicol.* **1993**, 105, 435–442.
- 67 G. FÁBIÁN, S. BENYHE, J. FARKAS, M. SZUCS, *J. Recept. Signal. Transduct. Res.* **1996**, 16, 151–168.
- 68 J. G. LI, R. B. RAFFA, P. CHEUNG, T. B. TZENG, L.-Y. LIU-CHEN, *Eur. J. Pharmacol.* **1998**, 354, 227–237.
- 69 P. ZEMAN, G. TOTH, R. KVETNANSKY, *Gen. Physiol. Biophys.* **1987**, 6, 237–248.
- 70 P. A. MAGUIRE, G. H. LOEW, *Eur. J. Pharmacol.* **1996**, 318, 505–509.
- 71 J. MOITRA, H. A. OKTEM, A. BORSODI, *J. Neurochem.* **1995**, 65, 798–801.
- 72 R. HITZEMANN, M. MURPHY, J. CURELL, *Eur. J. Pharmacol.* **1985**, 108, 171–177.
- 73 K. D. WILD, S. K. YAGEL, R. B. RAFFA, *Intl. Narcotics Res. Conf.*, Garmisch, **1998**.
- 74 R. B. RAFFA, K. D. WILD, H. I. MOSBERG, F. PORRECA, *Eur. J. Pharmacol.* **1993**, 244, 231–238.
- 75 K. D. WILD, F. PORRECA, H. I. YAMAMURA, R. B. RAFFA, *Proc. Natl. Acad. Sci. (USA)* **1994**, 91, 12018–12021.
- 76 R. B. RAFFA, K. D. WILD, H. I. MOSBERG, F. PORRECA, *J. Pharmacol. Exp. Ther.* **1992**, 263, 1030–1035.

4

The Biophore Concept

S. PICKETT

4.1

Introduction

The principles governing the binding of a ligand to its biological target can be understood in terms of intermolecular forces such as van der Waals interactions, hydrogen bonding, electrostatic interactions, and aromatic π - π interactions [1]. The detailed modeling and application of these principles are covered in detail elsewhere in this volume. In this chapter we employ a more general description of ligand-receptor binding that is an abstraction of these fundamental principles.

Complementary groups on the protein target recognize key features of the ligand. The three-dimensional arrangement of these features is commonly referred to as a pharmacophore. The Medicinal Chemistry Section of IUPAC has published a glossary of terms used in medicinal chemistry that includes an entry for the concept “pharmacophore” or “pharmacophoric pattern.”

“A pharmacophore is the ensemble of steric and electronic features that is necessary to ensure the optimal supramolecular interactions with a specific biological target structure and to trigger (or to block) its biological response.” [2]

In a clarification of this statement the definition goes on to state:

“A pharmacophore does not represent a real molecule or a real association of functional groups, but a purely abstract concept that accounts for the common molecular interaction capacities of a group of compounds towards their target structure. The pharmacophore can be considered as the largest common denominator shared by a set of active molecules. This definition discards a misuse often found in the medicinal chemistry literature which consists of naming as pharmacophores simple chemical functionalities such as guanidines, sulfonamides, or dihydroimidazoles (formerly imidazolines), or typical structural skeletons such as flavones, phenothiazines, prostaglandins, or steroids.”

The implication within this definition is that compounds, which share a pharmacophoric pattern pertinent to a particular target, are likely to bind to the said target. There are several important caveats to this statement that should be kept in mind. First, as defined, a pharmacophore is necessary but not sufficient for activity. The binding energy will be determined by other properties of the molecule not implied by the pharmacophore, e.g., additional groups could lead to a steric clash with the

protein site. As we see below, the inclusion of shape within the pharmacophore is an active area of research. Second, the pharmacophore is pertinent to one binding site and one binding mode. For example, the archetypal serine protease inhibitor has a basic group in the P1 pocket and for a long time this was thought to be essential for activity. However, inhibitors have been discovered recently that possess a neutral group that interacts with a tyrosine at the base of the P1 pocket [3, 4]. In other words, absence of the pharmacophore is not necessarily predictive of absence of binding. With these caveats in mind, the pharmacophore concept has proven to be very useful in the identification and optimization of drug candidates.

The term “pharmacophore” was first employed by Ehrlich early in the twentieth century [5], and the evolution of the concept has been reviewed by Gund [6]. The link between chemistry and biology implicit in the definition of a pharmacophore makes it an ideal descriptor for a number of design tasks. The application of pharmacophore methods in drug discovery is well established, and a number of reviews are available on this topic [7–10]. The concept of a “biophore” extends beyond the definition of key interacting features to include other forms of (3-D) molecular representations that capture the essence of the protein-ligand interaction. Thus, in this review we consider both pharmacophore methods and other methods taking account of the shape and surface electrostatic properties of ligands that ultimately determine the binding affinity for a protein site.

Section 4.2 provides an overview of methods for the generation of pharmacophores from both a ligand-based and structure-based perspective and their use in database searching. Section 4.3 describes methods for describing molecules by the ensemble of possible pharmacophores they possess and other methods of structural representation involving shape and electrostatic properties. Section 4.4 presents applications of these methods, ranging from lead generation, using a single well-defined pharmacophore, to library design, where the goal is to focus compounds towards pharmacophores of interest. Section 4.5 shows how the biophore concept is now being applied in the increasingly important area of ADME prediction. Section 4.6 summarizes the chapter.

4.2

Methodology for Pharmacophore Detection and Searching

The pharmacophore is an attempt to capture the essential features of the ligand-protein interaction. As such, it needs to be specific enough to be useful for a particular target and at the same time general enough that the information can be used to identify new molecules or chemotypes that are also likely to bind the target. In this section we look at the different stages of the application of a pharmacophore based methodology in drug discovery. In Section 4.3 we examine how these ideas can be expanded to give a full molecular descriptor based on the pharmacophore concept useful in diversity and library design.

Fig. 4.1 presents an overview of the different tasks involved in pharmacophore elucidation and searching. Each of these steps is covered in more detail below. To begin

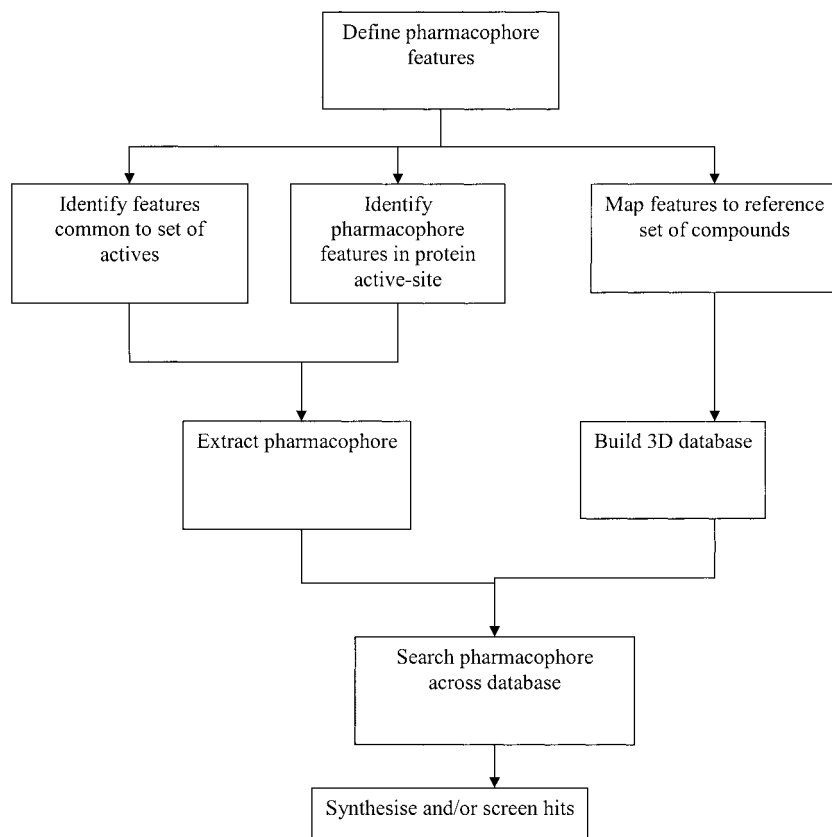


Fig. 4.1 Overview of the pharmacophore elucidation and searching process.

with, we need to arrive at a definition of pharmacophore groups that can be applied to a large number of compounds. Next we need to use these definitions in conjunction with molecular modeling software to define the pharmacophore based on a series of active compounds, derive it from knowledge of the structure of the target protein, or better to combine the two. Finally, the pharmacophore is used to design new compounds for synthesis or select compounds from databases for screening, again using bespoke software. As a preliminary to this, it will of course be necessary to generate an appropriately formatted database of compounds to search.

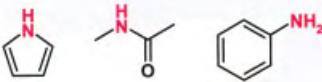
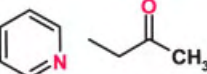
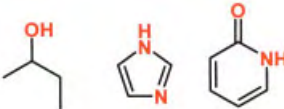
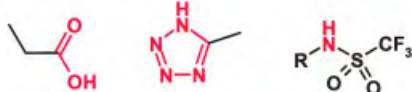
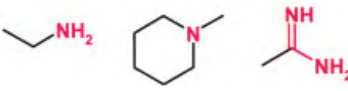
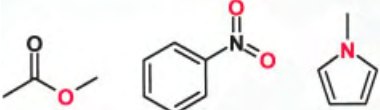
4.2.1

Definition of Pharmacophoric Groups

The pharmacophore concept is based upon the premise that different chemical groups can have the same types of interaction with a protein. For example, carboxylic acids, certain sulfonamides, and tetrazoles are acidic and so can be made equivalent in a pharmacophore context. In order to achieve this equivalence, how-

ever, the user must have a means for defining and identifying pharmacophorically similar groups. Pharmacophores are usually defined using six standard features: hydrogen-bond acceptor; hydrogen-bond donor; hydrophobic, aromatic, positive ionizable (base); and negative ionizable (acid). Some examples of functional groups giving rise to these features are given in Tab. 4.1. The commercially available product Catalyst [11, 12] identifies pharmacophore features via a features dictionary containing fragments that define the pharmacophore with exclusions being defined similarly so that, for example, the hydroxyl of a carboxylic acid is not included as a donor. The database system UNITY [13] utilizes Sybyl Line Notation (SLN) [14], a language specially developed for representing chemical structure and queries, to define donors and acceptors. Hydrophobic region definition is potentially more complex. Using a fragment-based approach would require defining a very large number of substructures and presents problems with regard to halogens and so on. Hence, within UNITY the default definition only identifies 5- or 6-membered rings. Catalyst utilizes a more complex algorithm where an atom is defined a hydrophobicity value based on the neighboring atoms and the atom's surface accessibility. Groups of hydrophobic atoms are then identified where the sum of the individual hydrophobicity values is above a user-definable minimum value.

Tab. 4.1 Examples of commonly defined functional group equivalences and the corresponding pharmacophoric definition.

Pharmacophore group	Example structures
Donor	
Acceptor	
Donor and acceptor	
Acid	
Base	
Atoms excluded ^{a)}	

a) Highlighted atoms in these groups are generally not considered pharmacophoric.

An important feature of the Catalyst and UNITY systems and pharmacophore perception programs such as DISCO [15] and GASP [16] is that they allow the inclusion of receptor site points within the feature definition for hydrogen-bond donor and acceptor groups. The nature of hydrogen bonding and electrostatic interactions means that a receptor atom or functional group can interact strongly with ligand atoms in different positions. Thus, for a ligand hydrogen-bond acceptor atom, the approximate position of the receptor donor atom can be identified from the position of the lone pair on the acceptor. If there are two lone pairs, as in a carbonyl group, then there are two possible positions for the receptor donor atom. For a ligand donor atom such as a hydroxyl, the approximate position of the receptor acceptor atom is defined along the O-H bond vector. Account needs to be taken of possible rotation of the O-H group. Fig. 4.2 illustrates the principle of re-

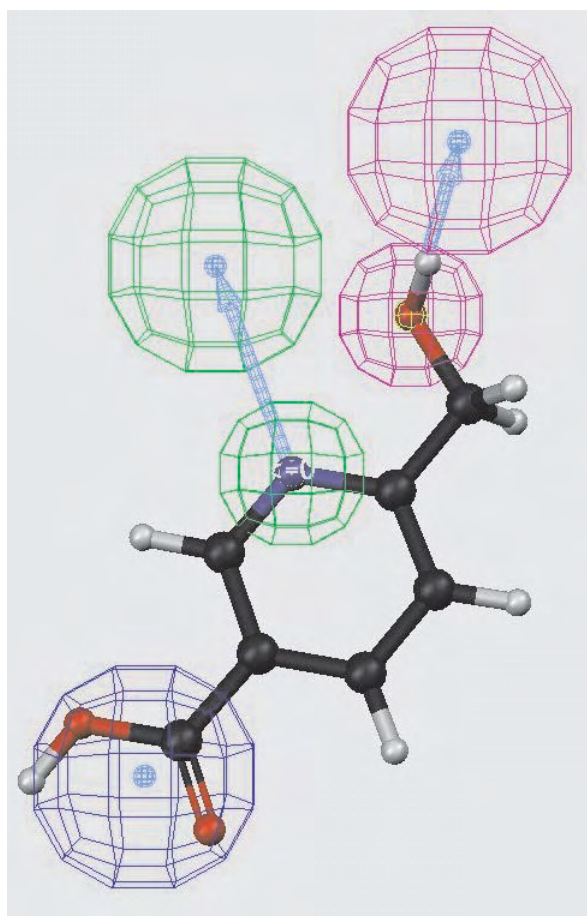


Fig. 4.2 Catalyst feature definition applied to a simple molecule, illustrating the inclusion of receptor site-points in the pharmacophore definition.

ceptor site-point definition through the application of Catalyst feature definition to a simple molecule. Its importance was evidenced by the work of Marshall and co-workers in deriving a pharmacophore for angiotensin-converting enzyme (ACE) inhibitors [17].

The appropriate definition of pharmacophore features allows ligand-protein interactions to be modeled in a general manner. The importance of this step should not be underestimated. For example, the parameter sets derived at Rhone-Poulenc in the early 1990s [18, 19] involved many man-weeks to obtain acceptable definitions for a range of groups. Nitrogens can be particularly troublesome in this regard, as they can be acceptors (pyridine), bases (primary amines), acids (certain sulphonamides), donors (amides), or have no feature (N-substituted pyrrole). The sp^3 oxygen of esters also needs consideration, as it is generally considered not to be a good acceptor. Thus, software systems need to be flexible enough to allow the user to input his/her own bias into the parameterization and also to allow for special cases. For example, when working on proteins containing metals, it may be appropriate to define a feature to represent a zinc-binding group.

4.2.2

Ligand-based Methods for Pharmacophore Perception

Having defined the appropriate pharmacophore features, the expert must next derive the pharmacophore of interest. A prerequisite, in the absence of a protein structure, is a series of active molecules that are presumed to bind in the same way. The pharmacophore may then be derived from examining the disposition of pharmacophore features within the molecules to locate common distances and then generating a superposition of the molecules. The key elements of this process are pharmacophore feature perception, described above, conformational analysis to explore the conformational space of the ligands, and identification of the common features.

The active analogue approach pioneered by Marshall and coworkers [20] attempts to identify common pharmacophoric features among a set of active compounds. This is a user-driven process whereby a set of potentially interacting atoms is selected in one molecule (usually the most rigid) and their pair-wise distances are recorded during a systematic search of the sterically allowed regions of conformational space. This distance map can then be used to constrain the conformational analysis on the next molecule, where corresponding atoms are identified. In general, this will lead to a more tightly defined set of distances that can be applied to the next molecule and so on. The pharmacophore can be identified from the final set of distances once all training set molecules have been processed. A number of successful applications of this method have been published [21, 22], and recent work has made significant improvements to the conformational sampling aspects of the method [23].

DISCO, developed by Yvonne Martin's group at Abbott, and the HipHop [24] module of Catalyst, rely upon pre-calculated conformers. For DISCO these can be generated using a standard modeling package. The Catalyst suite includes a mod-

ule, Confirm, that uses a novel poling algorithm to generate diverse sets of conformers [25–27]. In this method, a penalty function is added to the force field at a previously visited region of conformational space to bias the algorithm away from revisiting that region. This leads to broader conformational coverage. DISCO and HipHop utilize different methods for identifying the common pharmacophore features between the training set of molecules. DISCO employs a clique-detection algorithm [28] to locate common feature distances between conformers of the training set molecules. Martin has recently reviewed her experiences with DISCO [29]. HipHop employs an incremental buildup strategy, starting from one set of features (those features common to molecules in the training set) to consider feature pairs, triples, and so on, until no more correspondences can be found.

The program GASP [16, 30] adopts a different approach, employing a genetic algorithm to superimpose pharmacophore features between the training set of molecules. The GA also includes a conformational energy term in the scoring function so that conformations are changed as part of the mutation and crossover steps of the evolutionary process.

All of the automated methods will return multiple solutions (pharmacophores). A number of criteria can be used to manually select between these models: the conformational energies of the structures as they fit the pharmacophore, considering additional SAR on inactive compounds, and so on. If a number of active and inactive molecules are known, looking at retrieval rates of active versus inactive compounds from a database search using the pharmacophore can provide useful information on the quality of the pharmacophore. An example of such a procedure is shown in Fig. 4.3. Here the “enrichment” [31] as defined by the ratio (number of actives in the hit list/hit list size:number of actives in hit list/database size) is compared to the coverage, the ratio (number of actives in hit list:number of actives in the database) for a set of pharmacophores derived from a series of HIV protease inhibitors using Catalyst [32]. Each hypothesis was searched against a database of 150,000 compounds containing 647 known active compounds and the retrieval rate of actives recorded. From the plots in Fig. 4.3, it can be seen that in many cases the enrichment and coverage are in competition. Guner and Henry [33] have proposed a formula for assessing the “goodness-of-hit,” which builds upon some of these ideas:

$$GH = \left(\frac{H_a(3A + H_t)}{4H_tA} \right) \times \left(1 - \frac{H_t - H_a}{D - A} \right), \quad (\text{Eq. 4.1})$$

(reprinted with permission from [34])

where D is the number of compounds in the database, A is the number of actives, H_a is the number of actives in the hit list, and H_t is the total number of compounds in the hit list. Tab. 4.2 lists the values of GH for the hypotheses shown in Fig. 4.3.

Methods discussed so far are based on a few known active ligands. Several methods have been developed that attempt to use the activity data when generating the pharmacophore model. The HypoGen module of Catalyst uses a cost func-

Enrichment of hits and effectiveness of finding all possible hits.

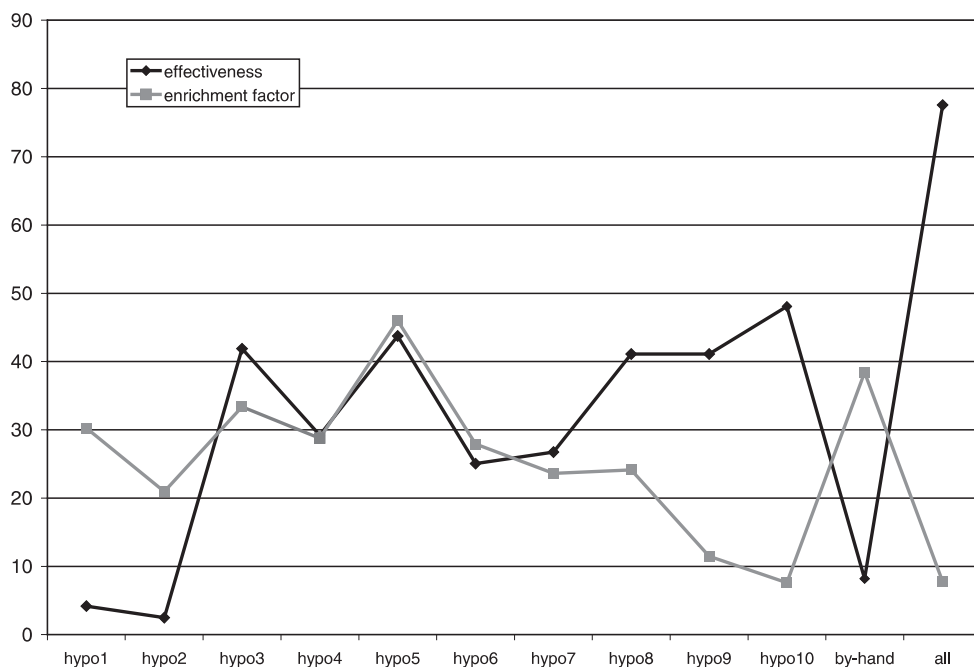


Fig. 4.3 Enrichment and coverage plots for hypotheses derived from a set of HIV protease inhibitors. Each hypothesis was used to

search a database containing active and inactive compounds.

Tab. 4.2 Goodness-of-hit (GH) scores for the hypotheses shown in Fig. 4.3^{a)}.

Hypothesis	H_a	H_t	%Y	%A	GH
Hypo1	27	207	0.13	4.17	0.11
Hypo2	16	177	0.09	2.47	0.07
Hypo3	271	1882	0.14	41.89	0.21
Hypo4	189	1523	0.12	29.21	0.16
Hypo5	283	1426	0.20	43.74	0.26
Hypo6	162	1346	0.12	25.04	0.15
Hypo7	173	1698	0.10	26.74	0.14
Hypo8	266	2555	0.10	41.11	0.18
Hypo9	266	5377	0.05	41.11	0.14
Hypo10	311	9481	0.03	48.07	0.14
By-hand	53	320	0.17	8.19	0.14
All ^{b)}	502	15061	0.03	77.59	0.20

a) H_a is the number of actives hit by the pharmacophore; H_t is the total number of hits; $\%Y = H_a/H_t$; $\%A = H_a/A$, where A is the total number of actives in the dataset and GH is given by Eq. 4.1. Results are for a database of 150,000 compounds containing 647 active molecules.

b) All represents the combination of all the hypotheses.

tion to optimize the selection of predictive pharmacophore models or hypotheses [35]. Each pharmacophore point is given a feature weight, and the cost function has three components: a Gaussian weight component that penalizes the deviation of the feature weight from an ideal value of two; an error component that increases as the differences between predicted and measured activities of the training set increase; and a configuration cost that depends on the complexity of the hypothesis space being searched. Compounds are scored based on their ability to fit a hypothesis. Such hypotheses have been used successfully as the basis for database searching, for example [36]. SCAMPI (statistical classification of activities of molecules for pharmacophore identification) [37] combines recursive partitioning, as implemented in SCAM [38], with a conformational search engine to dynamically evolve the pharmacophore. Recursive partitioning provides a statistical method for selecting pharmacophore features that are most significantly correlated with biological activity. A dendrogram or tree is generated where each branch point or split is determined by the presence or absence of a particular pharmacophore feature within the molecules of the training set. The conformational space of the compounds is continually refined under the constraints of the evolving model. Fig. 4.4a shows the SCAMPI tree derived from a series of CDK2 inhibitors [39]. An active structure contained in the highlighted node in Fig. 4.4b

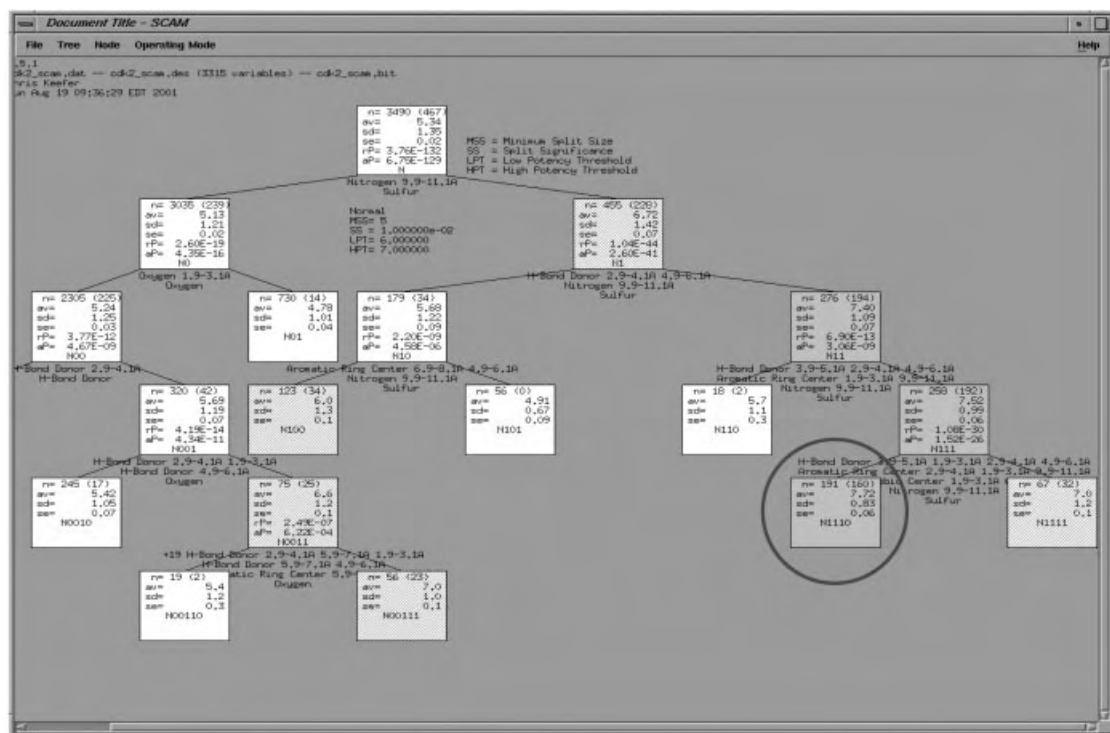


Fig. 4.4a

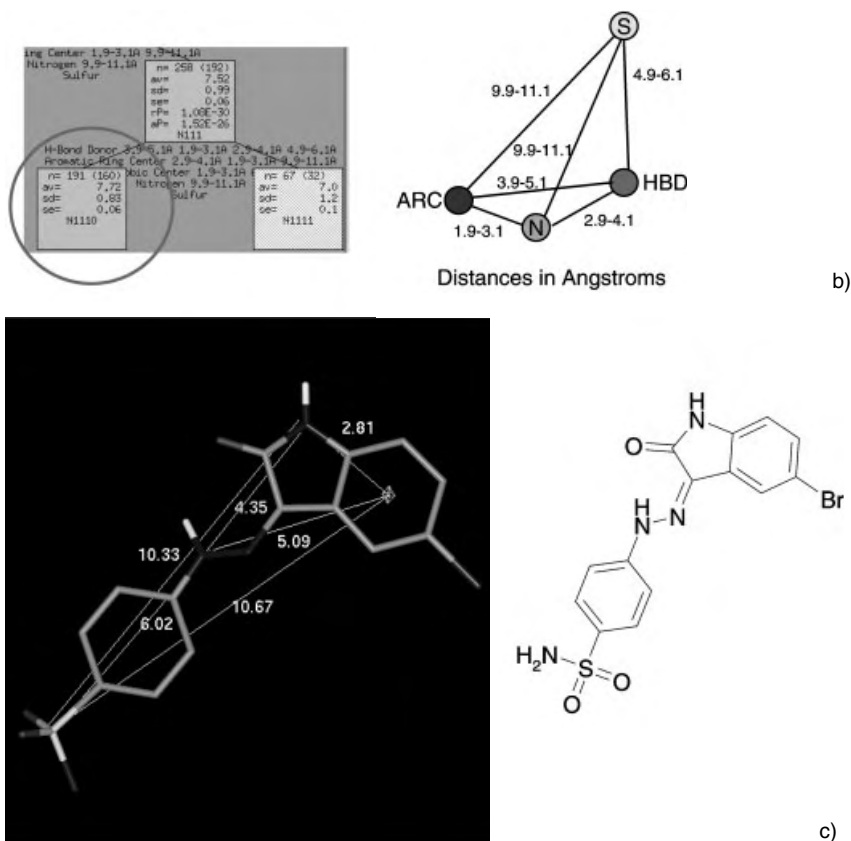


Fig. 4.4 SCAMPI analysis of a series of CDK2 inhibitors. (a) Full tree with one node highlighted, (b) enlarged picture of one node and

the resulting pharmacophore, (c) active molecule within the node with the pharmacophore highlighted.

is shown in Fig. 4.4c, with the node pharmacophore indicated. Compared to the other methods described here, SCAMPI can be applied to large datasets of several thousand compounds in a reasonable time (a day or so). The methodology has been shown to reproduce known literature pharmacophores.

A more detailed description of the ligand is obtained from field-based methods. Rather than the usual feature-based pharmacophore definition, the molecule is characterized by the electrostatic and steric fields at and beyond its surface. The interaction energy between a molecule and one or more probe fragments can be calculated on a uniform grid around the molecule, as in the well-known program GRID [40, 41]. The CoMFA method [42] uses a statistical methodology such as PLS to correlate the interaction energies with the biological activity for a series of molecules with a range of activities. Such an approach has been successfully ap-

plied in QSAR-type applications [43–45]. The important regions determining the biological activity can be identified from the contribution of individual grid points to the activity. Visualization of the resulting maps provides an insight into the regions of the molecules that are beneficial or detrimental to activity and can guide the design of new molecules. The time-consuming and difficult part of the process is to align the molecules and select a single conformation prior to field generation. As this step is user-defined and can be somewhat subjective, the resulting interaction maps do not necessarily represent a pseudo receptor. Several groups [46, 47] have used pharmacophore models to define the alignment for CoMFA and used the existence of a statistically relevant model as an indicator for reliability of the pharmacophore model.

Modifications to the original CoMFA methodology have been developed that improve the interpretability of the maps. Golpe [48, 49], developed by Clementi and coworkers at the University of Perugia, uses field maps from a number of sources including Sybyl [50] and GRID in combination with PLS. The program includes methodology for variable selection to focus on the grid points most important for describing the activity and a region selection algorithm [51] that ensures that the selected points are grouped together, making the resulting maps more interpretable.

A potential problem of the CoMFA method is that the descriptors are dependent on the compound alignment chosen and the position of the molecules within the grid used to calculate the interaction energies. The GRIND descriptors [52], available in the program ALMOND [53], were developed to produce a pharmacophoric representation of the fields in an effort to overcome these problems. The original field maps are transformed using a Maximum Auto-Cross Correlation (MACC) technique. The interaction energies of two grid nodes are multiplied together and recorded as a function of distance. The maximum interaction energy for a particular distance is recorded along with the location of the grid nodes giving rise to it. The resulting profile of energy and distance can then be analyzed using PCA or PLS. The pairwise interactions important for activity can be identified from the analysis. The methodology has been validated against several QSAR datasets. Fig. 4.5 shows a plot resulting from an analysis of butyrophenones with serotonergic affinities. The GRIND descriptors have subsequently been applied successfully to the derivation of a 3-D QSAR model for a series of dopamine transporter ligands [54], with the results being interpretable in a pharmacophoric sense. The 4-D QSAR methodology of Hopfinger and coworkers [55] includes multiple conformations and alignments of the molecules in the derivation of the QSAR model, thus removing the need for a user-defined alignment.

Other field-based methods have been developed for compound superposition [56]. In the ASP (Automated Similarity Package) procedure [57], the electrostatic field around the molecule is modeled using a Gaussian approximation [58]. This allows the use of analytic gradients and hence improved performance in optimizing the electrostatic overlap between two molecules. Conformational flexibility also can be included within the search. A similar procedure can be used for the “steric field.” Dean and coworkers [59] have used a simulated annealing approach in superimposing molecules by minimizing the difference in the intramolecular

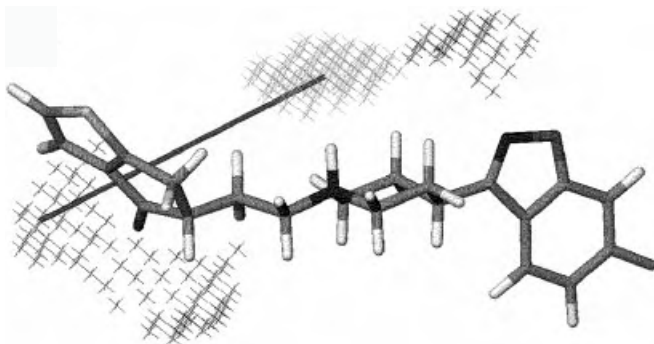


Fig. 4.5 Fields from a GRIND analysis on a series of butyrophenones with serotonergic affinities. Fields shown are for an O (acceptor)

probe and an N1 (donor) probe. Interactions at the distance shown are present in active compounds such as the one shown [52].

distance matrices between two molecules. This was extended to matching hydrogen bonding and hydrophobic regions [60]. The particular interest here is the use of hydrogen bonding probability maps [61] derived from an analysis of the Cambridge Crystallographic Database [62]. The SQ method [63, 64], developed at Merck, uses pharmacophorically relevant atomic properties (for example an atom is marked as a hydrogen-bond donor) in scoring alignments generated with a clique-detection algorithm. An innovation in this approach is to consider only atoms that have the same characteristics.

4.2.3

Protein Structure-based Pharmacophore Perception

Programs such as GRID [40], previously mentioned in the context of ligand-based analysis, were first developed to provide an understanding of a protein-binding site. By calculating the interaction energies of small molecular fragments with different binding properties such as a carbonyl oxygen (hydrogen-bond acceptor), anilinic nitrogen (hydrogen-bond donor), and Csp^3 carbon (hydrophobe), it is possible to identify regions within the binding site whereby such interactions are favored and disfavored. Such an approach was used successfully by von Itzstein and coworkers in the discovery of neuraminidase inhibitors [65].

The program LUDI [66] provides an alternative approach. Hydrogen bonding and hydrophobic interaction sites are identified within the protein using a series of rules [67] derived from interactions within the Cambridge small-molecule Crystallographic Database [62]. Fragments are searched to fit the interaction sites. An empirical scoring function has been derived by fitting to observed K_i values of protein ligand complexes to score designed structures [68]. A typical LUDI map for the bacterial enzyme DNA Gyrase [69, 70] is shown in Fig. 4.6. The figure shows a cut-through of the Connolly surface of the ATP-binding site of the DNA gyrase B subunit, with the non-hydrolyzable ATP analog ADPNP (5'-adenylyl β - γ -imidodiphosphate) bound. LUDI interaction sites are shown as blue and white (donor) or

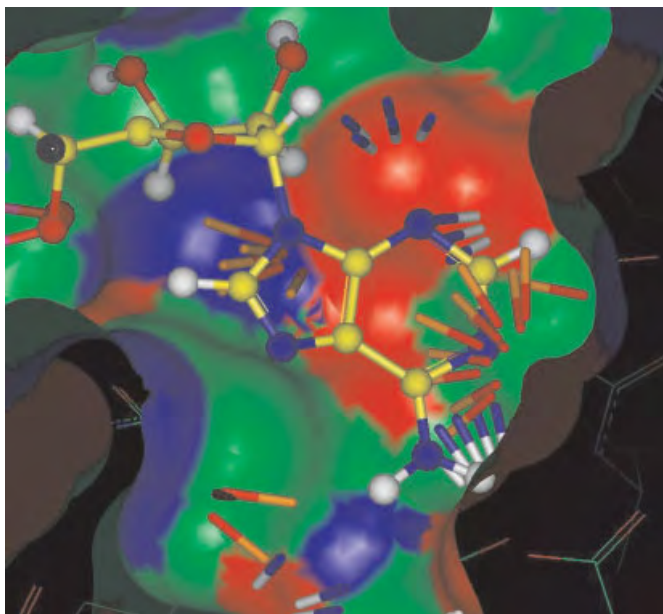


Fig. 4.6 Cut-through of the Connolly surface of the DNA Gyrase B subunit active site. The non-hydrolysable ATP analogue ADNP is shown in mainly yellow. LUDI interaction sites

are shown as blue and white rods to represent donor groups and orange and red rods to represent acceptor groups.

orange and red (acceptor) rods. It can be seen how predicted donor and acceptor sites map to the aromatic nitrogen and aromatic amine of the adenylyl ring. We will return to this particular system in the applications section. The CCDC (Cambridge Crystallographic Data Center) has developed the SuperStar program [71, 72] to predict interaction sites in the protein active site. SuperStar utilizes interaction maps derived with IsoStar [73] from the large amount of crystallographic information available within their databases.

Thus, analysis of GRID or LUDI maps or the application of other methods for feature extraction [74] allows the user to identify possible pharmacophores to use as a basis for database searching. Such searches are particularly powerful when the receptor site is included as an additional steric constraint [75]; this is discussed further in the next section. The Design in Receptor (DiR) module of Chem-X [76, 77] was developed specifically to generate possible pharmacophores from a receptor site. This functionality, in modified form, is now available through the THINK program [78]. Pharmacophore points complementary to receptor atoms are placed in the binding site. All possible three- or four-point pharmacophores are then available for searching.

A number of docking algorithms employ scoring functions similar in principle to those of LUDI [79–82]. Such functions tend to be faster to calculate than all-atom force field energies, and a good parameterization of hydrogen bonding geo-

metries and the empirical inclusion of hydrophobic interactions can lead to good results. The program DOCK [83] works by matching a small number of ligand atoms to spheres defined within the protein-binding site. The simplest scoring function within the program scores this binding mode based upon a contact score and considering the number of atoms having a steric overlap with the site. The concept of “sphere coloring” introduces a pharmacophoric element to the search by marking which atom types or substructures are allowed to match a particular sphere [84]. The atom-type definitions are general enough to allow definition of hydrogen-bond donors, acceptors, acids, bases, etc., or can be very specific, e.g., amide NH. The inclusion of such pharmacophoric constraints within the docking can significantly improve the results in database searching applications [10].

4.2.4

Methods for Pharmacophore Searching

Once a potential pharmacophore has been identified through analysis of active molecules or a ligand-receptor complex, it is necessary to search against a database of compounds. These compounds may be members of the corporate screening collection, where the aim is to identify compounds to screen. Alternatively, the database may be generated from a virtual library of compounds. The term “virtual” here covers a range of possibilities. These could be a set of compounds accessible via a particular combinatorial synthesis, compounds available for purchase from external suppliers, and so on. Whether the compounds are “real” or “virtual” the process described below is the same, the difference being that with virtual compounds they must be synthesized or otherwise acquired prior to testing.

We are now following the right-hand branch in Fig. 4.1. The compound database must be generated. Usually, compound structures are available in a 2-D or “flat” format such as an MDL Mol file [85] or SMILES [86, 87] notation. Several programs such as Corina [88] and Concord [89] are available to convert these 2-D formats to a 3-D structure. This is the first step in database building, though database-searching programs such as Catalyst can do this internally.

Following construction of a 3-D structure, the compound is registered in the database. All programs use some form of keying to speed up the search [90]. This keying is usually at several levels. For each molecule, the presence of certain atoms or pharmacophoric atom types, as defined above, is recorded in a bit-string. At search time, this bit-string provides a screen-out for all compounds that cannot possibly match the pharmacophore. At a secondary level, the distances between pairs of pharmacophoric atom types are also recorded. The distances are binned into ranges and stored in a binary key where a 1 will indicate the presence of that atom pair at a particular distance. Programs such as UNITY [13] and ISIS/3D [91, 92] store a single conformer upon registration and sample conformation space at search time. The distance screens are constructed at registration from the bond-path between the two atoms of interest. This allows definition of a max-min distance on an atom pair and provides a reasonable screen-out. The program Catalyst

[11] performs the conformational analysis at registration, storing conformational models within the database. Conformers are generated using the same poling algorithm described above. This allows a relatively small number of conformers to be generated that are representative of the full conformational space.

In order to search the database, it is first necessary to cast the query into a form suitable for the searching program. This usually involves sketching the required pharmacophore points, labeling the allowed types, and defining the distance constraints on the pairwise distances. More advanced pharmacophore models also may incorporate angle constraints, distances above planes, etc. It should be noted that tolerances are applied to all the constraints. The earliest 3-D searching software [93, 94] considered just a single conformer. A variety of approaches for including conformational models into the search have been studied for their efficiency in hit retrieval [95–97]. In the directed-tweak approach [91, 98], a starting conformer is generated and relevant torsions are driven towards satisfying the distance constraints of the query. However, satisfying distance constraints alone is not sufficient to satisfy the query [12], and a further step of superimposing onto the pharmacophoric points is required. Alternatively, a systematic search of conformational space can be performed at search time [99]. With such an approach, a van der Waals bump-check is employed in lieu of a full energy calculation because of the time constraints on searching tens of thousands of structures. On the other hand, if conformers are pre-calculated when building the database, at search time they can be retrieved and searched as in the Catalyst FAST mode. Catalyst BEST mode utilizes a combination of both methods, using the stored conformers as starting structures for further directed optimization to satisfy the query.

A pharmacophore model describes the minimum requirement for a ligand to bind to a receptor in a particular mode. One limit of such a model is that it takes no explicit account of molecule shape or size. That is, when the pharmacophore is used to search the database, it may match only a fragment of a molecule. The inclusion of shape into the query is thus an important area of research. Most searching software allows the inclusion of excluded volumes within the pharmacophore definition. It may be possible to identify proposed sterically inaccessible regions from a study of active and inactive molecules superimposed onto the pharmacophore. Hahn has developed a methodology for shape-similarity searching [100] that is based upon a receptor surface model (RSM) [101]. Shape indices are generated from the extents of the principal axes and the volume of a conformation. These indices act as pre-filters. Next, the overlapping volumes of the query and a candidate molecule are optimized. As an optional final step, the candidate molecule can be flexibly fit to the surface properties of the RSM, measuring complementarity to the hydrogen-bonding, hydrophobic, and electrostatic properties of the model. As an alternative to this latter step, the shape query can be merged with a pharmacophore query. A study of the effectiveness of this strategy suggests that pharmacophore searching is superior to shape-based searching in terms of both enrichment and the *GH* score (Eq. 4.1), while using a merged pharmacophore/shape query provides a very selective query but one with a low coverage [102].

The inclusion of the protein active site can add an important steric or shape constraint to the search, limiting the size of the selected molecules appropriately, as shown by application to the thyroid hormone receptor [75]. However, the real power of such methods comes from the ability to search receptor-based pharmacophores with only a limited user bias on the points to be used. This has been implemented via partial-match searching [103], available within UNITY, which marks a molecule as a hit if it matches 4 out of 20 possible pharmacophoric features in the site. Additional constraints on the solutions can be imposed by grouping the pharmacophore points and limiting matches to one or more points from each group, a feature of the THINK program [78]. In this way, molecules can be selected that explore several areas of the binding site.

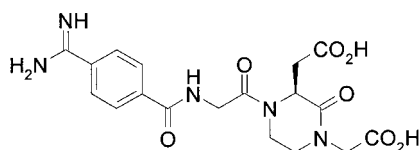
4.3

Pharmacophore Fingerprints

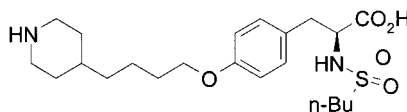
The successes achieved with “traditional” pharmacophore modeling have led many groups to look at ways of describing molecules in a similar way without the need for alignment or derivation of a single pharmacophore. The pharmacophore represents the key elements of a protein-ligand interaction, and, thus, the hope is to arrive at a descriptor that describes molecules based on their biology rather than their chemistry. The standard 2-D similarity measures based around the Daylight fingerprints [104] or ISIS keys [105] group compounds based on common chemistry. Pharmacophore-based descriptors attempt to move away from this chemistry-biased representation. Compounds similar in a pharmacophore space do not need to look similar in a chemical sense. As an example of this, Fig. 4.7 shows three potent fibrinogen antagonists [106]. In a 2-D sense, they are dissimilar, and yet they share the acid-base motif necessary for activity. Whole molecule descriptors based on pharmacophores were able to identify these compounds with a high degree of enrichment in a pool of inactive molecules, using the RGD tripeptide as a probe [107].

In a strict sense, the generation of a pharmacophore descriptor requires the generation of multiple 3-D conformations for a molecule and the accumulation of descriptors over all conformations. However, early work in this area built upon many of the 2-D similarity methods and did not require a 3-D structure at all. The relevance to the Biophore Concept is the way in which particular atoms are defined. In a Daylight fingerprint, for example, oxygen is different from nitrogen and sulfur, while in approaches such as the binding property pairs of Sheridan and coworkers [108], atoms are described according to pharmacophoric feature mappings similar to those in Section 4.2, $\log P$ and partial atomic charge. All these properties are important to ligand-receptor interactions. The importance of such descriptors is not that they are better per se at selecting actives from a database. 2-D measures usually do a good job at this, as chemists tend to synthesize series of compounds. Rather, the pharmacophore methods identify alternative series and thus can be useful for “scaffold-hopping.” Schneider and coworkers devel-

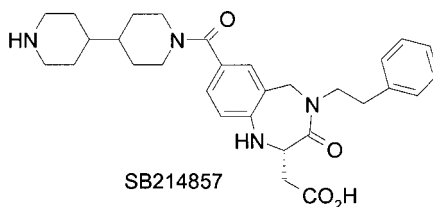
Fig. 4.7 Three potent fibrinogen antagonists that are 2-D dissimilar (different chemotypes) yet are similar in a pharmacophore sense (see text).



TAK029



MK383



SB214857

oped their CATS descriptor (chemically advanced template search) [109] for just such a purpose. For each pair of pharmacophoric features (donor, acceptor, acid, base, hydrophobe) in the molecule, the frequency of occurrence as a function of the number of bonds separating the features is accumulated into a pharmacophore pair vector P . These histograms are normalized by the number of heavy atoms in the molecule. As implemented, bond distances from 1 to 10 were considered over all 15 feature combinations to give a vector of size 150. The Euclidian distance between two molecules is used as the similarity measure (Eq. 4.2):

$$\text{Distance}(A, B) = \sqrt{\sum_{i=1}^{i=150} (p_i^A - p_i^B)^2}, \quad (\text{Eq. 4.2})$$

where p_i^A is the normalized count at position i for molecule A. As an application of this methodology, the T-type calcium channel blocker mibefradil (Fig. 4.8a) was used as a query molecule for a CATS search of the Roche corporate database. The 12 highest-ranking compounds were tested, and 9 had $\text{IC}_{50} < 10 \mu\text{M}$. One of these, clopimozid (Fig. 4.8b), had an $\text{IC}_{50} < 1 \mu\text{M}$. It is clear that the two structures are different in a 2-D sense, though common pharmacophore features, such as the basic nitrogen, are conserved.

We have already alluded to the problems associated with generating a full 3-D molecular descriptor, namely, the need to consider multiple conformations. In our

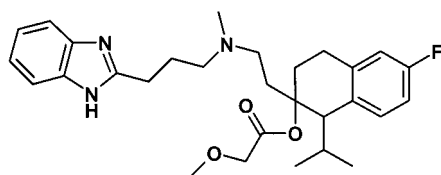
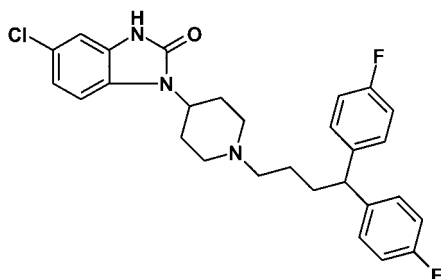


Fig. 4.8 (a) Calcium channel blocker mibefradil, (b) clopimozid, identified as a calcium channel blocker through application of the CATS methodology (see text).

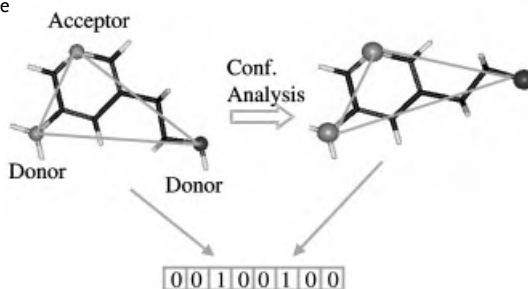
a)



b)

initial work in this area [18], we used the power of a pharmacophore-searching program (Chem-X) to address this problem. All geometrically allowed three-point pharmacophores were generated from combinations of six pharmacophore features (hydrogen-bond acceptor, hydrogen-bond donor, acid, base, aromatic ring centroid, and hydrophobe) and six distance ranges (2–4.5, 4.5–7, 7–10, 10–14, 14–19, and 19–24 Å). The resulting 5916 pharmacophores were then used to search the database of interest. The hit lists allowed a molecular descriptor to be generated as a bit-string marking the presence (1) or absence (0) of a particular pharmacophore within a molecule. Clearly this process was slow, as the conformational analysis is performed on each molecule for each query (provided the screens are passed; see Section 4.2). However, it was still possible to gain much useful information and to use the descriptor in profiling and designing compound libraries [18]. At about the same time, other groups developed similar descriptors based upon a single conformation [110]. However, the ChemDiverse module of Chem-X [76, 111] was also being developed at this time, and this provided an elegant solution to the problem. A conformational analysis is performed on each compound once. Each conformation is analyzed for the presence of pharmacophore triplets as shown in Fig. 4.9. The presence of a pharmacophore is marked by a 1 in a bit-string, as for the PDQ method. Theoretically, approximately 900,000 pharmacophores are possible, given 7 pharmacophore features and 32 distance ranges, though in practice fewer are used. The seventh pharmacophore feature can be used for atoms that are both hydrogen-bond donors and acceptors, such as hydroxyl and unsubstituted imidazole nitrogen. Alternatively, this seventh type can be used to define a particular point of reference in the molecules, which is useful for some focused design tasks, particularly when four-point pharmacophores are applied [112]. In this particular application, pharmacophore fingerprints were generated for known GPCR ligands and used to guide the de-

Fig. 4.9 Creation of a pharmacophore triplet fingerprint. As the conformation of the molecule changes, so do the distances between the pharmacophore features. The presence of a pharmacophore is indicated by a “1” in a bit-string.



sign of libraries containing GPCR “privileged” substructures. The privileged substructure was assigned to the seventh type so that diversity assessments could be made with respect to this motif.

Pharmacophore fingerprints tend to be very large (many bits) and are sparse (few bits are set). Such properties are not ideal for standard similarity methods [113]. In an interesting approach to overcome some of these limitations, Pozzan and coworkers have developed 3-D pharmacophoric hashed fingerprints [114]. The hashing algorithm first involves assigning a unique integer to each pharmacophore bit in the range 0–1016 and then folding these integers into a 1024-bit binary fingerprint. A similar procedure is followed by 2-D fingerprinting methods such as that used in the Daylight software [104]. Initial studies show promising results, particularly with regard to scaffold-hopping, when compared to 2-D fingerprints.

4.4

Applications of the Biophore Concept

4.4.1

Lead Generation

The earliest examples of the application of pharmacophore methods involved the use of such methods to derive pharmacophore models for compound design. This is a prerequisite to database searching as described in Section 4.2. Pharmacophore-based database searching is a proven method for lead identification [7, 8, 9] and continues to provide an alternative or a complement to high-throughput screening for this purpose. Several recent examples provide an overview of the methods discussed above. Endothelin antagonists have provided an important target for drug-discovery efforts. A group at Rhone-Poulenc derived a two-point pharmacophore from two published antagonists, a cyclic pentapeptide and a ligand based on the triterpene framework (Fig. 4.10) [115]. The triterpene structure was important to define the appropriate distances because of the flexibility of the cyclic peptide. A search of the corporate database retrieved 383 structures, and screening of these gave several active compounds of diverse structure. Building

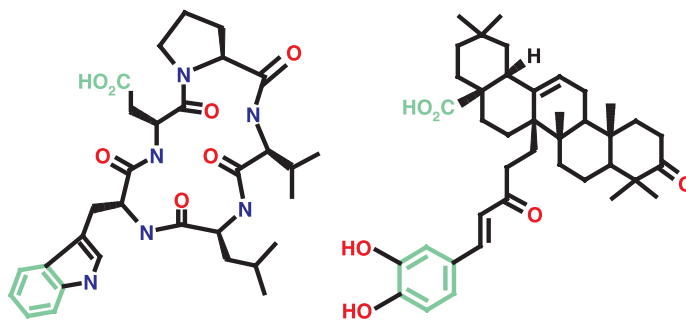


Fig. 4.10 Pharmacophore derived from two endothelin receptor antagonists. The common pharmacophore groups are highlighted in green.

upon this success, an active analogue-type approach was able to generate a unified model for these different series, as shown in Fig. 4.11 [116]. The important point to note about this model is the necessary inclusion of the site point representing a presumed basic center on the receptor that interacts with the acidic groups in ligands; these acidic groups do not overlap directly with each other. The use of this model guided the development of a series combining properties from the different lead series and, subsequently, led to the discovery of a development candidate RPR118031A [117].

Another recent example of pharmacophore-based lead discovery comes from the Zeneca (now AstraZeneca) group at Loughborough [118]. DISCO was used to generate several potential pharmacophore models for a set of muscarinic m_3 antagonists. Two models were selected and used to computationally screen the corporate database. Biological screening of 172 selected compounds gave three hits, which were structurally distinct from the compounds used to derive the pharmacophore models. The group at the NCI has had much success with the discovery of potential lead compounds against HIV integrase [119], HIV protease [120], protein kinase C [121], and HIV-RT [122].

4.4.2

Multi-pharmacophore Descriptors in Diversity Analysis and Library Design

A number of reviews are available on the theory and practice of molecular diversity analysis and library design [123–126]; therefore, we will only briefly summarize the role of pharmacophore methods in these areas. Multi-pharmacophore descriptors were first developed as a means to assess the diversity of a compound collection in a biologically relevant manner [18]. As a partitioning method, [127] it is straightforward to compare different sets of compounds, identifying common pharmacophores or pharmacophores deficient in one set of compounds (say a corporate collection) and highly populated in another set (say a group of compounds active against a particular target). These ideas also can be applied in the context of library design, where the aim is to select a subset of library products that can be

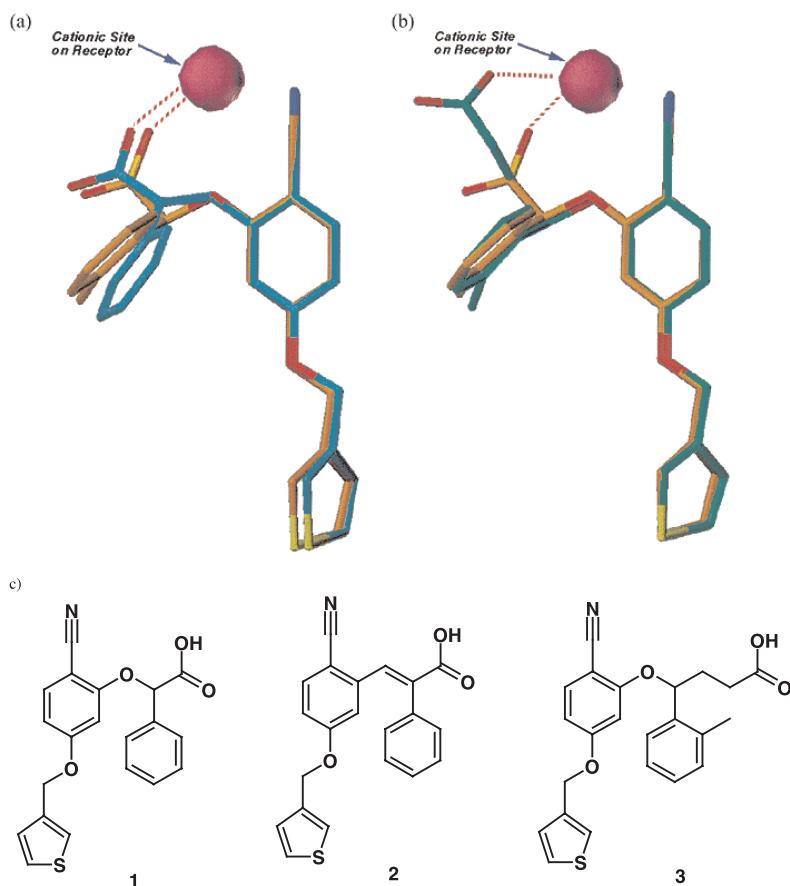


Fig. 4.11 Overlaps of the three potent endothelin antagonists shown in (c). (a) Overlap of 1 (beige) and 2 (cyan), showing the putative cationic site. (b) Overlap of 1 (beige)

and 3 (green), showing the putative cationic site. (a) and (b) (Reproduced from [117] with permission. Copyright 2000 American Chemical Society.)

synthesized in a combinatorial fashion and where the pharmacophores presented by the sub-library are representative of the full virtual library [128], cover the pharmacophore space, or complement that of another population. Library design is an optimization process where reagents are selected to optimize the required properties of the combinatorial products. Optimization methods such as genetic algorithms (GA) and simulated annealing have been applied in library design (see [125] for a review), and the same methods can be used with pharmacophore descriptors with minor modifications. Thus, a GA has been used to optimize the pharmacophore coverage of a library [129]. The HARPick procedure [130] utilizes simulated annealing to optimize a scoring function that includes terms for pharmacophore coverage, optionally matching to or complementing a pre-calculated

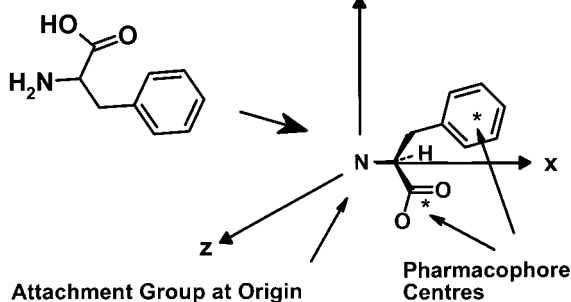
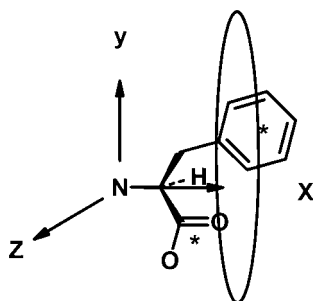
ALIGN**ROTATE & RECORD**

Fig. 4.12 An illustration of the GaP procedure for generating reagent-based pharmacophore fingerprints.

profile, in combination with other terms to optimize the property profile of the library ($\log P$, molecular weight, and so on). Mason and Beno [131] have developed a simulated annealing protocol that optimizes pharmacophore coverage in combination with another partitioning metric derived from BCUT descriptors [132]. The BCUTs contain information regarding the hydrogen-bonding capacity, polarizability, and size of molecules. McGregor and Muskal have developed the PharmPrint methodology for library design [133]. They demonstrate the method by optimizing a library selection with respect to a set of compounds from the MDDR (MDL Drug Data Report) [134].

Multi-pharmacophore descriptors can be slow to calculate for large sets of compounds. When specifically applied to library design, it is possible to calculate descriptors for reagents where the attachment bond to the scaffold can give a frame of reference. Several variants on this approach have been developed. In the OSPREY (Orientated Substituent Pharmacophore P_{RO}p_{ER}tY space) approach [135], two additional points are added to a substituent to represent the relationship with the scaffold. One-, two-, and three-point pharmacophore descriptors are then calculated for the substituents, including the distances to the two orienting points. The inclusion of the orienting points means that the descriptors are equivalent to

the use of a nine-point descriptor for the products of a three-component library. Because the number of substituents is considerably smaller than the number of combinatorial products, it is possible to perform a more detailed conformational sampling. In addition, a substituent similarity matrix can be calculated, allowing the use of more sophisticated experimental design methods for reagent selection. The GaP methodology was developed at GlaxoWellcome (now GlaxoSmithKline) primarily for the purposes of reagent acquisition [136]. The process of descriptor generation is outlined in Fig. 4.12, using phenylalanine as an example. The reagents are aligned with the bond to the reactive group aligned along the x-axis, and a conformational analysis is performed. For each conformation sampled, the position of each pharmacophore point is recorded on a grid, thus defining the orientation with respect to the attachment point. The process also can be applied in the context of the protein active site when searching for reagents. The attachment point is aligned with a ligand crystal structure or docked structure, and the active site represents a steric constraint to the conformational analysis. In both cases, the aim of reagent selection is to cover the available pharmacophores. A successful application of this to structure-based reagent selection has been reported [137].

4.4.3

Structure-based Design

Several programs have been developed to incorporate protein structure-derived pharmacophores into the design process. *De novo* design programs perhaps never lived up to their initial promise because of the difficulty of incorporating synthetic accessibility into the suggested compounds, even if only very simple reactions are used. The advent of combinatorial chemistry has given these methods a new lease of life, as the search space can be restricted to a small range of chemistries, though still with potentially many millions of products. The SPROUT suite of programs [138, 139] provides algorithms for identifying potential pharmacophore points, docking of small fragments to these points, and subsequent connection of the fragments. The space is searched using a tree-based algorithm and as such is exhaustive within the heuristics defined. A successful application of this methodology has been reported recently [140], where SPROUT was used to suggest frameworks as starting points in the design of Factor Xa inhibitors. Further extensions to the basic SPROUT functionality are under development to address the issue of synthetic tractability [141]. SYNSPROUT generates molecular structures that could be readily synthesized only by simple chemistry from a pool of readily available starting materials. VLSPROUT generates structures that can be synthesized by a particular combinatorial chemistry scheme. The *de novo* design program Skeldiv was developed at the University of Cambridge [142]. The program grows molecules within the constraints of a protein-binding site, using simulated annealing to optimize a number of user-defined parameters which include the frequency of occurrence of certain fragments, the fit to (a subset of) predefined pharmacophore points, the number of rotatable bonds, and other parameters to control the quality of structures. The program can grow molecules from a set of de

fault fragments, including rings and linker groups. Alternatively, several user-defined fragment sets can be used that represent potential reagent sets for a combinatorial library, thus allowing design of a library. The program has been applied successfully to the design of sub-micromolar Factor Xa inhibitors [143]. The LUDI program has been successfully applied to structure-based library design [144, 145].

A recent successful application of these methods was the work at Roche on DNA gyrase [146]. The LUDI maps for this site are shown in Fig. 4.6. Pharmacophores derived from these maps were used to search the corporate database using LUDI and Catalyst. Structures were deliberately chosen to be of low molecular weight and were screened at high concentration, so-called “needle screening.” The subsequent actives were then further developed via structure-based design techniques and medicinal chemistry to give several different series of active compounds.

The combination of ligand-derived and protein-structure-derived multi-pharmacophore descriptors provides a powerful technique for structure-based design, enabling the pharmacophore fingerprint of the protein active site to be compared against potential ligands. In a study of three related serine proteases, trypsin, thrombin, and Factor Xa, GRID was used to generate protein site points. A multi-pharma-

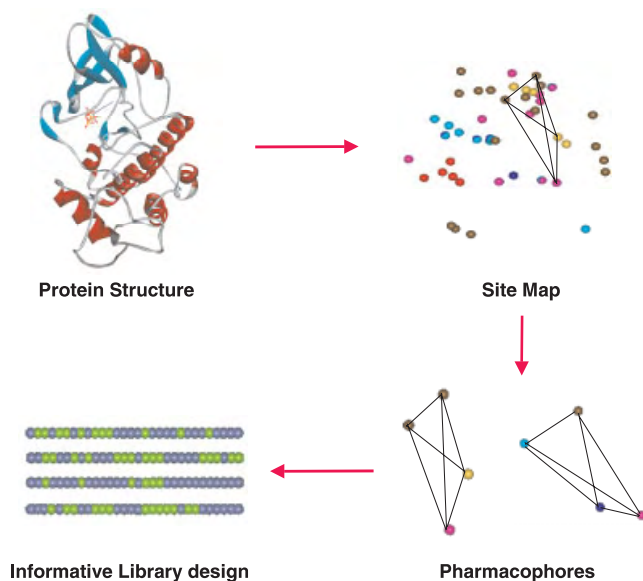


Fig. 4.13 Illustration of the protein site-directed library design protocol of Eksterowicz and coworkers [149]. Starting with the protein structure, a site map is generated from the active site. Pharmacophores are enumerated and used to define the space for library design. Compounds are then selected with the

informative design tool such that the resulting subset will interrogate the target in different but overlapping ways. The bit-strings for four sample compounds are illustrated. A green dot indicates a bit is turned on (pharmacophore is present in the molecule). (Reprinted with permission from [149]).

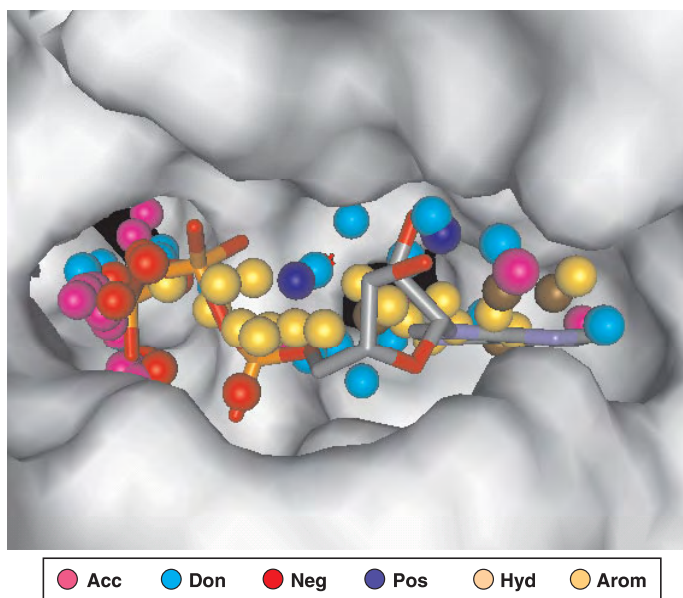


Fig. 4.14 Site map calculated for the ATP-bound structure of CDK2. The ligand is removed in the site map generation but is shown here to allow for comparison of the

site map feature positions with the ligand orientation. (Reprinted with permission from [149]).

cophore fingerprint of each active site was then compared in turn to ligand-derived descriptors for a series of selective ligands. With the use of a four-point pharmacophoric descriptor, the degree of overlap between the protein and ligand descriptors reflected the experimental selectivity [112, 147]. This work was extended with the design of a library against Factor Xa using the DiR approach [148, 131].

Eksterowicz and coworkers have recently reported on a similar approach that combines structure-based design and combinatorial chemistry through the use of site-derived pharmacophores [149]. Protein active site pharmacophores complementary to the protein residues are automatically generated and used to derive a protein pharmacophore fingerprint as shown in Fig. 4.13. Application of the site-map generation process to the ATP-bound structure of CDK2 is shown in Fig. 4.14. The ATP is removed during site-map generation but is shown here to illustrate the correspondence of the site points with pharmacophore groups on the ligand. Multiple ligand-bound protein structures can be used in the generation of a union fingerprint, thus allowing for receptor flexibility. Molecules are selected using the concept of informative design [150], i.e., the resulting subset samples the site pharmacophores in different but overlapping ways. The first iteration selects from a pool of compounds with known activity. This allows certain site pharmacophores that are hit by inactive but not active compounds to be removed from future designs.

4.5

The Biophore Concept in ADME Prediction

The application of pharmacophore-based methods now extends throughout the drug-discovery process. Traditionally, ADME (absorption, distribution, metabolism, excretion) considerations were applied at quite a late stage, but more recently there has been a trend to consider ADME properties at as early a stage as possible in an effort to reduce the attrition rate of compounds in development [151–154]. There is an increasing understanding of the molecular processes underlying ADME effects, and pharmacophore-based methods are being used widely to help understand the ligand properties with the aim of designing molecules with reduced ADME liabilities.

The most widely studied systems are the metabolizing enzymes of the cytochrome P450 superfamily [155], and much is known about the nature of their active sites and ligand requirements [156, 157]. Because this family of heme-containing enzymes shares a common oxidative mechanism [158], molecular orbital (MO) calculations have been used to identify likely sites of metabolism [159]. However, such calculations ignore the steric requirements of the protein active site. The enzymes metabolize a wide-range of xenobiotics, and as such the active sites must be fairly non-selective in the types of ligands they bind. Indeed, it has been shown that CYP3A4, the most ubiquitous of the human P450 isoenzymes, has at least two and possibly three distinct binding sites [160]. Nevertheless, pharmacophore-modeling methods are providing much insight into the structural requirements of these proteins. The first such report [161] related to CYP2D6, for which a requirement for a basic nitrogen and a flat hydrophobic (or aromatic) region up to 7.5 Å distant could be defined. The most potent inhibitors also possessed a hydrogen-bond acceptor feature about 7 Å from the nitrogen. Ekins and coworkers have employed Catalyst to study a number of P450 enzymes, including inhibitors [162] and substrates [163] of CYP3A4, inhibitors of CYP2D6 [164], substrates of CYP2B6 [165], and inhibitors of CYP2C9 [166]. However, as stated in the introduction, a pharmacophore is necessary but not sufficient for activity, and many inactive compounds will also be detected when performing a virtual screen. Similarly, absence of a pharmacophore is not an indicator of lack of activity. Thus, the best use of such methods in practice will be in suggesting how known substrates are metabolized rather than as screens for library design or similar. This is the approach taken by de Groot and coworkers, who combined pharmacophore modeling with homology modeling and MO calculations to generate a predictive system for determining the site of metabolism of CYP2D6 substrates [167, 168].

Until recently, protein structural information on human P450s could be obtained only by homology to prokaryotic enzymes for which several crystal structures exist. However, the structure of a mammalian P450 has now been solved: rabbit 2C5 [169]. de Groot and coworkers extended their combined pharmacophore- and protein-modeling approach to CYP2C9, utilizing the rabbit structure in their homology modeling to produce a more reliable structure [170]. Afzelius and coworkers have combined homology modeling, docking, and chemometric

methods in the derivation of a 3-D QSAR model for CYP2C9 inhibitors [171]. The docking program GOLD [81] was used to generate multiple conformations for compounds in the homology model based on the rabbit 2C5 structure. Information on compounds with known interaction modes and principal components analysis on the GRID [40] fields of the docked conformers was then used to select one conformer for each inhibitor to be used in a PLS analysis. The final model had a cross-validated q^2 of 0.73 on the training set and a predictive error of less than 0.5 log units on all eight compounds in the external test set. Interpretation of the model was in agreement with site-directed mutagenesis studies.

P-glycoprotein (P-gp) is a membrane transport protein that can limit oral bio-availability through drug efflux. P-gp substrates are also prevented from entering the brain. In the absence of protein structure information, pharmacophoric methods have been employed to gain a better understanding of the requirements for the substrates. Penzotti and coworkers have developed a computational ensemble pharmacophore model for P-gp substrates [172]. This methodology is similar to that for the active site modeling described in Section 4.4.3. All two-, three-, and four-point pharmacophores are generated for the substrates, and an information content measure is used to select the most informative pharmacophores. In this study the preferred ensemble contained 100 pharmacophores, with a substrate needing to match at least 20 of these. Application to an external test set gave a correct prediction for 53% of the substrates and 79% of the non-substrates. The ensemble model contained examples of the type I and type II patterns previously identified by Seelig as necessary for P-gp substrates [173]. The type I pattern contains two hydrogen-bond acceptors separated by about 2.5 Å, and the type II pattern contains two hydrogen-bond acceptors separated by 4–5 Å. Alternatively, the type II pattern consists of three hydrogen-bond acceptors separated by about 2.5 and 4–5 Å. Penzotti and coworkers propose using the ensemble pharmacophore model as a filter for enriching virtual libraries with non P-gp substrates.

Another system of key importance is that of the pregnane X receptor (PXR), identified recently [174], which is involved in the induction of P450s and P-gp, among others. As such, PXR ligands can be a major cause of severe drug-drug interactions. A pharmacophore model, derived using Catalyst, has been proposed for PXR based upon ligand-binding data [175]. The recently solved crystal structure of the ligand-binding domain of PXR [176] should help further validate and improve such models, as in the P450 case described above.

4.6

Summary

The biophore concept provides a framework for the understanding of ligand-protein interactions in terms of the key interacting groups on the ligand and protein. This abstraction enables the generation of models that are intuitive and at the same time broadly applicable. The applications discussed herein show the usefulness of this approach. The biophore concept, which first had an impact mainly in

lead optimization, rapidly became an important component of lead generation strategies with the availability of the first 3-D database-searching methods. The biophore concept has now become an integral part of modeling across the breadth of the drug-discovery process from lead discovery to understanding the pharmacokinetic properties of drugs.

4.7

References

- 1 H.-J. BÖHM and G. KLEBE, *Angew. Chem. Int. Ed. Engl.* **1996**, 35, 2588–2614.
- 2 C. G. WERMUTH, C. R. GANELLIN, P. LINDBERG and L. A. MITSCHER, *Pure Appl. Chem.*, **1988**, 70, 1129–1143.
- 3 W. C. LUMMA, K. M. WITHERUP, T. J. TUCKER, S. F. BRADY, J. T. SISCO, A. M. NAYLOR-OLSEN, S. D. LEWIS, R. M. FREIDINGER, *J. Med. Chem.* **1998**, 41, 1011–1013.
- 4 T. J. TUCKER, S. F. BRADY, W. C. LUMMA, S. D. LEWIS, S. J. GARDELL et al., *J. Med. Chem.* **1998**, 41, 3210–3219.
- 5 P. EHRLICH, *Chem. Ber.* **1909**, 42, 17; E. J. Artens, *Progr. Drug Res.* **1966**, 10, 429.
- 6 P. GUND, Evolution of the Pharmacophore Concept in Pharmaceutical Research. In *Pharmacophore Perception, Development and Use in Drug Design* O. GUNER Ed. International University Line, La Jolla CA 92038-2525 USA, **2000**, 5–11.
- 7 Y. C. MARTIN, *J. Med. Chem.* **1992**, 35, 2145–2154.
- 8 A. C. GOOD and J. S. MASON, *Reviews in Computational Chemistry* **1996**, 7, 67–117.
- 9 G. W. A. MILNE, M. C. NICKLAUS and S. WANG, *SAR and QSAR in Environmental Research*, **1998**, 9, 23–38.
- 10 A. C. GOOD, J. S. MASON, S. D. PICKETT, *Methods Princ. Med. Chem.* **10** (Virtual Screening for Bioactive Molecules), **2000**, 131–159.
- 11 Catalyst. Distributed by Accelrys Inc. 9685 Scranton Road, San Diego, CA 92121, USA.
- 12 J. GREENE, S. KAHN, H. SAVOJ, P. SPRAGUE, S. TEIG, *J. Chem. Inf. Comput. Sci.* **1994**, 34, 1297–1308.
- 13 UNITY® 4.2.1 Tripos, Inc., 1699 South Hanley Rd., St. Louis, Missouri, 63144, USA.
- 14 S. ASH, M. A. CLINE, R. W. HOMER, T. HURST, and G. B. SMITH, *J. Chem. Inf. Comput. Sci.*, **1997**, 37, 71–79.
- 15 Y. C. MARTIN, M. G. BURES, E. A. DANAHER, J. DELAZZER, I. LICO and P. A. PAVLIK, *J. Comput.-Aided Mol. Des.* **1993**, 7, 83–102.
- 16 G. JONES, P. WILLETT and R. C. GLEN, *J. Comput.-Aided Mol. Des.* **1995**, 9, 532–549.
- 17 D. MAYER, C. B. NAYLOR, I. MOTOC, G. R. MARSHALL, *J. Comput.-Aided Mol. Des.* **1987**, 1, 3–16.
- 18 S. D. PICKETT, I. M. MCLAY and J. S. MASON, *J. Chem. Inf. Comput. Sci.* **1996**, 36, 1214–1233.
- 19 J. S. MASON, I. MORIZE, P. R. MENARD, D. L. CHENEY, C. HULME and R. F. LABAUDINIERE, *J. Med. Chem.* **1999**, 42, 3251–3264.
- 20 G. R. MARSHALL, C. D. BARRY, H. E. BOSSHARD, R. A. DAMMKOEHLER and D. A. DUNN, in *Computer-Assisted Drug Design*, Vol. 112, E. C. OLSON and R. E. CHRISTOFFERSON eds. American Chemical Society, Washington D. C. **1979**, 205–226.
- 21 R. D. NELSON, D. I. GOTTLIEB, T. M. BALASUBRAMANIAN and G. R. MARSHALL, in *Opioid Peptides: Medicinal Chemistry* Vol. 69 R. S. RAPAKA, G. BARNETT and R. L. HAWKS eds. Rockville, NIDA Office of Science, **1986**, 204–230.
- 22 Y. TAKEUCHI, E. F. B. SHANDS, D. D. BEUSEN and G. R. MARSHALL, *J. Med. Chem.* **1998**, 19, 3609–23.
- 23 D. D. BEUSEN and G. R. MARSHALL, in *Pharmacophore Perception, Development*

- and Use in Drug Design, O. GUNER Ed., International University Line, La Jolla CA 92038-2525 USA. **2000**, 21–45.
- 24 D. BARNUM, J. GREENE, A. SMELLIE and P. SPRAGUE, *J. Chem. Inf. Comput. Sci.*, **1996**, 36, 563–571.
 - 25 A. SMELLIE, S. L. TEIG and P. TOWBIN, *J. Comput. Chem.* **1995**, 16, 171–187.
 - 26 A. SMELLIE, S. D. KAHN and S. L. TEIG, *J. Chem. Inf. Comput. Sci.* **1995**, 35, 285–294.
 - 27 A. SMELLIE, S. D. KAHN and S. L. TEIG, *J. Chem. Inf. Comput. Sci.* **1995**, 35, 295–304.
 - 28 A. T. BRINT and P. J. WILLETT, *J. Chem. Inf. Comput. Sci.* **1987**, 27, 152–158.
 - 29 Y. C. MARTIN, in *Pharmacophore Perception, Development and Use in Drug Design* O. GUNER Ed. International University Line, La Jolla CA 92038-2525 USA. **2000**, 49–68.
 - 30 G. JONES and P. WILLETT, In *Pharmacophore Perception, Development and Use in Drug Design* O. GUNER Ed. International University Line, La Jolla CA 92038-2525 USA. **2000**, 85–106
 - 31 K. S. HAKAKI, R. S. SHERIDAN, R. VENKATARAGHAVAN, D. A. DUNN and R. MCCULLOCH, *Tetrahedron Comp. Meth.* **1990**, 3, 565–573.
 - 32 S. D. PICKETT, J. KING-UNDERWOOD, I. WALL, B. S. SHERBORNE unpublished results. Presented at the UK QSAR & Cheminformatics Group Meeting, Hordsham, October 2001. http://www.iaimn.demon.co.uk/ukqsar_autumn2001.html
 - 33 O. F. GUNER and D. R. HENRY, in *Pharmacophore Perception, Development and Use in Drug Design* O. GUNER Ed. International University Line, La Jolla CA 92038-2525 USA. **2000**, 193–212.
 - 34 O. F. GUNER, R. HOFFMANN and H. LI, In Report by Wendy A. Warr on 217th ACS National Meeting and Exposition, Anaheim, California, March 12–25, **1999**. London: Wendy Warr & Associates, **1999**, 50–53.
 - 35 P. W. SPRAGUE, *Perspect. Drug Discovery Des.* **1995**, 3, 1–20.
 - 36 J. J. KAMINSKI, D. F. RANE, M. E. SNOW, L. WEBER, M. L. ROTHOFKY, S. D. ANDERSON and S. L. LIN, *J. Med. Chem.* **1997**, 40, 4103–4112.
 - 37 X. CHEN, A. RUSINKO III, A. TROPASHA and S. S. YOUNG, *J. Chem. Inf. Comput. Sci.* **1999**, 39, 887–896.
 - 38 D. M. HAWKINS, S. S. YOUNG and A. RUSINKO, *Quant. Struct.-Act. Relat.* **1998**, 16, 1–7.
 - 39 C. E. KEEFER, E. C. BIGHAM and S. S. YOUNG. Pharmacophore identification using massive distributed computing. Abstracts of Papers, 222nd ACS National Meeting, Chicago, IL, United States, August 26–30, 2001.
 - 40 P. J. GOODFORD, *J. Med. Chem.* **1995**, 28, 849–857.
 - 41 GRID 20 is available from Molecular Discovery Ltd. <http://www.moldiscovery.com/index.html>
 - 42 R. D. CRAMER III, D. E. PATTERSON and J. D. BUNCE, *J. Am. Chem. Soc.* **1988**, 110, 5959–67.
 - 43 Y. C. MARTIN, K.-H. KIM and C. T. LIN. *Adv. Quant. Struct.-Prop. Relat.* **1996**, 1, 1–52.
 - 44 K.-H. KIM, G. GRECO and E. NOVELLINO, *Perspect. Drug Discovery Des.* 12/13/14 (3D QSAR in Drug Design: Recent Advances), **1998**, 257–315.
 - 45 U. NORINDER, *Perspect. Drug Discovery Des.* 12/13/14 (3D QSAR in Drug Design: Recent Advances), **1998**, 25–39.
 - 46 R. D. CLARK, J. M. LEONARD and A. STRIZHEV, in *Pharmacophore Perception, Development and Use in Drug Design*, O. GUNER Ed. International University Line, La Jolla CA 92038-2525 USA. **2000**, 151–169.
 - 47 T. T. TALELE, S. S. KULKARNI and V. M. KULKARNI, *J. Chem. Inf. Comput. Sci.* **1999**, 39, 958–966.
 - 48 M. BARONI, G. COSTANTINO, G. CRUCIANI, D. RIGANELLI, R. VALIGI, and S. CLEMENTI, *Quant. Struct.-Act. Relat.* **1993**, 12, 9–20.
 - 49 G. CRUCIANI and S. CLEMENTI, *Methods Princ. Med. Chem.* 3 (Advanced Computer-Assisted Techniques in Drug Discovery), **1995**, 61–88.
 - 50 Sybyl, developed and distributed by Tripos Inc.
 - 51 M. PASTOR, G. CRUCIANI and S. CLEMENTI, *J. Med. Chem.* **1997**, 40, 1455–1464.
 - 52 M. PASTOR, G. CRUCIANI, I. MCLAY, S. PICKETT and S. CLEMENTI, *J. Med. Chem.* **2000**, 43, 3233–43.

- 53 ALMOND is available from Multivariate Infometric Analysis S.r.l.
<http://www.miasrl.com>.
- 54 P. BENEDETTI, R. MANNHOLD, G. CRUCIANI and M. PASTOR, *J. Med. Chem.* **2002**, *45*, 1577–84.
- 55 A. J. HOPFINGER, S. WANG, J. S. TOKARSKI, B. JIN, M. ALBUQUERQUE, P. J. MADHAV and C. DURAISWAMI, *J. Am. Chem. Soc.* **1997**, *119*, 10509–10524.
- 56 A. C. GOOD and W. G. RICHARDS, *Perspect. Drug Discovery Des.* 9/10/11 (3D QSAR in Drug Design: Ligand/Protein Interactions and Molecular Similarity), **1998**, 321–338.
- 57 ASP is a part of the TSAR3D package distributed by Accelrys Inc.
<http://www.accelrys.com>.
- 58 A. C. GOOD, E. E. HODGKIN and W. G. RICHARDS, *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 188–191.
- 59 T. D. J. PERKINS and P. M. DEAN, *J. Comput.-Aided Mol. Des.* **1993**, *7*, 155–172.
- 60 J. E. J. MILLS, I. J. P. DE ESCH, T. D. J. PERKINS and P. M. DEAN, *J. Comput.-Aided Mol. Des.* **15**, 81–96 (2001).
- 61 J. E. J. MILLS and P. M. DEAN, *J. Comput.-Aided Mol. Des.* **1996**, *10*, 607–622.
- 62 F. H. ALLEN and O. KENNARD, *Chemical Design Automation News*, **1993**, *8*(1), 1 & 31–37.
- 63 M. D. MILLER, R. P. SHERIDAN and S. K. KEARSLEY, *J. Med. Chem.* **1999**, *42*, 1505–14.
- 64 M. D. MILLER, E. M. FLUDER, L. A. CASTONGUAY, J. C. CULBERSON, R. T. MOSLEY, K. PRENDERGAST, S. K. KEARSLEY and R. P. SHERIDAN, *Med. Chem. Res.* **1999**, *9*, 513–534.
- 65 M. VON ITZSTEIN, W. Y. WU, G. B. KOK, M. S. PEGG, J. C. DYASON, B. JIN, T. V. PHAN, M. L. SMYTHE, H. F. WHITE et al., *Nature* **1993**, *363*, 418–423.
- 66 H. J. BÖHM, *J. Comput.-Aided Mol. Des.* **1992**, *6*, 61–78.
- 67 H. J. BÖHM, *J. Comput.-Aided Mol. Des.* **1992**, *6*, 593–606.
- 68 H. J. BÖHM, *J. Comput.-Aided Mol. Des.* **1998**, *12*, 1–15.
- 69 D. B. WIGLEY, G. J. DAVIES, E. J. DOBSON, A. MAXWELL and G. DODSON, *Nature*, **1991**, *351*, 624–629.
- 70 L. BRINO, A. URZHUMSTEV, M. MOUSLI, C. BRONNER, A. MITSCHLER, P. OUDET and D. MORAS, *J. Biol. Chem.* **2000**, *275*, 9468–9475.
- 71 M. L. VERDONK, J. C. COLE and R. TAYLOR, *J. Mol. Biol.* **1999**, *289*, 1093–1108.
- 72 M. L. VERDONK, J. C. COLE, P. WATSON, V. GILLET, P. WILLETT, *J. Mol. Biol.* **2001**, *307*, 841–859.
- 73 I. J. BRUNO, J. C. COLE, J. P. M. LOMMERSE, R. S. ROWLAND, R. TAYLOR and M. L. VERDONK, *J. Comput.-Aided Mol. Des.* **1997**, *11*, 525–537.
- 74 D. E. CLARK, D. R. WESTHEAD, R. A. SYKES, C. W. MURRAY, *J. Comput.-Aided Mol. Des.* **1996**, *10*, 397–416.
- 75 P. A. GREENIDGE, B. CARLSSON, L.-G. BLADH and M. GILLNER, *J. Med. Chem.*, **1998**, *41*, 2503–2512.
- 76 Chem-X was formally developed at Chemical Design (now part of Accelrys Inc.)
- 77 C. M. MURRAY and S. J. CATO, *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 46–50.
- 78 E. K. DAVIES and C. J. DAVIES, *J. Chem. Inf. Comput. Sci.*, submitted.
- 79 M. D. ELDRIDGE, C. W. MURRAY, T. R. AUTON, G. V. PAOLINI and R. P. MEE, *J. Comput.-Aided Mol. Des.*, **1997**, *11*, 425–445.
- 80 C. W. MURRAY, T. R. AUTON, and M. D. ELDRIDGE, *J. Comput.-Aided Mol. Des.* **1998**, *12*, 503–519.
- 81 G. JONES, P. WILLETT, R. C. GLEN, A. R. LEACH and R. TAYLOR, *J. Mol. Biol.* **1997**, *267*, 727–748.
- 82 M. RAREY, B. KRAMER, T. LENGAUER and G. KLEBE, *J. Mol. Biol.* **1996**, *261*, 470–489.
- 83 I. D. KUNTZ, J. M. BLANEY, S. J. OATLEY, R. L. LANGRIDGE and T. E. FERRIN, *J. Mol. Biol.* **1982**, *161*, 269–288.
- 84 B. K. SHOICHET and I. D. KUNTZ, *Protein Eng.* **1993**, *6*, 723–732.
- 85 A. DALBY, J. G. NOURSE, W. D. HOUNSHELL, D. L. GUSHURST, B. A. LELAND and J. LAUFER, *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 244–255.
- 86 D. WEININGER, *J. Chem. Inf. Comput. Sci.* **1988**, *28*, 31–36.
- 87 D. WEININGER, A. WEININGER and J. L. WEININGER, *J. Chem. Inf. Comput. Sci.* **1989**, *29*, 97–101.

- 88 J. GASTEIGER, C. RUDOLPH and J. SADOWSKI, *Tetrahedron Comp. Method.* **1990**, 3, 537–547.
- 89 R. S. PEARLMAN *CDA News*, **2**, 1–7 (1987). A. RUSINKO III, J. M. SKELL, R. BALDUCCI, C. M. MCGARITY and R. S. PEARLMAN. CONCORD, A program for the rapid generation of high quality approximate 3D molecular structures. Distributed by Tripos Inc., 1699 Hanley Road, Suite 303, St. Louis, MO 63144 U.S.A.
- 90 S. E. JAKES and P. WILLETT, *J. Mol. Graphics*, **1986**, 4, 12–20.
- 91 T. E. MOOCK, D. R. HENRY, A. G. OZKABAK and M. ALAMGIR, *J. Chem. Inf. Comput. Sci.* **1994**, 34, 184–189.
- 92 ISIS-3D developed and distributed by MDL Information Systems Inc., San Leandro, CA, USA.
- 93 R. P. SHERIDAN, R. NILAKANTAN, A. RUSINKO III, N. BAUMAN, K. S. HARAKI, and R. VENKATARAGHAVAN, *J. Chem. Inf. Comput. Sci.*, **1989**, 29, 255–260.
- 94 J. H. VAN DRIE, D. WEININGER, Y. C. MARTIN, *J. Comput.-Aided Mol. Des.* **1989**, 3, 225–251.
- 95 D. E. CLARK, P. WILLETT and P. W. KENNY, *J. Mol. Graphics* **1992**, 10, 194–204.
- 96 D. E. CLARK, P. WILLETT, and P. W. KENNY, *J. Mol. Graphics*, **1993**, 11, 146–156.
- 97 D. E. CLARK, G. JONES, P. WILLETT, P. W. KENNY, and R. C. GLEN, *J. Chem. Inf. Comput. Sci.* **1994**, 34, 197–206.
- 98 T. HURST, *J. Chem. Inf. Comput. Sci.* **1994**, 34, 190–196.
- 99 N. W. MURRALL and E. K. DAVIES, *J. Chem. Inf. Comput. Sci.* **1990**, 30, 312–316.
- 100 M. HAHN, *J. Chem. Inf. Comput. Sci.*, **1997**, 37, 80–86.
- 101 M. HAHN and D. ROGERS, *J. Med. Chem.* **1995**, 38, 2080–2090.
- 102 O. F. GUNER, M. WALDMAN, R. HOFFMANN and J.-H. KIM, in *Pharmacophore Perception, Development and Use in Drug Design* O. GUNER Ed. International University Line, La Jolla CA 92038-2525 USA. **2000**, 213–236.
- 103 T. HURST. Partial-match 3D searching. Book of Abstracts, 214th ACS National Meeting, Las Vegas, NV, 7–11 September 1997 COMP-037.
- 104 Daylight Chemical Information Systems Inc., 27401 Los Altos, Suite 370, Mission Viejo, Santa Fe, NM, USA. <http://www.daylight.com>.
- 105 SSKEYS Gateway, MDL Information Systems Inc., 14600 Catalina St., San Leandro, CA 94577 <http://www.mdli.com/>.
- 106 S. A. MOUSA and D. CHERESH, *Drug Discovery Today* **1997**, 2, 187–199.
- 107 S. D. PICKETT, I. M. MCLAY and D. E. CLARK, *J. Chem. Inf. Comput. Sci.* **2000**, 40, 263–272.
- 108 S. K. KEARSLEY, S. SALLAMACK, E. M. FLUDER, J. D. ANDOSE, R. T. MOSLEY, R. P. SHERIDAN, *J. Chem. Inf. Comput. Sci.* **1996**, 36, 118–127.
- 109 G. SCHNEIDER, W. NEIDHART, T. GILLER and G. SCHMID, *Angew. Chem. Int. Ed.* **1999**, 38, 2894–96.
- 110 R. D. BROWN and Y. C. MARTIN, *J. Chem. Inf. Comput. Sci.* **1996**, 36, 572–584.
- 111 E. K. DAVIES, in *Molecular Diversity and Combinatorial Chemistry: Libraries and Drug Discovery*, I. M. CHAIKEN and K. D. JANDA Eds. American Chemical Society, Washington D.C., **1996**, 309–316.
- 112 J. S. MASON, I. MORIZE, P. R. MENARD, D. L. CHENEY, C. HULME and R. F. LABAUDINIERE, *J. Med. Chem.* **1999**, 42, 3251–3264.
- 113 P. WILLETT, J. M. BARNARD and G. M. DOWNS, *J. Chem. Inf. Comput. Sci.* **1998**, 38, 983–996.
- 114 A. POZZAN, A. FERIANI, G. TEDESCO and A. M. CAPELLI, *Proceedings of QSAR 2000*, Prous Science S.A. **2000**.
- 115 P. C. ASTLES, T. J. BROWN, C. M. HANDSCOMBE, M. F. HARPER, N. V. HARRIS, R. A. LEWIS, P. M. LOCKEY, C. MCCARTHY, I. M. MCLAY, B. PORTER, A. G. ROACH, C. SMITH and R. J. A. WALSH, *Eur. J. Med. Chem.* **1997**, 32, 409–423.
- 116 P. C. ASTLES, C. BREALEY, T. J. BROWN, V. FACCHINI, C. HANDSCOMBE, N. V. HARRIS, C. MCCARTHY, I. M. MCLAY, B. PORTER, A. G. ROACH, C. SARGENT, C. SMITH and R. J. A. WALSH, *J. Med. Chem.* **1998**, 41, 2732–2744.
- 117 P. C. ASTLES, T. J. BROWN, F. HALLEY, C. M. HANDSCOMBE, N. V. HARRIS, T. N. MAJID, C. MCCARTHY, I. M. MCLAY, A. MORLEY, B. PORTER, A. G. ROACH, C. SAR-

- GENT, C. SMITH, and R. J. A. WALSH, *J. Med. Chem.* **2000**, *43*, 900–910.
- 118 D. P. MARRIOTT, I. G. DOUGALL, P. MEHGANI, Y.-J. LIU and D. R. FLOWER, *J. Med. Chem.* **1999**, *42*, 3210–3216.
 - 119 H. HONG, N. NEAMATI, S. WANG, M. C. NICKLAUS, A. MAZUMDER, H. ZHAO, T. R. BURKE JR., Y. POMMIER and G. W. A. MILNE, *J. Med. Chem.* **1997**, *40*, 930–936.
 - 120 S. WANG, G. W. A. MILNE, X. YAN, I. J. POSEY, M. C. NICKLAUS, L. GRAHAM and W. G. RICE, *J. Med. Chem.* **1996**, *39*, 2047–2054.
 - 121 S. WANG, D. W. ZAHAREVITZ, R. SHARMA, V. E. MARQUEZ, N. E. LEWIN, L. DU, P. M. BLUMBERG and G. W. A. MILNE, *J. Med. Chem.* **1994**, *37*, 4479–4489.
 - 122 D. W. ZAHAREVITZ, R. GUSSIO, A. WIEGAND, R. JALLURI, N. PATTABIRAMAN, G. E. KELLOGG, L. A. PALLANSCH, S. S. YANG and R. W. BUCKHEIT JR, *Med. Chem. Res.* **1999**, *9*, 551–564.
 - 123 D. K. AGRAFIOTIS, in *The Encyclopedia of Computational Chemistry*, P. v. R. SCHLEYER, N. L. ALLINGER, T. CLARK, J. GASTEIGER, P. A. KOLLMAN, H. F. SCHAEFER III, P. R. SCHREINER eds., John Wiley & Sons: Chichester, Vol. 1, **1998**, 742–761.
 - 124 R. A. LEWIS and P. M. DEAN eds. *Molecular Diversity in Drug Design*, Kluwer Academic Publishers, Dordrecht, Netherlands, **1999**.
 - 125 R. A. LEWIS, S. D. PICKETT and D. E. CLARK, in *Reviews in Computational Chemistry*, K. B. LIPKOWITZ and D. B. BOYD eds., Vol. 16, Wiley-VCH, John Wiley & Sons Inc., New York, **2000**, 1–51.
 - 126 A. K. GHOSE, V. N. VISWANADHAN and J. J. WENDOLOSKI, in *Combinatorial Library Design and Evaluation: Principles, Software Tools, and Applications in Drug Discovery*, A. K. GHOSE, V. N. VISWANADHAN eds. Marcel Dekker, Inc., New York, **2001**, 51–71.
 - 127 J. S. MASON and S. D. PICKETT, *Perspect. Drug Discovery Des.* **1997**, *7/8*, 85–114.
 - 128 S. D. PICKETT, C. LUTTMANN, V. GUERIN, A. LAOUI and E. JAMES, *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 144–150.
 - 129 R. A. LEWIS, A. C. GOOD and S. D. PICKETT, in *Computer-Assisted Lead Finding and Optimization: Current Tools for Medicinal Chemistry*, H. VAN DE WATERBEEMD, B. TESTA and G. FOLKERS eds., Wiley-VCH, Weinheim, **1997**, 135–156.
 - 130 A. C. GOOD and R. A. LEWIS, *J. Med. Chem.* **1997**, *40*, 3926–3936.
 - 131 J. S. MASON and B. R. BENO, *J. Mol. Graph. Mod.* **2000**, *18*, 438–451.
 - 132 R. S. PEARLMAN and K. M. SMITH, *Perspect. Drug Discovery Des.* **1998**, *9*, 339–353.
 - 133 M. J. MCGREGOR and S. M. MUSKAL, *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 117–125.
 - 134 MDL Information Systems, Inc. 14600 Catalina St., San Leandro, CA 94577. <http://www.mdli.com>
 - 135 E. J. MARTIN and T. J. HOEFFEL, *J. Mol. Graphics Modell.* **2000**, *18*, 383–403.
 - 136 A. R. LEACH, D. V. S. GREEN, M. M. HANN, D. B. JUDD and A. C. GOOD, *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 1262–1269.
 - 137 D. B. JUDD, A. R. LEACH, M. M. HANN, D. V. S. GREEN and A. C. GOOD, Book of Abstracts, 219th ACS National Meeting, San Francisco, CA, March 26–30, **2000**.
 - 138 V. GILLET, A. P. JOHNSON, P. MATA, S. SKE and P. WILLIAMS, SPROUT: A program for structure generation. *J. Comput.-Aided Mol. Des.* **1993**, *7*, 127–153.
 - 139 V. J. GILLET, G. MYATT, Z. ZSOLDOS and A. P. JOHNSON, *Perspect. Drug Discovery Des.* **1995**, *3*, 34–50.
 - 140 Q. HAN, C. DOMINGUEZ, P. F. W. STOUTEN, J. M. PARK, D. E. DUFFY, R. A. GALLEMMO JR., K. A. ROSSI, R. S. ALEXANDER, A. M. SMALLWOOD, P. C. WONG, M. M. WRIGHT, J. M. LUETTGEN, R. M. KNABB and R. R. WEXLER, *J. Med. Chem.* **2000**, *43*, 4398–4415.
 - 141 A. P. JOHNSON, K. BODA, J.-F. MARCHALAND and A. TING, Book of Abstracts, 219th ACS National Meeting, San Francisco, CA, March 26–30, **2000**.
 - 142 N. P. TODOROV and P. M. DEAN, *J. Comput.-Aided Mol. Des.* **1997**, *11*, 175–192.
 - 143 D. L. CHENEY, J. S. MASON, M. R. BECKER, R. W. EWING, H. W. PAULS, N. TODOROV, 215th ACS National Meeting, Dallas, March 29. April 2 1998.
 - 144 H. J. BÖHM, in *Computer-Assisted Lead Finding and Optimization*, H. VAN DE WATERBEEMD, B. TESTA and G. FOLKERS eds. Wiley-VCH, **1997**, 125–133.

- 145 H.-J. BÖHM, D.W. BANNER and L. WEBER, *J. Comput.-Aided Mol. Des.* **1999**, *13*, 51–56.
- 146 H.-J. BÖHM, M. BOEHRINGER, D. BUR, H. GMUENDER, W. HUBER, W. KLAUS, D. KOSTREWA, H. KUEHNE, T. LUEBBERS, N. MEUNIER-KELLER, and F. MUELLER, *J. Med. Chem.* **2000**, *43*, 2664–2674.
- 147 J.S. MASON and D.L. CHENEY, *Proc. Pac. Symp. Biocomput.* **1999**, *4*, 456–467.
- 148 J.S. MASON and D.L. CHENEY, *Proc. Pac. Symp. Biocomput.* **2000**, *5*, 576–587.
- 149 J. E. EKSTEROWICZ, E. EVENSEN, C. LEMMEN, G. P. BRADY, J. K. LANCTOT, E. K. BRADLEY, E. SAIAH, L. A. ROBINSON, P. D. J. GROOTENHUIS and J. M. BIANEY, *J. Mol. Graphics Mod.* **2002**, *20*, 469–477.
- 150 S. L. TEIG, *J. Biomol. Screening* **1998**, *3*, 85–88.
- 151 M. H. TARBIT and J. BERMAN, *Curr. Opin. Chem. Biol.* **1998**, *2*, 411–416.
- 152 D. E. CLARK and S. D. PICKETT, *Drug Discovery Today* **2000**, *2*, 49–58.
- 153 P. J. EDDERSHAW, A. P. BERESFORD and M. K. BAYLISS, *Drug Discovery Today* **2000**, *5*, 409–414.
- 154 H. E. SELICK, A. P. BERESFORD and M. H. TARBIT, *Drug Discovery Today* **2002**, *7*, 109–116.
- 155 S. A. WRIGHTON and J. C. STEVENS, *Crit. Rev. Toxicol.* **1992**, *22*, 1–21.
- 156 D. A. SMITH, M. J. ACKLAND and B. C. JONES, *Drug Discovery Today*, **1997**, *2*, 406–414.
- 157 D. A. SMITH, M. J. ACKLAND and B. C. JONES, *Drug Discovery Today*, **1997**, *2*, 479–486.
- 158 F. P. GUENGERICH and T. L. MACDONALD, *FASEB J.* **1990**, *4*, 2453–2459.
- 159 K. R. KORZEKWA, J. P. JONES and J. R. GILLETTE, *J. Am. Chem. Soc.* **1990**, *112*, 7042–46.
- 160 N. A. HOSEA, G. P. MILLER and F. P. GUENGERICH, *Biochemistry*, **2000**, *39*, 5929–5939.
- 161 G. R. STROBL, S. VON KRUEDENER, J. STOCKIGT, F. P. GUNGERICH and T. WOLFF, *J. Med. Chem.* **1993**, *36*, 1136–1145.
- 162 S. EKINS, G. BRAVI, S. BINKLEY, J. S. GILLESPIE, B. J. RING, J. H. WIKEL and S. A. WRIGHTON, *J. Pharmacol. Exp. Ther.*, **1999**, *290*, 429–438.
- 163 S. EKINS, G. BRAVI, J. H. WIKEL and S. A. WRIGHTON, *J. Pharmacol. Exp. Ther.*, **1999**, *291*, 424–433.
- 164 S. EKINS, G. BRAVI, S. BINKLEY, J. S. GILLESPIE, B. J. RING, J. H. WIKEL and S. A. WRIGHTON, *Pharmacogenetics* **1999**, *9*, 477–489.
- 165 S. EKINS, G. BRAVI, B. J. RING, T. A. GILLESPIE, J. S. GILLESPIE, M. VANDENBRANDEN, S. A. WRIGHTON and J. H. WIKEL, *J. Pharmacol. Exp. Ther.* **1999**, *288*, 21–29.
- 166 S. EKINS, G. BRAVI, S. BINKLEY, J. GILLESPIE, B. RING, J. WIKEL and S. WRIGHTON, *Drug Metab. Disp.* **2000**, *28*, 994–1002.
- 167 M. J. DE GROOT, M. J. ACKLAND, V. A. HORNE, A. A. ALEX, B. C. JONES, *J. Med. Chem.* **1999**, *42*, 1515–1524.
- 168 M. J. DE GROOT, M. J. ACKLAND, V. A. HORNE, A. A. ALEX and B. C. JONES, *J. Med. Chem.* **1999**, *42*, 4062–4070.
- 169 P. A. WILLIAMS, J. COSME, V. SRIDHAR, E. F. JOHNSON, and D. E. McREE, *Mol. Cell.* **2000**, *5*, 121–131.
- 170 M. J. DE GROOT, A. A. ALEX and B. C. JONES, *J. Med. Chem.* **2002**, *45*, 1983–1993.
- 171 L. AFZELIUS, I. ZAMORA, M. RIDDRESTROM, T. B. ANDERSSON, A. KARLEN and C. M. MASIMIMERMBAWA, *Mol. Pharmacol.* **2001**, *59*, 909–919.
- 172 J. E. PENZOTTI, M. L. LAMB, E. EVENSEN and P. D. J. GROOTENHUIS, *J. Med. Chem.* **2002**, *45*, 1737–1740.
- 173 A. A. SEELIG, *Eur. J. Biochem.* **1998**, *251*, 252–261.
- 174 J. M. LEHMAN et al., *J. Clin. Invest.* **1998**, *102*, 1016–1023.
- 175 S. EKINS and J. A. ERICKSON, *Drug Metab. Dispos.* **2002**, *30*, 1–4.
- 176 R. E. WATKINS et al., *Science* **2001**, *292*, 2392–2333.

5

Receptor-Ligand Interaction

M. M. HÖFLIGER, A. G. BECK-SICKINGER

5.1

Receptors

Receptors are the trigger molecules regulating a great variety of metabolic and physiological processes in cells as well as in more complex organisms. In the pharmacological sense, receptors are transducer proteins that selectively and reversibly bind an endogenous signaling molecule or its synthetic analogue, undergo a conformational change, and modify a cellular response as a consequence. This very general definition of the term receptor will be specified more precisely if we look at the different types of receptors that are present in mammalian cells. On one hand, there are the intracellular receptors of the nuclear receptor superfamily such as the steroid hormone receptors, the retinoid receptors, and the thyroid hormone receptors. The second category of receptor proteins is that of the membrane-bound receptors, which are by far more diverse. Those receptors serve the cells to transform a signal from the outside into a cellular answer. The superfamily comprises the ion-channel-coupled receptors, the kinase-coupled receptors, and the large family of G-protein-coupled receptors.

In this chapter we focus on the characterization of the interaction between G-protein-coupled receptors and their ligands. The presented techniques for the examination of ligand binding and receptor activation, however, are not limited to the analysis of this receptor group.

5.1.1

The G-Protein-Coupled Receptors

The G-protein-coupled receptors (GPCRs) are integral membrane proteins with the common characteristic of seven membrane-spanning helices. Their endogenous ligands can be monoamine messengers (epinephrine, acetylcholine, serotonin, histamine, dopamine, etc.), lipids (prostaglandins, endogenous lipids, etc), neuropeptides (neuropeptide Y [NPY], substance P [SP], cholecystokinin [CCK], opioids, etc.), and peptide hormones (glucagon, angiotensin, bradykinin), as well as small proteins (chemokines) and large proteins (glycoprotein hormones, thrombin, etc.). All GPCRs transduce their signals to the interior of the cell through the

interaction with heterotrimeric G-proteins. Additionally, important sensory proteins such as rhodopsin and the olfactory receptors belong to this family. The length of the different GPCRs varies considerably, from less than 300 amino acids for the smallest representatives, such as the adrenocorticotropin receptor, to more than 1100 amino acids for the metabotropic glutamate receptors. The receptors carry different posttranslational modifications such as glycosylation, which often leads to higher molecular weights of the receptors than expected from the amino acid sequence. Other modifications include disulfide bridges and palmitoylation at specific sites. Most frequently, the GPCRs are classified by primary sequence homology and subfamilies are named after well-characterized members. While only low homology is found in the loop segments, the seven transmembrane helices containing 20–25 hydrophobic amino acids show a higher degree of conservation [1].

So far our insight into the three-dimensional structure of GPCRs is rather limited, as it is very difficult and time-consuming to crystallize such complex integral membrane proteins. Therefore, our knowledge about the 3-D structure of GPCRs is mainly based upon the crystal structures of bacteriorhodopsin and rhodopsin, the only homologous proteins with an elucidated crystal structure. Mainly, the crystal structure of bovine rhodopsin, solved by Palczewski et al. in 2000 [2], in combination with molecular modeling provides valuable knowledge for the understanding of other GPCRs. Computer models based on the sequence homology between rhodopsin and other receptor proteins can now be calculated. However, these models have to be investigated experimentally and crystal structures do not necessarily represent a receptor's native state, which is to be considered as a dynamic equilibrium rather than a single solid state.

5.2

Ligand-binding Theory

Binding of a ligand can lead to agonistic, antagonistic, or inverse agonistic effects. Those effects are related to conformational changes in the receptor proteins and subsequent activation or deactivation of signal transduction cascades (Fig. 5.1). Due to a conformational change, receptors in the active state gain the ability to bind heterotrimeric G-proteins, a variety of which are displayed in Tab. 5.1. The G-proteins that are present in their inactive GDP-binding form then exchange GDP to GTP. Smaller subunits of the heterotrimeric G-protein complex dissociate. The GTP-binding subunit also dissociates from the receptor and activates a specific signal transduction cascade. This can mean activation or deactivation of the adenylate cyclase, ion channels, phospholipase C, or phosphodiesterase.

So far, mainly two general models have been suggested that describe the interaction between a ligand and its receptor. The first one, also known as the induced fit or conformational induction hypothesis, postulates the receptor to be there in an inactive conformation. Agonist and antagonist are thought to bind to the receptor in a similar way. Binding of an agonist induces a conformational change in

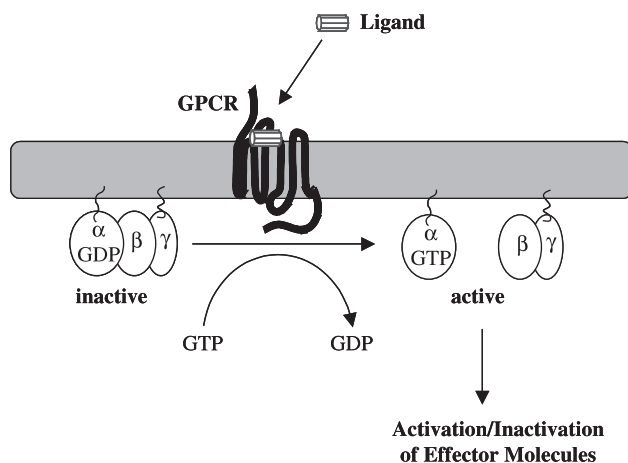


Fig. 5.1 General scheme of the activation and signal transduction cascade by a GPCR. After binding of a ligand, the receptor is stabilized in the active conformation. It then can bind to the heterotrimeric G-protein. GDP bound by the α -subunit is exchanged by GTP.

The α -subunit dissociates from the $\beta\gamma$ -complex. The signal can be transduced by the activated α -subunit as well as by the $\beta\gamma$ -complex, which both then interact with their respective effector molecules.

Tab. 5.1 G-protein subfamilies classified by their α -subunits.

Subfamily	Effector protein
G_s	Adenylate cyclase \uparrow Ca^{2+} channels \uparrow
G_i	K^+ channels \uparrow Ca^{2+} channels \downarrow cGMP specific phosphodiesterase \uparrow Adenylate cyclase \downarrow
G_q	Phospholipase $C\beta 4$
G_{12}	–

the receptor molecule, which then leads to the binding of G-proteins and activates the signal transduction cascade. In this model, antagonists bind to the same binding site as agonists, but do not induce the conformational change. The model can explain the action of agonistic or antagonistic ligands. However, some ligands have inverse agonistic effects, and, in fact, more and more substances formerly considered as antagonists are now classified as inverse agonists [3]. Their binding to a receptor not only inhibits the corresponding signal transduction pathway but even reduces it below its basic level. For example, this is the case for some therapeutic drugs such as cimetidine (histamine H_2 receptors), haloperidol (dopamine D_2 receptors), prazosin (α_1 -adrenoreceptors), timolol (β_2 -adrenoreceptors), and clo-

zapine (D_2 receptors and 5-HT_{2c} receptors) [4–8]. Furthermore, the phenomenon of constitutively active receptors has been observed where receptors are in an active state without ligand binding. This is, for instance, the case for the human β_2 -adrenergic receptor [9] and the human calcitonin receptor [10]. GPCRs have been expressed at different levels in cell culture, and a direct relationship between the level of expression and basal signaling could be shown [11]. Constitutive receptor activity has been found mostly as a result of mutational changes in the sequence. Both inverse agonism and constitutive activity suggest that not all receptor-ligand interactions can be explained by the induced fit hypothesis.

The second model is that of conformational selection, which has been developed more recently. According to this hypothesis, GPCRs exist in at least two conformations. At least one of them binds to G-proteins and therefore is considered as the active state (R^*). In other states the receptor cannot bind to G-proteins and is therefore referred to as the inactive, uncoupled receptor. There is an equilibrium between active and inactive conformations $R^* \rightleftharpoons R$. In the absence of a ligand, usually the inactive receptors represent the majority in this equilibrium. If a ligand is added, it may prefer a special conformation of the receptor for binding. In the case of an agonist, this will be the active state, whereas inverse agonists will prefer the inactive state. Binding of inverse agonists will therefore lower the level of basal signaling. Ligands that bind to active and inactive conformations of a receptor with the same affinity will not influence the equilibrium. They are competitive antagonists, as they can displace other ligands from the binding site but do not lead to changes in basal signaling. Ligands with only little preference for the active state will lead only to a small shift in the equilibrium towards the active conformation. The consequence is an only moderate increase in signaling, also referred to as low efficacy. Those compounds are partial agonists (Fig. 5.2).

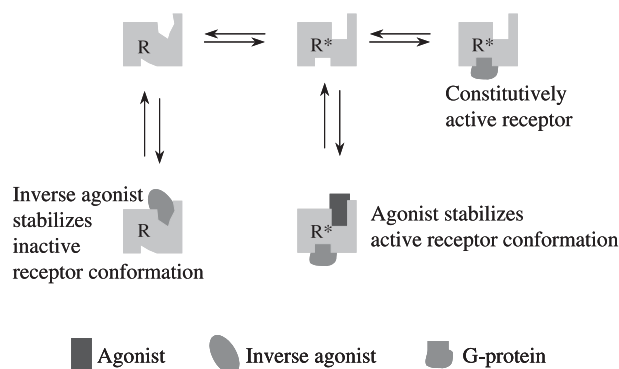


Fig. 5.2 Model of the receptor-ligand interaction. Active (R^*) and inactive (R) receptors are in a state of equilibrium. Whereas an agonist binds to the active conformation, an antagonist binds to the inactive conformation of the receptor. By stabilizing one conformation,

agonists and antagonists may shift the equilibrium to the respective side. Competitive antagonists will bind to both conformations with the same affinity and therefore not influence the equilibrium.

Between R and R^* there are only small differences in energy. This can be seen by single amino acid mutations leading to constitutive receptor activity [12–14]. Conformational selection and conformational induction, however, are not necessarily in contradiction. Both can be seen as two parts of one mechanism in which the active conformation of a receptor is induced during the binding process, whereas the ligand preferably binds to the active conformation [15].

5.3

Characterization of the Receptor-Ligand Interaction

The characterization of new receptors usually starts with the discovery of a new receptor-ligand pair. This often starts with knowledge about a pharmacologically active compound, the mode of action of which is unknown, as was the case for neuropeptide Y [16], a 36-amino-acid neuropeptide extracted from porcine brain in 1982. On the other hand, it is also possible that a receptor is discovered for which no ligand is known (orphan receptor). This was the case for the orexin receptors OX_1 and OX_2 [17]. For neuropeptide Y (NPY) in humans, until now at least four functional receptors have been identified, the $NPY-Y_1$, $NPY-Y_2$, $NPY-Y_4$, and $NPY-Y_5$ receptors. (For the $NPY-Y_4$ receptor, the endogenous ligand is the pancreatic polypeptide.) In the case of the orexin receptors, we know two peptide ligands, orexin A and orexin B, that differ significantly in their structure [18]. Once such a set of ligands and their corresponding receptors are known, the characterization of their interaction is necessary. On the one hand, it is desirable to understand how a ligand like NPY can distinguish between its receptors and whether there are differences in the binding mode. This will be especially important if it comes to the design of subtype-specific, pharmacologically active compounds. On the other hand, in cases like the orexin receptors – where we have the two ligands orexin A and B both binding to the OX_1 and the OX_2 receptor, respectively, but with different affinities [17] – it is important to find the causal connection between the ligand structure and its receptor selectivity.

5.4

Receptor Material

Usually, G-protein-coupled receptors and their ligands are first identified from primary tissue. However, it is difficult and also ethically questionable to get large enough amounts of primary tissue samples to closely investigate them. Furthermore, the stability of those samples and the accessibility of the embedded receptors can be a problem. There are different ways to circumvent those difficulties. Probably the easiest way is the isolation and cultivation of cancer cell lines that endogenously express the receptors that are of interest. This has been done with the SK-N-MC [19] cell line and the SMS-KAN cell line [20], which express the $NPY-Y_1$ and the $NPY-Y_2$ receptor, respectively. Another commonly used method is the cloning of the recep-

tors into either cancer cell lines or bacterial or yeast cells. Cancer cells have the advantage that they come closest to mammalian tissue cells, are able to make posttranslational modifications, and do not have a cell wall like yeast cells do. A number of such cell lines have been established, such as BHK (baby hamster kidney), CHO (Chinese hamster ovary), HEK (human embryonic kidney), or COS (SV40 transformed African green monkey kidney cells) cells. Those cell lines can be transfected with the DNA of GPCRs of interest [21–24]. The advantage is that this method allows studies with genetically modified receptors in which single or different amino acids are mutated or in which a receptor is fused to another protein [25]. The expression of fusion proteins is often used in combination with fluorescent proteins, such as green fluorescent protein (GFP) [26], to make receptors visible and to perform fluorometric assays (see Section 5.8). Usually, such transfections are transient. To make them stable, one can combine the receptor encoding DNA sequence with a sequence encoding for an antibiotic resistance. Cultivation of the respective cell line in an antibiotic-containing medium over a longer period of time then can lead to a stable transfection [27–29].

5.5

Binding Studies

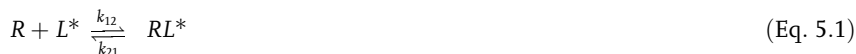
For new receptor-ligand pairs, usually the determination of binding parameters is the first step. Also, for new, artificial compounds designed as new ligands for a receptor, it is important to determine whether they are binders. Therefore, a variety of binding assays have been established that can be divided into two groups: separation assays and direct assays [30]. In separation assays either whole cells or membrane preparations containing the receptor of interest are used. They are incubated with a radioactive- or fluorescent-labeled ligand until the state of equilibrium is reached. The receptor-ligand complex is then separated by centrifugation or filtration and the amount of bound ligand is determined. The more recently developed direct assays, based on surface plasmon resonance (SPR) [31] or fluorescence correlation spectroscopy (FCS) [32], measure the receptor-ligand interaction in real time.

5.6

Binding Kinetics

Interactions between G-protein-coupled receptors and their ligands are reversible with the exception of rhodopsin. The binding parameters of a ligand can therefore be determined in a competition-binding assay with a labeled ligand. In such an assay, the displacement of labeled ligand from the receptor is measured in the presence of different concentrations of the examined ligand.

The general equation for a bimolecular association between a receptor (R) and a labeled ligand (L^*) is



where k_{12} is the association rate (on-rate) and k_{21} is the dissociation rate (off-rate) of the receptor-ligand complex. The association constant K (not to be confused with the association rate k_{12}) of the ligand binding reaction is $K = k_{12}/k_{21}$, whereas the dissociation constant K_d is $K_d = 1/K = k_{21}/k_{12}$. When $L^* = K_d = 1/K$, then $RL^* = R_t/2$, with R_t being the total receptor concentration. In other words, if the free concentration of the labeled ligand L^* reaches the value of K_d , the receptor-binding sites will be half-saturated with ligand. The value of K_d is one-half of the maximal specific binding B_{\max} . The free ligand concentration at 50% receptor saturation, the IC_{50} concentration, is a measure of K_d (or $1/K$).

The K_d and B_{\max} values of a ligand are determined in a saturation analysis. Cell membranes or whole cells are incubated with different concentrations of a labeled ligand. The resulting curve consists of specific binding to the receptor-binding sites and non-specific, non-receptor-binding sites. Each concentration of the labeled ligand should therefore be displaced by a 1000-fold excess of unlabeled ligand to distinguish between specific and non-specific binding. To get the values for the specific binding, the values for non-specific binding are subtracted from those for total binding. The resulting curve for specific binding should be saturable at sufficiently high concentrations and represent a hyperbolic function (Fig. 5.3a). As the K_d value is equal to the concentration of labeled ligand occupying 50% of the B_{\max} value, it can be determined from the saturation curve.

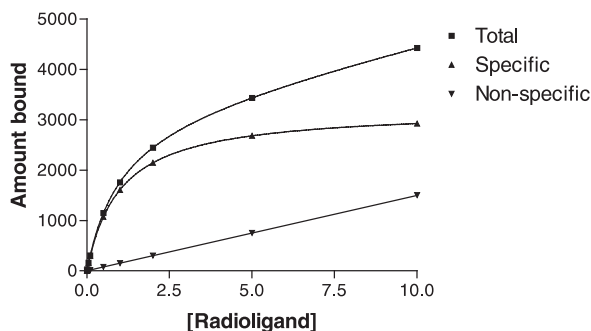
The data was formerly analyzed by Scatchard analysis where it is summarized in the Scatchard plot. The amount of bound ligand divided by the amount of free ligand in solution (y-axis) is plotted against the amount of bound ligand (x-axis) (Fig. 5.3b). In the case of a bimolecular interaction, this should lead to a straight line with a negative slope. The intercept point with the x-axis represents the B_{\max} value, whereas the absolute value of the slope represents the K_d value (Eq. 5.2). More recently, the Scatchard analysis is more and more replaced by computational non-linear regression analysis, a method that makes it possible to directly solve the equation of one-site ligand binding.

$$y = \frac{B_{\max} \cdot x}{K_d + x} \quad (\text{Eq. 5.2})$$

From the B_{\max} value, the number of binding sites per milligram membrane preparation or per cell, which corresponds to the number of receptors per cell, can be calculated.

The IC_{50} concentration of an unlabeled ligand is determined in a competition assay with a labeled ligand L^* . This has the advantage that not every new or uncharacterized ligand has to be labeled but can be tested against an already labeled and characterized compound. A constant concentration of L^* , usually at or below its K_d , is displaced by increasing concentrations of the unlabeled ligand. For a bimolecular reaction this results in a sigmoidal competition curve. The IC_{50} value is the concentration of unlabeled ligand, which displaces 50% of the specifically

A



B

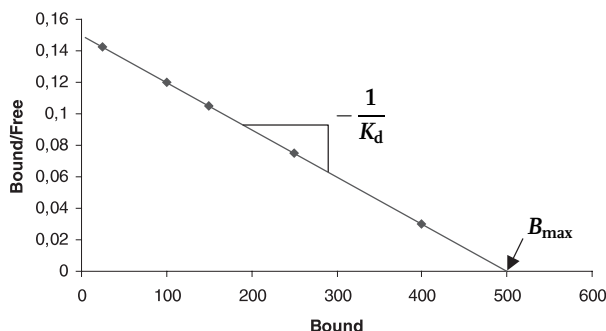


Fig. 5.3 (A) Saturation analysis for a radio-labeled ligand. The x-axis represents the concentration of the labeled ligand. The y-axis shows the dpm-values (decays per minute) for total binding, non-specific binding, and specific binding. The values for specific binding are obtained by subtraction of non-specific binding and total binding. (B) Scatchard

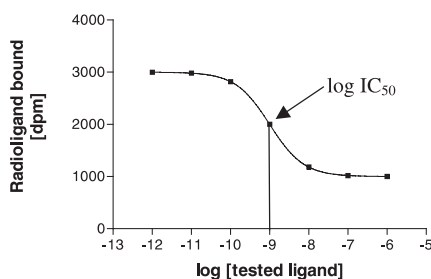
plot of a saturation analysis experiment. The x-axis shows the amount of bound ligand, whereas the y-axis represents the amount of bound ligand divided by the amount of free ligand in solution. B_{\max} is determined from the intercept point with the x-axis. K_d is the absolute value of the slope (both graphs present artificial, idealized data).

bound labeled ligand from the receptor-binding site (Fig. 5.4). The IC_{50} value, however, is dependent on the concentration of labeled ligand and may vary between experiments performed under different conditions.

With the constant value K_i , which is independent from the concentration of the labeled ligand, it is easier to compare datasets from different experiments. K_i is calculated according to the equation of Cheng and Prusoff [33]:

$$K_i = \frac{IC_{50}}{1 + \frac{L}{K_d}} \quad (\text{Eq. 5.3})$$

Fig. 5.4 Competition analysis (idealized data). The labeled ligand, which is held at a constant concentration at or below its K_d , is displaced by increasing concentrations of the examined ligand. The x value of the curve's inflection point represents $\log IC_{50}$.



5.7

Binding Assays

5.7.1

Separation Assays

In separation assays the receptors, usually in the form of membrane preparations or whole cells, are incubated with a labeled ligand. The incubation should be long enough to reach the state of equilibrium, which can be anything from seconds to hours. For thermodynamic reasons, the incubation should take place at room temperature or body temperature, which allows the receptors as well as the ligands to change conformation and therefore to interact with each other. As soon as the state of equilibrium is reached, the receptor-bound fraction of the labeled ligand is separated physically from the unbound fraction, which is still in solution. This can be done by either centrifugation or filtration. To avoid disturbance of the established equilibrium during the separation process and following washing steps, the sample should be cooled during these procedures. This significantly reduces the association and dissociation rate. Then the amount of receptor-bound labeled ligand can be determined [30].

5.7.2

Radioligand-binding Assay

One of the most common separation assays to test for a ligand's binding to a receptor is the radioligand-binding assay. Radioactively labeled compounds are used in this type of assay. The high sensitivity of the method allows for the detection and quantification of very small amounts of ligand. Tritium (^3H) and iodine-125 (^{125}I) are the most commonly used isotopes. The selection of the respective label should, besides the availability, consider the radiochemical and safety properties. Tritium has a long half-life period (more than 12 years), which means that no correction is needed for the decay during the experiment and a labeled tracer can be used over a long period of time (provided that it is chemically stable). Furthermore, its incorporation into a molecule has no or only few sterical effects. Its properties as a beta radiator with a low radiant energy and only short radiant distance (0.6 cm in air) make it more convenient for handling than the gamma emit-

ter iodine-125. However, the lower efficiency in measurement of ^3H (only 40% in liquid scintillation counting) may make it unsuitable for some applications. ^{125}I , with its higher radiant energy, higher maximum specific activity, and higher efficiency of measurement (75–90% in gamma counting), may be more suitable for assays in which very high sensitivity is required. One major disadvantage is, however, the size of the isotope ^{125}I , which can cause sterical changes in the tracer molecule and thereby influence its binding properties. Also, the low half-life period of ^{125}I makes it necessary to re-determine the specific activity of the labeled ligand before every assay.

The radioactive binding assay can be used in both saturation and competition assays (see Section 5.6). In the saturation analysis, a variety of ligand concentrations are used to get a saturation curve. Each concentration is then competed by a 1000-fold excess of the unlabeled ligand to determine the level of non-specific binding. In the competition assay, a single concentration of the radioligand is used. This concentration should preferably be at or below the K_d value of the respective ligand. It is then competed by a variety of concentrations of the unlabeled ligand. The concentrations of the unlabeled competitor should be chosen in a way that the resulting competition curve covers the complete range from no displacement at all to complete displacement of the labeled ligand from the receptor [30, 34].

5.8

Fluorometric Assays

5.8.1

Fluorescence Labels

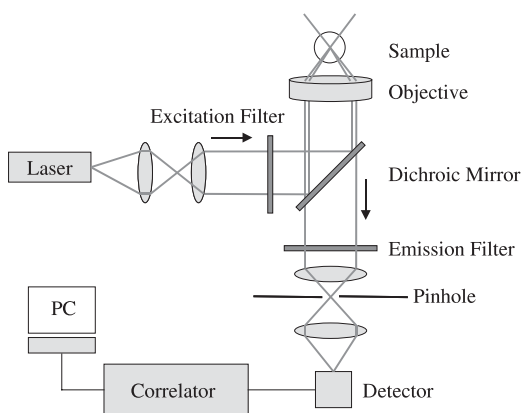
Examples of fluorescence labels for ligands are carboxyfluorescein, Cy3TM, a commercially available fluorescent marker based on a cyanine dye or tetramethylrhodamine. They are chemically introduced into a ligand. As with the radioactive labels, a possible influence of the labels on the binding behavior of the labeled ligands has to be considered, especially as the fluorescent dyes are complex molecules. Furthermore, the receptors themselves can be fluorescent labeled, which is done recombinantly. The respective receptors are expressed as fusion proteins with fluorescent proteins, e.g., green fluorescent protein (GFP) from *Aequorea victoria*, one of its mutant variants, or DsRed from *Discosoma striata* [26, 35].

5.8.2

Fluorescence Correlation Spectroscopy (FCS)

For a number of assays, fluorescent-labeled analogues of a ligand are used. Some of those assays are of a quantitative nature, such as in fluorescence correlation spectroscopy (FCS), where the fluorescent-labeled analogue is used to determine binding kinetics (Fig. 5.5) [36]. FCS allows the direct detection of molecular interactions in solution. FCS monitors the random motion of a fluorescent molecule

Fig. 5.5 Schematic setup of a fluorescence correlation spectrometer. A beam of parallel laser light is focused by an immersion objective onto a fluorescent sample. The fluorescence of molecules traversing the focus is collected by the same objective. It is filtered, focused, and detected. Signal autocorrelation is carried out by the computer (PC).



in a defined volume (~ 1 fL). Thereby, the diffusion rate of a particle can be detected, which is directly dependent on the particle's mass. Therefore, FCS can quantify the bound and the free fraction of the fluorescent-labeled ligand and can be used to determine binding parameters. In contrast to the radioligand-binding assay, this has the advantage that the receptor-ligand interaction can be directly monitored over a longer period of time. Recently, FCS was applied to study the receptor diversity of the neuropeptide galanin in cultured cells [37]. In this case three different diffusion times for rhodamine-labeled galanin (Rh-GAL) could be detected: one short diffusion time for unbound Rh-GAL and two different longer diffusion times for membrane-bound Rh-GAL. Those findings suggest that FCS allows one not only to determine the amounts of bound and unbound ligand for a receptor-ligand complex but also to distinguish between different receptor subpopulations or different receptor subtypes.

5.8.3

Fluorescence Microscopy

Other fluorometric assays will be of a more qualitative nature, if it comes to microscopic studies to characterize receptor expression on cells [38, 39] (Fig. 5.6) or receptor internalization upon the binding of a ligand [40].

5.8.4

Fluorescence Resonance Energy Transfer (FRET)

The use of either pairs of differently fluorescent-labeled receptors or fluorescent-labeled receptors and fluorescent-labeled peptides allows fluorescence resonance energy transfer (FRET) studies. The prerequisite is that the used chromophores form a FRET pair. This means, when in close proximity, that the so-called donor, excited at a certain wavelength, transfers its radiation energy non-radiatively to the closely located acceptor chromophore. Emission at the acceptor's emission wave-

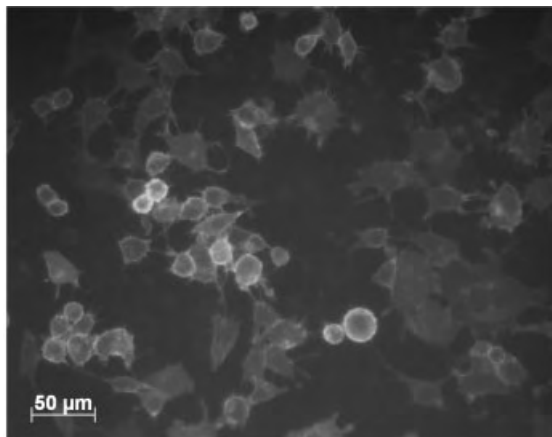


Fig. 5.6 Fluorescence microscopy image of BHK cells expressing a human NPY- Y_2 receptor-GFP fusion protein.

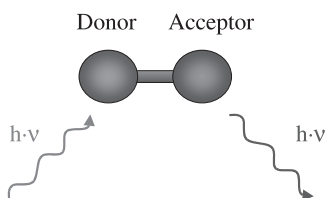


Fig. 5.7 Principle of FRET. A donor fluorophore is excited at a certain wavelength and transfers its radiation energy non-radiatively to a closely located acceptor fluorophore. Emission at the acceptor's wavelength can therefore be measured to determine whether the two chromophores are co-localized or not.

length can be detected (Fig. 5.7). A receptor-ligand FRET pair can be used to study the binding of a ligand to its receptor [41]. The co-transfection of cells with a receptor-receptor FRET pair can help one to decide whether a receptor is present in the cell membrane as a monomer or dimer [42, 43]. If two different types of receptors are used in a FRET assay, it may even help one to decide whether the receptors are present as heterodimers.

5.9

Surface Plasmon Resonance

Another technique used for the analysis of receptor-ligand interaction is surface plasmon resonance (SPR), with its first commercially available application in the BIAcore® instruments [44] (Fig. 5.8). Like FCS, it allows the determination of kinetics by monitoring the association and dissociation of a receptor-ligand complex in real time. The interaction partners do not necessarily have to be labeled, which is an advantage of the technique. The principle of SPR measurements is based on an optical phenomenon. The core unit in this technique is a sensor chip consisting of a thin gold film with a modified surface attached on one side. One reactant is attached to the modified sensor surface, whereas the other reaction partner flows past this surface in solution. When the two interaction partners form a com-

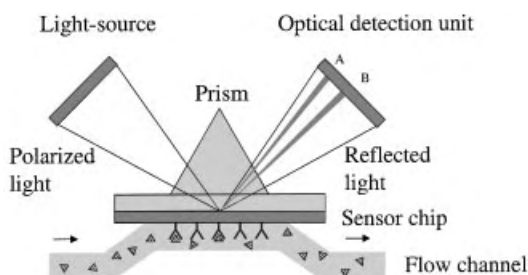


Fig. 5.8 Schematic setup of the surface plasmon resonance (SPR) detection unit in a Biacore® instrument. One interaction partner is immobilized on a modified gold surface, whereas the other flows by in solution. On the other side of the sensor chip, a beam of polarized light is reflected by the gold film. The optical phenomenon SPR leads to a re-

duction in the intensity of reflected light at a certain angle (A). This effect is dependent on the refractive index at the sensor surface and thereby on the mass bound to the chip. Therefore, the angle changes if a ligand is bound by the molecules immobilized on the chip (B). The response values are then displayed in resonance units (RU).

plex, the mass on the sensor chip surface increases; when the complex dissociates, the mass falls. This leads to changes in the refractive index in the aqueous layer close to the sensor chip surface that is measured by an optical detection unit on the dry side of the chip. The signal measured in arbitrary resonance units (RU) is approximately proportional to the change in mass with $1 \text{ RU} = 1 \text{ pg/mm}^2$ for proteins [45]. A variety of sensor chips with modified surfaces is commercially available, most of them based on a carboxymethyldextrane matrix. This can be used to directly couple one interaction partner via its functional groups by defined coupling chemistries for thiol, amine, or aldehyde coupling, or it can be further modified for capturing biotinylated or histidine-tagged interaction partners or for capturing liposomes. Finally, there is also a hydrophobic sensor surface available, composed of long-chain alkanethiol molecules, that can be used to construct lipid bilayers as membrane-like environments [46]. As recently described, liposome-capturing sensor chips can also be used to reconstitute G-protein-coupled receptors on them [47]. This was done with rhodopsin as the most frequently used model for G-protein-coupled receptors in general. Rhodopsin was immobilized and reconstituted in mixed micelles on the sensor chip surface. If the technique turned out to be suitable for other G-protein-coupled receptors as well, it might become a valuable tool to test for the functionality of solubilized receptors, as well as to screen for new ligands and to determine binding kinetics.

5.10

Molecular Characterization of the Receptor-Ligand Interaction

5.10.1

Antibodies

Antibodies can be used for a variety of applications in the molecular characterization of receptors and receptor-ligand interactions. Antibodies can be used for the detection of receptors in tissue slices, Western blot experiments [48], or ELISAs (enzyme-linked immunosorbent assays) [49]. They can also be used in competition experiments to map the binding epitope of a ligand [50]. Even though the use of antibodies is routine, there is no general protocol for their generation.

Polyclonal antibody sera contain many different antibodies with different affinities for the epitopes of an antigen. They are generated by immunization of animals, mostly chickens or rabbits [51], with the antigen (Fig. 5.9 b). To get good immunization results, the animal species should not be too closely related to the species where the antigen comes from. In the case of a G-protein-coupled receptor, immunization may be done with the whole protein or peptide segments of the receptor. Immunization with a whole receptor protein will create a serum with a high diversity of antibodies directed against all parts of the molecule – extracellular, intracellular, or transmembrane segments [51]. The production of sufficient amounts of purified receptor is often difficult and is connected with considerable loss of material. One possibility is the separation on a polyacrylamide gel, followed by blotting onto a nitrocellulose membrane. The membrane can then be introduced under the skin of the animal. Even more drastic is the direct injection of a portion of polyacrylamide gel containing the receptor. However, these methods not only are very stressful for the animal but also are applicable only if other methods to identify the respective receptors are known. As antibodies are often generated against not very well characterized proteins, this may be a problem.

The immunization of animals with synthetic peptides is a way to circumvent these problems. Furthermore, it has the advantage that it leads to antibodies directed against a specific and known part of the protein, a benefit in some applications. Such antibodies can recognize parts of a receptor involved in ligand binding or serve as ligands with agonistic [52, 53] or antagonistic properties themselves. They can also displace other ligands, which might help to localize their binding site at the receptor. The peptide sequences are available from GenBank entries of the respective receptor or can be determined from the coding gene sequence. To allow good accessibility in cell or membrane assays, the peptide segments should preferably be chosen from parts of the protein exposed to the surface, especially the extracellular domain [49]. For previously unknown proteins, this information can be estimated from hydropathicity blots. Antibodies generated against intracellular parts of the receptor can be used to explore signal transduction processes or G-protein binding [54].

To act as an antigen, the mass of a protein should be higher than 5–10 kDa. This often is not the case for the peptides of a receptor loop, so they have to be

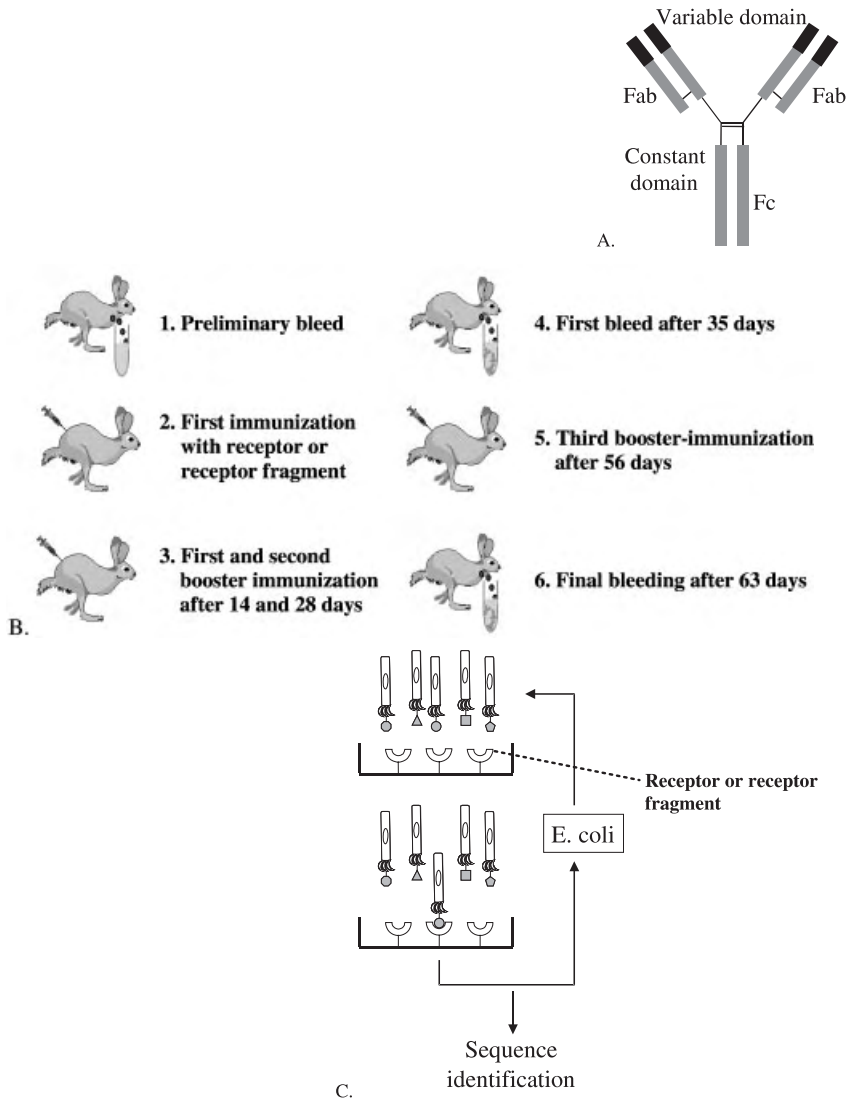


Fig. 5.9 (A) Model of an immunoglobulin molecule. The molecule consists of the Fc portion, containing two disulfide bridges and two Fab portions with one disulfide bridge each. The Fab portions carry the variable domains of the antibody. (B) Immunization scheme for rabbits. Before administration of the antigen, a preliminary bleed should be taken. The animal is then immunized with the antigen. After 14 and 28 days, the first and second booster immunizations are given. After

35 days a first bleed is taken from the animal. The third booster immunization follows after 56 days. Final bleeding is taken after 63 days. (C) Phage display for the generation of monoclonal antibodies. Different antibodies are displayed on the phages' surfaces. They are then selected via an antigen. The phages that display antibodies with affinity for the antigen are amplified in *E. coli*. Finally, the sequences of the respective antibodies are identified on the DNA level [51].

coupled to a larger protein carrier. Because the immunogenic reaction will be directed against parts of the carrier as well, it should not be relevant for future assays. Bovine serum albumin (BSA) or keyhole limpet hemocyanin (KLH) is commonly used. Because BSA is often used as a blocking agent, for instance in Western blotting, KLH might be a better choice in many cases.

Antibody sera obtained from immunized animals can in some cases be directly used for further experiments. Further purification and enrichment, however, are often necessary to increase the specificity and overall affinity of the purified antibodies. Simple enrichment of the antibodies can be achieved by precipitation with ammonium sulfate [55] or by chromatography with protein A [56] or protein G, bacterial cell wall proteins that specifically bind to the Fc portion of immunoglobulins. A more specific enrichment is achieved by affinity chromatography with an antigen column.

An alternative to the rather complex mixtures of polyclonal antibodies are monoclonal antibodies [51]. In the first step, as for polyclonal antibodies, an animal is immunized. The antibody-secreting lymphocytes are then isolated from lymphoid tissue and fused with plasmacytoma cells of a similar differentiation state. The new hybridoma cells can then be cultivated and selected, and supernatants can be screened for activity against the antigen. This procedure can be repeated until a clone is found, which produces antibodies with good affinity. Monoclonal antibodies have the advantage that they can theoretically be reproduced infinitely.

Another possibility for the generation of monoclonal antibodies is phage display [51] (Fig. 5.9c), an evolutionary technique in which antibody V genes are cloned for display of assembled heavy- and light-chain variable domains into filamentous bacteriophage. Phages binding to the antigen are selected and soluble antibody fragments are expressed by infected bacteria.

5.10.2

Applications of Antibodies

5.10.2.1 Receptor and Ligand Detection

Because there are a huge variety of applications for antibodies, in this section we focus on those concerning receptor characterization and receptor-ligand interaction. Antibodies represent a valuable tool to detect receptors as well as their ligands in tissue samples by direct staining of tissue slices. Another method is the separation of cell membranes by SDS polyacrylamide gel electrophoresis (SDS-PAGE), followed by Western blotting onto a nitrocellulose or PVDF membrane (Fig. 5.10). In both cases the receptors are incubated with the antibodies, non-bound antibodies are washed away, and a second antibody specific for the Fc fragment of the first antibody is added. The second antibody is labeled, either for direct detection with a fluorescent dye or radioactivity or for detection by a staining reaction, with enzymes like alkaline phosphatase or horseradish peroxidase.

Often nothing is known about new receptors but their amino acid sequence. In such cases antibodies can be used to screen for their production in various tissues either directly in tissue slices or after immunoblotting of membrane preparations

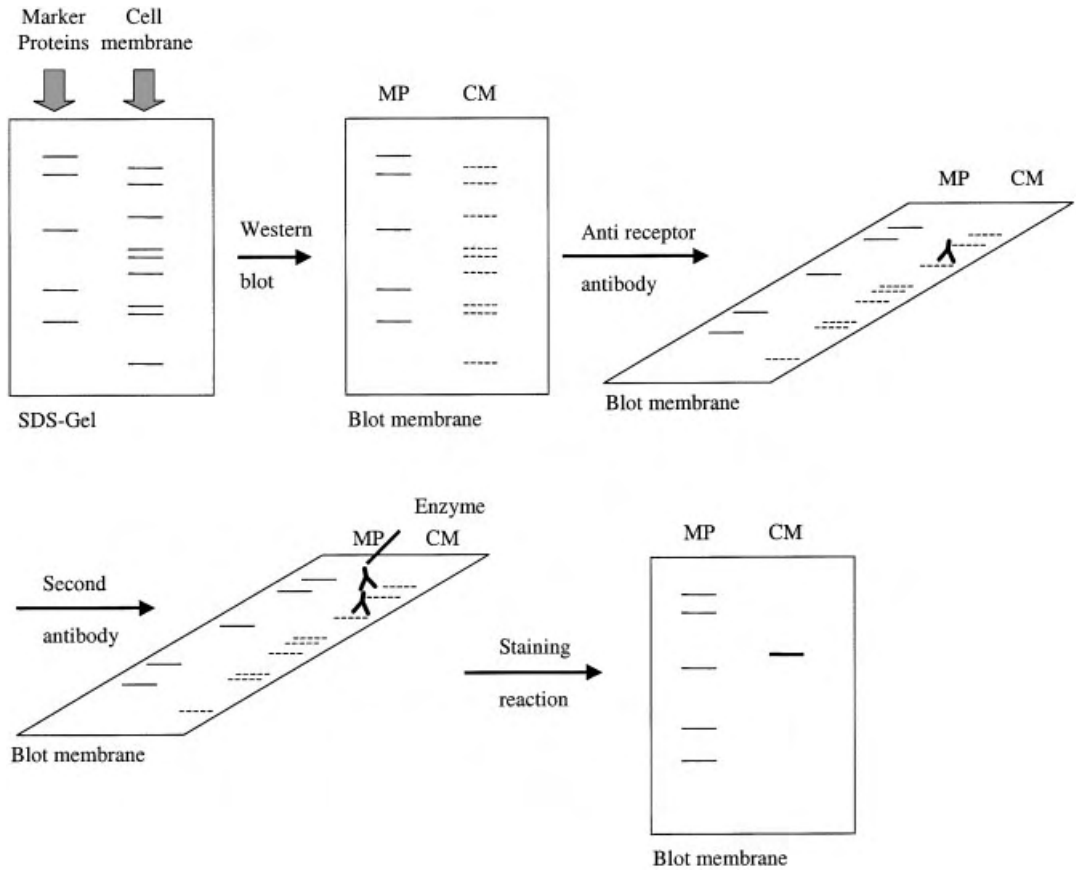


Fig. 5.10 Receptor characterization by SDS-PAGE and Western blotting. A cell membrane sample is separated by SDS polyacrylamide gel electrophoresis. The proteins are then blotted onto a nitrocellulose membrane. The membrane is incubated with an antibody that is specific for the receptor. After incubation

with a second enzyme-linked antibody, the receptor band is stained with an enzymatic dye reaction. The parallel separation of a standard protein mixture allows one to determine the molecular mass of the receptor (MP: marker proteins; CM: cellular membrane) [51].

[48, 49]. Such immunohistochemical studies were performed for the angiotensin II receptor subtypes AT_1 and AT_2 in rat adrenal gland [48] and in the heart of rabbits [57], for the metabotropic glutamate receptor in rat piriform cortex [58], for serotonin 5-HT_{2A} and 5-HT₃ receptors in inhibitory circuits of the primate cerebral cortex [59], and for the dopamine D_1 and D_2 as well as M_4 muscarinic receptor in striatonigral neurons [60].

5.10.2.2 Receptor Characterization

The detected receptor masses in immunoblotting experiments often differ significantly from those calculated from the amino acid sequence [48, 49]. Sometimes even several bands of different molecular masses are detected for one receptor. Higher receptor masses and different bands may be due to posttranslational modifications like glycosylation or lipid residues. However, such observations are often difficult to explain and have to be checked carefully in each case. To distinguish whether a receptor double band is due to differences in the degree of glycosylation, due to different splice variants, or simply due to receptor degradation during the work-up process, a deglycosylation experiment can be performed. For this the receptors are treated with enzymes that release oligosaccharides from glycosylation sites, followed by immunoblotting and staining with antibodies. If a receptor is glycosylated, this should result in a single band, which should be detectable at the calculated molecular weight of the receptor.

The most frequent posttranslational modifications for GPCRs are N-glycosylation at the N-terminus and external loop Asn-X-Ser/Thr sequences (human calcitonin receptor-like receptor [61] and α_{1B} -adrenergic receptor [62]), palmitoylation (human dopamine D₁ receptor [63]), and phosphorylation (β_2 adrenergic receptor [64]).

For many receptors different subtypes with a high degree of sequence homology are known. In cases such as the neuropeptide Y receptor family – with the Y₁, Y₂, Y₄, and Y₅ receptors functionally expressed in humans – or the angiotensin II receptor family – with the receptors AT_{1A}, AT_{1B}, and AT₂ in human – antibodies that distinguish between different receptor subtypes are a valuable tool to monitor the receptor expression on the protein level [48, 49].

5.10.2.3 Functional Characterization of the Receptor-Ligand Interaction

Antibodies generated against a peptide segment of a receptor can be used to map the binding site of the endogenous ligand at the respective receptor. This is especially the case if a variety of antibodies generated against different receptor peptides are available. The receptors are then simultaneously incubated with ligand and antibodies. In an ELISA the amount of antibody bound in the presence or absence of the ligand can be determined. Competition is expected for antibodies that have binding sites at the receptor that are overlapping with the ligand's binding site. In addition, competition assays with radiolabeled ligand can be performed. In that case the amount of ligand bound to the receptor is determined after simultaneous incubation with the antibody. The assay is performed in the same way as the radioactive competition assay (see Section 5.7.2). Finally, if photoactivatable analogues (see Section 5.10.5) of the ligand are available, they can be cross-linked to the receptor. Afterwards, the cross-linked receptor-ligand complex can be incubated with the anti-receptor antibodies to see whether their binding site is blocked. A combination of these techniques was used to characterize the binding site of NPY at the Y₁ receptor [50].

5.10.3

Aptamers

Aptamers (Latin *aptus*=fit) are RNA or DNA molecules isolated from combinatorial nucleic acid libraries by *in vitro* selection experiments termed SELEX: systematic evolution of ligands by exponential enrichment. A few binders with good affinities for the target are selected from up to 10^{15} different oligonucleotide sequences. The selection is performed by column chromatography or other enrichment techniques that are suitable to separate binders from non-binders. Non-binders are washed away, whereas the binders are regained, amplified, and subjected to a new cycle of selection. After several cycles the stringency of the binding conditions is increased, which allows the selection of good binders only [65]. Developed in 1990 [66, 67], until now aptamers have been isolated against more than 150 target molecules, among them small organic molecules, amino acids, peptides, and proteins [68].

The mechanism of target recognition by aptamers is adaptive [69]. Whereas they are predominantly unstructured in solution, aptamers fold upon association with their ligands. The ligand becomes an intrinsic part of the nucleic acid structure, and the 3-D structures of aptamer complexes form highly optimized scaffolds for specific ligand recognition. Thus, aptamers, like receptors, seem to be able to distinguish not only between different ligand molecules but also between different conformations of a single molecule. This was nicely demonstrated for an aptamer selected against the 36-amino-acid peptide NPY [70]. NPY is the endogenous ligand for the Y_1 , Y_2 , and Y_5 receptors. Receptor-selective, conformationally constrained, synthetic analogues of the peptide are known, suggesting that NPY binds to its receptors in different conformations (see Section 5.10.4.2). The generated aptamer was tested against a variety of these peptides and showed preference for Y_2 -receptor-selective analogues of NPY. Furthermore, a competition experiment was performed in which aptamer and radiolabeled NPY were simultaneously incubated with the respective receptors. Again, the aptamer showed higher competition for NPY at the Y_2 receptor.

Recently, intramers, aptamers that can be expressed inside cells and retain their function [71], were developed. Such intramers might also be generated against biomolecules that are part of a signal transduction pathway. In the future they might be used to elucidate signal transduction cascades triggered by a receptor.

5.10.4

Receptor Mutation and Ligand Modification

A way to gain insight into the interaction between a receptor and its ligand is to mutate single or several positions in the receptor sequence or to modify the ligand. Whereas in the receptor molecule an amino acid is usually substituted by another, the modifications of the ligand can be various. Different ligands, endogenous or synthetic, for a receptor do not necessarily have the same chemical structure. They do not even have to be members of the same chemical class; therefore,

the determination of a pharmacophore (defined orientation of functional groups being the basis of biological action) may be difficult. Another problem is the determination of the ligand's binding site at the receptor, especially because membrane receptors are difficult to crystallize and our structural models of G-protein-coupled receptors are based mainly on the X-ray structure of rhodopsin.

5.10.4.1 Receptor Mutagenesis

One way to locate domains of a receptor that are involved in ligand binding is the creation of chimeric receptors. Chimeric receptors are fusion proteins in which one part originates from one receptor and the other part originates from another receptor. In most cases, chimeric receptors are used to obtain an initial picture of the location of interesting segments involved in ligand binding [72]. Usually, such chimeric constructs are made of two closely related receptors, such as the neurokinin NK₁ and NK₂ receptor [73], to obtain functional receptors. There are, however, examples of chimeric constructs from two distantly related receptors, such as the muscarinic M₃ and the adrenergic α_2 receptor [74, 75]. In the respective chimeric constructs, the amino terminal five transmembrane domains (TMs) originated from the α_2 adrenergic receptor and the carboxy terminal two TMs originated from the M₃ receptor (or *vice versa*). Whereas the single chimeric constructs were not functional, the co-expression of both constructs leads to functional activity corresponding to both an adrenergic and a muscarinic receptor.

Another way to gain more detailed information on important residues for ligand recognition is site-directed mutagenesis of single amino acid residues of a receptor sequence. Before receptor mutations are introduced, one has to select amino acids that are likely to be of relevance for receptor function and ligand binding. One approach is to search for conserved residues of a receptor. This can be done by alignment either of sequences of the same receptor from different species or of sequences of different subtypes of a receptor family. Conserved amino acids, especially in the extracellular regions but also in the transmembrane regions, are likely to play a role in the mechanism of ligand binding.

Amino acids in a transmembrane helix often interact with amino acids from another helix in a way that stabilizes the inactive conformation of a receptor [76]. Disruption of such an interaction by an agonistic ligand or by mutation of one interaction partner leads to a change in conformation and to activation of the receptor. This was demonstrated for the α_{1B} -adrenergic receptor in which residues Asp125 in transmembrane domain 3 and Lys331 in transmembrane domain 7 apparently form a salt bridge that holds the receptor in the inactive conformation [77]. The mutation of Lys331 to Ala led to a six-fold increase in recognition of the endogenous ligand epinephrine without influencing the binding behavior of selective antagonists. Furthermore, the mutation led to an increased level of basal receptor signaling. This explains why some receptors upon mutation of one single amino acid residue are constitutively active. In combination with modeling studies, an assumption for the second interaction partner can be made, which then has to be confirmed by creating the respective mutant.

Conserved residues in the intracellular parts of a receptor are more likely to be involved in G-protein binding and signal transduction. This is the case for the DRY motif, a highly conserved amino acid sequence that, with the exception of some conservative mutations, is present in all rhodopsin-like receptors [1].

Another way to identify residues that might be important for ligand binding is to take amino acids from the putative binding region of a receptor and to search for possible interaction partners in the ligand. This approach was performed with the human NPY-Y₁ receptor, where a number of negatively charged aspartic acid and glutamic acid residues in the extracellular domains of the NPY-Y₁ receptor were systematically mutated to alanines [78]. Those residues are possible partners for electrostatic interactions with positively charged amino acids (arginine, lysine, and histidine) in the N- and C-termini of NPY. In radioligand-binding assays, a number of residues were shown to be essential for ligand binding (D104, D194, D200, D287), whereas others had no or only moderate effects.

5.10.4.2 Ligand Modification

To investigate which residues of the ligand are important for receptor recognition, analogues of a known ligand often are synthesized and tested for binding at the respective receptors. This can be done by rational or combinatorial approaches or by a mixture of both. In combinatorial methods the ligand is modified systematically, and up to several hundred slightly different compounds are prepared. Such methods have become possible with the introduction of robotic techniques for chemical synthesis that allow the simultaneous synthesis of many compounds at the same time. Such a set of chemical entities is also called a library. The single compounds are then tested for binding and for their biological activities. Thus, more important residues can be distinguished from less important ones. If analogues with antagonistic properties are detected, structural comparison to the endogenous ligand may help to detect elements that are necessary for receptor activation. For instance, aromatic side chain residues like Tyr, Trp, His, and Phe can be found to play an important role in receptor activation [79]. Also, the reduction of backbone amide bonds converts some peptide agonists to antagonists. This was demonstrated for the C-terminal tetrapeptide of gastrin, where the respective pseudo-peptide analogues had antagonistic properties [80].

Other approaches are more rational. For peptide ligands, a fast way to locate the binding site on the ligand is to create truncated analogues, representing only parts of the original molecule. The truncation does not necessarily have to be N- or C-terminal; also, centrally truncated analogues have been reported to act as ligands. In the case of NPY, a highly Y₂-receptor-selective analogue was reported in which the central amino acids 5–24 have been replaced by a single aminohexanoic acid (Ahx) molecule [81]. This leads to the assumption that a conformation in which the N- and the C-termini of the peptide are closely associated is relevant for binding at this receptor.

As with the receptors, for peptide ligands, chimera can be created in which one part of the molecule originates from one peptide, whereas the other part origi-

nates from a second peptide. If the original peptides are ligands of the same receptor family, the chimera can be used to determine structural characteristics that are the precondition for subtype specificity. Chimeric peptides have been used to study the interaction between galanin and its receptors [82].

A more rational approach is the alanine scan (Fig. 5.11), a method frequently used to screen the amino acid sequence of peptide ligands for the contribution of each residue to the receptor-ligand interaction. In an alanine scan, every single amino acid of a peptide's natural sequence is replaced by L-alanine. Alanine residues in the natural sequence are usually replaced by glycine. For example, a complete alanine scan of neuropeptide Y revealed residues that are important for NPY's interaction with the NPY-Y₁ and NPY-Y₂ receptors [83]. Testing of the analogues at the respective receptors revealed that parts of the ligand, such as the C-terminal pentapeptide amide and especially R33, were necessary for the recognition of both receptors. Other residues were needed only for the binding to one receptor (P2, the NPY loop and R33 for the Y₁ receptor and the C-terminal helix, and Y36 for the Y₂ receptor), whereas their exchange was tolerated by the other receptor.

Based on the findings from an alanine scan, further peptides can be synthesized in which residues involved in the receptor recognition are substituted by homologous amino acids. Interesting in this context is also the use of conformationally constrained analogues in which amino acid residues such as the helix breaker proline or the turn-inducing motive alanine-aminoisobutyric acid are introduced into the natural peptide sequence.

In the case of NPY, a ligand that has a distinct conformation in solution [84, 85], it is questionable whether the observed effects are due to direct interactions between residues of the ligand and side chains of the receptor or whether they

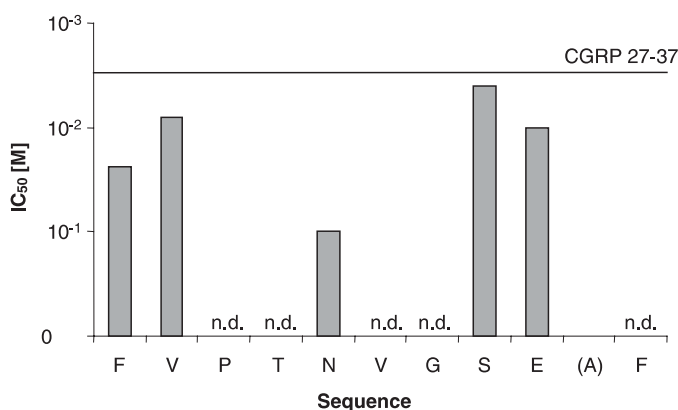


Fig. 5.11 L-Alanine scan of CGRP Y⁰-28-37, a selective antagonist for the human calcitonin gene-related peptide 1 receptor (CGRP₁). The analogues were tested in binding-competition assays against (¹²⁵I-iodohistidyl)-CGRP. Whereas the replacement of some residues

leads to complete loss of binding (P29, T30, V32, and G33), other replacements have a more moderate effect on the receptor recognition (n.d.=no displacement; (A36) was not replaced) [100].

are caused by an altered conformation of the ligand. Thus, instead of solely determining the binding affinities of modified ligands, one has to investigate their conformation as well. Usually, it is easier by far to gain structural information on the ligands than on G-protein-coupled receptors. Many ligands can be synthesized in sufficient quantity and purity, most of them are water-soluble, and some of them can also be crystallized. A number of techniques allow structural insight into ligand conformation, such as X-ray crystallography, solid-state and solution NMR, and circular dichroism studies for peptides. Well-characterized sets of ligands, especially when containing agonists as well as antagonists for a specific receptor, can be used for computer modeling and structure affinity relationship studies. Ligands for receptors that are not members of the same chemical class and therefore do not share structural similarities at first sight are especially helpful in creating a pharmacophore hypothesis.

5.10.4.3 Combination of Receptor Mutation and Ligand Modification

Some amino acids of a receptor's sequence, when mutated, may have effects on the recognition of an agonist but not on the recognition of an antagonist and *vice versa*, a fact that supports the conformational selection hypothesis for the interaction between ligands and their receptors (see Section 5.2). Based on the mentioned mutagenesis study at the NPY-Y₁ receptor and the results of the alanine scan of NPY, a model for the interaction of NPY with the hY₁ receptor was designed [86]. In the meantime, the non-peptide compound BIBP 3226 was created and shown to act as a competitive, specific, and selective Y₁ receptor antagonist [87, 88]. A second receptor mutation study at the NPY-Y₁ receptor, based on those findings, showed that the agonist NPY and the antagonist BIBP 3226 share an overlapping, but not identical, binding site [89]. Whereas some mutations affected the binding of the endogenous ligand NPY only, others lead to decreased binding of both NPY and BIBP 3226, and one mutation affected the binding of the antagonist BIBP 3226 only.

A similar approach was done with the NK-1 receptor and its peptide ligand substance P [90]. Important residues for the recognition of substance P were identified in the N-terminal extension, just outside the transmembrane domain 1 (TM-I), in the first extracellular loop outside TM-II and at the top of TM-III. Substitutions of these residues lead to a dramatic loss of binding of substance P but did not affect non-peptide antagonist binding. The residues important for non-peptide antagonist binding were identified in the outer parts of TM-IV and TM-V. As in the case of the NPY-Y₁ receptor, substance P and the antagonists are competitors at the NK-1 receptor but seem to bind differently. Strikingly, one of the antagonist compounds CP96,345, seems to share no interaction point at the receptor with substance P, despite being a competitive antagonist. This again may be explained by the hypothesis of conformational selection, with the antagonists stabilizing an inactive conformation of the receptor and thereby inhibiting the recognition of substance P. The introduction of some mutations interestingly did not affect the

ability of radiolabeled substance P to bind to the mutant receptors but impaired its ability to compete with the radiolabeled non-peptide antagonists.

The combination of receptor mutagenesis and ligand modification helps to elucidate specific interactions between residues of the ligand and residues within the receptor sequence.

5.10.5

Cross-linking

Modification of the ligand and receptor mutagenesis studies are indirect approaches for the analysis of molecular interactions between a receptor and its ligand. It can be difficult to determine whether a change in affinity is caused by an altered conformation of either the receptor or the ligand or whether it is due to changed direct molecular interactions. An alternative method to circumvent this problem is to covalently link a ligand to its receptor after incubation. The interaction site can then be determined. In principle two different proceedings are known for cross-linking. One is the use of an additional bifunctional photoactivatable linker molecule. This has the advantage that the ligand does not have to be modified. However, a limitation of this method is that bifunctional reagents often cross-link the ligand with the receptor at 14–16 Å from the binding site [1].

One way to circumvent this problem is to introduce photoactivatable groups into the ligand itself (Fig. 5.12a). For peptide ligands a number of photoactivatable amino acids are known that upon irradiation with UV light form highly reactive species, for example, a carbene, a nitrene, or a diradical [1]. The binding behavior of the respective photoactivatable analogues of a native ligand of course has to be characterized before performing a cross-linking experiment, e.g., in a radioligand-binding assay [91].

Provided the photoactivatable ligand shows binding behavior similar to the native ligand, it can be used for cross-linking studies. To enable detection of successful cross-linking, the introduction of a second label is favorable [92]. This can be a radioactive isotope, which has the advantage of a very low detection limit, a fluorescent dye, or a group with a defined interaction pattern such as biotin or a histidine-tag. The latter ones have the advantage that they allow the specific purification of the cross-linked receptor-ligand complex [93]. A standard procedure is to incubate the photoactivatable ligand with the receptor as in a standard binding assay [91]. The ligand's photoactivatable group is then activated by irradiation with UV-light and thus is covalently cross-linked to close residues in the receptor sequence [91]. At this point it is of crucial importance to check for the specificity of the cross-linking, as most photoactivatable groups will react with any residue in their close environment [92]. Therefore, displacement assays with unlabeled ligands should be performed in parallel [92]. If the cross-linking was specific, it should be significantly competed by an unlabeled ligand. Because photoaffinity labeling studies are usually performed with whole cells or membrane preparations, purification of the cross-linked complex is necessary before further characterization of the interaction site is possible. This can be done by SDS-PAGE or via spe-

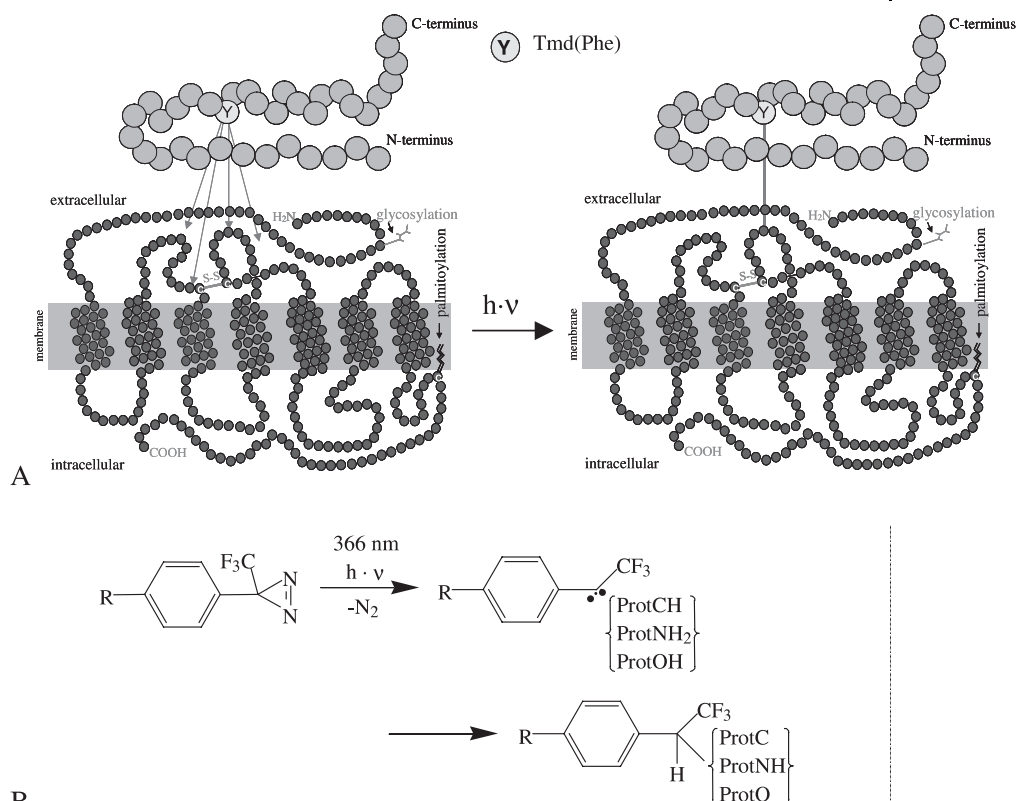


Fig. 5.12 (A) Model of photo-cross-linking of a peptide ligand to its receptor by the photoactivatable amino acid *p*-(3-trifluoromethyl)-diazirino-phenylalanine (Tmd(Phe)). In this example Tmd(Phe) is used to replace its homologue natural amino acid tyrosine. (B) Reaction

scheme of Tmd(Phe). After irradiation with UV light, Tmd(Phe) forms a reactive carbene, which then reacts with residues of the receptor sequence that are closely located in the receptor-ligand complex.

cific interactions of additional labels such as biotin or a histidine-tag. If radio or fluorescent labeling is used for detection, the cross-linked receptors will be directly identified from the gel by phosphorimaging [94]. If labels such as biotin or antibodies are used for detection, the gel will have to be blotted onto a membrane first, followed by staining with enzyme-coupled streptavidin or second antibodies (see Section 5.10.2.1). The corresponding bands on a silver or coomassie stained gel can be identified via their mass. The identified bands of the cross-linked receptors can then be excised from the gel, de-stained, and submitted to in-gel digestion by a proteolytic enzyme with specific cleavage sites, for example, trypsin [95, 96]. Trypsin specifically cleaves after lysine and arginine and therefore leads to defined peptide fragments of the photoaffinity-labeled receptors. Before trypsin digestion, sulfur atoms of the receptor protein can be reduced and alkylated, e.g., with iodoacetamide, to avoid the formation of peptide dimers via disulfide bonds.

Among the digestion fragments should then be a peptide that is cross-linked to the ligand. It can be identified by mass spectrometry (MS) analysis, either of the crude peptide mixture (MALDI-TOF-MS) [95] or after further purification (ESI-MS), [97] or other methods. To facilitate the identification of digestion products, a number of online devices for *in silico* digestion are available (e.g., at URL: www.expasy.org). The experimentally found fragments can then be compared to the theoretical ones. The cross-linked fragment can be identified by its mass, which is the mass of one receptor fragment plus that of the label.

A number of photoactivatable ligands have been used to identify binding sites of ligands at their receptors [92]. The choice of the photoactivatable group is strongly dependent on the ligand into which it is introduced. It should alter the binding properties and the conformation of the ligand as little as possible. Examples of photoactivatable amino acids are *p*-(3-trifluoromethyl)diazirinophenylalanine (Tmd(Phe)) and *p*-benzoylphenylalanine (Bpa), two analogues of the amino acid phenylalanine [98]. They can be used to replace Phe or its homologue Tyr, as was shown in cross-linking experiments with the NPY-Y₂ receptor [91]. Whereas Tmd(Phe) reacts via formation of a carbene, Bpa forms a biradical upon irradiation with UV light (Fig. 5.12b).

Examples for the identification of G-protein-coupled receptor binding sites by photoaffinity labeling are the renal V₂ vasopressin receptor [99] and the human brain cholecystokinin_B receptor [97].

5.11

Conclusion

Despite the lack of crystallographic structural data for all G-protein-coupled receptors except rhodopsin, the information collected from the various experiments described here can be used to create models how GPCRs function and how the interaction between a ligand and its receptor takes place on the molecular level. The processing of these data not only requires careful evaluation by the scientist but also would be impossible without advanced computing technology. In combination, they lead to an evolving process of the formulation of new hypotheses and their experimental proofs. Nevertheless, direct structural determination of the receptor-ligand interaction remains a great aim for the future. The crystallization of rhodopsin was one step in this direction, and most likely other receptors' crystallization will follow. Maybe other techniques for the elucidation of transmembrane receptors' structures will be developed. However, as the interaction between a receptor and its ligand is always a dynamic process, techniques that monitor different aspects of this interaction will retain their importance.

5.12

References

- 1 BECK-SICKINGER, A.G., *Drug Discovery Today*, **1996**, 1, 502–513.
- 2 PALCZEWSKI, K., KUMASAKA, T., HORI, T., BEHNKE, C.A., MOTOSHIMA, H., FOX, B.A., LE TRONG, I., TELLER, D.C., OKADA, T., STENKAMP, R.E., YAMAMOTO, M. and MIYANO, M., *Science*, **2000**, 289, 739–745.
- 3 STRANGE, P.G., *Trends Pharmacol. Sci.*, **2002**, 23, 89–95.
- 4 SMIT, M.J., LEURS, R., ALEWIJNSE, A.E., BLAUW, J., VAN NIEUW AMERONGEN, G.P., VAN DE VREDE, Y., ROOVERS, E. and TIMMERMAN, H., *Proc. Natl. Acad. Sci. USA*, **1996**, 93, 6802–6807.
- 5 HALL, D.A. and STRANGE, P.G., *Br. J. Pharmacol.*, **1997**, 121, 731–736.
- 6 ROSSIER, O., ABUIN, L., FANELLI, F., LEONARDI, A. and COTECCHIA, S., *Mol. Pharmacol.*, **1999**, 56, 858–866.
- 7 CHIDIAC, P., HEBERT, T.E., VALIQUETTE, M., DENNIS, M. and BOUVIER, M., *Mol. Pharmacol.*, **1994** 45, 490–499.
- 8 WESTPHAL, R.S. and SANDERS-BUSH, E., *Mol. Pharmacol.*, **1994**, 46, 937–942.
- 9 MILANO, C.A., ALLEN, L.F., ROCKMAN, H.A., DOLBER, P.C., MCMINN, T.R., CHIEN, K.R., JOHNSON, T.D., BOND, R.A. and LEFKOWITZ, R.J., *Science*, **1994**, 264, 582–586.
- 10 COHEN, D.P., THAW, C.N., VARMA, A., GERSHENGORN, M.C. and NUSSENZVEIG, D.R., *Endocrinology*, **1997**, 138, 1400–1405.
- 11 ARVANITAKIS, L., GERAS-RAAKA, E. and GERSHENGORN, M.C., *Trends Endocrinol. Metabol.*, **1998**, 9, 27–31.
- 12 TURNER, P.R., BAMBINO, T. and NISSENSON, R.A., *Mol. Endocrinol.*, **1996**, 10, 132–139.
- 13 REN, Q., KUROSE, H., LEFKOWITZ, R.J. and COTECCHIA, S., *J. Biol. Chem.*, **1993**, 268, 16483–16487.
- 14 WESTPHAL, R.S., BACKSTROM, J.R. and SANDERS-BUSH, E., *Mol. Pharmacol.*, **1995**, 48, 200–205.
- 15 TUNG, M.F., *Cellular Signalling*, **1996**, 8, 217–224.
- 16 TATEMOTO, K., CARLQUIST, M. and MUTT, V., *Nature*, **1982**, 296, 659–660.
- 17 SAKURAI, T., *Regul. Pept.*, **1999**, 85, 25–30.
- 18 SÖLL, R. and BECK-SICKINGER, A.G., *J. Pept. Sci.*, **2000**, 6, 387–397.
- 19 BIEDLER, J.L., HELSON, L. and SPENGLER, B.A., *Cancer Res.*, **1973**, 33, 2643–2652.
- 20 REYNOLDS, C.P., BIEDLER, J.L., SPENGLER, B.A., REYNOLDS, D.A., ROSS, R.A., FRENKEL, E.P. and SMITH, R.G., *J. Natl. Cancer Inst.*, **1986**, 76, 375–387.
- 21 ALALUF, S., MULVIHILL, E.R. and MCILHINNEY, R.A.J., *J. Neurochem.*, **1995**, 64, 1548–55.
- 22 JORDAN, B.A. and DEVI, L.A., *Nature*, **1999**, 399, 697–700.
- 23 SAITO, Y., NOTHACKER, H.P., WANG, Z., LIN, S.H., LESLIE, F. and CIVELLI, O., *Nature*, **1999**, 400, 265–269.
- 24 DUROCHER, Y., PERRET, S., THIBAudeau, E., GAUMOND, M.-H., KAMEN, A., STOCO, R. and ABRAMOVITZ, M., *Analytical Biochemistry*, **2000**, 284, 316–326.
- 25 GUJER, R., ALDECOA, A., BUEHLMANN, N., LEUTHAEUSER, K., MUFF, R., FISCHER, J.A. and BORN, W., *Biochemistry*, **2001**, 40, 5392–5398.
- 26 MILLIGAN, G., *Br. J. Pharmacol.* **1999**, 128, 501–510.
- 27 FUKUNAGA, K., ISHII, S., ASANO, K., YOKOMIZO, T., SHIOMI, T., SHIMIZU, T. and YAMAGUCHI, K., *J. Biol. Chem.* 276, 43025–43030.
- 28 TARASOVA, N.I., STAUBER, R.H., CHOI, J.K., HUDSON, E.A., CZERWINSKI, G., MILLER, J.L., PAVLAKIS, G.N., MICHEJDA, C.J. and WANK, S.A., *J. Biol. Chem.*, **1997**, 272, 14817–14824.
- 29 MOSER, C., BERNHARDT, G., MICHEL, J., SCHWARZ, H. and BUSCHAUER, A., *Can. J. Physiol. Pharmacol.*, **2000**, 78, 134–142.
- 30 CHO, W., BITTOVA, L. and STAHELIN, R.V., *Anal. Biochem.*, **2001**, 296, 153–161.
- 31 JOENSSON, U., FAEGERSTAM, L., IVARSSON, B., JOHNSON, B., KARLSSON, R., LUNDH, K., LOEFAAS, S., PERSSON, B., ROOS, H., SJOLANDER, S., STAHLBERG, R., STENBERG, E., URBANICZKY, C., MALMQUIST, M., OSTLIN, H. and RONNBERG, I., *Bio-Techniques*, **1991**, 11, 620–627.
- 32 HESS, S.T., HUANG, S., HEIKAL, A.A. and WEBB, W.W., *Biochemistry*, **2002**, 41, 697–705.

- 33 CHENG, Y. and PRUSOFF, W.H., *Biochem. Pharmacol.*, **1973**, 22, 3099–3108.
- 34 LAZARENO, S., *J. Recept. Signal Transduct. Res.*, **2001**, 21, 139–165.
- 35 SCHULZ, R., WEHMEYER, A. and SCHULZ, K., *J. Pharmacol. Exp. Ther.*, **2002**, 300, 376–384.
- 36 RIGLER, R., PRAMANIK, A., JONASSON, P., KRAITZ, G., JANSSON, O.T., NYGREN, P., STAHL, S., EKBERG, K., JOHANSSON, B., UHLEN, S., UHLEN, M., JORNVALL, H. and WAHREN, J., *Proc. Natl. Acad. Sci. USA*, **1999**, 96, 13318–13323.
- 37 PRAMANIK, A., OLSSON, M., LANGE, U., BARTFAI, T. and RIGLER, R., *Biochemistry*, **2001**, 40, 10839–10845.
- 38 INGENHOVEN, N. and BECK-SICKINGER, A.G., *J. Recept. Signal Transduct. Res.*, **1997**, 17, 407–418.
- 39 FABRY, M., CABRELE, C., HOCKER, H. and BECK-SICKINGER, A.G., *Peptides*, **2000**, 21, 1885–1893.
- 40 FABRY, M., LANGER, M., ROTHEN-RUTISHAUSER, B., WUNDERLI-ALLENSPACH, H., HOCKER, H. and BECK-SICKINGER, A.G., *Eur. J. Biochem.*, **2000**, 267, 5631–5637.
- 41 TURCATTI, G., NEMETH, K., EDGERTON, M.D., MESETH, U., TALABOT, F., PEITSCH, M., KNOWLES, J., VOGEL, H. and CHOLLET, A., *J. Biol. Chem.*, **1996**, 271, 19991–19998.
- 42 OVERTON, M.C. and BLUMER, K.J., *Current Biology*, **2000**, 10, 341–344.
- 43 DEVI, L.A., *Trends Pharmacolog. Sci.*, **2000**, 21, 324–326.
- 44 MYSZKA, D.G., *Curr. Opin. Biotechnol.*, **1997**, 8, 50–57.
- 45 STENBERG, E., PERSSON, B., ROOS, H. and URBANICZKY, C., *J. Colloid Interface Sci.*, **1991**, 143, 513–526.
- 46 PLANT, A.L., BRIGHAM-BURKE, M., PETRELLA, E.C. and O'SHANNESSY, D.J., *Anal. Biochem.*, **1995**, 226, 342–348.
- 47 KARLSSON, O.P. and LOFAS, S., *Anal. Biochem.*, **2002**, 300, 132–138.
- 48 FREI, N., WEISSENBERGER, J., BECK-SICKINGER, A.G., HÖFLIGER, M., WEIS, J. and IMBODEN, H., *Regul. Pept.*, **2001**, 101, 149–155.
- 49 ECKARD, C.P., BECK-SICKINGER, A.G. and WIELAND, H.A., *J. Recept. Signal Transduct. Res.*, **1999**, 19, 379–394.
- 50 WIELAND, H.A., ECKARD, C.P., DOODS, H.N. and BECK-SICKINGER, A.G., *Eur. J. Biochem.*, **1998**, 255, 595–603.
- 51 ECKARD, C.P. and BECK-SICKINGER, A.G., *Curr. Med. Chem.*, **2000**, 7, 897–910.
- 52 LEBESGUE, D., WALLUKAT, G., MIJARES, A., GRANIER, C., ARGIBAY, J. and HOEBEKE, J., *Eur. J. Pharmacol.*, **1998**, 348, 123–133.
- 53 LEIBER, D., HARBON, S., GUILLET, J.G., ANDRE, C. and STROSBERG, A.D., *Proc. Natl. Acad. Sci. USA*, **1984**, 81, 4331–4334.
- 54 KUSAKABE, Y., ABE, K., TANEMURA, K., EMORI, Y. and ARAI, S., *Chem. Senses*, **1996**, 21, 335–340.
- 55 DUNBAR, B.S. and SCHWOEBEL, E.D., *Methods Enzymol.*, **1990**, 182, 663–670.
- 56 EY, P.L., PROWSE, S.J. and JENKIN, C.R., *Immunochimistry*, **1978**, 15, 429–436.
- 57 FU, M.L., SCHULZE, W., WALLUKAT, G., ELIES, R., EFTEKHARI, P., HJALMARSON, A. and HOEBEKE, J., *Receptors and Channels*, **1998**, 6, 99–111.
- 58 BENITEZ, R., FERNANDEZ-CAPETILLO, O., LAZARO, E., MATEOS, J.M., OSORIO, A., ELEZGARAI, I., BILBAO, A., LINGENHOEHL, K., VAN DER PUTTEN, H., HAMPSON, D.R., KUHN, R., KNOPFEL, T. and GRANDES, P., *J. Comp. Neurol.*, **2000**, 417, 263–274.
- 59 JAKAB, R.L. and GOLDMAN-RAKIC, P.S., *J. Comp. Neurol.*, **2000**, 417, 337–348.
- 60 INCE, E., CILIAK, B.J. and LEVEY, A.I., *Synapse*, **1997**, 27, 357–366.
- 61 KAMITANI, S. and SAKATA, T., *Biochim. Biophys. Acta*, **2001**, 1539, 131–139.
- 62 BJOERKLOEF, K., LUNDSTROEM, K., ABUIN, L., GREASLEY, P.J. and COTECCHIA, S., *Biochemistry*, **2002**, 41, 4281–4291.
- 63 JIN, H., ZASTAWNY, R., GEORGE, S.R. and O'DOWN, B.F., *Eur. J. Pharmacol.*, **1997**, 324, 109–116.
- 64 FRASER, I.D.C., CONG, M., KIM, J., ROLLINS, E.N., DAAGA, Y., LEFKOWITZ, R.J. and SCOTT, J.D., *Current Biology*, **2000**, 10, 409–412.

- 65 KLUG, S.J. and FAMULOK, M., *Mol. Biol. Rep.*, **1994**, 20, 97–107.
- 66 TUERK, C. and GOLD, L., *Science*, **1990**, 249, 505–510.
- 67 ELLINGTON, A.D. and SZOSTAK, J.W., *Nature*, **1990**, 346, 818–822.
- 68 FAMULOK, M. and MAYER, G., *Curr. Top. Microbiol. Immunol.*, **1999**, 243, 123–136.
- 69 HERMANN, T. and PATEL, D.J., *Science*, **2000**, 287, 820–825.
- 70 PROSKE, D., HÖFLIGER, M., SÖLL, R.M., BECK-SICKINGER, A.G. and FAMULOK, M., *J. Biol. Chem.*, **2002**, 277, 11416–11422.
- 71 FAMULOK, M., BLIND, M. and MAYER, G., *Chem. Biol.*, **2001**, 8, 931–939.
- 72 SCHWARTZ, T.W., *Curr. Opin. Biotechnol.*, **1994**, 5, 434–444.
- 73 YOKOTA, Y., AKAZAWA, C., OHKUBO, H. and NAKANISHI, S., *EMBO J.*, **1992**, 11, 3585–3591.
- 74 MAGGIO, R., VOGEL, Z. and WESS, J., *FEBS Lett.*, **1993**, 319, 195–200.
- 75 MAGGIO, R., VOGEL, Z. and WESS, J., *Proc. Natl. Acad. Sci. USA*, **1993**, 90, 3103–3107.
- 76 LEURS, R., SMIT, M.J., ALEWIJNSE, A.E. and TIMMERMAN, H., *Trends Biochem. Sci.*, **1998**, 23, 418–422.
- 77 PORTER, J.E., HWA, J. and PEREZ, D.M., *J. Biol. Chem.*, **1996**, 271, 28318–28323.
- 78 WALKER, P., MUNOZ, M., MARTINEZ, R. and PEITSCH, M.C., *J. Biol. Chem.*, **1994**, 269, 2863–2869.
- 79 MARSHALL, G.R., *Biopolymers*, **2001**, 60, 246–277.
- 80 MARTINEZ, J., BALI, J.P., RODRIGUEZ, M., CASTRO, B., MAGOUS, R., LAUR, J. and LIGNON, M.F., *J. Med. Chem.*, **1985**, 28, 1874–1879.
- 81 CABRELE, C. and BECK-SICKINGER, A.G., *J. Pept. Sci.*, **2000**, 6, 97–122.
- 82 BARTFAI, T., LANGE, U., BEDECS, K., ANDRELL, S., LAND, T., GREGERSEN, S., AHREN, B., GIROTTI, P., CONSOLO, S., CORWIN, R. et al., *Proc. Natl. Acad. Sci. USA*, **1993**, 90, 11287–11291.
- 83 BECK-SICKINGER, A.G., WIELAND, H.A., WITTNEBEN, H., WILLIM, K.D., RUDOLF, K. and JUNG, G., *Eur. J. Biochem.*, **1994**, 225, 947–958.
- 84 MONKS, S.A., KARAGIANIS, G., HOWLETT, G.J. and NORTON, R.S., *J. Biomol. NMR*, **1996**, 8, 379–390.
- 85 BADER, R., BETTIO, A., BECK-SICKINGER, A.G. and ZERBE, O., *J. Mol. Biol.*, **2001**, 305, 307–329.
- 86 SAUTEL, M., MARTINEZ, R., MUNOZ, M., PEITSCH, M.C., BECK-SICKINGER, A.G. and WALKER, P., *Mol. Cell. Endocrinol.*, **1995**, 112, 215–222.
- 87 RUDOLF, K., EBERLEIN, W., ENGEL, W., WIELAND, H.A., WILLIM, K.D., ENTZEROTH, M., WIENEN, W., BECK-SICKINGER, A.G. and DOODS, H.N., *Eur. J. Pharmacol.*, **1994**, 271, R11–R13.
- 88 DOODS, H.N., WIENEN, W., ENTZEROTH, M., RUDOLF, K., EBERLEIN, W., ENGEL, W. and WIELAND, H.A., *J. Pharmacol. Exp. Ther.*, **1995**, 275, 136–142.
- 89 SAUTEL, M., RUDOLF, K., WITTNEBEN, H., HERZOG, H., MARTINEZ, R., MUNOZ, M., EBERLEIN, W., ENGEL, W., WALKER, P. and BECK-SICKINGER, A.G., *Mol. Pharmacol.*, **1996**, 50, 285–292.
- 90 ROSENKILDE, M.M., CAHIR, M., GETHER, U., HJORTH, S.A. and SCHWARTZ, T.W., *J. Biol. Chem.*, **1994**, 269, 28160–28164.
- 91 INGENHOVEN, N., ECKARD, C.P., GEHLERT, D.R. and BECK-SICKINGER, A.G., *Biochemistry*, **1999**, 38, 6897–6902.
- 92 DORMAN, G. and PRESTWICH, G.D., *Trends Biotechnol.*, **2000**, 18, 64–77.
- 93 BAYER, E.A. and WILCHEK, M., *J. Chromatogr.*, **1990**, 510, 3–11.
- 94 BOLT, M.W. and MAHONEY, P.A., *Anal. Biochem.*, **1997**, 247, 185–192.
- 95 SHEVCHENKO, A., WILM, M., VORM, O. and MANN, M., *Anal. Chem.*, **1996**, 68, 850–858.
- 96 DIHAZI, H., KESSLER, R. and ESCHRICH, K., *Anal. Biochem.*, **2001**, 299, 260–263.
- 97 ANDERS, J., BLUGGEL, M., MEYER, H.E., KUHNE, R., TER LAAK, A.M., KOJRO, E. and FAHRENHOLZ, F., *Biochemistry*, **1999**, 38, 6043–6055.
- 98 WEBER, P.J. and BECK-SICKINGER, A.G., *J. Pept. Res.*, **1997**, 49, 375–383.
- 99 KOJRO, E., EICH, P., GIMPL, G. and FAHRENHOLZ, F., *Biochemistry*, **1993**, 32, 13537–13544.
- 100 RIST, B., ENTZEROTH, M. and BECK-SICKINGER, A.G., *J. Med. Chem.*, **1998**, 41, 117–123.

6

Hydrogen Bonds in Protein-Ligand Complexes

M.A. WILLIAMS, J.E. LADBURY

6.1

Introduction

6.1.1

The Importance of Hydrogen Bonds

Hydrogen bonding plays a significant role in many chemical and biological processes, including ligand binding and enzyme catalysis. Consideration of hydrogen-bonding properties in drug design is important because of their strong influence on specificity of binding, transport, adsorption, distribution, metabolization, and excretion properties of small molecules. Their ubiquity and flexibility make hydrogen bonds the most important physical interaction in systems of biomolecules in aqueous solution. Because hydrogen atoms comprise approximately one-half of the atoms within biological macromolecules and two-thirds of the atoms of the solvating water, hydrogen atoms, or protons, are found between almost every pair of non-covalently bonded heavy atoms in a biological system. Since the basic necessary condition for a hydrogen bond being present is that a proton lies between the electron clouds of two other atoms and modifies their interaction in a manner that is not explicable in terms of the van der Waals (dispersion-repulsion) effect, hydrogen bonds almost rival van der Waals interactions in number. Because van der Waals interactions occur unavoidably and with similar strength between all atoms, their contribution to selectivity of interactions largely lies in the shape selection caused by the repulsive component of the interaction. Consequently, from both an evolutionary and design perspective, modification of local hydrogen-bonding potential is the principal mechanism available for favorably enhancing the interactions between pairs of molecules. The popular notions of “hydrophobic” or “lipophilic” forces being important are merely a result of a non-atomic perspective. The hydrophobic forces, while being a simplifying concept, are a complex compound phenomenon resulting from redistribution and change in strength of water-water hydrogen bonds as solvent is released upon close approach of two apolar chemical groups. This may seem to be a somewhat enthusiastic view of the importance of hydrogen bonding, yet it is a consequence of regarding all forces from an atomic perspective and extrapolates a developing trend

of the past decade, which has seen an increasing diversity of hydrogen-mediated interactions being considered significant in modulating the behavior of biological molecules.

6.1.2

Defining the Hydrogen Bond

The presently accepted definition of a hydrogen bond is due to Pimentel and McClellan's classic 1960 book *The Hydrogen Bond* [1]. A hydrogen bond exists between a donor functional group (D-H) and an atom or group of atoms (A) able to accept the bond, when there is evidence of association between the groups and that this is due to, or enhanced by, the presence of the hydrogen atom already covalently linked to D. This is clearly a rather broad definition, and the evidence for the bond itself can come from a wide variety of sources, e.g., X-ray crystallography, infrared spectroscopy, and calorimetry. Since 1960 a growing number of interactions have been characterized and encompassed by this definition. In the most familiar case of hydrogen bonding, an electronegative donor group (e.g., O, N, S) is considered capable of withdrawing electrons from the proton in a D-H covalent bond, leaving the proton partially de-shielded, yielding a net partial positive charge, and resulting in the possibility of an electrostatic interaction between the proton and another electronegative group. This group has its electron density enhanced through induced polarization that is augmented by charge transfer to the proton. This simple electrostatic view of hydrogen bonding of electronegative atoms is originally due to Pauling [2] and is still the one embodied in almost all computational models of hydrogen bonding in biomolecular systems. However, hydrogen atoms have been subsequently shown to mediate many other interactions. Non-electronegative atoms such as carbon and silicon are known to act as donors, and in addition to the p-orbitals of electronegative atoms, de-localized π -orbitals of unsaturated or aromatic systems and negatively charged ionic species can act as acceptors (Tab. 6.1).

Theoretical efforts toward understanding the hydrogen bond have led to a progressive revision of the simple electrostatic picture from the 1970s onward [3, 4]. It is presently believed that in addition to the electrostatic contribution, there is a similarly sized covalent (quantum-mechanical) component to hydrogen bonds, which lies in the interaction between the empty σ^* anti-bonding orbital of the donated hydrogen and highest occupied molecular orbital of the acceptor, which form a new shared orbital that is the dominant contribution to charge transfer in the interaction. This inherently quantum-mechanical view of the hydrogen bond seems to have received substantial confirmation from the recent experimental observation of peaks corresponding to the $\text{O} \cdots \text{H}$ (1.72 Å) and $\text{O} \cdots \text{O}$ (2.85 Å) distances in the anisotropic inelastic scattering of X-rays from the valence electrons of ice, indicating that electrons are being shared by atoms separated by these distances [5].

Tab. 6.1 Potential hydrogen bond donor and acceptor groups classified according to their strength of interaction.

	<i>Donor^{a)}</i>	<i>Acceptor</i>
Very strong	N ⁺ H ₃ , X ⁺ -H, F-H	COO ⁻ , O ⁻ , N ⁻ , F ⁻
Strong	O-H, N-H, Hal-H	O=C, O-H, N, S=C, F-H, Hal ⁻
Weak	C-H, S-H, P-H, M-H	C=C, Hal-C, π , S-H, M, Hal-M, Hal-H, Se

a) X is any atom, Hal is any of the lighter halogens, and M is a transition metal.

6.2

Physical Character of Hydrogen Bonds

6.2.1

Crystallographic Studies of Hydrogen Bonds

High-resolution X-ray and neutron crystallography have provided a great deal of information on the geometry of hydrogen bonds in both small molecule and protein crystals. X-rays or neutrons are scattered by collision with atomic electrons or nuclei, respectively, and in the ordered environment of a crystal produce diffraction patterns that can be interpreted to provide atomic positions. Neutron diffraction data are preferred for the study of hydrogen bonding, as the uncharged neutrons are scattered by direct collision with a proton (or deuteron) almost as efficiently as from heavier atoms. This is in contrast to the scattering of X-rays, which is very weak from the low electron density of hydrogen, rendering it invisible in all except the very highest-resolution (better than 1 Å) X-ray structures. However, neutron data have been historically much more difficult to obtain, and only 0.4% of small molecule and 0.05% of protein structures to date have been derived from neutron data. Advances in neutron production and detection technology and sample preparation promise to alleviate these experimental restraints, and many more neutron structures are expected in the future [6]. An example of the potential of the technique is the recent neutron diffraction structure of myoglobin [7], which shows in rich detail the extensive hydrogen bonding, the protonation states of histidines, and the orientation of water molecules in and around the heme-binding site. However, it will be some time before many such protein-ligand complexes are available for study.

X-ray crystal structures numerically dominate the available molecular structural data, and our knowledge of hydrogen bonding geometry, particularly in proteins, is largely derived from them. Because the hydrogen atom is usually invisible in protein X-ray structures, in the vast majority of cases the positions of hydrogen atoms and the presence of a hydrogen bond is inferred from the proximity and expected covalent geometry of donor and acceptor groups. Inherent errors in uncertainty of the donor and acceptor positions are also often larger than with neutron data, and biases in structure refinement procedures can lead to further uncer-

tainty in local geometry. Clearly, without accurate atomic positions or even evidence of the presence of a hydrogen atom, in some cases the inferred hydrogen bonding may be erroneous and could lead to errors of interpretation.

Small molecule X-ray crystal structures are usually determined by direct methods using little knowledge of the chemistry of the molecule other than its chemical composition and atom valances. Thus, the accuracy of the positions of heavy atoms depends only on the quality of the experimental data. Consequently, analysis of small molecule crystal structures provides unbiased information on both the relative orientation and separation of donor and acceptor groups (with errors in position typically 0.1 Å). However, the location of the hydrogen atom is not well determined even at very high resolution, as the electron density is not centered on the hydrogen atom but in the D-H bond. D-H bond lengths must be “normalized,” i.e., made equal to the average observed in chemically similar neutron-determined structures. This may produce small errors (<0.05 Å), as the presence and nature of a neighboring H-bond acceptor will affect the actual D-H bond length, but such fine detail is not usually significant in comparison to other errors in structures.

In the case of protein crystallography, considerably greater use is made of prior knowledge, of average covalent geometry, of torsion angle preferences, and of van der Waals radii of atoms in building a model whose properties are then compared to, and refined against, the experimental diffraction data until the agreement between the two is satisfactory. Consequently, some biases from the parameters representing the covalent geometry and non-bonded interactions are found in X-ray protein structures. The intrinsic uncertainty in the position of atoms in protein crystal structures is also usually greater and depends on whether the atom is part of the polypeptide backbone or in an amino acid side chain, and whether it is found in the interior or on the surface of the protein, where dynamic averaging of the position of atoms can be substantial. The average error in a set of coordinates derived from a 2.0–2.8 Å resolution electron density map is about 0.2–0.4 Å. Thus, for an interaction between a donor and acceptor, the error in measurement of a hydrogen bond distance can easily be 0.5 Å. Consequently, in some cases it can be very difficult to determine whether or not a hydrogen bond is present.

6.2.2

The Geometry of Hydrogen Bonds

Individual uncertainties in local structure are overcome in statistical surveys of large numbers of small molecule or protein structures, which are able to give a consistent and reliable picture of hydrogen bonding. Superposition of a particular donor-acceptor pair from many structures gives an anisotropic three-dimensional distribution of observed geometries. Differences between such distributions and those expected for a van der Waals interaction (more dispersed and isotropic) are diagnostic of the hydrogen-bonding capability of the selected groups where this may otherwise be in doubt. This is particularly true of weak hydrogen bond donors and acceptors, where deviation from van der Waals behavior may only be ascertained by superposition of large numbers of structures. The sheer weight of

numbers means that the X-ray data are more informative, despite the poor visibility of hydrogen, and no significant differences from analysis of the much smaller number of neutron structures have been observed [8, 9]. Since hydrogen bond restraints are not used directly in structure refinement, even protein structures at relatively low resolution reveal unbiased information on the relative frequency of occurrence of particular types of hydrogen bond and orientation information on donor and acceptor groups (although distance information is less reliable due to the van der Waals parameters in the refinement force fields). Overall, studies of small molecule structures have greater accuracy and generality, which is particularly useful in thinking about a wide variety of potential ligands [10], and studies of protein structures give more specific information on amino acids and the significance of a particular type of hydrogen-bond interaction among all the competing interactions in proteins.

The classic survey of hydrogen bonding in proteins was conducted by Baker and Hubbard [11]. This highlighted the importance of hydrogen bonds in forming not only helices and sheets but also the structure of loops, in binding water molecules, in recognition processes, and in catalysis. The most common bond in biological chemistry is the $\text{C}=\text{O} \cdots \text{H}-\text{N}$ peptide backbone bond, which is usually substantially shorter than the van der Waals distance for O and N and is easily identified even in relatively low-resolution structures. Indeed, in the great majority of biological hydrogen bonds, the donating and accepting groups are either N or O atoms; however, weaker interactions including $\text{C}-\text{H} \cdots \text{O}$ or N, $\text{S}-\text{H} \cdots \text{O}$ and $\text{N}-\text{H} \cdots \pi$ electrons are observed [9]. In protein structures there is a systematic tendency toward a greater number of observed hydrogen bonds with increasing resolution. At 3.0 Å resolution, 85% of C=O acceptors and 75% of N-H donors form at least one hydrogen bond according to standard geometric criteria. However, at 1.5 Å resolution more than 95% of C=O and 90% of N-H are satisfied according to the same criteria, and virtually all are satisfied if weaker criteria are used that leave a greater margin for positional errors [12]. This means that inaccuracies in the vast majority of protein structures cause underestimation of the number of hydrogen bonds present. It is important to be aware of this in considering any particular protein-ligand structure. Indeed, there is a good case for forcing satisfaction of the weak criteria for classic hydrogen bond donors and acceptors in structure refinement and/or modeling studies of proteins, although this is not routinely done.

Statistical analyses of both proteins and small molecules reveal that hydrogen bond stereochemistry is influenced by two major factors: the electronic configuration of the acceptors and the steric accessibility of the acceptors and donors. Considering the most common hydrogen bond in biological chemistry, there is a distinct preference for $\text{N}-\text{H} \cdots \text{O}=\text{C}$ bonds to form in the $\text{O}=\text{C}-\text{RR}'$ plane and in the directions of the conventionally viewed sp^2 lone pairs [13], with the proton lying within 30° of the plane and at $30\text{--}60^\circ$ to the $\text{O}=\text{C}$ axis in the majority of cases. Note that in relation to the simple model of a hydrogen bond as an electrostatic dipole-dipole interaction, which is embodied in most modeling software, the D-H group does not lie along the dipole but points to the lone pairs. Note also that there is no absolute requirement for linearity of the acceptor-proton-donor system,

as the σ^* anti-bonding orbital of the H-D, which is responsible for the covalent/charge transfer component of the bond, is a spherically symmetric orbital and imposes no directionality. The principal contributions to the observed directional preferences are the total electrostatic fields (accurately modeled as higher-order multipoles [14]) and steric effects of donor, acceptor, and adjacent groups.

The effect of electronic configuration of the acceptor is particularly apparent in the contrast of Ser/Thr and Tyr hydroxyl groups; the phenolic hydroxyl of tyrosine has a preference for near-plane position for donors and/or acceptors, as its sp^2 hybridization leaves the lone pair electrons in the plane of the ring [15, 16], whereas serine and threonine hydroxyls have sp^3 hybridization with two acceptor and one donor position at 120° spacing (the donating proton is usually *trans* to the carbon three covalent bonds away). The observed spatial distributions for the principal amino acid donor and acceptor groups are illustrated schematically in Fig. 6.1.

Not all of the hydrogen-bonding potential of a particular amino acid is necessarily fulfilled. Although virtually all strong hydrogen-bonding groups form at least one hydrogen bond, and the very strong charged groups usually use all their capacity, R-O-H, R-S-H, C=O and glutamate COO^- form only one hydrogen bond [12]. It seems either that once the first interaction is made the second is less favorable energetically or that crowding makes it difficult to surround a group with the full complement of partners. The O-H and S-H groups tend either to donate or accept, and the main-chain carboxyl and glutamate COO^- acceptors tend to be bifurcated. A bifurcated hydrogen bond occurs when a donated proton is close to two acceptor sites or a single acceptor is close to two donors. Analysis of high-resolution protein X-ray crystal structures shows that a quarter of all main-chain N-H donors are bifurcated, i.e., they make more than the expected number of hydrogen bonds. Bifurcated donors occur systematically in both alpha-helices (most hydrogen bonds are of this type; the dominant component is $\text{N-H} \cdots \text{O}=\text{C}_{(i-4)}$, but longer distance $\text{N-H} \cdots \text{O}=\text{C}_{(i-3)}$, are also present) and in a substantial minority of β -sheet N-H groups [17]. The possibility of such bifurcation should not be neglected in protein-ligand structural studies, e.g., making one hydrogen bond to a bidentate acceptor may gain most of the binding energy.

The above-mentioned hydrogen bond surveys considered only the classical strong hydrogen bond donors and acceptors O, N, and S, but recently the importance of weaker hydrogen bond groups has come to be more widely recognized. Weak-strong and weak-very-strong pairs of donor and acceptor are sufficiently energetically favored to compete with other biomolecular interactions and have an impact on protein structure and function.

It has long been known that π -orbitals of aromatic ring systems can act as hydrogen bond acceptors (see the monumental review of weak hydrogen bonding by Desiraju and Steiner [18] for a detailed history), but it was not until the mid-1980s that they were unambiguously observed in several very high-resolution neutron and X-ray structures, including the observation by Perutz et al. [19] of a hydrogen bond from an asparagine side chain N-H to an aromatic ring of the anti-sickling agent bezafibrate bound to hemoglobin. This complex also contained several other weak hydrogen bonds involving C-H groups that help determine drug binding.

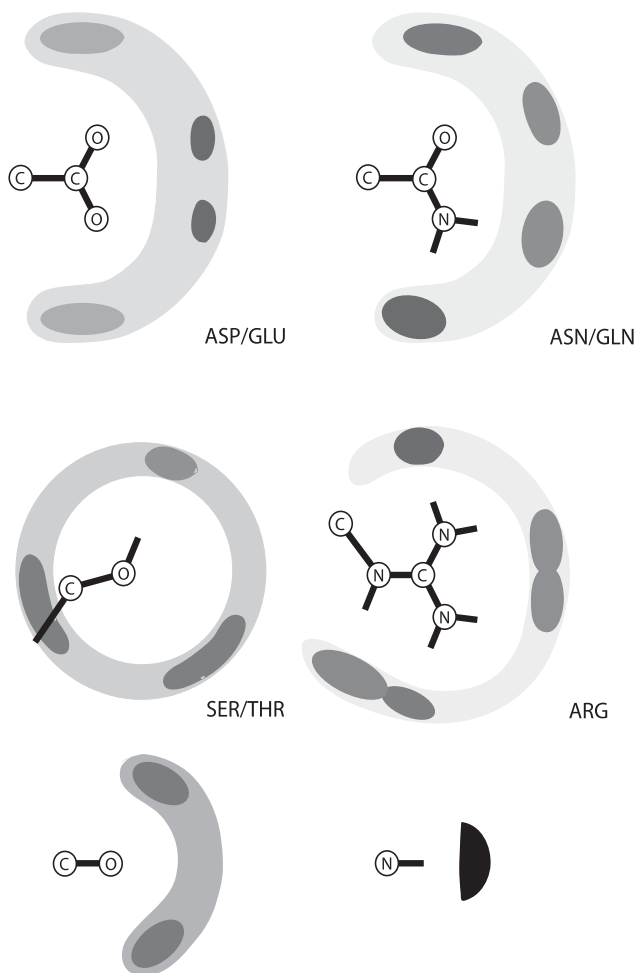


Fig. 6.1 Schematic illustration of the observed spatial distributions of hydrogen bond partners for several amino acids [13, 16]. Darker shading indicates an increased likelihood of finding a hydrogen bond partner at a particular position. Most groups have maximum likelihood of having their partner located in

the plane of R'RD or R'RA; only Ser/Thr/Cys/Lys have maxima out of plane. The Cys and Lys distributions are similar to Ser/Thr but with a greater and shorter D-A distance, respectively. The N-H of the amide/His/Trp are also similar, and the O-H group of tyrosine has a distribution similar to C=O.

The significance of these observations was not immediately appreciated, but a few further important examples, including the observation of several amino hydrogen bonds to the ring of the phosphotyrosine in its recognition by SH2 domains, as shown in Fig. 6.2 [20], and a later review by Perutz [21] stimulated both structural analyses and theoretical studies of D-H $\cdots\pi$ -acceptor systems in proteins [22, 23]. It has been shown that such bonds are relatively rare compared to the usual

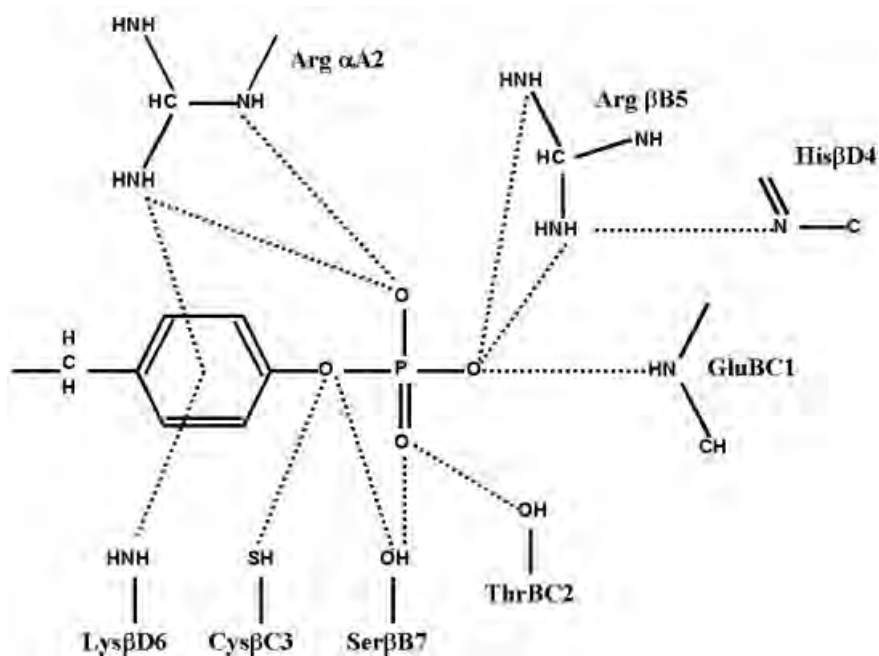


Fig. 6.2 Schematic of the observed hydrogen bonding to phosphotyrosine in a complex with the *v-src* SH2 domain [20]. Both very

strong hydrogen bonds to the phosphate and weak bonds to the aromatic ring are important to binding.

strong bonds. However, numerous examples are found where water and peptide $D-H \cdots \pi$ interactions are functional in stabilization of helix termini, strand ends, strand edges, beta-bulges, and regular turns. Side chain $D-H \cdots \pi$ hydrogen bonds are also formed in considerable numbers in α -helices and β -sheets. A recent survey [24] identified approximately 1 such weak bond for approximately every 11 aromatic residues (increasing to 1 per 6 residues for tryptophan because of its double-ring structure) in protein structures. There has been no systematic survey of such hydrogen bonds in protein-ligand complexes, but because many drugs contain a high proportion of atoms in ring structures, it is expected that there are many unrecognized cases in which they are important.

Although the peptide C_α -H group has historically not been thought to form hydrogen bonds within proteins, recent experimental evidence and *ab initio* quantum calculations show it to be an effective proton donor, as it is activated by the neighboring peptide groups. Experimentally, C_α -H \cdots O contacts observed in neutron crystal-structure determinations of amino acids show the lengthening of the C_α -H bond with decreasing H \cdots O separation that is characteristic of hydrogen bonds [25]. Its binding energy to a water molecule has been calculated to lie in the range between 7.9 and 10.5 kJ mol⁻¹, comparable to the interactions between water molecules themselves, and a hydrogen bond to a charged lysine residue is

significantly stronger than a conventional O-H...O interaction [26]. Both experimental and theoretical equilibrium C_α...O bond lengths are about 3.3 Å, somewhat longer than the 2.7–3.0 Å of strong-donor/strong-acceptor pairs.

6.2.3

Infrared Spectroscopy of Hydrogen Bonds

The formation of a hydrogen bond changes the electron distribution in donor and acceptor groups, resulting in changes in the depth and shape of the potential wells corresponding to the covalent bonds in system. The changes in the covalent bonds are minor structurally and difficult to detect, as discussed above, but are readily observable via the change in the vibrational (infrared) spectrum of the system. Infrared spectroscopy is particularly useful in the study of homogeneous liquids and solids or dilute solutions of small molecules, where the simplicity of the system allows peaks or bands in the spectrum to be assigned to particular bonds via theoretical methods. Shifts in the D-H stretching frequency are often the principal evidence of the reality of a group's hydrogen-bonding ability, particularly in weak bonding cases and where insufficient statistical evidence is available from structural databases. Such shifts are also diagnostic of whether or not a hydrogen bond exists in an individual case, which can be difficult to decide on the basis of a single structure. Infrared spectroscopy is the primary method for monitoring the extent of hydrogen bond formation of particular groups in dilute solution in an inert solvent, the concentration and temperature dependence of which allows determination of the intrinsic thermodynamics of the formation of particular hydrogen bonds. The degree of change in a vibrational frequency can also be empirically related to hydrogen bond length and the energy of the bond [27, 28].

With these characteristics, infrared spectroscopy sounds like a highly appropriate technique for definitive characterization of individual hydrogen bonds in protein-ligand structures. However, the difficulty of assigning a given infrared band to a particular hydrogen bond in a spectroscopically crowded heterogeneous biomolecular system has severely limited its use so far. Novel infrared techniques for spectral simplification such as ultraviolet resonance Raman spectroscopy, which selects polarizable aromatic groups [29], and Raman optical activity, which selects chiral centers [30], may see an increased use in understanding protein-ligand interactions in the future. For now, absorption spectroscopy remains a tool for fundamental, not protein-specific, investigation. In particular, it has provided considerable insight into apolar hydration processes via monitoring the signals from water (see Section 6.3.1).

6.2.4

NMR Studies of Hydrogen Bonds

Because of the ability to identify all individual C, N, and H atoms in the NMR spectra of proteins via isotopic labeling and multi-dimensional NMR techniques [31], NMR has become the key technique for investigation of biomolecular struc-

ture in solution. Several NMR spectral features are affected by hydrogen bonding, allowing hydrogen bonds to at least be identified and in some cases more accurately characterized than in high-resolution crystal structures. First, the chemical shift of a hydrogen-bonded proton is usually higher than an otherwise equivalent non-hydrogen bonded one, as the proton is de-shielded when it is withdrawn from the donor group. This can cause shifts of up to 6 ppm in the case of a very strong bond or bonds (e.g., a salt bridge). This effect is useful in comparative studies (i.e., with and without ligand) for the detection of hydrogen-bonding interactions on a ligand or protein. Because the H shift depends on many other conformational and environmental factors, however, it can be difficult to definitely ascribe the cause of a particular shift in the absence of a structure. If a series of closely related ligands is available (i.e., with and without the donors or acceptors in particular positions), comparison of spectra may determine which specific ligand and protein groups are interacting. Interpretational difficulties are, however, exacerbated by changes in solvent structure upon ligand binding. The change being measured is not simply due to the creation, or not, of a hydrogen bond but rather to the exchange of a hydrogen bond to water for the interaction with the ligand. This means that shifts in either direction can occur, reflecting changes in hydrogen bond strength. However, when a structure is available or in other cases where the binding partners are definitely identified, the hydrogen bond interaction and particularly its local energetic environment can be investigated in great detail [32]. Titratable groups (pK_a near 7) are often of considerable importance to protein function and ligand binding. In particular, change of ionization state upon binding can have profound effects on the binding constant. NMR is the standard method of determining pK_a of most groups in biomolecular complexes, by direct observation of changes in the apparent proton population or indirect observation of the donor atom chemical shift with pH [33]. Additional information on bond lengths (and energies) can be found from the quantitation of populations of protonated and deuterated hydrogen bonds in protein-ligand system in mixed H_2O/D_2O solvent [34].

Recently the direct coupling of nuclear magnetic energy levels across a hydrogen bond via their shared electrons has been detected as splitting of the NMR peaks, so-called 2J (proton-acceptor) and 3J (donor-acceptor) couplings. Such coupling is expected given the experimental observation and theoretical expectation of electron sharing (Section 6.1.2). A single coupling experiment allows direct identification of the hydrogen bonding partners (D, H, A), provided that they are isotopically (^{13}C , ^{15}N) labeled, and gives a very accurate determination of hydrogen bond length and angle based on a correlation with neutron structural data [35]. The requirement for isotope labeling and the general low sensitivity of the experiments have restricted their use so far to monitoring ligand-induced changes to the protein's own hydrogen-bonding patterns [36], but with increased awareness of their utility and the resources to make labeled ligands, they could become a mainstay of future protein-ligand hydrogen-bonding studies.

6.2.5

Thermodynamics of Hydrogen Bonding

Observed crystal geometries represent a compromise between many competing forces, and the 3-D distribution observed for a particular donor-acceptor pair roughly (and through the blur of the positional uncertainties discussed in Section 6.2.1) represents a Boltzmann distribution on the potential energy surface of the hydrogen bond interaction. These distributions are generally rather broad (see Fig. 6.1), implying a shallow energy minimum or minima with depths of the same order of magnitude as thermal energies. Measurements of the energetics of individual hydrogen bonds can be made calorimetrically or spectroscopically in a dilute solution of the hydrogen-bonding moieties in an otherwise inert medium. Such experimental systems do not replicate the polarizing environment found in biological systems, but do allow comparison of the strengths of different types of hydrogen bonds and assessment of the ability of quantum theoretical methods for the prediction of those strengths. The success of theoretical methods in this simplified chemical milieu underpins confidence in their use in understanding the greater complexity of the situations found in biological systems.

The diverse nature of the geometric and electronic character of the hydrogen bond is reflected in the range of its reported energetic value (Tab. 6.2). Intrinsic hydrogen bond strength differs by more than two orders of magnitude, depending upon the interacting partners. Even when the donor and acceptor atoms are the same, the dependence of hydrogen bonding on local electron distributions creates a variation of a factor of 10 for $\text{N-H}\cdots\text{O}=\text{C}$, depending on the nature of the substituents attached to donor and acceptor (see [37] and references therein for a summary of substituent effects on donor and acceptor ability in drug-like molecules).

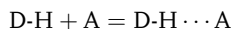
Tab. 6.2 Examples of intrinsic hydrogen bond enthalpies measured for dilute solutions in inert solvents.

Type	Example	$-\Delta H$ (kJ mol ⁻¹)	Reference
Very strong...very strong	$\text{F-H}\cdots\text{F}^-$	163	18
Strong...very strong	$\text{H-O-H}\cdots(\text{OH})^-$	96	18
Strong...strong	$\text{H-O-H}\cdots\text{OH}_2$	21	18
	$\text{R'RN-H}\cdots\text{O}=\text{CR}_1\text{R}_2$	14.5–18	28
	$\text{R-O-H}\cdots\text{O}=\text{CR}_1\text{R}_2$	14–32	28
	$\text{R-O-H}\cdots\text{SR}_1\text{R}_2$	17.5	28
Weak...strong	$\text{R'RN-H}\cdots\pi$	4.5–16	21, 38
	$\text{RR'C-H}\cdots\text{OH}_2$	9.2	18
Weak...weak	$\text{RR'C-H}\cdots\pi$	5.8	18
	$\text{H-S-H}\cdots\text{SH}_2$	4.5	18
	$\text{CH}_4\cdots\text{SH}_2$	1.7	18

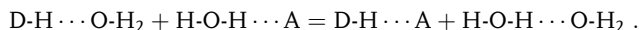
6.2.6

Experimental Thermodynamics of Biomolecular Hydrogen Bonds

Numerous studies have been performed to quantify the energetic contribution of hydrogen bonding in protein systems. Despite this significant experimental effort, many apparent discrepancies remain between studies on different systems. The main difficulty is that of isolating the contribution to binding from one or several defined hydrogen bonds. In aqueous solvent, hydrogen bond formation between the interacting moieties involves first breaking similar bonds with solvent water. This increases the complexity of the deconvolution of the components of the interaction that is required to give the energetic value for the formation of the protein-ligand hydrogen bond. In other words, instead of looking at the interaction



the actual event includes



Rather than measuring the formation of one hydrogen bond in isolation, the experiment measures a rearrangement of already existing hydrogen bonds. This can have some counterintuitive consequences, e.g., the S-H group is regarded as almost as effective a contributor to protein-ligand interactions as O-H, since the energetic difference between being hydrated and forming an interfacial hydrogen bond is thought to be similar for both groups [39].

The most amenable way to measure the thermodynamic effect of hydrogen bonding in protein systems is to measure the differences between wild-type protein and a mutant form in which a residue is substituted, or a subtle chemical modification of a ligand is made, in order to remove selected hydrogen-bonding moieties. Interpretation of these data are complicated by the fact that hydrogen bonds affect, and are affected by, their local chemical environment, particularly in such adaptable and flexible entities as proteins. This causes an inherent difficulty in measuring the thermodynamic parameters for the contribution of a hydrogen bond to protein stability or protein-ligand interaction. Any experiment (e.g., mutagenesis) that creates a change in the disposition 1) of chemical groups surrounding the potential hydrogen bond site, 2) of the solvent, or 3) of the propensity of protein structures to conformationally adapt to binding, may affect the thermodynamic value obtained [40]. In the absence of structural information on each mutant, or modification, interpretation of such thermodynamic experiments is risky. As an alternative to these complexities, many studies have been carried out on model dilute solutions in media that supposedly mimic properties of the physiochemical environments relevant to protein-protein and protein-ligand interactions.

To emphasize the variation in reported experimental quantification of hydrogen bonds, we list several studies of the energetics (enthalpy or free energy) of hydrogen bond formation in biological interactions and related model systems. Early

work by Schellman [41] on urea solutions arrived at a value of -6.3 kJ mol^{-1} for the ΔH of formation of an amide hydrogen bond (i.e., a hydrogen bond between the N-H and C=O groups from amide linkages) in aqueous solution. Subsequently, the 1962 study of Klotz and Franzen [42] using dimerization of *N*-methylacetamide in aqueous solution suggested that the value of ΔH for an amide hydrogen bond in water was near zero and that the overall free energy was unfavorable. However, later studies on δ -valerolactam [43, 44] calculated values of between -8 and -13 kJ mol^{-1} for the formation of a hydrogen bond in water. Scholtz et al. [45] also arrived at a favorable value for the ΔH of formation of a hydrogen bond in water. Each of these studies is regularly quoted as the typical strong...strong hydrogen bond energy for a solvent-exposed situation.

Fersht et al., in a very well-known study of the binding of tyrosyl-tRNA synthetase to its substrate [39], found that deletion of a strong hydrogen bond donor or acceptor from the enzyme reduced the free energy of binding by only 2.1 – 6.3 kJ mol^{-1} , whereas removal of the partner of a very strong donor or acceptor weakened binding by an additional 12 kJ mol^{-1} . Data on mutants of ribonuclease T1 by Shirley et al. [46] similarly showed that a buried strong...strong bond contributed -5.4 kJ mol^{-1} to conformational stability in the folded form. Here a different physical process is operating from that in the physiochemical studies of hydrogen bond formation in water because the groups are buried upon ligand binding or folding. In addition to loss of hydrogen bonds, buried groups lose long-range interactions with the surrounding water and gain long-range interactions with the protein. Several groups have tried to create physiochemical models of the burial process that allow partitioning of the two processes (bond formation and burial). The ΔG for the transfer of a hydrogen-bonded peptide group from water to octanol (taken as a model of the protein interior) was determined as being about 4.6 kJ mol^{-1} , i.e., desolvation of the hydrogen-bonded pair is unfavorable [47]. In conjunction with the above mutagenesis data, this implies that the ΔG contribution made by the hydrogen bond formed in water is favorable but is made less so by burial. However, it was suggested on the basis of model systems that the associated unfavorable ΔH of dehydrating polar groups on burial might result in an overall unfavorable ΔH for the burial of hydrogen-bonded polar groups [48, 49]. This is a contentious hypothesis, as dissolution of crystalline cyclic dipeptides [50] demonstrates that amide-hydroxyl hydrogen bonds give a favorable contribution to interaction from both $\Delta H = -7.1 \pm 1.3 \text{ kJ mol}^{-1}$ and $\Delta G = -2.4 \pm 0.8 \text{ kJ mol}^{-1}$. It is apparent from the above data that despite the many attempts to quantify the ΔH and ΔG for the formation of hydrogen bonds in biological systems, there is no convergence to a single value. This is not surprising, since the strength of hydrogen bonds depends on the local environment, which itself can be highly variable in biological systems. More data on ΔH , ΔG , and structural changes in series of similar complexes, categorized by type of hydrogen-bonding partners and local environmental change, are needed in order to provide a set of case histories from which it may be possible to predict the effect of a proposed change by similarity to a previous case.

An important feature of all these observed energy changes resulting from rearrangement of hydrogen bonds is that they are of the same order of magnitude as

the thermal energies at room (or body) temperature ($RT=2.48 \text{ kJ mol}^{-1}$ at 298 K). The biological importance of hydrogen bonds lies not only in their specific structuring geometry but also in the possibility of rapid reorganization of hydrogen bonds (and consequent structural change) in response to environmental changes, such as ligand binding, assisting specificity of recognition.

6.3

Interactions with Water

6.3.1

Bulk and Surface Water Molecules

Thinking about the hydration of protein complexes is simplified by dividing water molecules into four classes: bulk water molecules that are not directly in contact with the biomolecules, surface water hydrogen bonded to the protein or ligand, surface water associated with apolar biomolecular groups, and buried water molecules that have no direct connection to the bulk solvent.

Water molecules have a strong tendency to interact with each other and consequently cause the association of compounds with which they cannot interact as strongly. A bulk water molecule makes between four and five hydrogen bonds with neighboring water molecules, each contributing approximately -10 kJ mol^{-1} to the energy of bulk water. Bulk water molecules have another important property – they are easily reoriented in response to an electrostatic field. This reorientation serves to attenuate the attractive links between charged groups, which form the strongest hydrogen bonds found in protein-ligand systems. Such strong interactions could otherwise impose unacceptable rigidity on the molecules.

Surface water molecules are distributed over the entire surface of the interacting molecules, are not usually particularly restricted in their motion, and exchange with the bulk solvent on a time scale of 10–300 picoseconds. Water molecules hydrogen bonded to surface polar groups are generally thought to have energetic behavior similar to bulk water, with subtle differences dependent on the strength of their hydrogen-bonding partner. A significant minority of external waters make more than one hydrogen bond with the macromolecular surface, and those water molecules hydrogen bonded to the main chain C=O and N-H groups generally appear at better-defined sites than those bound to amino acid side chains [51]. These structural observations probably reflect underlying thermodynamic preferences for particular polar sites to be hydrated. In ligand binding, water molecules at particularly favorable sites are usually retained or replaced by strong hydrogen-bonding groups of the ligand in order to enhance binding affinity.

Water at apolar surfaces has rather distinctive thermodynamic properties, in particular an unusually low entropy and high heat capacity. Displacement of water from apolar surfaces to bulk during protein folding and ligand binding dominates observed heat-capacity changes of the whole system. Consequently, apolar surface hydration has received much more attention than that of polar surfaces. Water

molecules associated with exposed apolar groups make only relatively weak direct interactions; the discontinuity in hydrogen-bonding options presented at the apolar surface means that they tend to become organized. Many biochemistry textbooks propagate the view that in order to satisfy their hydrogen bond potential, waters become ordered in rigid cages around the apolar groups in a similar manner to clathrate structures found in crystalline hydrates. However, this picture is being progressively challenged by recent studies in solution. Neutron diffraction studies of aqueous solutions imply that the layer of water around apolar groups is much less well organized and much more dynamic than the clathrate model would suggest [52]. Total internal reflection vibrational sum frequency spectroscopy (an infrared technique that selectively probes molecules at boundaries) has revealed that at an apolar surface, hydrogen bonding between adjacent water molecules is weakened relative to those in bulk water. It also shows that a substantial number of hydrogen bonds are lost (i.e., the H points into the surface) and confirms that this anisotropy of the environment results in substantial orientation of the waters at the interface [53]. This general weakening of hydrogen bonding at apolar surfaces is almost diametrically opposed to the clathrate picture of a cage of water surrounding hydrophobic groups. However, it makes sense of the observed, yet otherwise anomalous, increase in heat capacity of protein systems upon unfolding, i.e., as more apolar groups are exposed to solvent, more hydrogen bonds are weakened, the vibrational energy level spacing decreases, and the heat capacity increases [54].

Surface waters are displaced on formation of the protein-ligand complex and thus provide a favorable entropic contribution to the free energy of complex formation. In particular, displacement of the water found interacting on apolar surfaces makes a large contribution to the ΔG and provides the driving force for many interactions (the hydrophobic effect).

6.3.2

Buried Water Molecules

High-resolution crystal structures have revealed the presence of water molecules in many cavities within proteins [55] and in the interfaces between proteins and ligands [56, 57]. Such water molecules usually make at least one strong hydrogen bond and in many cases fulfill their total hydrogen-bonding capacity. In cases where the water hydrogen bond capacity is not fulfilled by strongly interacting partners, waters are systematically found to take part in weaker interactions with C_{α} -H groups [58]. The functional role of buried water molecules is often difficult to ascertain, but in some cases their positioning is highly specific and is crucial to binding or function (this is seen particularly in catalytic active sites). On the other hand, water molecules can act as a rather general mechanism to extend protein structure and/or to increase the promiscuity of the binding site. They are able to help accommodate polar hydrophilic side chain groups in what is often a largely hydrophobic interface. They can overcome steric problems of pairing of donor and acceptor groups at the interface. Networks of these waters can provide flexibility

in recognition, and the additional hydrogen bonds formed may contribute toward the stability of the overall interaction.

Since the buried water molecules usually form extensive interactions with the relatively rigid protein, they have greater difficulties in rearranging their hydrogen bonding than do surface water molecules. Consequently, they are usually kinetically trapped in the interface and have many orders of magnitude lower mobility than at the surface. There has been a widely held view that the entropic cost associated with this entrapment means that such buried water molecules are usually unfavorable to the overall free energy of binding and are found merely as a byproduct of the non-optimal shape complementarity of ligand and protein. As a result, currently available drug-design programs work on dehydrated binding surfaces, assuming that liberating any water molecules that may have been visualized in the structure of the target and substrate (or cognate ligand) will provide a favorable contribution to ΔG . This general approach has proven successful in several cases. Perhaps the best described of these is that of inhibitors to the HIV protease [59].

However, recent observations have suggested that this generally accepted treatment of interfacial water molecules is not always appropriate. For example, the binding of tripeptides with the general sequence of LysXxxLys to the protein OppA results in the entrapment of different numbers of water molecules in an isolated cavity depending on the side chain on the amino acid represented by Xxx [60, 61]. For example, if Xxx=Trp, the high-resolution X-ray structure reveals three water molecules that form a network hydrogen bonded into the cavity. Substituting Trp with Ala results in seven immobilized water molecules being found in the cavity. Although this would widely be expected to result in a less favorable ΔG , the LysAlaLys peptide is in fact seen to bind significantly more tightly [60]. This has been rationalized by considering that although an entropic penalty is paid for the inclusion of additional water molecules, they are able to make optimal hydrogen-bonding arrangements in the interface, producing a sufficiently favorable ΔH as to overcome this. This view is supported by inspection of the crystal structures for OppA-LysXxxLys complexes, which have been solved for the cases where Xxx is any one of the 20 naturally occurring and for a few non-natural amino acids [61, 62]. In these structures, a subset of water molecules is seen to adopt the same position no matter which residue is in the Xxx position. Furthermore, on assessing the potential energy surface of the binding cavity, it appears that the water molecules adopt the positions of lowest enthalpy.

These data and other recent studies (e.g., the extensive water-mediated hydrogen bonding in SH2 domain recognition of its specific peptide [57, 63], shown in Fig. 6.3) show that water molecules have to be incorporated into any rigorous ligand-design program. Effective ways of doing this have yet to be achieved and form a new frontier in the drug-development process.

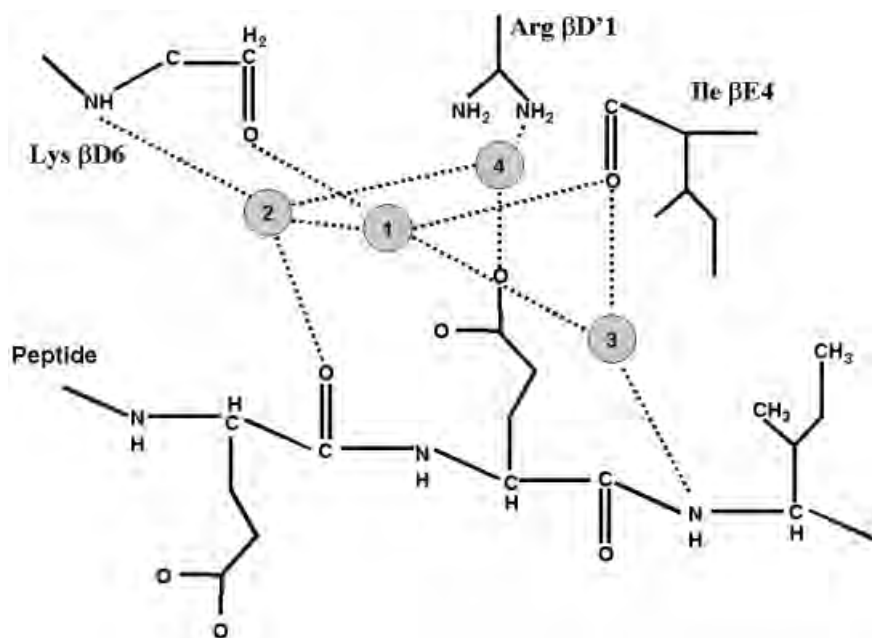


Fig. 6.3 The extensive water-mediated hydrogen bond network that mediates recognition of the specific cognate peptide pYEEI for ν -src SH2. The illustrated bonds are thought to be particularly favorable based on modeling

studies, and efforts to remove these waters by appropriately placed hydrogen-bonding groups in peptidomimetic inhibitors persistently reduce binding affinity [57, 63].

6.4

Hydrogen Bonds in Drug Design

6.4.1

Diverse Effects of Hydrogen Bonding on Drug Properties

As we have seen above, single hydrogen bonds between ligand and protein or water have a substantial effect on binding affinity (5.7 kJ mol^{-1} is a factor of 10 and 11.4 kJ mol^{-1} is a factor of 100). However, good binding affinity is only one aspect of successful drug design. The rather specific geometric requirements of hydrogen bonding are able to create specificity of binding as well, which is important in avoiding binding to proteins other than the target. Hydrogen bonding also strongly influences the transport, adsorption, distribution, metabolism, and excretion properties of molecules. Lipinski's analysis of drugs and development candidates [64] suggests that there is a finite limit to the number and nature of non-covalent interactions that a drug is expected to make with its environment. In particular two of the "rules" state that drugs should contain no more than 5 hydrogen bond donors and 10 acceptors. These rules arise because of the need to balance absorption and distribution properties with binding specificity within a rela-

tively small drug molecule. However, it is not clear why there is a numerical difference between donors and acceptors. The rules are based only on counting strong donors and acceptors (O, N, NH, OH) and make no modifications for variation in hydrogen bond strength due to activating neighbors or for the presence of other types of hydrogen-bonding groups. It seems, from the discussion that has gone on before concerning the variability of hydrogen-bonding groups, that the “rules” are not sufficiently subtle.

A more quantitative measure of expected lipophilicities and permeabilities than Lipinski’s rules should probably be used in design strategies. Raevsky [65] has carried out an analysis of the donor and acceptor strength of many thousands of molecules, leading to a simple scoring system that describes these strengths. It was demonstrated that relatively accurate predictions of lipophilicity could be made from any structure on the basis of this scoring system. Similarly, in addition to steric bulk effects, both the H-bond donor and acceptor strength play an important role in explaining differences in permeability and absorption of neutral chemical compounds and drugs [66].

6.4.2

Optimizing Inhibitor Affinity

If we consider only the process of improving the affinity of existing ligands or developing a high-affinity ligand *de novo*, then the incorporation of maximal numbers of hydrogen bonds appears likely to convey an advantage. Of course, a particular binding site will offer only a limited number of hydrogen-bonding opportunities based on the number, type, and disposition of the amino acids forming the site. A structure-based drug-design approach will take into account the structure of the complex of the target protein and a lead compound (often the enzyme substrate). Based on this information, the potential to incorporate additional or alternative groups on the ligand to enhance hydrogen bonding can be assessed. There are numerous programs that may facilitate this process, ranging from simple scoring functions to molecular dynamics simulations (Section 6.4.3).

There are a plethora of examples in the literature whereby the involvement of hydrogen bonds has been modified to enhance ligand affinity. It is not the purpose of this chapter to provide a comprehensive review of these; instead, we select a few instructive examples (suitable for further reading) where hydrogen bonding has been demonstrated to have significant effect on the structure-based design process.

In an early classic structural and thermodynamic study, the high-resolution crystal structures of a pair of thermolysin inhibitors revealed that they bound identically except for the appearance of one hydrogen bond that made a specific interaction [67, 68]. The comparison was specifically designed to produce an intrinsic hydrogen-bonding energy. The hydrogen bond donor group of the tighter-binding inhibitor was replaced by an acceptor in order to maintain its hydrogen-bonding capability with water in the unbound state. The difference in affinity between the two interactions suggested that this hydrogen bond alone contributed a total of $-16.8 \text{ kJ mol}^{-1}$ to the free energy of binding, i.e., a factor of 840.

Significant effort has been directed at the design of drugs to inhibit specific proteins involved in intracellular signal transduction pathways by targeting SH2 domains of relevant proteins. A number of studies report a rational approach to developing tyrosyl phosphate-peptidomimetic ligands that have properties suitable for drugs. One of the greatest challenges in this venture has been to design a suitable replacement for the phosphotyrosine (pY) residue. The phosphotyrosine contributes approximately 60% of the free energy of binding of peptide-based ligands to SH2 domain interactions [69, 70]; therefore, failure to adequately replace this group substantially compromises affinity. The phosphotyrosine moiety is involved in a high concentration of hydrogen bonds (see Fig. 6.2). Substitution of the pY by the non-hydrolyzable phosphomethyl phenylalanine (i.e., $>\text{C}-\text{O}-\text{PO}_3\text{H}_2$ for $>\text{C}-\text{CH}_2-\text{PO}_3\text{H}_2$) showed good resistance to phosphatase activity, but the loss of at least two hydrogen bonds to the phosphate oxygen compromised binding [71]. The best substitute for the phosphate group appeared to be via a sulfur group, $-\text{OPSO}_2\text{H}$, the sulfur group presumably able to sustain the hydrogen bonds of the substituted oxygen [63, 72].

Calorimetric studies on both the free energy and the enthalpy of binding can be informative as to the source of binding energy because of the distinct thermodynamic characteristics of different processes, i.e., desolvation, water entrapment, or direct hydrogen bonding [73]. Work based on the interaction of FKBP and the immunosuppressive drug FK506 [74] suggested that a hydroxyl-carbonyl hydrogen bond itself was enthalpically unfavorable even though the overall free energy for the bond formation is favorable because of a favorable ΔS term for dehydration of the hydroxyl group. Structure-activity relationships for series of tricyclic inhibitors to farnesyl protein transferase revealed an interesting correlation between the enthalpic contribution to binding and the increase of nonpolar surface resulting from addition of halogen atoms to the compounds. Nonetheless, the majority of the dominant enthalpy term appeared to be derived from hydrogen bonding, which incorporated a crucial water-mediated interaction [75].

An exhaustive study on novel serine protease inhibitors revealed the role of a multi-centered short hydrogen-bonding network in ligand recognition [76]. The appearance of eight inhibitor-enzyme or enzyme-water-inhibitor hydrogen bonds at the active site is a common feature of serine protease inhibitor binding observed in a large number of crystal structures of trypsin, thrombin, and urokinase-type plasminogen activator complexes. The short hydrogen-bonding networks were estimated to contribute approximately 7.1 kJ mol^{-1} to the free energy of binding. This seems rather small but is a differential effect, as many bonds to water of similar strength exist in the apo-enzymes. This work also emphasized the importance of the pK_a values of groups in the binding site and the potential effects that these can have on hydrogen-bonding capability. The interaction of some of these inhibitors via hydrogen-bonding networks that incorporate water molecules also begs for the inclusion of interfacial water molecules into the drug-design process (as discussed in Section 6.3.2), where they can improve binding and specificity properties [56, 57].

6.4.3

Computational Tools for Hydrogen Bond Analysis and Design

There are several computational tasks that occur frequently in structure-based drug design, including high-throughput, structure-based searching for lead compounds that complement a given binding site; suggesting modifications to a known ligand in order to improve affinity or specificity; and *post hoc* rationalization of trends in experimentally well-characterized protein-ligand complexes. In each of these tasks, proper account needs to be taken of the effects of changes to hydrogen bonding as the ligand binds. Each of the three tasks requires similar structural and energetic computations, although the first is often carried out with a lower level of detail for reasons of speed.

If we first consider the latter tasks, where the structure of a protein-ligand complex is already known, it is usually necessary to build the hydrogen atoms into the structure, determine which hydrogen bond donor and acceptor sites are present, bearing in mind possible bifurcation and steric accessibility issues, and then correctly score the contributions of hydrogen bonding and all other aspects of the structure to the thermodynamics of binding. It should be emphasized that no single accurate automated procedure yet exists to carry out all these computations, but many tools are available that can assist visualization and thinking.

In building hydrogens into a structure, although it is fair to assume that all strong acceptors and donors are satisfied to a first approximation, it is often difficult to decide between alternative pairings and geometries. Glick and Goldblum [77] have described a strategy of ordered hydrogen placement that begins by adding non-rotatable hydrogens such as those of the peptide backbone according to known covalent geometry. Then water protons, polar side chain protons, and the C- and N-termini of a protein are added in such a manner as to maximize the number and strength of hydrogen bonds. Since there are many possible combinations, a sophisticated mixed stochastic/hierarchical search is employed to find the optimal configuration. The program was benchmarked successfully against several neutron structures of proteins. An alternative approach is to build all hydrogen according to known covalent geometry and then subject the system to a short period of molecular dynamics simulation, which allows the system to find a favorable low-energy configuration for the hydrogen bonding. This can be accomplished in most commercial biomolecular modeling packages (Sybyl, Insight, etc.), but note that molecular dynamics is essential in order to search thoroughly for alternatives to the initially built structure; energy minimization alone does not have sufficient flexibility. Molecular dynamics also has the potential advantage of allowing a flexible response of the protein in response to modifications of the ligand. Both methods require prior knowledge from NMR data or from the structure itself [12] or computation [78] of the pK_a of all groups, which profoundly affect hydrogen-bonding patterns and consequent binding affinity [79]. When relying on computation of pK_a values – since hydrogen bonding can significantly affect the pK_a of His, Glu, and Asp – it may be necessary to iteratively cycle the pK_a and hydrogen-bond calculations until a steady state is reached.

The stochastic search and molecular dynamics approaches are both limited by how well hydrogen bonding is described within them. Generally, lone pair geometry and the interactions of weak acceptors are poorly represented in computational models of protein-ligand interactions. Accurate models of the local energetics of strong donor-weak acceptor pairs would be particularly useful, as the very existence of the bond depends on a fine balance of energetic considerations. Furthermore, it can be extremely difficult to decide on the basis of a medium-resolution X-ray structure alone whether or not such a bond is present or significant in binding.

An alternative or additional procedure to attempting to model the actual position of every hydrogen bond is to create a 3-D contour map of the probability of finding a donor or acceptor at a particular position in the binding site. Taking advantage of the ready availability of 3-D modeling/visualization software, binding site probability maps are created by superimposing the individual probability maps for known donor and acceptor groups (which have been derived from structural databases, Section 6.2.2) on each donor and acceptor in the binding site. The 3-D site mapping idea has been incorporated into several programs: XSITE (based on data from the PDB and therefore actual protein-ligand complexes [80]) and ISOSTAR/SUPERSTAR (based on the CSD data and therefore on broader chemistry [81, 82]). These templates can be used to predict the potential positions at which a ligand could interact via a hydrogen bond and therefore could be used in conjunction with molecular similarity studies, pharmacophore query searching of databases, or *de novo* design algorithms [13]. An interesting variation on this approach is to create a hydrogen-bond probability map for a given ligand or set of ligands that contains all the feasible positions at which a complementary protein atom could be found. This can be useful in creating a model of a receptor where the structure of the receptor is not known from experiment [83].

The success of structure-based lead development is hampered by the inability to accurately relate changes in structure to the detailed energetics of binding. *Ab initio* quantum theory is quite successful at modeling hydrogen bond geometry and energetics, but it is too computationally demanding for routine use in ligand design. Consequently, simplified force fields in which the energy is expressed as a sum of Lennard-Jones and electrostatics interaction are widely used to model macromolecular systems. None of the standard force fields accurately model polarizability. Instead, parameters are averaged over a number of the configurations considered by quantum mechanics and are modified to account for an average polarizable environment. In the case of homogenous liquids such as water, this has been a successful strategy for the reproduction in calculation of the bulk properties of the liquid. However, the success of this temporal and spatial averaging strategy is unlikely to be extendable to discrete binding sites for water or other ligands that possess properties distinct from the bulk phase. Another defect of common force fields is that they place partial charges at atom positions and consequently model only the dipole character of local interactions. As we have seen, the preference for hydrogen bonding to lone pairs means that the dipole model does not have appropriate geometry. It has been found that more sophisticated electrostatic models (e.g., using distributed multi-poles) can produce correct orientation

information [84], but even then polarizability is not correctly included. Although such molecular-simulation-based rationalization of inhibitor binding energies can provide some insight and is a very active area of research (e.g., [85, 86]), it seems that it will be some time before sufficiently accurate models of hydrogen bonding are available to allow accurate physics-based predictions of binding free energy.

High-throughput scoring of ligands is possible with lower levels of structural detail and greater margins of error in calculated binding affinity. Such scoring schemes partition the interaction energy of a protein-ligand system in a simple way, e.g., counting strong and very strong hydrogen bonds, contacts between, or surface area of, buried apolar groups and ignore the explicit water molecules. They are considered in more detail in Chapter 1 of this book. They can be quite successful in ranking compounds' binding affinities (e.g., [87]) and consequently useful in the drug-design process. However, they are subject to a substantial margin of error and are usually very poor at identifying the thermodynamic (enthalpy or entropy) and structural source of changes in binding affinity [88]. We believe that improvements can be made by more detailed consideration of the strength of hydrogen bonds [37, 65] and the role of buried waters [56] and by parameterizing the scores against enthalpy and entropy data in addition to binding affinities.

6.5

Conclusion

Hydrogen bonds are crucial to the recognition of ligands by proteins. We have learned much structurally and energetically about proteins' hydrogen-bonding capacity over the past 20 years, and this is beginning to make an effective contribution to drug-design strategies. However, hydrogen bonding remains a very active area of research, with new insights promised by the determination of more protein-ligand structures of better quality, by emerging spectroscopic techniques, and the possibility of building a greater experimental databank of thermodynamic knowledge through advances in microcalorimetry. The integration of this knowledge into theoretical and computational molecular models will be an exciting and rewarding challenge in the coming decade.

6.6

References

- 1 G. C. PIMENTEL, A. L. MCCLELAN, *The hydrogen bond*, Freeman, San Francisco 1960.
- 2 L. PAULING, *The nature of the chemical bond and the structure of molecules and crystals*, Cornell University Press, Ithaca, NY 1939.
- 3 K. MOROKUMA, *Acc. Chem. Res.*, **1977**, *10*, 294–300.
- 4 D. A. SMITH, A brief history of the hydrogen bond, in: D. A. SMITH (Ed.) *Modeling the hydrogen bond*, American Chemical Society, Washington DC 1994, 1–5.

- 5 E.D. ISAACS, A. SHULKA, P.M. PLATZMAN, D.R. HAMANN, B. BARBIELLINI, C.A. TULK, *Phys. Rev. Lett.*, **1999**, 82, 600–603.
- 6 N. NIIMURA, *Curr. Opin. Struct. Biol.*, **1999**, 9, 602–608.
- 7 F. SHU, V. RAMAKRISHNAN, B.P. SCHOENBORN, *Proc. Nat. Acad. Sci. USA*, **2000**, 97, 3872–3877.
- 8 R. TAYLOR, O. KENNARD, W. VERSICHEL, *J. Am. Chem. Soc.*, **1983**, 105, 5761–5766.
- 9 G.A. JEFFERY, W. SAENGER, *Hydrogen bonding in biological structures*, Springer, Berlin **1991**.
- 10 J.P. LOMMERSE, R. TAYLOR, *Journal of Enzyme Inhibition*, **1997**, 11, 223–243.
- 11 E.N. BAKER, R.E. HUBBARD, *Prog. Biophys. Molec. Biol.*, **1984**, 44, 97–197.
- 12 I.K. McDONALD, J.M. THORNTON, *J. Mol. Biol.*, **1994** 238, 777–793.
- 13 J.E.J. MILLS, P.M. DEAN, *J. Computer-Aided Molecular Design*, **1996**, 10, 607–622.
- 14 J.B.O. MITCHELL, S.L. PRICE, *J. Comp. Chem.*, **1990**, 11, 1217–1233.
- 15 N. THANKI, J.M. THORNTON, J.M. GOODFELLOW, *J. Mol. Biol.*, **1988**, 202, 637–657.
- 16 J.A. IPPOLITO, R.S. ALEXANDER, D.W. CHRISTIANSON, *J. Mol. Biol.*, **1990**, 215, 457–471.
- 17 R. PREISSNER, U. EGNER, W. SAENGER, *Febs Lett.*, **1991**, 288, 192–196.
- 18 G.R. DESIRAJU, T. STEINER, T., *The Weak Hydrogen Bond-In Structural Chemistry And Biology*, Oxford Science Publications, Oxford **1999**.
- 19 M.F. PERUTZ, G. FERMI, D.J. ABRAHAM, C. POYART, E. BURSAX, *J. Am. Chem. Soc.*, **1986**, 108, 1064–78.
- 20 G. WAKSMAN, D. KOMINOS, S.C. ROBERTSON, N. PANT, D. BALTIMORE, R.B. BIRGE, D. COWBURN, H. HANAFUSA, B. J. MAYER, M. OVERDUIN, M.D. RESH, C. B. RIOS, L. SILVERMAN, J. KURIYAN, *Science*, **1992**, 358, 646–653.
- 21 M.F. PERUTZ, *Philos. Trans. Royal Soc. London A*, **1993**, 345, 105–112.
- 22 J.B.O. MITCHELL, C.L. NANDI, I.K. McDONALD, J.M. THORNTON, S.L. PRICE, *J. Mol. Biol.*, **1994**, 239, 315–331.
- 23 K. FLANAGAN, J. WALSHAW, S.L. PRICE, J.M. GOODFELLOW, *J.M., Protein Eng.*, **1995**, 8, 109–116.
- 24 T. STEINER, G. KOELLNER, *J. Mol. Biol.*, **2001**, 305, 535–557.
- 25 T. STEINER, *J. Chem. Soc.-Perkin Transactions*, **1995**, 2, 1315–1319.
- 26 S. SCHEINER, T. KAR, Y.L. GU, *J. Biol. Chem.*, **2001**, 276, 9832–9837.
- 27 G.A. JEFFERY, *An introduction to hydrogen bonding*, Oxford University Press, Oxford **1997**.
- 28 S.N. VINOGRADOV, R.H. LINNELL, *Hydrogen bonding*, Van Nostrand Reinhold, New York **1971**.
- 29 V.W. COULING, P. FISCHER, D. KLENERMAN, W. HUBER, *Biophys. J.*, **1998**, 75, 1097–1106.
- 30 L.D. BARRON, L. HECHT, E.W. BLANCH, A. F. BELL, *Prog. Biophys. Mol. Biol.*, **2000**, 73, 1–49.
- 31 G.M. CLORE, A.M. GRONENBORN, *Curr. Opin. Chem. Biol.*, **1998**, 2, 564–570.
- 32 P.M. NIETO, B. BIRDSALL, W.D. MORGAN, T.A. FRENKIEL, A.R. GARGARO, J. FEENEY, *FEBS Lett.*, **1997**, 405, 16–20.
- 33 J. FEENEY, *Angew. Chem. Int. Ed. Engl.*, **2000**, 39, 290–312.
- 34 T.K. HARRIS, Q. ZHAO, A.S. MILDVAN, *J. Mol. Struct.*, **2000**, 552, 97–109.
- 35 A.J. DINGLEY, F. CORDIER, S. GRZESIEK, *Concepts Magn. Res.*, **2000**, 13, 103–127.
- 36 F. CORDIER, C.Y. WANG, S. GRZESIEK, L.K. NICHOLSON, *J. Mol. Biol.*, **2000**, 304, 497–505.
- 37 E. GANCIA, J. G. MONTANA, D. T. MANALLACK, *J. Molec. Graph. Modelling*, **2001**, 19, 349–362.
- 38 H. ADAMS, K.D.M. HARRIS, G.A. HEMBURY, C.A. HUNTER, D. LIVINGSTONE, J.F. MCCABE, *Chem. Comm.*, **1996**, 22, 2531–2532.
- 39 A.R. FERSHT, J.-P. SHI, J. KNILL-JONES, D.M. LOWE, A.J. WILKINSON, D.M. BLOW, P. BRICK, P. CARTER, M.M.Y. WAYE, G. WINTER, *Nature*, **1985**, 314, 235–238.
- 40 S.J. TEAGUE, A.M. DAVIES, *Angew. Chem. Int. Ed. Engl.*, **1999**, 38, 736–749.
- 41 J.A. SCHELLMAN, *C. R. Trav. Carlsberg Ser. Chim.*, **1955**, 29, 223–229.
- 42 I.M. KLOTZ, J.S. FRANZEN, *J. Am. Chem. Soc.*, **1962**, 84, 3461–3466.
- 43 H. SUSI, S.N. TIMASHEFF, J.S. ARD, *J. Biol. Chem.*, **1964**, 239, 3051–3054.

- 44 S. J. GILL, L. NOLL, *J. Phys. Chem.*, **1972**, 76, 3065–3068.
- 45 J. M. SCHOLTZ, S. MARQUSEE, R. L. BALDWIN, E. J. YORK, J. M. STEWART, M. SANTORO, D. W. BOLEN, *Proc. Natl. Acad. Sci. USA.*, **1991**, 88, 2854–2858.
- 46 B. A. SHIRLEY, P. STANSSENS, U. HAHN, C. N. PACE, *Biochemistry*, **1992**, 31, 725–732.
- 47 K. A. SHARP, A. NICHOLLS, R. FRIEDMAN, B. HONIG, *Biochemistry*, **1991**, 30, 9686–97.
- 48 A. S. YANG, K. SHARP, B. HONIG, *J. Mol. Biol.*, **1992**, 227, 889–900.
- 49 G. I. MAKHATADZE, P. L. PRIVALOV, *J. Mol. Biol.*, **1993**, 232, 639–659.
- 50 S. M. HABERMANN, K. P. MURPHY, *Protein Sci.*, **1996**, 5, 1229–1239.
- 51 N. THANKI, J. M. THORNTON, J. M. GOODFELLOW, *Protein Eng.*, **1990**, 3, 495–508.
- 52 J. L. FINNEY, A. K. SOPER, *Chem. Soc. Reviews*, **1994**, 23, 1–10.
- 53 L. F. SCATENA, M. G. BROWN, G. L. RICHMOND, *Science*, **2001**, 292, 908–912.
- 54 J. M. STURTEVANT, *Proc. Natl. Acad. Sci. USA*, **1977**, 74, 2236–2240.
- 55 M. A. WILLIAMS, J. M. GOODFELLOW, J. M. THORNTON, *Protein Sci.*, **1994**, 3, 1224–35.
- 56 J. E. LADBURY, *Chem. & Biol.*, **1996**, 3, 973–980.
- 57 D. A. RENZONI, M. J. J. M. ZVELEBIL, T. LUNDBÄCK, J. E. LADBURY. In: J. E. LADBURY, & P. R. CONNELLY (Eds.) *Structure-based drug design: Strategy, thermodynamics and modelling*, Springer, Berlin **1997**, 161–180.
- 58 T. STEINER, *Acta Cryst D*, **1995**, 51, 93–97.
- 59 P. Y. S. LAM, P. K. JADHAV, C. J. EYERMANN, C. N. HODGE, Y. RU, L. T. BACHELIER, J. L. MEEK, M. J. OTTO, M. M. RAYNER, Y. N. WONG, C. -H. CHANG, P. C. WEBER, D. A. JACKSON, T. R. SHARPE, S. ERICKSON-VIITANEN, *Science*, **1994**, 263, 380–384.
- 60 J. R. H. TAME, S. H. SLEIGH, A. J. WILKINSON, J. E. LADBURY, *Nature Struct Biol.*, **1996**, 3, 998–1001.
- 61 S. H. SLEIGH, P. R. SEEVERS, A. J. WILKINSON, J. E. LADBURY, J. R. H. TAME, *J. Mol. Biol.*, **1999**, 291, 393–415.
- 62 T. G. DAVIES, R. E. HUBBARD, J. R. H. TAME, *Protein Sci.*, **1999**, 8, 1–13.
- 63 D. A. HENRIQUES, J. E. LADBURY, *Arch. Biochem. Biophys.*, **2001**, 390, 158–168.
- 64 C. A. LIPINSKI, F. LOMBARDI, B. W. DOMINY, P. J. FEENEY, *Adv. Drug Delivery Rev.*, **1997**, 23, 3–25.
- 65 O. A. RAEVSKY, *J. Phys. Organic. Chem.*, **1997**, 10, 405–413.
- 66 O. A. RAEVSKY, K. J. SCHAPER, K. J. EUR. *J. Med. Chem.*, **1998**, 33, 799–807.
- 67 D. E. TRONRUD, H. M. HOLDEN, B. W. MATTHEWS, *Science*, **1987**, 235, 571–574.
- 68 P. A. BARTLETT, C. MARLOWE, *Science*, **1987**, 235, 569–571.
- 69 J. M. BRADSHAW, V. MITAXOV, G. WAKSMAN, *J. Mol. Biol.*, **1999**, 293, 971–985.
- 70 D. A. HENRIQUES, J. E. LADBURY, R. M. JACKSON, *Protein Sci.*, **2000**, 9, 1975–1985.
- 71 T. R. BURKE JR, M. S. SMYTH, A. OTAKA, M. NOMIZU, P. P. ROLLER, G. WOLF, R. CASE, S. E. SHOELSON, *Biochemistry*, **1994**, 33, 6490–6494-73.
- 72 T. GILMER, M. RODRIGUEZ, S. JORDAN, R. CROSBY, K. ALLIGOOD, M. GREEN, M. KIMERY, C. WAGNER, D. KINDER, P. CHARIFSON, A. M. HASSELL, D. WILLARD, M. LUTHER, D. RUSNAK, D. D. STERNBACH, M. MEHROTRA, M. PEEL, L. SHAMPINE, R. DAVIS, J. ROBBINS, I. R. PATELL, D. KASSEL, W. BURKHART, M. MOYER, T. BRADSHAW, J. BERMAN, *J. Biol. Chem.*, **1994**, 269, 31711–31719.
- 73 G. KLEBE, H. J. BÖHM, *J. Receptor Signal Transduction Res.*, **1997**, 17, 459–473.
- 74 P. R. CONNELLY, R. A. ADALPE, F. J. BRUZZESE, S. P. CHAMBERS, M. J. FITZGIBBON, M. A. FLEMING, S. ITOH, D. J. LIVINGSTON, M. A. NAVIA, J. A. THOMSON, K. P. WILSON, *Proc. Natl. Acad. Sci. USA.*, **1994**, 91, 1964–1968.
- 75 C. L. STRICKLAND, P. C. WEBER, W. T. WINDSOR, Z. WU, H. V. LE, M. M. ALBANESE, C. S. ALVAREZ, D. CESARZ, J. DEL ROSARIO, J. DESKUS, A. K. MALLAMS, F. G. NJOROGO, J. J. PIWINSKI, S. REMISZEWSKI, R. R. ROSSMAN, A. G. TAVERAS, B. VIBULBHAN, R. J. DOLL, V. M. GIRIJAVALLABHAN, A. K. GANGULY, *J. Med. Chem.*, **1999**, 42, 2125–2135.
- 76 B. A. KATZ, K. ELROD, C. LUONG, M. J. RICE, R. L. MACKMAN, P. A. SPRENGELER, J. SPENCER, J. HATAYE, J. JANC, J. LINK, J. LITVAK, R. RAI, K. RICE, S. SIDERIS,

- E. VERNER, W. YOUNG, *J. Mol. Biol.*, **2001**, 307, 1451–1486.
- 77 M. GLICK, A. GOLDBLUM, *Proteins: Struct. Funct. Gen.*, **2000**, 38, 273–287.
- 78 D. BASHFORD, M. KARPLUS, *J. Phys. Chem.*, **1991**, 95, 9556–9561.
- 79 F. DULLWEBER, M. T. STUBBS, D. MUSIL, J. STURZEBECKER, G. KLEBE, *J. Mol. Biol.*, **2001**, 313, 593–614.
- 80 R. A. LASKOWSKI, J. M. THORNTON, C. HUMBLET, J. SINGH, *J. Mol. Biol.*, **1996**, 259, 175–201.
- 81 I. J. BRUNO, J. C. COLE, J. P. M. LOMMERSE, R. S. ROWLAND; R. TAYLOR; M. L. VERDONK, *J. Computer-Aided Molec. Design*, **1997**, 11, 525–537.
- 82 M. BÖHM, G. J. KLEBE, *J. Med. Chem.*, **2002**, 45, 1585–1597.
- 83 J. E. MILLS, T. D. J. PERKINS, P. M. DEAN, *J. Computer-Aided Molec. Design*, **1997**, 11, 229–242.
- 84 J. B. O. MITCHELL, C. L. NANDI, J. M. THORNTON, S. L. PRICE, J. SINGH, M. SNAREY, *J. Chem. Soc. Faraday Trans.*, **1993**, 89, 2619–2630.
- 85 P. KALRA, T. V. REDD, B. JAYARAM, *J. Med. Chem.*, **2001**, 44, 4325–4338.
- 86 R. C. RIZZO, J. TIRADO-RIVES, W. L. JORGENSEN, *J. Med. Chem.*, **2001**, 44, 145–154.
- 87 H. J. BÖHM, *J. Computer-Aided Molec. Design*, **1998**, 12, 309–323.
- 88 J. R. H. TAME, *J. Computer-Aided Molec. Design*, **1999**, 13, 99–108.

7

Principles of Enzyme-Inhibitor Design

D. W. BANNER

7.1

Introduction

The field of enzyme inhibition is one of the most fruitful sources of experimental information on the interaction of small chemical ligands with proteins. It is well known that the majority of pharmaceutical companies have a range of drug-development projects where the active principle is an enzyme inhibitor. The reason for this is clear: many enzymes are well-characterized, soluble, stable proteins with an established assay suitable for either high-throughput screening or precise measurement of inhibition constants. Perhaps the most widely studied enzyme is thrombin. In this chapter, we will use the example of active site inhibition of thrombin to illustrate a range of principles of enzyme-inhibitor design. It will be left to the reader to perceive when the terms “enzyme” and “inhibitor” may be generalized to “receptor” and “ligand.” The modern drug-design and -development process is extremely complex. Here we will concentrate only on the molecular recognition aspects. Comprehensive surveys of the thrombin inhibitor patent literature have been made by Wiley and Fisher (pre-1997) [1] and Coburn (1997–2000) [2], and the clinical use of direct thrombin inhibitors has been extensively reviewed [3–11]. Thrombin residue numbering follows throughout the chymotrypsinogen convention.

The first principle of enzyme inhibitor design is “Use all the available information”. This information can be biological, functional, structural, chemical, or theoretical. There is such an immense amount of biological information on thrombin that it cannot be surveyed here: we focus on thrombin as a serine protease of the trypsin family and take fibrinogen to be its primary substrate. A convenient way to look at the information available is from the more general to the very specific. For thrombin, we may take four levels: the general catalytic mechanism; the particular substrate types processed; the structure of the protein; and, often forgotten, the flexibility of the protein required to achieve this function.

Enzymes are biological catalysts: the active site exists to correctly position a substrate molecule so that functional groups on the enzyme may perform “chemistry” on it. All trypsin-like enzymes have a “catalytic triad” of aspartic acid 102, histidine 57, and serine 195 in which the serine O γ is activated so that it may attack a

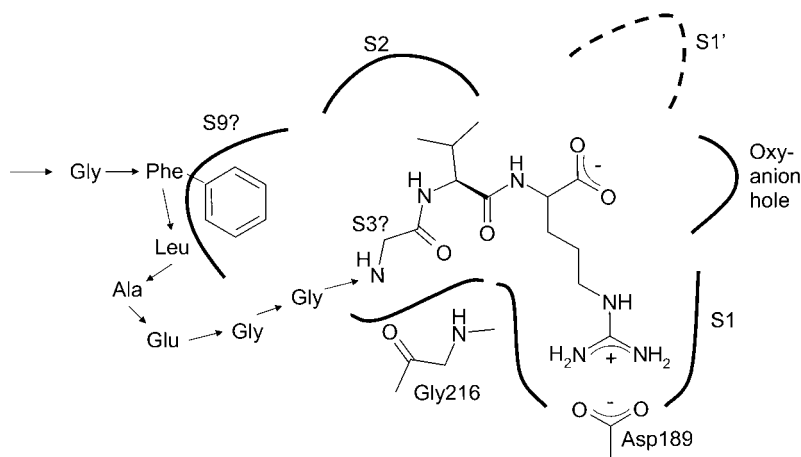


Fig. 7.1 Thrombin active site regions as defined by the binding of fibrinopeptide of sequence D-F-L-A-E-G-G-G-V-R (from [1bbr]).

suitably positioned substrate carbonyl carbon atom to form an acyl intermediate which is, in turn, attacked by water to release products.

Thrombin substrates are normally peptidic, with nucleophilic attack of the serine O γ being on the carbonyl carbon of an amide bond. The tetrahedral intermediate so formed is stabilized by two enzyme backbone hydrogen bonds from the –NHs of Gly193 and Ser195, which form the “oxyanion hole.” This catalytic mechanism is positioned next to a “recognition pocket” that has an aspartic acid (Asp189) at the bottom and is highly arginine specific (Fig. 7.1). Substrates normally have to be positioned quite accurately in the active site of an enzyme for catalysis to proceed quickly. In the case of thrombin, this is further achieved via an anti-parallel beta interaction between substrate peptide and enzyme residues 214–216.

There is perhaps more three-dimensional structural information available on thrombin than any other enzyme. We will study selected examples of the use of such information, indicating the relevant Protein Data Bank (<http://www.rcsb.org/pdb/>) entries [1abc] for those who wish to visualize the structures in 3-D.

Enzymes are intrinsically mobile. Catalysis proceeds stepwise, first by formation of a complex between enzyme and substrate(s), next by passage through one or more transition states, and finally by release of product(s). There is therefore a range of low-energy structures that may be regarded as possible targets for inhibition. These, when known, should be considered in the inhibitor design process. In the case of thrombin, no large structural changes take place, but in other systems – particularly where multiple substrates, products, and cofactors are involved – it may be necessary to document and analyze a significant number of very different structural states. This may be done by determining X-ray structures of complexes with different combinations of functional and non-functional analogues of substrates, products, and cofactors. In particular, “transition state analogues” are of great value and should be synthesized whenever possible.

7.2

The Active Site

It has proved extremely useful for thrombin and many other enzymes to provide a standard nomenclature to describe the active site. The notation of Shechter and Berger is widely used for enzymes whose substrates are polymers: the positions of the polymer are named -P4-P3-P2-P1/P1'-P2'-P3'-P4'-, where / is the cleavage site, and the sequence for polypeptides runs from the N- to the C-terminus [12]. The corresponding pockets on the protein that are responsible for the recognition of these polymer elements are called “sub-sites” and are labeled ... S2, S1, S1', S2'...

For thrombin, with fibrinogen as defining substrate, this is inconvenient. Inspection of complexes of thrombin with fibrinopeptide analogues [*1bbr*, *1dm4*, *1fph*, *1ucy*, and *1ycp*] shows the fibrinopeptide, the N-terminal fibrin cleavage product, to have a folded structure as illustrated in Fig. 7.1. (It is more convenient to use the notation of Fig. 7.2. [13]).

7.3

The Heuristic Approach

It could be argued that in the ideal case and given the power of modern computational methods, one single X-ray crystal structure of thrombin should suffice to design thrombin inhibitors with the desired properties using “virtual screening” techniques. Certainly, examples of success using this approach are known and are

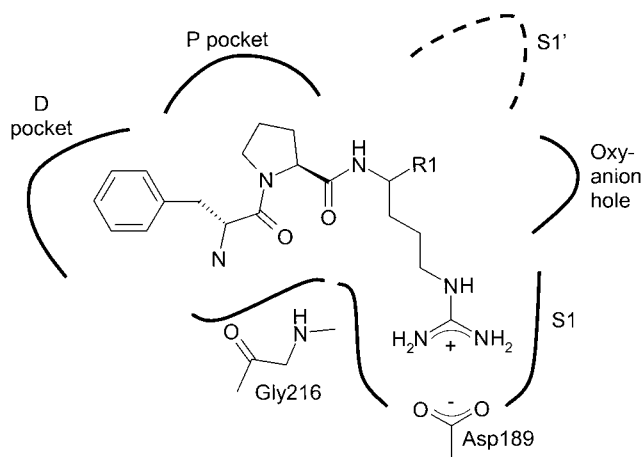


Fig. 7.2 Thrombin active site regions as defined by the binding of D-Phe-Pro-Arg analogues, e.g., PPACK. The recognition pocket (S1) is clear. The proximal (P) hydrophobic pocket binds the proline side chain and thus corresponds to S2. The distal (D) hydrophobic pocket binds the D-Phe side chain. For PPACK 1 R1 = CO.CH₂Cl.

given below (Section 7.5). The majority of thrombin inhibitors reported to date, however, were produced by classical medicinal chemistry, either alone or in combination with X-ray structural information.

To facilitate the dialogue among the chemist, the modeler, and the crystallographer, it has proved most useful not only to define terms, as in Fig. 7.2, but also to develop “design rules.” Examples might be: “A terminal basic group is required to fit into the S1 pocket,” “Two hydrophobic groups are required to fill the P- and D-pockets,” or “A hydrogen-bond acceptor has to be positioned over the –NH of Gly 216.” Such rules may give valuable direction to design, particularly if they capture some aspect of the active site that is not particularly obvious but that is indicated by experiment. There is, however, a clear problem with the approach, namely, that it may not be necessary or desirable to obey all the rules at once. Thrombin, for example, has two large hydrophobic pockets (D and P) as well as the S1 pocket. It is thus relatively easy to generate molecules that bind tightly to the active site, that is, with inhibition constants in the low nanomolar range. The practical issue here (at least in the pharmaceutical industry) is not how to obtain better inhibition but rather how to produce compounds with optimal biological properties. There are many examples of thrombin inhibitors where, for instance, a less basic P1 group has been introduced in the attempt to improve oral availability, resulting in a “non-optimal” interaction with the S1 pocket.

A further handicap associated with any rule-based approach to inhibitor design is that it tends strongly to lead to just one class of very similar molecules: the process often converges to a single (local) minimum. The best approach is “Try to extract helpful rules from the available data – but be prepared to break them!”

7.4

Mechanism-based Covalent Inhibitors

The penicillins are one of the most successful classes of drug. They are mechanism-based inhibitors of beta lactamases and penicillin-binding proteins (PBPs) involved in bacterial cell-wall synthesis. In brief, these enzyme inhibitors contain a lactam ring that opens on acylation of the active site serine; ring opening is followed by structural rearrangement of the inhibitor, and the best inhibitors are those where the rearrangement is such that the water attack required for de-acylation is hindered. This results in very slow off-rates: the enzyme is covalently inhibited for a very long time. It might be thought that this principle could easily be transferred to other enzymatic systems, such as thrombin. This has been the subject of much effort, but the results have been generally disappointing. It turns out that lack of selectivity is a serious problem. Chemical compounds with suitable reactive centers can inhibit a wide variety of similar enzymes, with the risk of severe toxicity effects *in vivo*. For the penicillins this is not a problem, as humans do not possess homologues of beta lactamases and PBPs. We do, however, have many important thrombin-like serine proteases involved in vital functions, which must not be significantly inhibited.

Can we, nevertheless, use mechanism-based inhibitors to study the molecular interactions in active sites? Unfortunately the simple answer appears to be negative: on- and off-rates and structural rearrangements are difficult to interpret in terms of the energetics of specific interactions. Although detailed spectroscopic studies have begun to shed light on these complex mechanisms (e.g., [14, 15]), much more work will be required before all the enthalpic and entropic effects can be unscrambled.

Knowledge of nothing more than the catalytic mechanism and the P1 residue can indeed be used directly to design thrombin inhibitors. All that is required is an arginine analogue with an electrophilic center in the correct position. The simplest of these is APPA (Fig. 7.3).

The seminal work of Bode and Huber produced crystal structures of both benzamidine and APPA bound to trypsin [*3ptb*, *1tpp*] [16]. Thrombin has a very similar primary sequence to trypsin, with amino acid identities (similarities) of about 40% (55%), depending on species. The structures of the enzymes are also very similar, with 200 Ca positions superimposing with about 0.75 Å rms deviation. The remaining ~60 thrombin residues are in surface loops which are much shorter in trypsin, where there are only ~30 corresponding residues.

Given this similarity, it is no surprise that both benzamidine and APPA show little selectivity among thrombin, trypsin, and the large number of closely related enzymes.

About half of the thrombin structures in the PDB are active site complexes with covalent inhibitors and with other molecules bound to so-called “exo-sites” [17]. The majority of these contain PPACK 1 (Fig. 7.2) [18], which has been very widely used as a tool. This inhibitor forms a stable complex with thrombin, making covalent bonds to both Ser 195 and His 57. Here a word of warning must be given: a protein structure is distorted somewhat by inhibitor binding, and for covalent inhibitors, particularly doubly covalent ones, this distortion may be significant. For thrombin this is not a serious problem: in the PPACK complex, the Tyr-Pro-Pro-Trp lid of the P pocket adjusts its position so that the Trp side chain moves down by about 1 Å to better pack over the PPACK proline [13]. It is not unknown in other enzymes that the acyl-enzyme relaxes to a structure significantly different from the initial Michaelis complex. For investigation in structural detail of the acylation of elastase, a serine protease with the same active center as thrombin, consult [19–24].

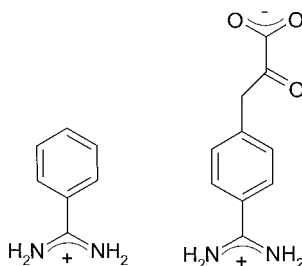


Fig. 7.3 The simplest thrombin inhibitors, benzamidine **2** (left) and APPA, **3** *p*-amidino phenyl pyruvate.

While no structure of a true Michaelis complex of this kind of inhibitor has been reported, Skordalakes et al. have made a fascinating observation with a phosphonate tripeptide thrombin inhibitor of the structure of a trapped pentacovalent intermediate state which precedes the covalent intermediate [25].

Many research groups have started with PPACK and produced inhibitors of a less peptidic nature (to improve *in vivo* stability), and/or with conformational restriction (to tackle the entropy loss problem), and/or with a variety of “serine trap” functionalities, for example, aldehydes, boronic acids, α -keto amides and acids, α -keto heterocycles, polyfluorinated ketones, and phosphonates. Structures of many of these are known in complex with thrombin but will not be reviewed here (see e.g., publications of C. A. Kettner and coworkers). As well as lack of specificity, these potential drugs suffer from slow on-rates and have not progressed to the market.

7.5

Parallel *de novo* Design of Inhibitors

As asserted above, it should easily be possible these days to progress from a crystal structure to a useful lead inhibitor quite quickly using *in silico* screening. A few years ago we used thrombin inhibition to test a conceptually simple *de novo* approach combining combinatorial docking and combinatorial chemistry [26]. Reductive amination was chosen as a convenient synthetic chemistry: a set of aldehydes and a set of amines were chosen on the criteria of size and availability, and all aldehyde-amide reductive amination products were “synthesized” *in silico* and docked into the thrombin active site, and binding affinities were estimated using modified Ludi algorithms [27]. Ten of the predicted best inhibitors were then synthesized chemically and assayed for thrombin inhibition.

The best compound **4**, shown in Fig. 7.4, had a K_i for thrombin of 95 nM, an encouraging result for a compound of molecular weight 317. The “amine moiety,” p-amino-benzamidine, which serves as the arginine mimetic, has on its own a K_i of 34 μ M for thrombin and 5.7 μ M for trypsin, that is, the “needle,” as we have named such entities [28], is selective for trypsin by 5 \times . The full compound, however, has a K_i for trypsin of 520 nM and is thus 5 \times selective for thrombin. To check the binding mode, an X-ray crystal structure of the complex was determined. This confirmed the binding mode predicted but with some significant differences in detail. The terminal phenyl group fits into the D pocket, but the ether oxygen is almost 2 Å further out into solution than expected. This is because the Tyr-Pro-Pro-Trp peptide sequence that forms the “lid” of the hydrophobic P pocket moves down over the central phenyl group in a way already seen for other inhibitors, notably PPACK.

A number of principles are clear from this experiment. The first is that a straightforward approach to obtaining easy-to-synthesize inhibitors may be successful, so is worth trying. A second is that empirical estimates of binding energy such as implemented in Ludi are capable of giving useful results. A third is that the protein may adapt to the inhibitor rather than vice versa.

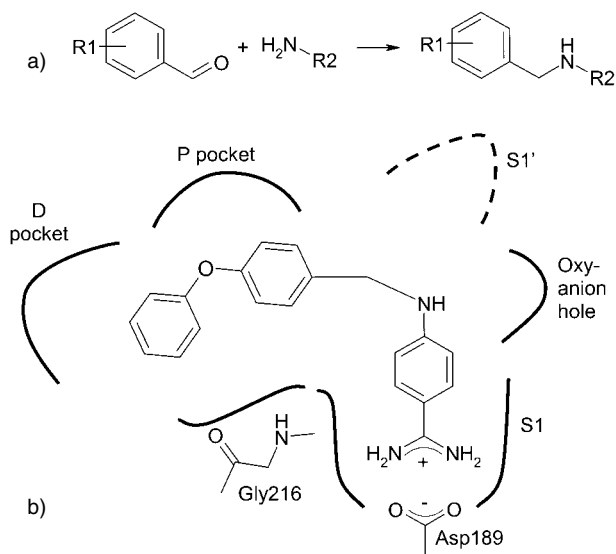


Fig. 7.4 (Above) The reductive amination scheme. (Below) The best compound **4** from combinatorial docking

This last principle may be generalized as “The system moves to the structure with the lowest free energy.” This is, of course, a well-known, fundamental principle of thermodynamics, so it is surprising that it is often overlooked. The problem arises because it is not yet possible to compute the behavior of a system consisting of protein, ligands, water, and, quite possibly, a variety of other solvent molecules for a time period long enough for the system to reach an energy minimum significantly different from that of the starting conformation. As a result, we take many approximations, which may not be valid. In particular, we know that we can express the change in free energy on binding as

$$\delta G = \Delta E - T\Delta S \quad (\text{Eq. 7.1})$$

where ΔE is the enthalpy change, T is the temperature, and ΔS is the entropy change, but in practice we slip into the easier way of thinking of binding energy as simply enthalpic and ignore the entropic effects altogether.

7.5.1

Evolution of Inhibitors

Inhibitors may also be generated by the use of “molecular evolution.” This technology has been successfully applied to thrombin [29]. In brief, a synthetic scheme is chosen that potentially enables the synthesis of a very large combinatorial library: in this case, a three-component Ugi-type reaction was selected. A small set of molecules is synthesized with a random choice of components and tested

for thrombin inhibition. This “first generation” is “evolved” to a second and further generations. After about 20 or so generations, the “population” normally converges to a set of closely related thrombin inhibitors (for a detailed analysis of the method applied to thrombin, see [30]). In this way it is possible to generate low nanomolar thrombin inhibitors, potentially using different synthetic schemes where each scheme would give a different inhibitor series with a different chemical backbone (scaffold). The libraries for thrombin were “biased,” and the synthetic scheme was chosen so that the final molecules contain two hydrophobic groups (targeted to the D and P pockets), a basic group (targeted to the S1 pocket), and potential hydrogen-bonding groups (targeted at Gly216). It is therefore no surprise that the best inhibitors look “familiar” as thrombin inhibitors.

The relevance for understanding molecular interactions is that the selection criterion for “survival of the fittest” is solely the measured K_i for thrombin. It might thus be expected that the best inhibitors would indicate optimal interactions with the protein-binding site. In practice the information obtained is limited by the geometrical restraints imposed by the synthetic scheme chosen and by the restricted choice of building blocks. Use of only “affinity” as the selection criterion also drives the process towards larger and more hydrophobic compounds, which may not be “drug-like.” There is, of course, the possibility to include a bias towards lower molecular weight, or indeed towards any other property of the inhibitor that can be rapidly computed.

A very informative practical exercise would be to use the approach with a number of synthetic schemes, to observe the binding modes of the best inhibitors by use of X-ray crystal structure determinations, and to superimpose the resulting structures. This would then enable a 3-D mapping of the active site on the assumption that inhibitor features occurring most frequently at a particular spatial location indicate the most favorable functional group to place at that point. Further, the range of relevant low-energy structures of the protein would also be mapped.

The important principle here is that although a single ligand-protein complex contains much useful information, the overlay of a series of complexes – as independent of each other as possible – gives a much more objective picture of which interactions contribute most to affinity (and, in principle, to selectivity).

7.6

Inhibitors from Progressive Design

An alternative approach to finding good inhibitors is to start from some kind of small “anchor” building block and then “grow” the inhibitor. In principle, the anchor can be covalent; indeed, it has been suggested that, in the absence of any other starting point, a cysteine might be cloned into a protein of interest. Inhibitors could then be grown starting from a disulfide bridge as anchor. Once sufficient affinity is reached, the anchor could be abandoned altogether or replaced by a substituent targeted at the wild-type molecule. To my knowledge, such an approach has not been

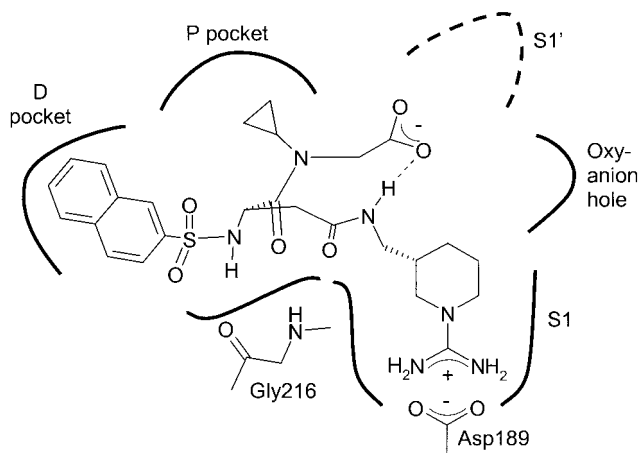


Fig. 7.5 The Roche thrombin inhibitor, napsagatran **5**.

realized¹⁾, but there is a large literature on metalloproteases where an anchor, such as hydroxamate, binding tightly to the catalytic zinc ion has been used.

It has not, however, been particularly fruitful to progressively develop inhibitors starting from small, covalently bound thrombin inhibitors. As indicated in Section 7.3, PPACK-based thrombin inhibitors have been widely varied, but in most cases the D and P pocket groups were left alone and the groups interacting with Gly216, Ser195, or Asp189 were modified to improve biological properties. For more-or-less linear inhibitors, there always exists the possibility of “shuttle” optimization, i.e., working from one end to the other and back again.

A very successful progressive design approach for thrombin is to start from a small building block, known or expected to bind non-covalently in a precise position in the active site, and expand this by progressive addition of substituents. We have called such building blocks “needles” and have described the discovery of a thrombin-specific needle and its evolution to a full-blown thrombin inhibitor with the use of sequential X-ray structural analyses (see [28]). The details will not be repeated here. The inhibitor, napsagatran, **5** is shown in Fig. 7.5.

Napsagatran is, with K_i for thrombin of 270 pM and K_i for trypsin 1.9 μ M, one of the most potent and selective thrombin inhibitors known and will be used here to illustrate a number of principles.

First, the needle itself, amidino-piperidine, has K_i for thrombin of 150 μ M and for trypsin 360 μ M and thus is 2.4 \times selective for thrombin. This contrasts with the classical needle benzamidine, which has K_i for thrombin of 300 μ M and 31 μ M for trypsin and thus is 10 \times selective for trypsin. This is perhaps unexpected, as it could be argued that benzamidine, being planar, is a much better analogue of the substrate arginine guanidinium group. The width of the recognition pocket (measured from

1) *Note added in proof:* This concept has been published by D.A. Erlanson, A.C. Braisted, D.R. Raphael, M. Randal, R.M. Stroud, E.M. Gor-

don, J.A. Wells, *Proc. Natl. Acad. Sci. U.S.A.* **2000**, 97, 9367–9372.

Gly216 N to Cys191 carbonyl carbon and taken from high-resolution X-ray structures) is 7.6 Å for PPACK, 8.0 Å for benzamidine, and 8.4 Å for napsagatran. The recognition pocket is thus not rigid and in thrombin is able to expand more easily than in trypsin. It is very difficult to see why this is so by inspecting X-ray structures. The only sequence difference between thrombin and trypsin in the whole of the recognition pocket is that thrombin has Ala at 190, whereas trypsin has Ser, but the effect of this difference on needle binding is obscure (see Section 7.9.1). It is quite possible that residues forming a second layer around the recognition pocket help determine the structural variability. Perhaps needle binding is a realm better covered by experiment than theory at the present time.

Secondly, napsagatran bound to thrombin shows very good intramolecular interactions, and the bound conformation was observed to be very similar to that found in crystals of napsagatran alone. This points to the general principles of a “lock-and-key” interaction between enzyme and inhibitor being favorable [31], which is equivalent to saying that the entropy loss of the inhibitor on binding should be low. The two hydrophobic substituents, naphthyl- and cyclopropyl-, pack well together. Such “hydrophobic collapse” of the inhibitor structure in solution is presumed to help pre-form the “key” to fit in the thrombin “lock” [32, 33]. Further, the carboxylate provides a “cap” to the needle, being positioned over the hydrophobic part of the piperidine, protecting it from interaction with water, and the conformation is stabilized by an intramolecular hydrogen bond from an –NH to the carboxylate.

An important principle, much neglected, is that the conformation of a potential inhibitor in solution is also relevant. The observed inhibition constant is a measure of the free energy change in the whole system on mixing enzyme and inhibitor. It is very unfavorable if the inhibitor prefers a conformation in solution of substantially lower energy than when bound to the enzyme.

In general, it is to be expected that, after complex formation, both enzyme and inhibitor will be in low-energy conformations. It seems unlikely, and also unnecessary, that both will be simultaneously in their lowest energy conformations. Just as the topological and geometrical constraints on protein folding lead not infrequently to amino acids – mostly but not always proline – being found in the *cis* rather than the energetically more favorable *trans* conformation, the global energy minimum attained in inhibitor binding may sometimes distribute local energy unevenly.

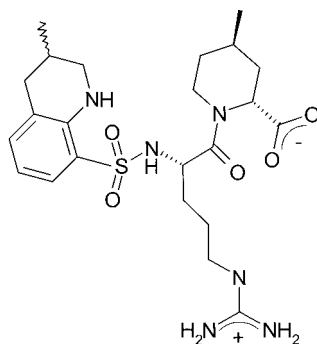
A fundamental principle of inhibitor design, not always easy to achieve in practice, is that when docking enzyme and inhibitor, all reasonably low energy conformations have to be taken into account.

7.7

Lessons from Classical Inhibitors

As early as 1981, chemists at the Mitsubishi chemical company synthesized the thrombin inhibitor MD-805, **6** (Fig. 7.6), which has been extensively tested in humans under the name argatroban [34]. The compound, as might be seen from its

Fig. 7.6 The Mitsubishi compound MD-805 (argatroban) **6**.



structure, was produced by the classical medicinal chemistry approach starting with arginine-containing tripeptides. Secondary amines, ultimately the piperidine, were introduced to prevent processing as substrate. The structure in complex with alpha thrombin has been published by myself [13] and at higher resolution in complex with epsilon thrombin by Bode and Brandstetter [35].

Another early inhibitor is Napap, **7** (Fig. 7.7) [36]. The compound was synthesized as the racemate. The stereochemistry of the binding species was demonstrated by determining the structure in complex with alpha thrombin and was published by myself [13] and, at higher resolution in complex with epsilon thrombin, by Bode and Brandstetter [35].

This inhibitor can be considered a tetra-peptide analogue if the piperidine is taken as replacing a cyclized amino acid.

Fig. 7.8 shows, in simplified form, the binding of the three inhibitors **5–7** to thrombin. All may be considered peptide analogues. All occupy roughly the same volume. Most of the interactions with thrombin are similar or identical. There is, nevertheless, a fundamental difference in the way they bind – the “binding modes” are not all the same. I take here a definition of “binding mode” as being determined by which substituents are in which pockets. If we take the arginine or analogue to be P1 and stretch the definitions somewhat, we have the following binding pattern according to where the residue side chains are found (Tab. 7.1).

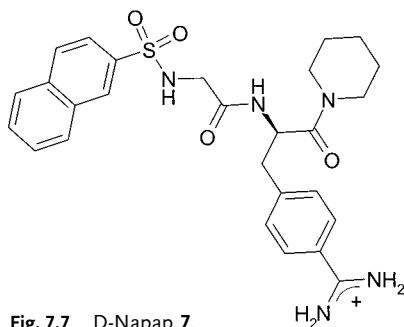


Fig. 7.7 D-Napap **7**.

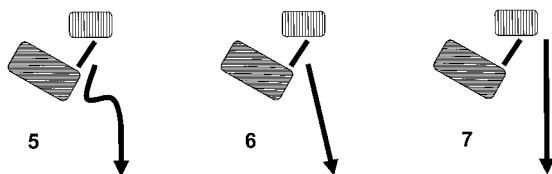


Fig. 7.8 Sketch of the binding modes of napsagatran, 5, MD-805 6, and Napap 7.

Tab. 7.1 Binding patterns of thrombin inhibitors.

	<i>S₁</i>	<i>D pocket</i>	<i>P pocket</i>
Napsagatran	P1	P2	P1'
MD-805	P1	P2	P1'
Napap	P1	P3	P1'

Two different binding modes are seen. For Napap the P1 residue lies substrate-like towards Ser 195, but the D conformation (as opposed to the substrate L conformation) brings the piperidine corresponding to P1' into the P pocket; the glycine P2 acts as a spacer and makes the hydrogen bonds expected for a P3 residue, positioning the P3 naphthyl-sulfonyl correctly to fit into the D pocket. For MD-805 the P1 arginine tilts towards Gly216 and makes there the hydrogen bonds expected for a P3 residue, allowing the P1' piperidine to enter the P pocket and the P2 substituent the D pocket. Surprisingly, napsagatran formally has the same binding mode as MD-805, despite their different origins.

MD-805 originated as a tripeptide with P2, P1, and P1' substituents. To prevent cleavage of the P1-P1' peptide bond, a secondary amide was introduced and optimized to the piperidine shown. What was not appreciated at the time was that the P1 arginine effectively jumped to P3. This is possible, as the structures show, but only if the arginine needle turns so that only one guanidine -NH_2 interacts with Asp189 (Fig. 7.9).

In retrospect, this example reinforces a number of important principles:

1. Expect the unexpected – the smallest change in an inhibitor can cause it to bind totally differently.
2. The tail does not wag the dog – in this case, the guanidine does not make “optimum” interactions with Asp189 but has to settle for “second best” in order to allow a large number of other favorable interactions.
3. As indicated above for napsagatran, “hydrophobic collapse” is a powerful driving force. Models that placed the P1' piperidine in the rather hydrophilic S1' region did not take this into account.
4. A water molecule at a well-defined position helps stabilize the bound conformation (Fig. 7.9).

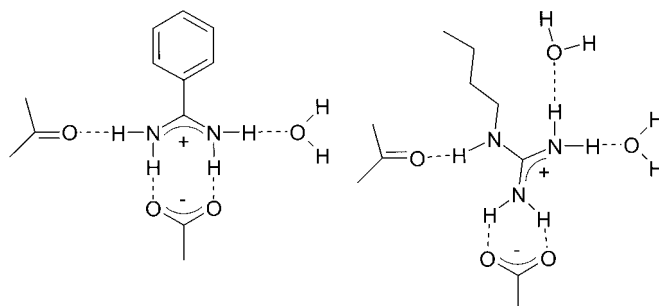


Fig. 7.9 Interactions at the bottom of the S_1 (recognition) pocket. (Left) The canonical benzamidine hydrogen-bonding scheme with two hydrogen bonds to Asp189, one to the carbonyl of Gly219 and one to the conserved

water molecule. (Right) The guanidine of MD-805 makes interactions with the same hydrogen bond acceptors plus an extra water molecule.

5. The carboxylate substituent on the piperidine ring has the equatorial conformation and not the expected axial conformation that has lower energy taken in isolation. As previously discussed in detail [13], this allows better packing in the P pocket than any other conformation of this or any other stereoisomer. Modeling suggests that packing of the quinoline against the piperidine also gives preference to this conformation, so hydrophobic collapse may be the primary driving force. At any rate, there is a local conformation, which is not in its lowest energy state.

Napsagatran was evolved progressively from the 3-substituted amidino-piperidine needle. The exit vector from the needle is such that amide extension first interacts with the rim of the S_1 pocket by accepting a hydrogen bond from the $-NH$ of Gly217 at the front of the pocket. This excursion of the “extended needle” allows the central amino acid (regarding the extended needle as side chain) to achieve an “ideal P3” hydrogen-bonding interaction with Gly216, being in plane with Gly216 and close to the position of the P3 glycine in the fibrinopeptide structures. This repositioning of the central amino acid backbone then requires a smaller P pocket substituent. The possible energy/entropy losses associated with an extended needle are compensated by the fact that all torsional angles are close to ideal values and that the amide makes two good hydrogen bonds – to Gly217 and to the inhibitor carboxylate.

Superimposition of the three lead inhibitor and fibrinopeptide structures indicates the following design rules for thrombin inhibitors:

1. There has to be a basic group in close interaction with Asp189.
2. There is a small volume deep in the P pocket that must be occupied by a hydrophobic group.
3. There is a small volume deep in the D pocket that must be occupied by a hydrophobic group (over CE3 of Trp215).
4. Non-aromatic residues are preferred in the P pocket.

5. Aromatic residues are preferred in the D pocket.
6. Both “anti-parallel beta” hydrogen bonds to Gly216 must be made.
7. A carboxylate or carbonyl is preferred near Ser195.
8. It is not a requirement that the oxy-anion hole be occupied.

These heuristic rules are there to be broken but have to be kept in mind, as most will have to be satisfied most of the time. To progress further, we need to try to quantitate the energetics of the enzyme-inhibitor interaction.

7.8

Estimating the Energies of Interactions

There are many approaches to the energetics of intermolecular interactions (*vide infra*). Here we document some cases where thrombin inhibitors have been used to provide energy estimates.

Obst, Diederich, and coworkers generated by rational design a series of thrombin inhibitors with rigid, bicyclic core structures [37]. These were further extended to tricyclic structures and modified specifically “to generate detailed information on the strength of individual intermolecular bonding interactions and their contribution to the overall free energy of complexation” [38]. The general formula of the inhibitors is given in Fig. 7.10. For details of synthesis and stereochemistry, please refer to the original articles. Here, two pairs of inhibitors will be presented for which there is evidence from high-resolution X-ray structures that the pairs bind in exactly the same way. This is most important, since, as discussed above, quite similar inhibitors might bind differently and also the protein might adapt to the inhibitor.

The reference inhibitors **8a** have carbonyl oxygens at both R1/R2 and R3/R4 and bind with the lower carbonyl oxygen accepting a hydrogen bond from the –NH of Gly216 (as modeled) and the upper carbonyl oxygen in the P pocket.

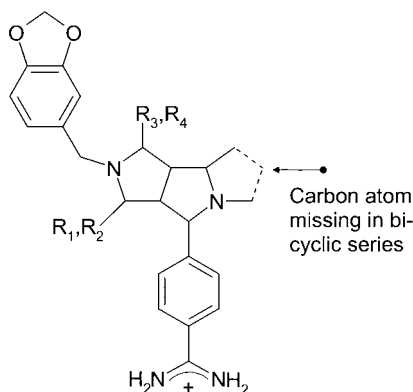


Fig. 7.10 General formula of the bi- and tricyclic thrombin inhibitors **8** from Obst et al. [37, 38] (simplified representation).

8a (R1/R2)=O, (R3/R4)=O; **8b** R1=H, R2=H, (R3/R4)=O; **8c** (R1/R2)=O, R3=H, R4=CH₂(CH₃)₂.

A first inhibitor **8b** was generated (in the bicyclic series) with R1=H, R2=H, i.e., the lactam carbonyl was replaced by a methylene group. The only difference in the X-ray structures of this and the reference bicyclic inhibitor was the lack of the carbonyl oxygen atom. The energy difference between the two compounds as derived from K_i measurements is $\Delta\Delta G=0.8 \text{ kcal mol}^{-1}$. We can directly equate this with the energy of the lost hydrogen bond, since there are no other differences apparent. This is at the lower end of the generally accepted range (0.5–1.5 kcal mol^{-1}). The authors propose a number of reasons for this. First, residue glycine 216 is planar, and the phenylamidinium residue of the inhibitor stacks parallel to it so that its π -electrons will tend to stabilize a partial positive charge on the –NH of Gly216. The resultant antiparallel dipoles of the glycine –NH and C=O will then tend to stabilize each other. Second, because Gly216 is planar, repulsion between the inhibitor carbonyl group and the Gly216 carbonyl group will prevent the hydrogen bond from having optimal geometry (and thus optimal energy). Finally, the interaction between the methylene group on the inhibitor and the methylene group of Trp215 are assessed as positive, albeit small.

There is a complementary way of looking at this problem. We may simply ask, does a water molecule prefer to bind here when no inhibitor is present? The answer seems to be “No,” which supports the above argument.

A second inhibitor pair was synthesized, this time in the tricyclic series with R1/R2 as carbonyl, with R3 as H, and with variation of R4 with R stereochemistry. This was done since it was clear that a carbonyl group in the P pocket does not obey the rules for optimum binding. The reference tricyclic **8a** compound is, in fact, quite potent, with a K_i for thrombin of 90 nM, but with R4 as isopropyl **8c**, the K_i improved to 13 nM. This corresponds to an improvement in binding energy of $\Delta\Delta G=1.1 \text{ kcal mol}^{-1}$. This is quite substantial, but given that the structure closely resembles the S_2 valine of the natural substrate, it is perhaps less than might have been expected. Close inspection of the X-ray structures shows that the isopropyl substituent is fractionally too large and pushes the tricycle out by about 0.5 Å without a significant change in the protein structure. Cyclopropyl and ethyl are slightly better substituents, with K_i 's of 10 nM and 8 nM, respectively. Nevertheless, the carbonyl group does better than expected, probably because the P pocket is not fully closed. There is a smear of residual electron density between this carbonyl oxygen and the –NH₃ of Lys60F, which can be interpreted as a poorly ordered water molecule partially solvating the carbonyl oxygen.

Another inhibitor series has been used to estimate the value of P pocket interactions – the Boehringer Mannheim diaryl sulfonamides [39] (also reported by 3-Dimensional Pharmaceuticals [40–43]) (Fig. 7.11).

The diaryl sulfonamide inhibitors were discovered by a screening exercise aimed at finding less basic thrombin inhibitors [44]. A crystal structure of the complex of thrombin with the R=CH₃ compound **9b** (BM14.1248) shows the phenyl group in the D pocket, the central tolyl group in the P pocket, and the 4-aminopyridine in the S1 pocket (but not interacting directly with Asp189) [45] [*luvt*]. **9b** has a K_i of 23 nM for thrombin, whereas the R=H compound **9a** has a K_i of 300 nM [39]. This corresponds to $\Delta\Delta G=1.5 \text{ kcal mol}^{-1}$. As discussed by the

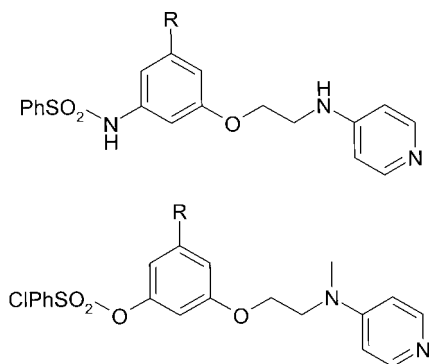


Fig. 7.11 (Above) Diaryl sulfonamide inhibitors from Boehringer Mannheim, **9a** $R=H$, **9b** $R=CH_3$. (Below) A similar inhibitor from 3-Dimensional Pharmaceuticals, **10** $R=CH_3$.

authors [45], this is certainly lower than the energetic cost of a “hole” left by deleting a methyl group ideally packed deep into a hydrophobic pocket. The P pocket is sufficiently flexible to compensate somewhat for the loss of the methyl group: the Tyr-Pro-Pro-Trp loop moves down by 1.0–1.5 Å (R. Engh, personal communication).

The corresponding $R=CH_3$ compound **10** from 3-Dimensional Pharmaceuticals has a K_i for thrombin of 11 nM and shows the same binding mode [40, 43].

In a similar way, using 4-TAPAP as template [35], a methyl group deep in the P pocket was shown to produce an affinity gain of 17× with no change to the position of inhibitor binding [46].

We conclude from the above examples that up to ~ 2 kcal mol $^{-1}$ of binding energy may be obtained by placing a methyl or similar small hydrophobic group correctly in the P pocket. Something similar must be true of the D pocket, although examples with X-ray validation are missing.

We also conclude that placing a hydrogen-bond acceptor correctly above the $-NH$ of Gly 216 is worth only <1 kcal mol $^{-1}$.

We further observe that it is possible to obtain low nanomolar inhibition without making the canonical “benzamidine” hydrogen bonds (Fig. 7.9). For a detailed discussion of “non-canonical needles” and recognition pocket and P pocket flexibility, the reader should consult [41, 42, 45].

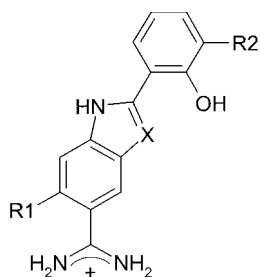
7.9

Water and Solvent

Hydrogen bonding in general is reviewed in the previous chapter. Here, the discussion will be limited to the role of water and other solvent molecules in inhibitor binding.

The structure of the solvent around biological macromolecules has been reviewed in detail by Mattos and Ringe [47]. Serine proteases of the trypsin family have 21 conserved buried water molecules, as first reported by Sreenivasan and Axelsen [48]. These may be regarded as integral to the protein structure, and it

Fig. 7.12 Inhibitors from Axy: **11a** APC-8696 R1=H, R2=, X=CH; **11b** APC-10302 R1=Cl, R2= ϕ , X=CH, **11c** APC-1144 R1=H, R2=H, X=N.



might be expected that they would be difficult to displace. Dunitz [49] estimates the entropic cost of “freezing” in a water molecule as part of the protein structures as up to 2 kcal mol^{-1} , which is a quite significant penalty.

High-resolution X-ray structures of thrombin in the Protein Data Bank show rather variable total numbers of water molecules, presumably according to the preferences of the depositors. A generally accepted number is around one water per amino acid [50], i.e., ~ 300 for thrombin. Of these, the only ones of direct interest here are the conserved water at the bottom of the S1 pocket (Fig. 7.9); possible waters hydrogen bonding to the -NHs of Gly216 and Gly219; and whatever waters are in the S1, P, and D pockets and are normally displaced by inhibitors.

Most of water molecules found around the active site do not appear to be particularly difficult to displace. In particular, there is no highly conserved water structure around the oxyanion hole or around Ser 195. These catalytically important features are not “frozen” in “ice-like” water but rather are intrinsically able to adapt to substrate, transition state, and product structures.

An elegant, detailed structural description of an unusual multi-centered short hydrogen-bonding network, induced by the binding of Axy inhibitors of the **11c** APC-1144 type (Fig. 7.12), is given by Katz et al. [51]. Those interested in the possibilities for interaction with the residues responsible for the catalytic mechanism and in the roles of water and pH should consult this reference.

7.9.1

Displacing a Tightly Bound Water

The conserved S1 water (Fig. 7.9) is one of the best defined, as judged by the low B -values (thermal disorder parameters) observed with benzamide- or guanidine-type inhibitors. It was long regarded as simply part of the protein until Katz et al. [52] produced inhibitors that displaced it. The objective was to improve selectivity between those serine proteases that have serine at position 190, e.g., uPA, trypsin, trypsinase, and those that have alanine at 190, e.g., tPA, thrombin, factor X. The side chain of residue 190 is spatially close to the conserved water, and this region can be accessed by benzamidine substituents ortho- to the amidine (Fig. 7.12).

Katz et al. [52] report binding constants for **11a** APC-8696 of $K_i=130 \text{ nM}$ for trypsin and $K_i=320 \text{ nM}$ for thrombin. The compound is thus trypsin selective by

2.5 \times . Upon introduction of the chlorine substituent to give **11b** APC-10302, the binding constant for trypsin becomes $K_i=230$ nM and $K_i=60$ μ M (60,000 nM!) for thrombin. **11b** is thus 260 \times trypsin selective. Katz et al. [52] provide a wealth of both binding data and X-ray structural data on uPA, thrombin, and trypsin and discuss in detail how the protein structure responds to inhibitor binding. This excellent set of high-resolution structures [1gjb, 1gjc, 1gj7, 1gj8, 1gj9, 1gja, 1gjd, 1gj4, 1gj5, and 1gj6] has recently been made public (May 2002) and will be a profitable subject for analysis. Only a preliminary summary can be given here.

It might have been expected that the chlorine would interact favorably with the alanine 190 side chain in thrombin and less well with the serine 190 –OH, giving selectivity for thrombin, but the inhibition constants reveal just the opposite. Trypsin is favored by $\Delta\Delta G=3$ kcal mol⁻¹, although the conserved water is indeed removed from both enzymes and the general binding mode is the same.

The conserved water donates hydrogen bonds to the carbonyl oxygen of residue 227 and to the π -electrons of Tyr228. Both of these interactions are lost when the water is displaced, but the energy change will be similar for both enzymes.

The most important difference seems to be that in thrombin there is no fourth hydrogen bond partner for the amidino group, and this hydrogen bond is totally lost. Further, the chlorine is not quite large enough to fill the pocket left by the water, so a “hole” is generated, which as we have already seen, costs energy.

In trypsin the nearer –NH₂ of the inhibitor amidino group donates a hydrogen bond to the Ser190 –OH, which in turn donates a hydrogen bond to the –OH of Tyr228. The side chain of Tyr228 moves inward slightly and thus contacts the chlorine atom, filling the potential “hole.” The chlorine can thus displace the trypsin water at no net energy loss, as all hydrogen-bonding capabilities of the amidino group and the Ser190 O γ are satisfied and good close packing is achieved.

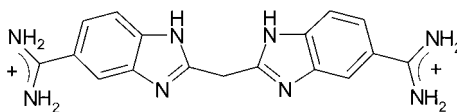
7.9.2

Binding of Solvent Molecules

Besides water, proteins also can bind a wide range of other molecules on their surface. Many of those found in structures in the Protein Data Bank have clearly been introduced to promote crystal formation. In recent years, nearly all crystal structures reported have been analyzed in the frozen state, and in many cases glycerol or another cryoprotectant has been added to aid in the crystal-freezing process. While the binding of such molecules from crystallization or freezing buffers may not be of direct biological relevance, specific binding sites are often identified that can deliver information on the preferred binding of small ligands, which then has predictive value for inhibitor design. A logical extension of this observation is to actively produce crystal structures in the presence of high concentrations of small “probe” molecules and thus produce an experimental binding map of the protein surface. This can then be used instead of, or in combination with, the theoretically derived functions used for *in silico* screening.

Mattos and Ringe have analyzed protein surfaces [53], reviewed “proteins in organic solvents” [54], and discussed the use of such information in inhibitor design [55].

Fig. 7.13 Bis(5-amidino-2-benzimidazolyl) methane (BABIM) **12**.



English et al. have studied thermolysin in high concentrations of organic solvents [56, 57]. It is to be expected that this kind of experimental approach will be extensively used in the future, both to give design ideas in specific cases and to help improve prediction methods in general. Bartlett et al., for instance, have used increasing ethanol concentrations to help estimate the hydrophobic contribution to inhibitor binding [58].

Solvent or water molecules can be “recruited” by inhibitors to enable them to bind better. For example, the early thrombin-inhibitor complexes MD-805 and PPACK have water strongly bound to the guanidine group in the recognition pocket, in MD-805 it makes a bridge to the inhibitor carboxylate, while in PPACK it makes a bridge to the inhibitor N-terminal -NH_3 .

This, and the recruitment of common ions such as chloride, should be regarded as the norm and taken into account in the inhibitor design process. Ladbury [59] has analyzed the way inhibitors recruit water molecules but concluded that at present it is difficult to predict such behavior.

An unusual observation is the recruitment of a zinc ion by serine protease inhibitors, reported by Katz et al., which is analyzed structurally in fine detail in [60–63]. The simplest compound, **12**, showing this “delta effect” (greater potency in physiological buffers or plasma than in EDTA-containing assay buffers, i.e., an inverse plasma shift) is bis(5-amidino-2-benzimidazolyl) methane (BABIM) (Fig. 7.13). The “delta effect” gives increases of affinity of greater than $1000\times$ in the presence of Zn^{2+} . It will be interesting to see whether this paradigm can be extended to the recruitment of other (physiological) molecules.

7.9.3

Screening

While “screening” and “design” are commonly seen as opposite approaches to drug finding, it has to be pointed out here that screening by physical methods is an extremely useful way of mapping an active site. Abbott has developed this approach in extensive studies, principally on urokinase, using X-ray [64–73] and Raman screening methods [74] in addition to their SAR by NMR approach [75–78]. As higher throughput X-ray and NMR technologies are developed, it is to be expected that this kind of experimental approach will be used more and more.

7.10

Structure-Activity Relationships (SAR)

In Section 7.7, the binding modes revealed by some crystal structures of thrombin-inhibitor complexes were discussed. Inhibitor studies on thrombin have been

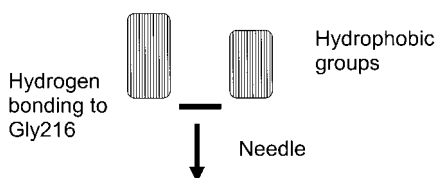


Fig. 7.14 Symmetry of a typical thrombin inhibitor.

complicated by the tendency of small changes in the inhibitor backbone to change the binding mode [28, 79]. This occurs partly because the inhibitors have some internal symmetry, as depicted in Fig. 7.14

As this behavior is to be expected in other projects, it is perhaps worth commenting on a general principle – the use of structure-activity relationships (SAR). If even a few variants of a molecule are available, it is normally possible to identify the binding mode by inspecting the SAR. If there is any doubt, it may be worth deliberately making a few test compounds: normally, there is a position where substituents of one type are allowed in one binding mode but not in the other.

A particular issue in thrombin inhibitors has been stereochemistry. The SAR of MD-805 and NAPAP and analogues, for example, was hard to understand without knowing that only the D isomer of NAPAP binds to thrombin. It has been very helpful on many occasions to determine X-ray structures by soaking racemic mixtures into crystals and observing which stereoisomer binds.

It is now possible to produce a high-resolution structure of a thrombin-inhibitor complex in a day or so, which is quicker and easier than chiral separation. Where selectivity against related enzymes is an issue, which it certainly is for thrombin, it has been repeatedly found that this can only be understood, and thus improved, if the binding modes to these other enzymes can also be determined. This is particularly true where diastereomers are involved.

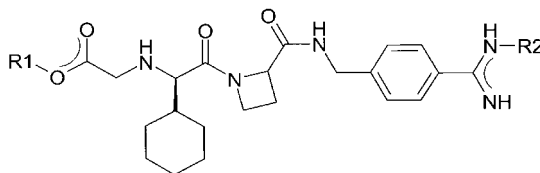
7.11

Present Clinical Status of Thrombin Inhibitors

Other than argatroban, the only non-peptidic thrombin inhibitors to have reached phase III clinical trials are the simple tripeptide analogue melagatran and its orally available pro-drug ximelagatran from AstraZeneca [5] (Fig. 7.15). For a recent clinical status review, see [80].

Given the enormous worldwide efforts over many years to develop thrombin inhibitors as anti-thrombotic drugs, this is rather disappointing. Until more results of early-phase clinical trials are published, it will not be clear whether the problem is that thrombin is a difficult drug target, e.g., because of bleeding risks, or whether the inhibitors proposed as clinical candidates simply do not have sufficiently drug-like properties. The suspicion is that tripeptide analogues with a strongly basic group have intrinsically poor pharmacokinetic and pharmacodynamic properties [8]. If this is true, new and different inhibitors are needed, and

Fig. 7.15 Melagatran **13** and its pro-drug ximelagatran **14**. Melagatran is the unsubstituted parent compound. Ximelagatran has R1 = CH_2CH_3 , R2 = OH .



here a good understanding of the interactions between the inhibitor and the enzyme active site can contribute significantly to the identification of inhibitor molecules suitable for development as drugs.

7.12

Conclusions

Thrombin inhibition is a fruitful source of raw data for the study of molecular recognition. Several groups have determined, published, and deposited coordinates for sets of high-resolution X-ray crystal structures. In combination with binding and kinetic data, it is now possible to “map” the thrombin active site in some detail in terms of both structural changes and the energies of interactions.

Heuristic models based on the binding of peptidomimetic inhibitors pointed to hydrophobic interactions in the D and P pockets and optimal hydrogen bonding to Gly216 and Asp189 as being vital for good inhibition.

More recent experience, coming from a wide variety of sources, shows that for acceptable affinity the hydrophobic interactions have to be maintained and good surface complementarity is essential (no holes), but it is not a requirement that all possible hydrogen bonds are made to Gly216 and Asp189.

Given the progress made so far, it is to be hoped that systems such as thrombin will continue to be actively employed to further our understanding of inter-molecular interactions.

7.13

Acknowledgments

I would like very much to thank all those excellent colleagues with whom I have had the great privilege to work and publish on thrombin, in particular Ulrike Obst for checking this manuscript. I thank also Fritz Winkler and Hans-Joachim Böhm for continued support and encouragement.

7.14

References

- 1 WILEY, M.R. and FISHER, M.J., *Exp. Opin. Ther. Pat.*, **1997**, 7(11), 1265–1282.
- 2 COBURN, C.A., *Expert Opin. Ther. Pat.*, **2001**, 11(5), 721–738.
- 3 ANAND, S., *Haemostasis*, **1999**, 29 (Suppl. 1), 76–78.
- 4 FRENCH, J.K. and WHITE, H.D., *Curr. Card. Rep.*, **1999**, 1(3), 184–191.
- 5 GUSTAFSSON, D., et al., *Thrombosis Res.*, **2001**, 101(3), 171–181.
- 6 HAUPTMANN, J., *Nova Acta Leopold.*, **1999**, 80(311), 61–82.
- 7 HAUPTMANN, J. and STURZEBECHER, J., *Thromb. Res.*, **1999**, 93(5), 203–241.
- 8 HAUPTMANN, J., *Eur. J. Clin. Pharmacol.*, **2002**, 57(11), 751–758.
- 9 HIRSH, J., *Am. Heart J.*, **2001**, 142(2, Suppl.), S3–S8.
- 10 MENEAR, K., *Expert Opin. Invest. Drugs*, **1999**, 8(9), 1373–1384.
- 11 SHEN, G.X., *Curr. Drug Targets: Cardio-vasc. & Haematol. Disord.*, **2001**, 1(1), 41–49.
- 12 SCHECHTER, I. and BERGER, A., *Biochem. Biophys. Res. Commun.*, **1968**, 32(5), 898–902.
- 13 BANNER, D.W. and HADVARY, P., *J. Biol. Chem.*, **1991**, 266(30), 20085–20093.
- 14 WILKINSON, A.S., et al., *Biochemistry*, **1999**, 38(13), 3851–3856.
- 15 CHITTOCK, R.S., et al., *Biochem. J.*, **1999**, 338(Pt 1), 153–159.
- 16 MARQUART, M., et al., *Acta Cryst.*, **1983**, B39, 480–490.
- 17 BANNER, D.W., *Nature*, **2000**, 404(6777), 449–450.
- 18 KETTNER, C. and SHAW, E., *Thrombosis Res.*, **1979**, 14(6), 969–973.
- 19 MATTOS, C., et al., *Biochemistry*, **1995**, 34(10), 3193–3203.
- 20 DING, X., et al., *Biochemistry*, **1994**, 33(31), 9285–9293.
- 21 DING, X., et al., *Biochemistry*, **1995**, 34(23), 7749–7756.
- 22 MATTOS, C., et al., *Nat. Struct. Biol.*, **1994**, 1(1), 55–58.
- 23 PEISACH, E., et al., *Science*, **1995**, 269(5220), 66–69.
- 24 KATONA, G., et al., *J. Biol. Chem.*, **2002**, 14, 14.
- 25 SKORDALAKES, E., et al., *J. Mol. Biol.*, **2001**, 311(3), 549–555.
- 26 BOHM, H.J., D.W. BANNER, and WEBER, L., *J. Comput. Aided Mol. Des.*, **1999**, 13(1), 51–56.
- 27 BOHM, H.J., *J. Mol. Recognit.*, **1993**, 6(3), 131–137.
- 28 HILPERT, K., et al., *J. Med. Chem.*, **1994**, 37(23), 3889–3901.
- 29 WEBER, L., et al., *Angew. Chem. Int. Ed. Engl.*, **1995**, 107, 2453–2454.
- 30 ILLGEN, K., et al., *Chem. Biol.*, **2000**, 7(6), 433–441.
- 31 VERLINDE, C.L. and HOL, W.G., *Structure*, **1994**, 2(7), 577–587.
- 32 HART, P.A. and RICH, D.H., *Pract. Med. Chem.*, **1996**, 393–412.
- 33 RICH, D.H., *Perspect. Med. Chem.* **1993**, 15–25.
- 34 KIKUMOTO, R., et al., *Biochemistry*, **1984**, 23(1), 85–90.
- 35 BRANDSTETTER, H., et al., *J. Mol. Biol.*, **1992**, 226(4), 1085–1099.
- 36 STURZEBECHER, J., et al., *Thromb. Res.*, **1983**, 29(6), 635–642.
- 37 OBST, U., et al., *Angew. Chem. Int. Ed. Engl.*, **1995**, 34, 1739–1742.
- 38 OBST, U., et al., *Chem. Biol.*, **1997**, 4(4), 287–295.
- 39 WEBER, I.R., et al., *Bioorg. Med. Chem. Lett.*, **1998**, 8(13), 1613–1618.
- 40 LU, T., et al., *Bioorg. Med. Chem. Lett.*, **1998**, 8(13), 1595–1600.
- 41 LU, T., et al., *Bioorg. Med. Chem. Lett.*, **2000**, 10(1), 83–85.
- 42 LU, T., et al., *Bioorg. Med. Chem. Lett.*, **2000**, 10(1), 79–82.
- 43 BONE, R., et al., *J. Med. Chem.*, **1998**, 41(12), 2068–2075.
- 44 VON DER SAAL, W., et al., *Bioorg. Med. Chem. Lett.*, **1997**, 7(10), 1283–1288.
- 45 ENGH, R.A., et al., *Structure*, **1996**, 4(11), 1353–1362.
- 46 BERGNER, A., et al., *J. Enzyme Inhib.*, **1995**, 9(1), 101–110.
- 47 MATTOS, C. and RINGE, D., *International Tables for Crystallography*, **2001**, 623–647.
- 48 SREENIVASAN, U. and AXELSEN, P.H., *Biochemistry*, **1992**, 31(51), 12785–12791.
- 49 DUNITZ, J., *Science*, **1994**, 264, 670–670.

- 50 CARUGO, O. and BORDO, D., *Acta. Cryst. D: Biol Cryst*, **1999**, D55(2), 479–483.
- 51 KATZ, B.A., et al., *J. Mol. Biol.*, **2001**, 307(5), 1451–1486.
- 52 KATZ, B.A., et al., *Chem. Biol.*, **2001**, 8(11), 1107–1121.
- 53 RINGE, D. and MATTOS, C., *Med. Res. Rev.*, **1999**, 321–331.
- 54 MATTOS, C. and RINGE, D., *Curr. Opin. in Struct. Biol.* **2001**, 11(6), 761–764.
- 55 MATTOS, C. and RINGE, D., *Nature Biotechnol.*, **1996**, 14(5), 595–599.
- 56 ENGLISH, A.C., GROOM, C.R., and HUBBARD, R.E., *Protein. Eng.*, **2001**, 14(1), 47–59.
- 57 ENGLISH, A.C., et al., *Proteins: Struct. Funct. Gen.*, **1999**, 37(4), 628–640.
- 58 BARTLETT, P.A., et al., *J. Am. Chem. Soc.*, **2002**, 124(15), 3853–3857.
- 59 LADBURY, J.E., *Chem. Biol.*, **1996**, 3(12), 973–980.
- 60 KATZ, B.A., et al., *Nature*, **1998**, 391(6667), 608–612.
- 61 KATZ, B.A. and LUONG, C., *J. Mol. Biol.*, **1999**, 292(3), 669–684.
- 62 JANC, J.W., et al., *Biochemistry*, **2000**, 39(16), 4792–4800.
- 63 KATZ, B.A., et al., in *Special Publication – Royal Society of Chemistry*, **2001**, 211–222.
- 64 NIENABER, V.L., et al., in U.S. **2001**, (Abbott Laboratories, USA): US. p. 33 pp., Cont.-in-part of U.S. Ser. No. 36, 184.
- 65 MCCLELLAN, W.J., et al., in *Abstr. Pap. – Am. Chem. Soc.* **2001**. p. MEDI-294.
- 66 ROCKWAY, T.W., et al., in *Abstr. Pap. – Am. Chem. Soc.* **2000**. p. MEDI-030.
- 67 NIENABER, V.L., et al., *Structure (London)*, **2000**, 8(5), 553–563.
- 68 NIENABER, V., et al., *J. Biol. Chem.*, **2000**, 275(10), 7239–7248.
- 69 NIENABER, V.L., et al., in PCT Int. Appl. **1999**, (Abbott Laboratories, USA): WO. p. 57.
- 70 NIENABER, V.L., MERSINGER, L.J., and KETTNER, C.A., *Biochemistry*, **1996**, 35(30), 9690–99.
- 71 NIENABER, V.L., et al., *Nature Biotech.*, **2000**, 18(10), 1105–1108.
- 72 HAJDUK, P.J., et al., *J. Med. Chem.*, **2000**, 43(21), 3862–3866.
- 73 NIENABER, V.L., BOXRUD, P.D., and BERLINER, L.J., *J. Prot. Chem.*, **2000**, 19(4), 327–333.
- 74 DONG, J., et al., *Biochemistry*, **2001**, 40(33), 9751–9757.
- 75 HAJDUK, P.J., et al., *J. Am. Chem. Soc.*, **1997**, 119(25), 5818–5827.
- 76 FESIK, S.W., et al., *Protein Eng.*, **1997**, 10(Suppl.), 73.
- 77 SHEPPARD, G.S., et al., in Book of Abstracts, 213th ACS National Meeting, San Francisco, April 13–17. **1997**. p. MEDI-081.
- 78 PETROS, A.M. and FESIK, S.W., *Methods Enzymol.*, **1994**, 239 (Nuclear Magnetic Resonance, Pt. C), 717–739.
- 79 BANNER, D., et al., *Perspect. Med. Chem.* **1993**, 27–43.
- 80 STEINMETZER, T., HAUPTMANN, J., and STURZEBECHER, J., *Expert. Opin. Investig. Drugs.*, **2001**, 10(5), 845–864.

8

Tailoring Protein Scaffolds for Ligand Recognition

A. SKERRA

8.1

Introduction

A major goal of medicinal chemistry is the design of low-molecular-weight ligands that bind to target proteins in a tight and specific manner. In the case of enzymes, these ligands act as inhibitors or allosteric effectors, while in the case of transmembrane receptors, they serve as agonistic or antagonistic signaling molecules. Ligands of these types have conventionally been derived from natural compound libraries and, more recently, via combinatorial synthesis. The quickly growing number of proteins with known three-dimensional structure and the significant methodological improvements in the structural elucidation of proteins during the past decade [1] – employing X-ray crystallography or nuclear magnetic resonance (NMR) techniques – have strongly promoted the computer-aided drug-design approach. Especially enzyme inhibitors can now be readily constructed on the basis of structural information about the target macromolecule [2]. Nevertheless, in the case of receptor targets, the rational prediction of cognate compounds is still hampered due to the inherent difficulties associated with their crystallization or NMR study.

An inverse task is given when there is demand for a macromolecule that specifically binds a small ligand. This question has only recently been addressed by peptide chemistry. For example, antiparallel bundles of four α -helices, which were assembled on a cyclic peptide structure as template, have been used to create hydrophobic cavities for heme as a low-molecular-weight compound [3]. The specific complexation of Fe^{III} · protoporphyrin IX was facilitated by the proper positioning of liganding His residues. While this approach could be interesting from the perspective of rational protein design, it may be limited to special applications, and detailed structural information about the complex is not yet available.

Deeper mechanistic insight into the molecular recognition of small molecules has been gained from antibodies, a class of natural proteins that have traditionally served as specific binding agents for a variety of “hapten” ligands [4]. Numerous practical applications exist for such antibodies in the fields of medical diagnostics as well as bioanalytics, where so-called immunochemical methods provide a quick, inexpensive, and reliable method for the sensitive detection of metabolites

and even xenobiotic compounds [5]. There are also examples of hapten-binding antibodies with clinical use, e.g., in the therapeutic treatment of poisoning with cardiac steroids like digoxin. In this case an F_{ab} fragment of an antibody with high affinity towards the small molecule is administered, which removes the free compound from blood circulation, prevents it from binding to the cell surface receptor, and makes it amenable to renal filtration or degradation in the liver [6].

With respect to the generation of cognate ligand-receptor proteins, one disadvantage is that low-molecular-weight compounds as such cannot be directly used for the immunization of animals. Rather, these haptens must be conjugated to macromolecular carriers in order to elicit an effective immune response. Nevertheless, attempts to generate antibodies with high affinities and specificities against small ligands have often remained unsuccessful. Two potential problems need to be considered in this context. First, when antibodies are to be raised against metabolically occurring substances, they might interfere with physiological processes. Furthermore, if the compounds are toxic, immunization may not be possible at all. Second, and more generally, antibodies were probably evolved by the immune system mainly for the recognition of proteins or other macromolecular targets (like nucleic acids or oligosaccharides) rather than low-molecular-weight compounds.

This notion is supported by the crystal structures of various antibody fragments in complex with either antigens or haptens. In the first case an extended interface is formed between the antigen-binding site of the antibody (the paratope) and the macromolecular target. Typically, a surface of approximately 800 \AA^2 is buried and at least five of the six hypervariable loops (complementarity-determining regions, CDRs) – possibly even together with residues from the structurally conserved framework regions of the antibody variable domains – are involved [7, 8]. The shape of the combining site is often flat but also can be slightly concave or convex.

In contrast, in the case of haptens the mode of interaction with the paratope is much more restricted because a pocket for the ligand needs to be formed in order to provide a sufficient number of interactions that ensure tight complex formation. This pocket is usually located at the interface between the pair of variable domains from the light and heavy chains of the antibody. Hence, a cleft must be formed whose shape is mainly determined by the two CDR-3 loops protruding from the V_H and V_L domains, which are related by a pseudo C_2 -symmetry axis. Because a minimal hydrophobic contact area between V_H and V_L is required in order to maintain the non-covalent domain association, the size of this pocket is limited, and many ligands therefore become just partially buried when bound to an antibody (see Section 8.4).

In addition to this structural consideration, there are empirical observations from recombinant antibody technology indicating that it is difficult to generate an antibody fragment with exquisite specificity towards a hapten using combinatorial libraries cloned from unimmunized donor gene pools or derived from synthetic genes. In contrast, the *in vitro* selection of high-affinity antibodies against proteins is nowadays a routine procedure. Consequently, the recruitment of alternative protein classes for the generation of small ligand receptors has attracted attention [4].

To this end, the concept of using a scaffold – which means a protein architecture with high intrinsic stability – to create a binding site for the specific interaction with the target molecule has gained interest. An appropriate protein scaffold should provide a rigid folding unit that spatially brings together several exposed loops that form a continuous and extended interface such that multiple interaction with the target and hence tight binding are ensured. Ideally, such a scaffold should have structurally partitioned the generic information and stability of its polypeptide fold on the one hand and the local shape and molecular recognition function of its active site on the other (Fig. 8.1).

Initially, this approach has had remarkable success in the generation of artificial binding proteins towards “protein antigens” (for a general review, see [9]). Several single domain proteins that belong to the generic immunoglobulin (Ig) fold, thus supporting a set of two or three hypervariable peptide loops on one end of a sandwich of β -sheets, have proven to be suitable for the recognition of such macromolecular targets. Prominent examples include an individual fibronectin III domain [10] as well as certain V_H domains derived from camel or llama Ig [11], which constitute soluble globular proteins even in the absence of a cognate V_L domain.

Typically, these scaffold proteins exhibit a wedge-shaped structure with the set of variable loops located at the tip in close mutual neighborhood (Fig. 8.1). Therefore, they seem to be particularly suited for complex formation at a groove on the surface of the target protein. In many cases this corresponds to the active site, and, indeed, effective enzyme inhibitors have been generated using the camel V_H -domain approach. In contrast, the structural complexation of small molecules is difficult to achieve with this scaffold. So far only cameloid antibodies recognizing rather large azo dye compounds as haptens have been described [12].

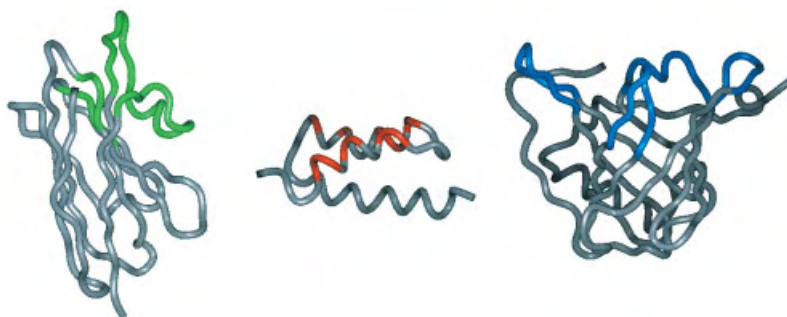


Fig. 8.1 Three types of scaffolds, with a convex, flat, and concave interface, respectively. (Left) Camel V_H domain, here presenting its extended CDR-3 loop towards the active site of an enzyme (PDB entry 1MEL). (Middle) Protein A with its interface made of two α -helices directed against an Ig constant domain (PDB entry 1BDD). (Right) Bilin-binding pro-

tein with its pocket formed by four loops at the open end of the β -barrel structure for the complexation of biliverdin IX $_\alpha$ (PDB entry 1BBP). Loops or amino acid positions that are important for the molecular recognition of the respective target and that may be amenable to side chain exchanges are colored.

Another type of scaffold that has been successfully used for the recognition of prescribed target proteins originates from the bacterial immunoglobulin receptor protein A. So-called “affibodies” were obtained by reshaping the natural Ig-binding interface of the Z domain of protein A, which is formed by a side-by-side pair of α -helices [13]. An essentially flat surface is generated in this manner (Fig. 8.1), which can probably pack against a patch with low curvature on the target protein.

Compared with the generation of recombinant receptor proteins against macromolecular targets, the recognition of small, hapten-like compounds obviously poses a greater challenge. To this end, a pocket with complementary shape must be created in order to enable the burial of a significant area of hydrophobic surface and to provide a sufficient number of protein-ligand interactions – van der Waals contacts, hydrogen bonds, and possibly salt bridges – such that practically useful dissociation constants in the nanomolar range result. In fact, these stringent demands still make it rather difficult to apply rational design principles to the creation of cognate receptor proteins; instead, their construction has to rely on the powerful methodology of combinatorial biochemistry that is available today.

Nevertheless, there exist only a few protein families in nature whose function lies in the plain complex formation with small molecules – as opposed to their biochemical conversion by enzymes or to the triggering of cellular signals via membrane or nuclear receptors. One example is given by the periplasmic nutrient-binding proteins that are found in *Escherichia coli* and other Gram-negative bacteria, comprising a variety of proteins with specificities for sugars, amino acids, and essential inorganic ions such as phosphate and sulfate [14]. These proteins serve for the transient complexation of their cognate compounds, followed by controlled delivery to transporter proteins that reside in the inner bacterial plasma membrane. Yet, despite their similarity in function, their sizes vary considerably and the mechanism of ligand complexation usually involves several distinct globular domains.

Similarly, streptavidin from *Streptomyces avidinii* – and also its eukaryotic counterpart avidin, which occurs in chicken egg white – has evolved only in order to tightly complex biotin, a small vitamin compound [15]. In this case the complexation is kinetically almost irreversible, which makes sense for its role as a bacterial antibiotic protein and has led to its widespread use as a biochemical reagent (for references, see [16]).

The lipocalins constitute another family of secretory ligand-binding proteins, which are typical for higher organisms. Initially, they were discovered in vertebrates, such as the retinol-binding protein (RBP) in man [17], but in fact lipocalins are found in a variety of eukaryotes and even in bacteria [18–20]. Generally, they serve for the transport or storage of poorly soluble or chemically sensitive compounds. Although their primary structures mostly lack detectable homology, structural analyses revealed a common fold for these proteins, comprising a rigid β -barrel as the central element of the lipocalin architecture [21]. The ligand is bound at the open end of this supersecondary structure, where a set of four loops forms the entrance to a structurally well-defined pocket. Consequently, lipocalins have emerged as an attractive scaffold with potential for the engineering of artificial ligand-binding proteins.

8.2

Lipocalins: A Class of Natural Compound Carriers

The first lipocalin whose 3-D structure was solved and refined at high resolution was the human plasma retinol-binding protein (RBP) [22, 23]. RBP acts as a natural transporter of vitamin A (retinol) in the blood of vertebrates. Upon complexation in a hydrophobic cavity with complementary shape, the poorly soluble terpenoid alcohol becomes packaged by the protein and protected from oxidation or double-bond isomerization. RBP is synthesized in the liver and directly loaded with the ligand in the hepatocyte, where retinol is stored. Furthermore, the holo-RBP forms a structurally defined ternary complex with transthyretin [24], also known as prealbumin. After delivery of the retinol ligand to a target tissue, the complex decomposes and the monomeric apo-RBP becomes filtered out by the kidney and degraded.

In the crystal structure, RBP folds into a single globular domain of approximately 40 Å in diameter (Fig. 8.2) whose central part is made of an eight-stranded, up-and-down β -barrel. At the amino-terminal end, the β -sheet region is flanked by a coiled peptide segment, and at the carboxy-terminal end, it is followed by an α -helix and an amino acid stretch in a more or less extended conformation. Within the β -barrel the anti-parallel strands (assigned A to H) are arranged in a $(+1)_7$ topology. They wind in a right-handed and conical manner around a central axis such that part of the first strand A is hydrogen bonded via its backbone to the last strand H again.

One end of the β -barrel is closed by the amino-terminal peptide segment that runs across its bottom between the two short loops connecting strands B/C and F/G, respectively, before it enters into β -strand A. Dense packing of side chains in this region and within the adjacent interior of the barrel structure leads to the formation of a hydrophobic core. The other end of the β -barrel is open to the solvent and forms a characteristic pocket. In the case of RBP, retinol is encapsulated as a ligand and protrudes into the barrel by almost half of its depth. The entrance to the pocket is formed by a set of four loops, which connect the eight antiparallel β -strands in a pairwise fashion. Because of the chalice-like shape of the protein (Fig. 8.2) and since many members of this family complex lipophilic compounds, the term “lipocalins” was proposed [25].

Several other lipocalins whose tertiary structures have been elucidated adopt a very similar fold. These were dubbed “prototypic” lipocalins [21] in order to distinguish them from more distantly related members of the family [18]. Within this subset, especially the β -barrel with the attached α -helix is highly conserved. In contrast, the four loops that form the entrance to the ligand pocket vary considerably in sequence, conformation, and length, thus effecting the differing ligand specificities (Fig. 8.2).

However, not all lipocalins need to complex a small ligand in order to fulfill their physiological role. In aphrodisin, for example, which acts as a strong pheromone on male hamsters, the polypeptide itself seems to be responsible for the biological activity, thus requiring transfer of the non-volatile macromolecule by

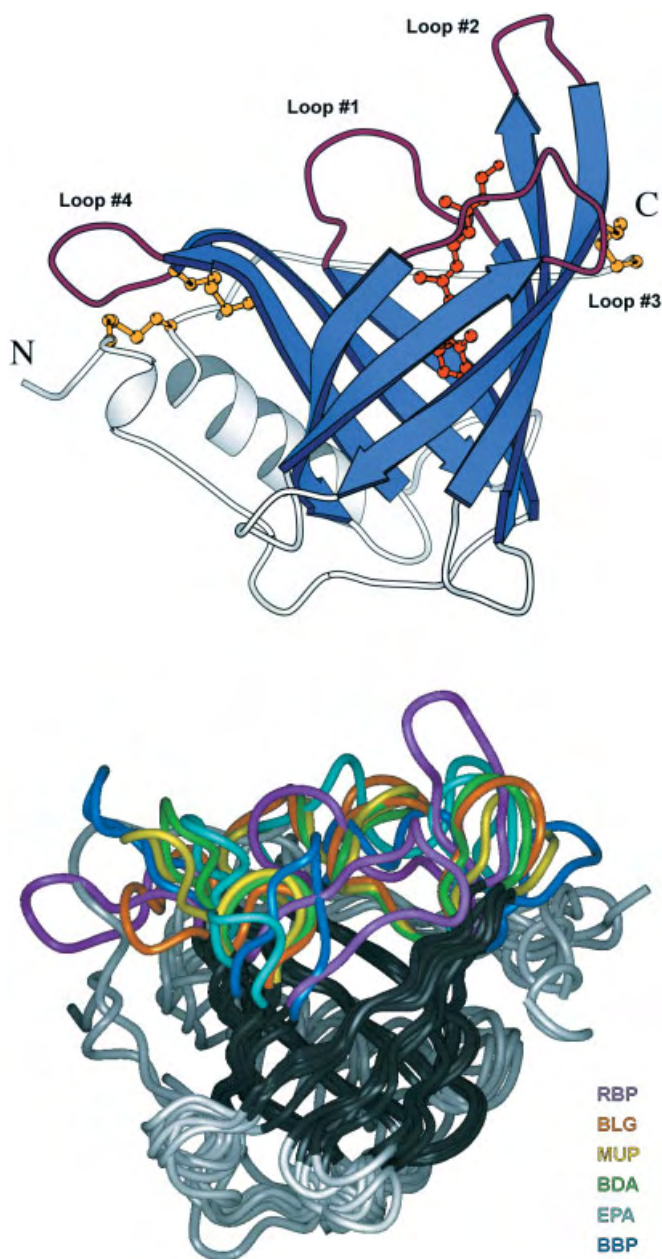


Fig. 8.2 Generic structure of lipocalins. (Top) Ribbon representation of the retinol-binding protein with the bound vitamin A (PDB entry 1RBP). The four loops are shown in dark red at the open end of the β -barrel, and the three characteristic disulfide bonds of the RBP are

highlighted. (Bottom) Superposition of six natural lipocalins with diverse ligand specificities (PDB entries 1BBP, 1BEB, 1BJ7, 1EPA, 1MUP, 1RBP). The β -barrel is colored black (for detailed description, see [21]).

physical contact [26]. Even though a potential ligand pocket was recently detected in its crystal structure [27], attempts to identify a cognate low-molecular-weight molecule have failed.

For some other lipocalins promiscuous binding of hydrophobic ligands was assumed [18, 28]. In the case of human apolipoprotein D (ApoD), for example, which occurs as a peripheral protein of the high-density lipoprotein (HDL) complex in serum, a whole series of potential ligands has been discussed [29]. Yet, thorough binding studies with the rigorously purified recombinant protein revealed just two ligands to be complexed with approximately micromolar dissociation constants: progesterone and arachidonic acid [30]. Because progesterone is well discriminated by ApoD against related steroids, such as pregnenolone and testosterone, it seems likely that arachidonic acid is recognized at a different binding site.

Lipocalins are typical secretory proteins containing disulfide bonds. Human RBP possesses the maximal number of three disulfide cross-links that was observed so far. One of them joins the carboxy-terminal end of the polypeptide chain to the β -barrel (Cys⁷⁰–Cys¹⁷⁴). Another one fixes the amino-terminal segment of the protein to the carboxy-terminal end of the α -helix (Cys⁴–Cys¹⁶⁰). The third disulfide bond (Cys¹²⁰–Cys¹²⁹) links the two neighboring strands G and H just underneath loop #4 at the open end of the β -barrel (cf. Fig. 8.2). The latter two disulfide bridges are characteristic for RBP. Although a Cys residue close to the amino-terminus is also found in several other lipocalins, it usually forms a disulfide bond with a Cys residue in strand G of the β -barrel, as in the bilin-binding protein (BBP), which carries two disulfide bonds [31, 32].

The C-terminal disulfide bond is obviously conserved in the lipocalin family, especially in those members that possess just one of them, e.g., the human neutrophil gelatinase-associated lipocalin (hNGAL) [33, 34]. Yet, there are certain local deviations as in the BBP, where this link is made to a Cys residue in strand B instead of strand D, as in RBP. Some lipocalins do not possess disulfide bonds at all, e.g., the bacterial lipocalin [20]. Hence, it seems that stabilization of the lipocalin architecture does not generally necessitate disulfide cross-links, contrasting with the immunoglobulin fold [35].

Many lipocalins are abundant in serum or tissue fluids. However, their glycosylation status varies. Human RBP, for example, is not glycosylated, whereas ApoD from human plasma was shown to be glycosylated at both of its potential N-glycosylation sites, Asn⁴⁵ and Asn⁷⁸ [36] – possibly in contrast with other tissues where this lipocalin is also expressed. Nevertheless, when synthesized as a recombinant protein in *E. coli*, via secretion into the bacterial periplasm, the unglycosylated ApoD can be isolated as a soluble and functional protein [30]. Similarly, hNGAL, which is normally glycosylated at a single position, can be obtained as an unglycosylated protein from *E. coli* and adopts its proper tertiary structure, as elucidated by NMR analysis [33].

Finally, many lipocalins exist as soluble monomers. RBP, for example, which participates in a reversible complex formation with transthyretin, can be isolated as a fully stable monomer, either in complex with retinol or in the absence of the ligand [37]. Interestingly, several mammalian lipocalins appear to be linked via a

disulfide bond to other functional protein complexes. For example, human ApoD carries a fifth Cys residue in addition to those giving rise to its two intra-chain disulfide bonds. Its unpaired thiol side chain is connected to a Cys residue of apolipoprotein A-II, which is an integral lipoprotein of the HDL particle. When this residue is removed by site-directed mutagenesis, the recombinant ApoD can be isolated in a soluble monomeric state [30]. Moreover, human ApoD is naturally produced as an individual protein in some other tissue fluids [29], and in other organisms, such as rabbits, the unpaired Cys residue is even missing. Similarly, hNGAL normally occurs cross-linked with gelatinase B (known as matrix metalloproteinase 9), but also as a monomeric or homodimeric serum protein, and it was successfully produced in a monomeric state for structural studies [33].

Taken together, lipocalins provide attractive candidates in order to engineer novel ligand specificities. Features like their small size (typically between 150 and 180 residues), monomeric polypeptide composition, dispensable posttranslational modification, and robust protein fold not only facilitate protein-engineering studies but also provide advantages for practical applications.

8.3

Anticalins: Lipocalins Reshaped via Combinatorial Biotechnology

In a first attempt to tailor the ligand pocket of a lipocalin, the bilin-binding protein (BBP) served as a biochemically well-characterized model protein. The BBP originally occurs as a secretory protein in the butterfly *Pieris brassicae*, where it complexes biliverdin IX₇, a metabolic oxidation product of protoporphyrin IX. Hence, it serves for coloration as well as photoprotection, especially at the larval state. Natural BBP is found in two isoforms [32]. BBP-I forms a dimer in solution, whereas BBP-II, which likely arises from deamidation of the amino-terminal Asn residue of BBP-I, adopts a stable monomeric state. Genetic analysis revealed that only BBP-I is encoded on the insect chromosome [38]. After fusion of the mature part of the polypeptide chain to a bacterial leader peptide, the apo-BBP could be produced in the periplasm of *E. coli* as a recombinant protein in a functional state. For this purpose the amino-terminal Asn residue was directly exchanged by Asp at the genetic level and a monomeric BBP with full ligand-binding capability was obtained [38].

The crystal structure of the natural holo-BBP was elucidated at high resolution [31, 32]. It revealed the characteristic β -barrel fold with the tetrapyrrole ligand bound in a helical conformation (Fig. 8.3, see p. 196). Compared with human RBP, the four loops give room to a wider and shallower pocket for biliverdin. Consequently, BBP appeared to be a promising candidate for the reshaping of its ligand-binding site towards a variety of target compounds. To this end, the methodology of combinatorial biochemistry was applied (Fig. 8.3, see p. 196), comprising steps of (1) directed random mutagenesis of the loop regions in order to create a molecular library and (2) selection of cognate binding proteins from this library against a prescribed ligand.

Based on the 3-D structure of BBP, a total of 16 amino acid positions were identified within the four loop segments – as well as adjoining regions of the β -strands – that dominate the interface with the natural ligand (Fig. 8.3). These positions were chosen to fulfill two criteria: first, they could be expected to tolerate both small and large side chain substitutions, and second, they appeared to reach as deeply as possible into the binding pocket. Hence, one could assess whether the hydrophobic core in the deeper part of the β -barrel would still be functional with respect to the stable folding of corresponding mutants.

The 16 positions in the cloned BBP cDNA [38] were subjected to site-directed random mutagenesis using a two-step assembly polymerase chain reaction (PCR) with the help of primer oligodeoxynucleotides that were synthesized with mixed bases at the mutagenic codon positions. In order to introduce unique restriction sites at both ends of the amplified central fragment of the BBP structural gene (both for *Bst*XI, but with mutually non-compatible overhangs), two amino acids had to be exchanged at positions belonging to the β -barrel: Asn²¹ \rightarrow Gln and Lys¹³⁵ \rightarrow Met. In addition, the recombinant BBP carried the mutation Asn¹ \rightarrow Asp mentioned above and Lys⁸⁷ \rightarrow Ser, which was introduced in order to remove a proteolytic cleavage site [39]. Thus, there were altogether four fixed amino acid replacements in addition to the randomized side chains. The mutagenized gene cassette was then inserted into an appropriate *E. coli* vector and a genetic library comprising 3.7×10^8 variants was prepared [40].

The phagemid-display technique [41] was employed in order to select BBP variants with novel binding specificities from the resulting library [39]. For this purpose the BBP variants were produced as fusion proteins with a bacterial signal peptide at the amino-terminus and with the *Strep*-tag II, followed by a truncated pIII phage coat protein, at the carboxy-terminus [40]. In this case the amino acids 217 to 406 of the gene III product from filamentous phage M13 were used. The whole fusion gene was cloned on a phasmid vector under the tight transcriptional control of the chemically inducible tetracycline promoter [42] so that phagemid particles displaying BBP variants on their surface were efficiently produced under appropriate conditions.

Fluorescein, a well known immunological hapten [43] with many applications in biochemistry and biophysics and a collection of commercially available derivatives, served as the prescribed ligand for BBP variants in the first selection study. The phagemid random library was used for panning on a plastic surface coated with a covalent conjugate of fluorescein with bovine serum albumin (BSA). After six cycles of adsorption, acid elution, and phagemid re-amplification, the specific enrichment of a mutant phagemid fraction was observed.

From DNA sequence analysis of 10 arbitrarily chosen clones, it became apparent that just four different BBP variants were still present in this population. Three of them – dubbed FluA, FluB, and FluC – gave rise to strong signals for the binding of several fluorescein conjugates when produced as soluble proteins and investigated in an ELISA. In each of these variants, all 16 randomized amino acids had been exchanged when compared with the wild-type BBP [40]. Interestingly, four substitutions were identical among the three selected variants: Arg⁵⁸,

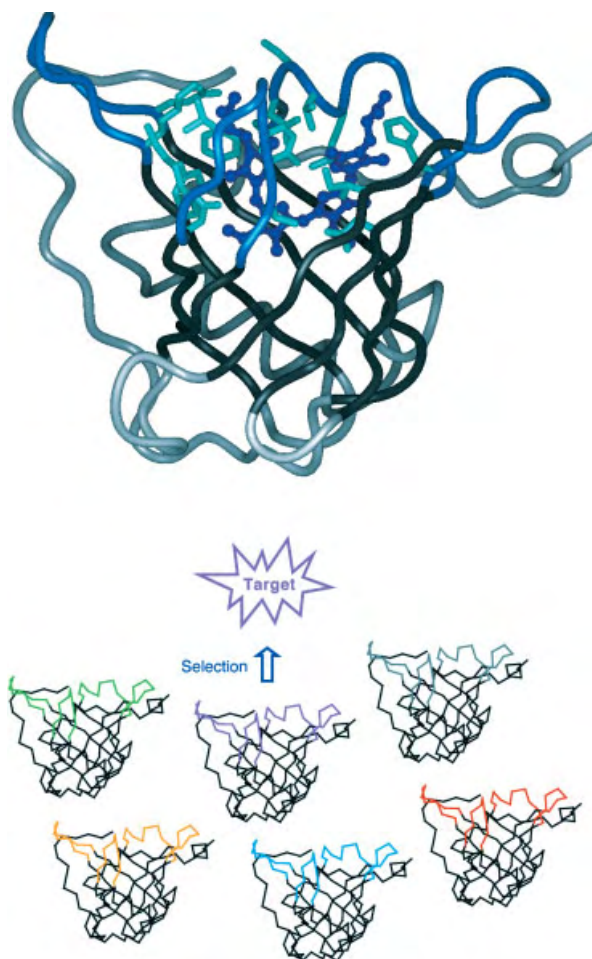


Fig. 8.3 Generation of anticalins by randomization and selection. (Top) Randomized amino acid positions (corresponding side chains are shown in light blue) in the BBP, depicted

here with the bound biliverdin in dark blue. (Bottom) Principle of the molecular library selection.

Arg⁹⁵, Arg¹¹⁶, and His¹²⁷. The corresponding preponderance of positively charged side chains was in agreement with the several negative charge centers of the fluorescein derivative that was used in the selection. A similar effect had been observed in attempts to select recombinant antibody fragments against the same hapten from a semi-synthetic combinatorial library [44].

The BBP variant FluA was subjected to detailed biochemical characterization. The engineered lipocalin could be produced at high yield in the periplasm of *E. coli* (9.1 mg per 2 L culture compared with 1.2 mg for wild-type BBP) and isolated to homogeneity in one step via the *Strep*-tag method [45]. According to the relative shift in electrophoretic mobility between oxidized and reduced state of the pro-

tein, the two disulfide bonds of the BBP scaffold were correctly formed. Furthermore, the far UV circular dichroism spectrum of FluA revealed no marked differences from the wild-type protein. Consequently, the BBP had tolerated 16 amino acid exchanges within its ligand pocket – plus the four rationally introduced mutations mentioned above – without losing its folding properties as a lipocalin.

The novel ligand-binding activity of FluA was studied in ELISA experiments with fluorescein coupled to different carrier proteins. In each case steep saturation curves were observed with half-maximal concentrations in the nanomolar range, so that the recognition of the hapten appeared to be independent of the macromolecular context. Thermodynamic dissociation constants for the complex formation between FluA and fluorescein and some related compounds were determined by fluorescence titration in solution, measuring the emission of the protein's Tyr and Trp residues. As a result, fluorescein was bound slightly stronger than its two derivatives 4-aminofluorescein and 4-glutarylamidofluorescein, the compounds that actually had been used in the synthesis of the protein conjugates for the selection experiments. In contrast, pyrogallol red, a chemically similar triphenylmethane compound, was bound two orders of magnitude less tightly, while no binding at all could be detected for the related dyes phenolphthalein or rhodamine B.

Interestingly, when the titration was performed such that the hapten's own characteristic fluorescence was measured, almost complete quenching was observed. From this very accurate titration experiment, a K_D value of 35.2 ± 3.2 nM was determined for the FluA · fluorescein complex [40]. Using the same spectroscopic effect, the association kinetics between fluorescein and FluA could be measured by rapid mixing, yielding a K_{on} value of $5.28 \pm 0.05 \times 10^6$ M⁻¹ s⁻¹ (G. Beste and A. Skerra, unpublished).

The phenomenon of almost complete fluorescence quenching upon complex formation between fluorescein and the engineered lipocalin FluA was elucidated in a series of time-resolved light-absorption measurements after pulse activation of the bound ligand [46]. These experiments revealed an ultrafast electron transfer between the fluorescein and an aromatic side chain (Tyr or Trp) in its close proximity. The excited fluorescein dianion within the ligand pocket abstracts an electron from the neighboring amino acid at a rate of 400 fs. The resulting radical trianion is deactivated in a radiationless process – with a larger time constant of 4 ps – to the spectroscopic ground state of fluorescein by back transfer of the electron. The observed monoexponentiality in the formation of the excited state, and also of its subsequent decay, points toward a high structural definition of the hapten-binding site and explains the highly efficient quenching effect that becomes apparent under stationary conditions.

Clearly, this spectroscopic phenomenon was a serendipitous event because there was no corresponding selection applied during the generation of FluA. Also, the other mutants that were selected along with it did not show fluorescence quenching to the same high extent. Nevertheless, it is remarkable that such an efficient electron transfer process, which is even faster (by a factor 3–4) than the one measured between bacteriochlorophyll and bacteriopheophytin in the bacterial reac-

tion center of *Rhodobacter sphaeroides*, can be achieved by combinatorial protein design (see the discussion in [46]). The fluorescence-quenching effect observed with FluA is also significantly more pronounced than for antibodies that were raised against fluorescein by immunization. Hence, the engineered lipocalin may be of interest as a reagent in biophysical studies where the specific masking of fluorescein groups is desired.

The same random library of BBP variants was used in selection experiments with several other haptens. The cardiac steroid digoxigenin served as a molecular target of practical relevance [47]. In this case the library was screened by combining phagemid display with a filter-sandwich colony-screening assay in order to rapidly identify individual BBP variants with corresponding ligand-binding activity. As a result, one variant with specificity for digoxigenin was isolated whose dissociation constant was determined to be 295 ± 37 nM by means of protein fluorescence titration [47]. In an attempt to further improve the ligand affinity of this engineered lipocalin, dubbed DigA, an *in vitro* affinity maturation was performed.

Inspection of the primary sequence of the BBP variant revealed that the first of the four loops mainly had charged side chains acquired during the selection. However, these amino acids did not appear to be optimal for the complexation of the hydrophilic, though uncharged, steroid. In this respect it should be noted that the BBP random library with its complexity of 3.7×10^8 was by far too small to represent all possible combinations of the 16 randomized amino acids. Therefore, in principle, considerable room remained for further sequence optimization in parts of the binding pocket, once molecular recognition of a specific ligand was achieved. Consequently, six amino acid positions within loop^{#1} of DigA were selectively subjected to oligodeoxynucleotide-directed random mutagenesis, again followed by selection for the binding of digoxigenin groups via phagemid display and colony screening.

In this way, the variant DigA16 was obtained, which binds digoxigenin significantly tighter, with a K_D value of 30.2 ± 3.6 nM [47]. Remarkably, the glycosylated natural compound digoxin, which has three sugar molecules attached to C-3 of the steroid system [48], is bound with precisely the same affinity. Likewise, digoxigenin conjugates with several different carrier proteins – bovine serum albumin, ovalbumin, or ribonuclease A – that were covalently linked via an aliphatic spacer to the same steroid ring position gave rise to indistinguishable binding signals in solid-phase assays. Thus, the BBP variant DigA16 recognized the digoxigenin group as a true hapten, without detectable context dependence.

Subsequent ligand-binding studies [47] revealed that a chemically similar cardiac glycoside, digitoxin, which differs from digoxin just by a single missing hydroxyl group, is bound stronger still, with a K_D value equal to 3.2 ± 0.54 nM. In contrast, complex formation with the related steroid ouabain, which often shows cross-reactivity with antibodies raised against digoxin [48], was not detectable. In addition, no complex formation was observed with the steroid testosterone or with 4-aminofluorescein, the ligand that had served before in the selection of the BBP variant FluA.

Hence, DigA16 represents an engineered lipocalin with high affinity and pronounced specificity towards a rather hydrophilic steroid. The dramatic alteration in the ligand-binding function of this lipocalin was achieved by exchanging a total of 17 amino acids [47], which form most of the pocket in the natural BBP, together with the four site-directed amino acid replacements that were introduced into the scaffold in order to make it better amenable to protein-engineering experiments (see above).

Attempts were made to raise the affinity of DigA16 for the digoxigenin group even further by applying additional cycles of targeted random mutagenesis at loops #3 and #4 [49]. The resulting variant DigA16/19, which carries several new mutations in loop #4, exhibits improved affinity for digoxigenin, with $K_D = 12.4 \pm 1.3$ nM. In addition, DigA16/19 possesses enhanced ligand specificity and also recognizes part of the linker that was used for fixing the steroid group to the carrier protein.

During these experiments the randomized residues were still restricted to the original set of positions chosen within the four loops. However, from recent structural analyses (see Section 8.4) it appeared that there are additional, so far non-mutated amino acids that contribute to the shape of the ligand pocket and may therefore govern affinity and specificity for the steroids. Thus, future improvement in molecular recognition by this engineered lipocalin may be guided by rational principles.

Nevertheless, the successful construction of a digoxigenin-binding lipocalin provides a novel and useful tool in biochemistry. Digoxigenin and digitoxigenin are medically important compounds, either as potentially poisonous substances or as drugs with therapeutic value – as long as they are applied at a precisely adjusted dose [50]. Furthermore, the digoxigenin group has acquired recent popularity in biochemistry as a non-radioactive label for a variety of biomolecules. Several chemically activated derivatives are available for the selective labeling of proteins or nucleic acids so that digoxigenin can be used independently from the commonly employed biotin group, with the advantage of very low background-staining activity [51].

The generation of BBP variants with novel binding specificities for fluorescein or digoxigenin, respectively, has demonstrated for the first time that a lipocalin can be tailored to recognize non-natural ligands. In order to illustrate the antibody-like binding function of the engineered lipocalins, this new class of proteins was termed “anticalins” [40].

8.4

Structural Aspects of Ligand Recognition by Engineered Lipocalins

The 3-D structures of the fluorescein- and digoxigenin-binding BBP variants have recently been analyzed by X-ray crystallography and compared with the original bilin-binding protein. The crystal structures were determined in different space groups and, for one variant, in both the presence and absence of the hapten, thus

giving insight into the structural mechanisms of specific ligand recognition by the engineered lipocalins.

In the case of the fluorescein-binding variant FluA, crystals were obtained in the presence of the ligand at pH 8.1 with two FluA · fluorescein complexes in the asymmetric unit, which were refined to a resolution of 2.0 Å [52]. The two molecules were highly similar in structure, with a root mean square difference (rmsd) of 0.33 Å for 173 mutually superimposed C_α positions. The overall topology of the β -barrel with the α -helix attached to it, both of which are characteristic features of the lipocalin architecture (see Section 8.2), was found to be conserved (Fig. 8.4). Both disulfide bonds of the BBP scaffold, one between Cys¹⁸ and Cys¹¹⁵ and one between Cys⁴² and Cys¹⁷, were also clearly visible. Upon superposition with the BBP crystal structure (molecule A from the Protein Data Base entry 1BBP [32]), an rmsd of 1.2 Å was calculated for 159 superimposed C_α positions.

The largest structural differences were seen at the four loops that form the entrance to the binding site. The most prominent conformational changes occurred in loops #1 and #2. Loop #1 had adopted a more extended conformation and moved away from the center of the ligand-binding site towards the bulk phase of the solvent. The C_α position of the mutated residue Asn³⁶ (Val in BBP) at its tip was concomitantly displaced by approximately 8 Å. Loop #3, which was also involved in the contacts with the structural neighbor in the asymmetric unit, had moved away from the barrel axis, thus opening the cavity for the bound ligand, with the C_α position of the non-mutated residue Gly⁹² at its tip shifted by about 6.6 Å.



Fig. 8.4 Crystal structure of the anticalin FluA with the bound fluorescein (green). C_α positions of the 16 randomized amino acids are shown as gray spheres. Trp¹²⁹ is depicted with its side chain in magenta.

Fluorescein is bound at the bottom of the cleft that harbors biliverdin IX_α in the wild-type BBP structure (Fig. 8.2). Its xanthenolone moiety is located close to the center of the β -barrel, while the carboxyphenyl group is oriented towards the entrance of the pocket. The para ring position (with respect to the central carbon atom of the triphenylmethane dye), which carried the linker group during the selection experiments for this anticalin [40], is accessible from the solvent via a narrow channel. An area of 454 Å², corresponding to 91% of the solvent-accessible surface of fluorescein, became buried in the complex.

Approximately 50% of the buried area from the side of the protein is contributed by 6 of the 16 residues that were mutated in the generation of FluA from BBP. The remaining buried surface belongs to 10 residues that were not mutated. Unexpectedly, when compared with the complexation of biliverdin by BBP, fluorescein was found to be inserted even more deeply into the hydrophobic core of the β -barrel. In the central region of the protein, where no mutations had been introduced, the necessary space was created by the movement of loop^{#3} and by rearrangement of several side chains.

In particular, there is a ladder of residues comprising His⁸⁶, Phe⁹⁹, His¹²⁷, and Trp¹²⁹, which have undergone a concerted reorientation of their aromatic side chains. Of those, only His¹²⁷, which participates in a packing interaction with one of the phenolic rings of fluorescein, was mutated in the generation of FluA from BBP. The non-mutated residue Trp¹²⁹ is located directly underneath. It has shown major side chain reorientation and gives rise to an extended π stacking interaction at the middle of the xanthenolone ring system of fluorescein via coplanar arrangement in van der Waals distance. It seems that the introduction of the imidazole side chain at position 127 has triggered the whole movement, including those of the non-mutated residues His⁸⁶ and Phe⁹⁹ at the bottom of the pocket.

As a result of these changes, original residues of BBP form a significant part of the reshaped ligand pocket. Hence, it appears that in addition to the loop region, the hydrophobic core of the lipocalin displays considerable plasticity as well. Residues close to the hydrophobic core that had been thought to be crucial for proper folding of the protein have adopted completely new side chain orientations in order to allow for the binding of the new ligand. Future design of anticalins with even higher affinities for small haptens may therefore also include residues from the central part of the β -barrel.

When considering the side chains that make up the binding site for fluorescein, both mutated and non-mutated, it is found that they are predominantly polar in nature. Thus, the lipocalin pocket is by no means restricted to lipophilic ligands, as was anticipated before and as the name of this protein family may suggest. Furthermore, the crystal structure confirms the crucial role of the mutated basic residues Arg⁵⁸ and His¹²⁷ in FluA for the tight binding of fluorescein, which was previously demonstrated by site-directed mutagenesis experiments [40].

However, the mechanism of interaction is different from the earlier assumption that was based on the crystal structure of the anti-fluorescein antibody 4-4-20 [53]. There, the xanthenolone group is oriented such that an Arg and a His residue (with C α distance similar to BBP) are each in contact with one of the phenolic

moieties and form hydrogen bonds. In contrast, Arg⁵⁸ and His¹²⁷ of FluA contact the same phenolic group, but from opposite sides. This arrangement is made possible by the extensive structural reorganization at the open end of the β -barrel (see above). Three Arg residues at positions 88, 95, and 116, which were shown to have a favorable effect on the ligand affinity via their positively charged side chains, do not form direct contacts with the bound fluorescein and should therefore mainly exert an electrostatic influence. Yet, the reasons for their peculiar arrangement at the entrance to the ligand pocket and their strong conservation among the several selected fluorescein-binding BBP variants [40] are not obvious at present.

Taken together, the crystallographic analysis of the anticalin FluA, which was generated by combinatorial design from a prototypic lipocalin, reveals that mutated residues within the loop region and adjoining parts of the β -barrel can give rise to three different effects. First, they can contribute direct contacts with the bound ligand or at least provide an appropriate electrostatic environment. Second, they may induce novel backbone conformations in the loops and thus lead to the formation of a pocket with generic shape complementarity with the prescribed ligand. Finally, there are certain amino acids that influence the side chain conformations of neighboring residues and thus reshape the pocket in an indirect manner. Similar phenomena are known from antibodies where, apart from amino acids that contact the antigen or hapten, key residues within the hypervariable loops are responsible for their canonical backbone conformations [54] and framework residues indirectly fine-tune the shape of the combining site [55].

The crystal structure of the FluA · fluorescein complex also provides a structural explanation for the strong quenching effect of this particular anticalin that was mentioned before. Tight coplanar packing of the indole ring of Trp¹²⁹ against the xanthenolone system of the bound fluorescein was observed (Fig. 8.4). Consequently, this aromatic residue is the likely candidate for the highly efficient electron-transfer process and is optimally positioned in this respect.

In the case of the anticalin DigA16, crystal structures were solved not only for the bound digoxigenin but also for the complex with the related steroid digitoxigenin and for the uncomplexed apo-protein [56]. The crystals, which were grown at pH 7.6–8.0 and whose structures were refined to resolutions between 1.8 and 1.9 Å, were essentially isomorphous and contained one monomer per asymmetric unit. In addition, crystals were obtained in another space group for the uncomplexed original BBP variant DigA, although with poorer diffraction quality.

Again, the overall topology of the lipocalin, comprising the β -barrel with the α -helix attached to it, remained conserved, whereas the set of four loops at the entrance to the ligand pocket revealed clear structural differences in comparison with the BBP. The most prominent conformational change was observed for loop^{#1}, where an α -helical segment of seven amino acids appeared in DigA16 (residues 33 through 39), in both the presence and absence of the steroid ligand (Fig. 8.5). Notably, the non-mutated residue Tyr³⁹ within the new helix – which faces the solvent in the BBP – has shifted in all three DigA16 structures such that it packs with its side chain against the bound steroid, if present. This dramatic

conformational change results in a displacement of its C α atom by 9 Å with respect to the position in the BBP.

The α -helical loop conformation seems to be essentially stabilized by two specific interactions. First, the side chains of the mutated residues Arg⁵⁸ and Ser⁶⁰ in loop^{#2} form hydrogen bonds with the carbonyl oxygen of Tyr³⁹ at the carboxy-terminal end of the helix. Second, the newly introduced side chain of His³⁵ at the amino-terminal end of the helix becomes packed between this Tyr residue and the side chain of Leu¹²⁷ in loop^{#4}, which also resulted from the mutagenesis. Thus, a small pocket for the imidazole group is formed, which is closed at the bottom by the side chains of Trp¹²⁹ (already present in the BBP) and Gln²⁸. The latter residue was introduced during the affinity maturation from DigA to DigA16, leading to a 10-fold improved ligand affinity (see Section 8.3).

In the complex with DigA16, digoxigenin is bound at the bottom of the cleft that otherwise harbors biliverdin IX γ in the wild-type BBP, and it roughly replaces the space previously occupied by one of the tetrapyrrole rings [32]. An area of 514 Å², corresponding to 95% of the solvent-accessible surface of digoxigenin have thus become buried. Approximately half of the buried surface is contributed by 9 of the 17 residues that were mutated in the generation of DigA16, while the remainder is due to 10 residues that have not been mutated. Similarly, as deduced for the anticalin FluA, specific recognition of the steroid ligand is achieved by preformed shape complementarity of the ligand pocket in apo-DigA16 as a result of (1) side chain replacements and (2) indirect effects of mutated positions on wild-type residues.

The latter effect can especially be seen for the shifted residue Tyr³⁹, which is part of the newly formed helix in loop^{#1}, and for altered side chain conformations of Phe⁹⁹ and Trp¹²⁹ at the bottom of the ligand pocket. These two conserved residues have rotated their aromatic side chains by approximately 120° compared with BBP, thus enabling accommodation of the new bulky steroid ligand. Their conformation appears to be similar in all three DigA16 structures, even in the absence of the ligand. Apparently, the side chain rotation of Trp¹²⁹ is in concert with the extensive rearrangement in the upper part of the binding cleft, particularly within loop^{#1} due to the mutated residues.

The steroid ligand is bound mainly via van der Waals interactions but also via hydrogen bonds with its polar substituents. Several water molecules have become buried in the ligand pocket as well and participate in hydrogen bond interactions with the hydroxyl and lactone groups of the hydrophilic steroid. Of the buried protein surface in the digoxigenin complex, 34% is provided by non-hydrocarbon groups (for comparison, 33% in the BBP). Hence, the binding site has considerable polar character, which is in contrast with the almost entirely hydrophobic nature of the ligand pocket in other lipocalins, such as the retinol-binding protein (with a corresponding value of 16%). Similarly, the steroid ligand is almost fully trapped within the binding site, with a remaining accessible surface of 5%. Indeed, there is a small gap in the protein shell that permits accessibility of the steroid position C-3, which carried the linker group during the selection procedure [47]. Thus, as a result of the combinatorial protein design experiment, the corresponding steroid derivative must have been almost perfectly shielded from solvent.

From a comparison between the DigA16 complexes with digoxigenin versus digitoxigenin, the structural mechanism of the fine specificity in the steroid recognition – and discrimination between these closely related ligands – became apparent. The non-mutated residue His⁸⁶ at the bottom of the ligand pocket, which is close to the hydrophobic core within the β -barrel, plays a crucial role in this respect and is involved in an induced fit during ligand complexation. Upon binding of digoxigenin, the side chain of His⁸⁶, which points into the empty cavity in apo-DigA16, is displaced by the ligand and forms a hydrogen bond with the hydroxyl group HO-12 of the steroid. Thus, it becomes rotated towards Tyr²², which is part of a loop at the closed end of the β -barrel. The side chain of Tyr²² itself rotates away from its original hydrogen-bonding partner Thr¹⁰⁴ in apo-DigA16 into the direction of His⁸⁶ and forms a hydrogen bond with the imidazole side chain instead.

Digitoxigenin, which is bound essentially at the same position and with the same orientation as digoxigenin, lacks the OH group at the steroid position C-12. Consequently, a hydrogen bond between His⁸⁶ and digitoxigenin is missing in the complex with DigA16, and its imidazole side chain packs closer to the steroid ring system. Thus, compared with the digoxigenin complex it is partially rotated back into the position that it has assumed in the apo-protein. Accordingly, the side chain of Tyr²² still forms a hydrogen bond with Thr¹⁰⁴. Instead, a water molecule appears at the position that is occupied by the phenolic hydroxyl group of Tyr²² in the DigA16 · digoxigenin complex, which is weakly hydrogen-bonded to His⁸⁶ in the case of the bound digitoxigenin.

The set of available crystal structures also provides an explanation for the effect of the affinity maturation that led from the original DigA anticalin to the DigA16 mutant. During this step, several amino acids were randomized in loop #1 [47]. While most of the corresponding side chains are solvent exposed in the tertiary structure, two residues are probably relevant for the loop conformation and improved ligand affinity of DigA16. First, His³⁵ remained conserved with respect to the DigA sequence, which underlines its role in the helix conformation as described above, together with the invariant residue Tyr³⁹. Second, Glu²⁸ was consistently replaced by Gln (in several mutants that were selected along with DigA16; see [47]). In the apo-DigA structure, the side chain of Glu²⁸ adopts a different conformation compared with Gln²⁸ in apo-DigA16 and forms a salt bridge with the Arg⁵⁸ guanidinium group. As a consequence, the carboxylate moiety of Glu²⁸ may sterically interfere with the position of the lactone substituent at ring D of the bound steroid ligand. In addition, due to a corresponding shift of the Arg side chain, the helix – which covers the bound ligand – probably undergoes minor repositioning. Hence, it seems that the introduction of Gln²⁸ is mainly responsible for the 10-fold enhanced affinity of DigA16.

Notably, the mechanism of molecular recognition by the anticalin is different from the manner in which digoxigenin is bound by immunoglobulins. Two crystal structures have been described for F_{ab} fragments of monoclonal antibodies that were raised against digoxin as a hapten: the F_{ab} “26-10” in complex with digoxin (PDB accession code 1IGJ; [57]) and the F_{ab} “40-50” in complex with ouabain

(PDB accession code 1IBG; [58]). In both cases only between 60% and 70% of the steroid became buried from solvent during complex formation (Fig. 8.5). The ring system is covered by protein residues mostly from one side, probably due to the limited capacity of the Ig architecture to create a deep pocket at the interface of the V_H and V_L domains. Even though high affinities have been achieved by these antibodies, their specificities are poor because both significantly cross-react with ouabain, a cardiac steroid related to digoxin [48, 58]. In contrast, the engineered lipocalin DigA16 clearly distinguishes between these two compounds [47], thus providing a functional advantage.

In summary, a phenomenon of pronounced structural plasticity was observed in the engineered BBP variants FluA and DigA or DigA16, which means that the backbone conformation of the lipocalin loop region was strongly influenced by the side chain replacements (Fig. 8.6, see p. 207). However, in the context of a given amino acid sequence, the conformational flexibility of these loops seems to be rather low because no significant differences were observed in the two independently refined FluA · fluorescein complexes or when comparing the DigA16 structures in the absence or presence of the ligand. This effect was not entirely expected *a priori*, because from a superposition of natural lipocalins with known tertiary structures [21], it was not clear to which extent the loop conformation is governed by its own distinct sequence versus individually variable features of the β -barrel structure that provides the support.

8.5

Prospects and Future Applications of Anticalins

The functional and structural data that have been gathered during the engineering of lipocalins for the recognition of two unrelated low-molecular-weight ligands clearly demonstrate the potential of this scaffold for the generation of artificial receptor proteins with high affinity and specificity for prescribed target molecules. Our findings confirm that the β -barrel architecture of lipocalins constitutes a remarkably stable scaffold. Even though amino acids were replaced in at least 20 different positions, most of them within the binding site of the BBP, the overall topology and the β -barrel structure itself were retained. Structural changes were merely observed at a local level and essentially restricted to the loop regions.

It seems that lipocalins indeed provide a partitioned protein architecture, wherein the β -barrel – together with the fixed loops at its closed end and the α -helix attached to it – provides a rigid framework that is structurally conserved among the “prototypic” members of this family [21], while the set of four loops at its open end can be hypervariable. This situation is reminiscent of antibodies, where a set of six CDRs presented on top of a largely constant framework region is responsible for the specific binding of the antigen or hapten. However, compared with recombinant antibody fragments, engineered lipocalins should provide significant benefits because they are composed of one instead of two polypeptide chains, they have a much smaller size, and their set of four loops can be easily manipulated

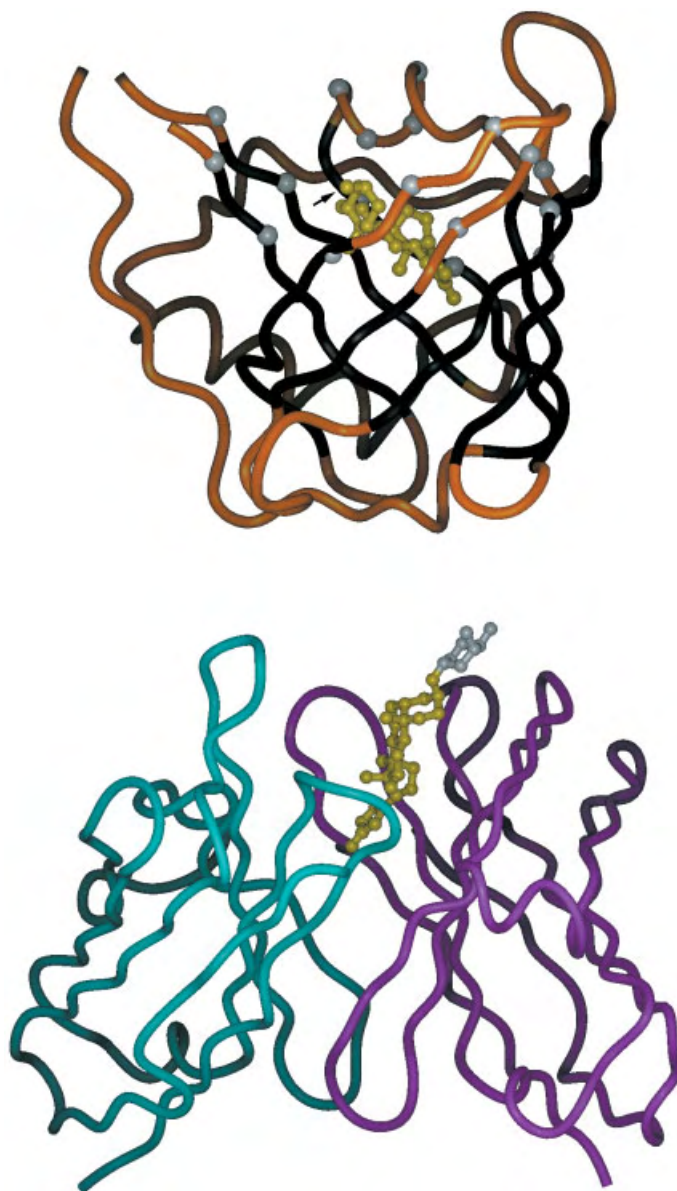


Fig. 8.5 Molecular recognition of haptens by engineered lipocalins versus antibodies. (Top) Crystal structure of the anticalin DigA16 with the bound digoxigenin (yellow). The arrow points to the hydroxyl substituent of steroid ring A, which had served for covalent attachment – via a flexible spacer – to a solid support during the selection process for this BBP variant. The C_{α} positions of the 16 initially ran-

domized amino acids are shown as gray spheres. (Bottom) Crystal structure of the anti-digoxigenin F_{ab} fragment 26-10 (PDB entry 1IGJ) with the V_L and V_H domains colored cyan and magenta, respectively. The bound digoxigenin group is shown in yellow, while the digitoxose sugar attached to it (at the same ring position as above) is colored light gray.

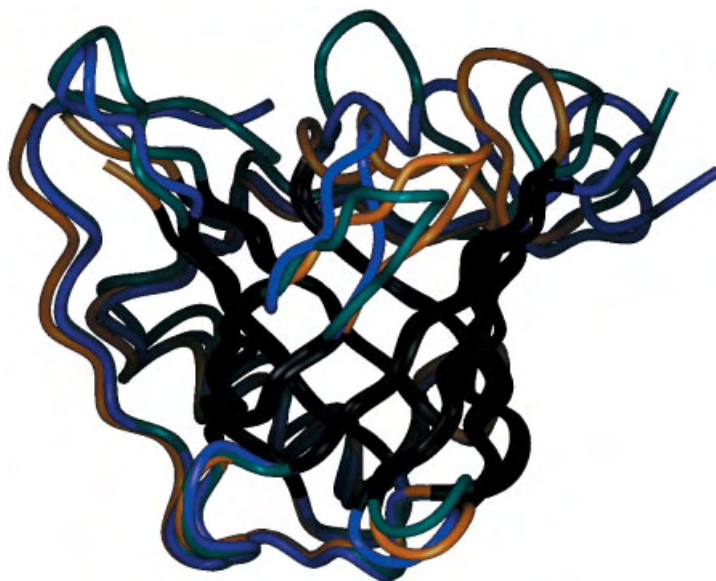


Fig. 8.6 Plasticity of the engineered lipocalin-binding site: superposition of the natural BBP with its variants FluA and DigA16 (with loops colored blue, green, and yellow, respectively).

simultaneously at the genetic level. Consequently, lipocalins provide a promising alternative to antibody fragments for the engineering of artificial ligand-binding proteins, therefore called “anticalins,” using the methods of combinatorial biochemistry [21, 40].

The most striking property of the lipocalin scaffold is its ability to provide a well-defined and conformationally rigid cavity for the ligand. The ligand pocket is significantly deeper than the hapten-binding sites found in antibodies and may even reach down into the hydrophobic core of the lipocalin, as was seen in the case of the fluorescein-binding anticalin FluA. A similar mode of complexation would not be possible for an antibody fragment because of the detrimental effect on the non-covalent association between V_H and V_L . Hence, the lipocalin-bound ligand can be trapped from the solvent and becomes almost fully surrounded by protein residues, which explains the pronounced specificity, especially observed for the steroid ligands. Thus, an extended linker structure, which should also include hydrophilic groups, seems to be required for the functional immobilization of the target compound during the selection procedure for the cognate anticalin. It would be nice to devise selection techniques that no longer necessitate the covalent fixation of the target because under such conditions anticalins might be generated that fully encapsulate their ligands.

Remarkably, the anticalins obtained up to now recognize their low-molecular-weight ligands independently from the carrier – usually a stable globular protein like BSA or RNase – that was employed for target display. This was shown for

both the fluorescein group [40] and digoxigenin [47]. In this respect, anticalins distinguish themselves from many antibodies, especially when derived from synthetic libraries [4], and also from different protein scaffolds that have been tested for similar purposes. For example, when mutants of cytochrome *b*₅₆₂ with two randomized loop regions were selected against an organic target compound, the hapten was only recognized – with a weak micromolar affinity – as long as it remained linked to the original BSA carrier [59].

The well-defined binding properties for the small ligand in combination with the lack of cross-reactivity with the macromolecular conjugate partner are probably due to the choice of the randomized positions in the lipocalin within an inner zone at the open end of the β -barrel but below the exposed tips of the loop region (Fig. 8.3). Consequently, just a few of the randomized residues are potentially accessible from outside the pocket, most likely only after significant changes in the backbone conformation – as observed for loop^{#1}, which adopts an α -helical structure in DigA16 (see Section 8.4). However, the scope of molecular recognition by engineered lipocalins should not be restricted to low-molecular-weight compounds. Indeed, the more exposed side chains of the four loops may be specifically randomized for the generation of another sort of anticalin libraries that could be useful for selection towards macromolecular targets. Preliminary data from our laboratory suggest that mutants of the BBP can be obtained in this manner, which exhibit specific binding activity for prescribed proteins with dissociation constants in the nanomolar range.

The proof of concept for the generation of anticalins as a novel class of receptor proteins with defined ligand-binding properties was realized using the BBP as a model lipocalin. It has been shown that, because of their simple and robust architecture, these anticalins provide several practical advantages. For example, they are remarkably stable against denaturation. Thermal unfolding studies revealed a melting temperature of 61.3 °C for the recombinant BBP and an even higher *T*_m value of 72.8 °C for the anticalin DigA, although this variant had not been selected for enhanced folding stability [49]. Another advantage of the lipocalin architecture relates to the fact that both ends of the polypeptide chain are sterically accessible at the outside of the β -barrel and should normally not interfere with the structure of the ligand-binding site. Thus, anticalins are amenable to the construction of functional fusion proteins at both their amino- and carboxy-termini. This was demonstrated in the case of DigA16 for alkaline phosphatase, which could serve directly as a reporter enzyme for the detection of digoxigenin groups after fusion with either end of the anticalin [47]. Anticalins may even be fused to each other, leading to so-called “duocalins,” a novel class of bifunctional ligand-binding proteins [60].

The insights that were gained so far from the anticalin approach illustrate once again the huge potential of polypeptides to adopt diverse molecular shapes, seen here for the ligand pockets of engineered lipocalins. Given the high plasticity of the loop region, which probably constitutes a specific functional advantage of this protein family, the rational prediction of the influence of amino acid substitutions on the structure of the binding site will probably remain difficult in the near fu-

ture. But even when applying the powerful methods of combinatorial biochemistry, one should be aware of the still limited options for the realization of novel functional active sites, which is caused by the vast number of possible sequence combinations on the one hand and the restricted number of molecules that can be physically generated and applied to a selection experiment on the other.

We have tried to address this generic problem in protein engineering by making a careful choice of amino acids for random mutagenesis – in fact, just less than half of the positions that one could actually consider – in order to reduce the combinatorial complexity and thus create a potent molecular library in the functional sense. The fact that specific hapten-binding activities were immediately derived from this library confirms the validity of this concept. Nevertheless, additional steps of affinity maturation may be needed, as demonstrated for DigA16 [47, 49], in order to fine-tune the shape complementarity of the binding site after initial ligand recognition property for the ligand was imprinted. The rational choice of positions to be modified, in combination with repeated cycles of targeted randomization and selection – corresponding to a kind of molecular evolution – is probably the best general strategy for obtaining novel proteins with well-defined ligand-binding function, at least for the moment.

In conclusion, specifically engineered anticalins open numerous areas of application as ligand-binding proteins not only in bioanalytics and separation technology but also in medical diagnostics and possibly even therapy. Especially for the latter purpose, it could be advantageous to generate anticalins based on a human lipocalin framework [21]. Hence, immunogenic side effects will be minimized upon repeated administration to patients. The preparation of appropriate fusion proteins should permit the introduction of useful effector functions – as already demonstrated with enzymes or certain binding modules, such as the albumin-binding domain [47].

The field of engineered protein scaffolds for molecular recognition has rapidly emerged during the past few years (for reviews, see [9, 61]). Among the several scaffold structures that are currently being exploited, immunoglobulins and lipocalins certainly stand out. Both families are utilized by nature itself in order to provide specific binding proteins based on a stable tertiary fold that supports hypervariable loops. In the case of immunoglobulins, hundreds of millions of different antibodies are constantly being created in each individual's immune system using mechanisms of genetic recombination and somatic hypermutation. In contrast, lipocalins are much smaller in number and have been stably evolved in many organisms in order to serve more specialized physiological functions. Whereas antibodies must predominantly recognize macromolecular antigens – such as proteins and carbohydrates – in their defense against microbial or viral pathogens, lipocalins seem to mainly serve for the transport and storage of low-molecular-weight compounds.

Especially in this respect, lipocalins are distinct from other protein scaffolds that have been subject to protein engineering and successfully used for the generation of binding modules against protein targets. The promising results obtained here from the tailoring of a natural lipocalin for the recognition of haptens

like ligands emphasizes the unique potential of this protein family for the generation of corresponding receptor proteins. Apart from the interesting practical applications, this field of research will also offer conceptual insight into the mechanisms of molecular recognition between proteins and small molecules in general.

8.6

References

- 1 R. C. STEVENS, S. YOKOYAMA, I. A. WILSON, *Science* **2001**, 294, 89–92.
- 2 H. GOHLKE, G. KLEBE, *Angew. Chem. Int. Ed.* **2002**, 41, 2644–2676.
- 3 H. K. RAU, N. DEJONGE, W. HAEHNEL, *Angew. Chem. Int. Ed. Engl.* **2000**, 39, 250–253.
- 4 G. A. WEISS, H. B. LOWMAN, *Chem. Biol.* **2000**, 7, R177–R184.
- 5 B. HARRIS, *Trends Biotechnol.* **1999**, 17, 290–296.
- 6 M. R. UJHELYI, S. ROBERT, *Clin. Pharmacokinet.* **1995**, 28, 483–493.
- 7 D. R. DAVIES, S. CHACKO, *Acc. Chem. Res.* **1993**, 26, 421–427.
- 8 E. A. PADIAN, *Mol. Immunol.* **1994**, 31, 169–217.
- 9 A. SKERRA, *J. Mol. Recognit.* **2000**, 13, 167–187.
- 10 A. KOIDE, C. W. BAILEY, X. HUANG, S. KOIDE, *J. Mol. Biol.* **1998**, 284, 1141–1151.
- 11 S. MUYLDERMANS, M. LAUWEREYS, *J. Mol. Recognit.* **1999**, 12, 131–140.
- 12 S. SPINELLI, M. TEGONI, L. FRENKEN, C. VAN VLIET, C. CAMBILLAU, *J. Mol. Biol.* **2001**, 311, 123–129.
- 13 K. NORD, E. GUNNERIUSSEN, J. RINGDAHL, S. STAHL, M. UHLÉN, P.-Å. NYGREN, *Nature Biotechnol.* **1997**, 15, 772–777.
- 14 F. A. QUIOCHO, P. S. LEDVINA, *Mol. Microbiol.* **1996**, 20, 17–25.
- 15 Y. LINDQUIST, G. SCHNEIDER, *Curr. Opin. Struct. Biol.* **1996**, 6, 798–803.
- 16 M. WILCHEK, E. A. BAYER, *Biomol. Eng.* **1999**, 16, 1–4.
- 17 M. E. NEWCOMER, D. E. ONG, *Biochim. Biophys. Acta* **2000**, 1482, 57–64.
- 18 D. R. FLOWER, *Biochem. J.* **1996**, 318, 1–14.
- 19 G. GUTIÉRREZ, M. D. GANFORNINA, D. SÁNCHEZ, *Biochim. Biophys. Acta* **2000**, 1482, 35–45.
- 20 R. E. BISHOP, *Biochim. Biophys. Acta* **2000**, 1482, 73–83.
- 21 A. SKERRA, *Biochim. Biophys. Acta* **2000**, 1482, 337–350.
- 22 M. E. NEWCOMER, T. A. JONES, J. ÅQVIST, J. SUNDELIN, U. ERIKSSON, L. RASK, P. A. PETERSON, *EMBO J.* **1984**, 3, 1451–1454.
- 23 S. W. COWAN, M. E. NEWCOMER, T. A. JONES, *Proteins: Struct. Funct. Genet.* **1990**, 8, 44–61.
- 24 H. L. MONACO, M. RIZZI, A. CODA, *Science* **1995**, 268, 1039–1041.
- 25 S. PERVAIZ, K. BREW, *Fed. Am. Soc. Exp. Biol.* **1987**, 1, 209–214.
- 26 W. J. HENZEL, H. RODRIGUEZ, A. G. SINGER, J. T. STULTS, F. MACRIDES, W. C. AGOSTA, H. NIALL, *J. Biol. Chem.* **1988**, 263, 16682–16687.
- 27 F. VINCENT, D. LÖBEL, K. BROWN, S. SPINELLI, P. GROTE, H. BREER, C. CAMBILLAU, M. TEGONI, *J. Mol. Biol.* **2001**, 305, 459–469.
- 28 D. R. FLOWER, *J. Mol. Recognit.* **1995**, 8, 185–195.
- 29 E. RASSART, A. BEDIRIAN, S. DO CARMO, O. GUINARD, J. SIROIS, L. TERRISSE, R. MILNE, *Biochim. Biophys. Acta* **2000**, 1482, 185–198.
- 30 M. VOGT, A. SKERRA, *A. J. Mol. Recognit.* **2001**, 14, 79–86.
- 31 R. HUBER, M. SCHNEIDER, O. EPP, I. MAYR, A. MESSERSCHMIDT, J. PFLUGRATH, H. KAYSER, *J. Mol. Biol.* **1987**, 195, 423–434.
- 32 R. HUBER, M. SCHNEIDER, I. MAYR, R. MÜLLER, R. DEUTZMANN, F. SUTER, H. ZUBER, H. FALK, H. KAYSER, *J. Mol. Biol.* **1987**, 198, 499–513.
- 33 M. COLES, T. DIERCKS, B. MUEHLENWEG, S. BARTSCH, V. ZÖLZER, H. TSCHESCHE, H. KESSLER, *J. Mol. Biol.* **1999**, 289, 139–157.

- 34 D. H. GOETZ, S. T. WILLIE, R. S. ARMEN, T. BRATT, N. BORREGAARD, R. K. STRONG, *Biochemistry* **2000**, 39, 1935–1941.
- 35 R. GLOCKSHUBER, T. SCHMIDT, A. PLÜCKTHUN, *Biochemistry* **1992**, 31, 1270–1279.
- 36 P. A. SCHINDLER, C. A. SETTINERI, X. COLLET, C. J. FIELDING, A. L. BURLINGAME, *Protein Sci.* **1995**, 4, 791–803.
- 37 H. N. MÜLLER, A. SKERRA, *J. Mol. Biol.* **1993**, 230, 725–732.
- 38 F. S. SCHMIDT, A. SKERRA, *Eur. J. Biochem.* **1994**, 219, 855–863.
- 39 A. SKERRA, *Rev. Mol. Biotechnol.* **2001**, 74, 257–275.
- 40 G. BESTE, F. S. SCHMIDT, T. STIBORA, A. SKERRA, *Proc. Natl. Acad. Sci. USA* **1999**, 96, 1898–1903.
- 41 G. P. SMITH, V. A. PETRENKO, *Chem. Rev.* **1997**, 97, 391–410.
- 42 A. SKERRA, *Gene* **1994**, 151, 131–135.
- 43 E. W. VOSS, JR., *Fluorescein Hapten: An Immunological Probe*, CRC Press, Boca Raton, FL, **1984**.
- 44 C. F. BARBAS III, J. D. BAIN, D. M. HOEKSTRA, R. A. LERNER, *Proc. Natl. Acad. Sci. USA* **1992**, 89, 4457–4461.
- 45 A. SKERRA, T. G. M. SCHMIDT, *Methods Enzymol.* **2000**, 326, 271–304.
- 46 M. GÖTZ, S. HESS, G. BESTE, A. SKERRA, M. E. MICHEL-BEYERLE, *Biochemistry* **2002**, 41, 4156–4164.
- 47 S. SCHLEHUBER, G. BESTE, A. SKERRA, *J. Mol. Biol.* **2000**, 297, 1105–1120.
- 48 M. MUDGETT HUNTER, M. N. MARGOLIES, A. JU, E. HABER, *J. Immunol.* **1982**, 192, 1165–1172.
- 49 S. SCHLEHUBER, A. SKERRA, *Biophys. Chem.* **2002**, 96, 213–228.
- 50 A. R. HASTREITER, E. G. JOHN, R. L. VAN DER HORST, *Clin. Perinatol.* **1988**, 15, 491–522.
- 51 T. MCCREERY, *Mol. Biotechnol.* **1997**, 7, 121–124.
- 52 I. KORNDÖRFER, G. BESTE, A. SKERRA, *in preparation*.
- 53 M. WHITLOW, A. J. HOWARD, J. F. WOOD, E. W. VOSS, JR., K. D. HARDMAN, *Protein Eng.* **1995**, 8, 749–761.
- 54 C. CHOTHIA, A. M. LESK, *J. Mol. Biol.* **1987**, 196, 901–917.
- 55 J. FOOTE, G. WINTER, *J. Mol. Biol.* **1992**, 224, 487–499.
- 56 I. KORNDÖRFER, S. SCHLEHUBER, A. SKERRA, *in preparation*.
- 57 P. D. JEFFREY, R. K. STRONG, L. C. SIEKER, CY. Y. CHANG, R. L. CAMPBELL, G. A. PETSKO, E. HABER, M. N. MARGOLIES, S. SHERIFF, *Proc. Natl. Acad. Sci. USA* **1993**, 90, 10310–10314.
- 58 P. D. JEFFREY, J. F. SCHILDBACH, CY. Y. CHANG, P. H. KUSSIE, M. N. MARGOLIES, S. SHERIFF, *J. Mol. Biol.* **1995**, 248, 344–360.
- 59 J. KU, P. G. SCHULTZ, *Proc. Natl. Acad. Sci. USA* **1995**, 92, 6552–6556.
- 60 S. SCHLEHUBER, A. SKERRA, *Biol. Chem.* **2001**, 382, 1335–1342.
- 61 P.-Å. NYGREN, M. UHLÉN, *Curr. Opin. Struct. Biol.* **1997**, 7, 463–469.

9

Small Molecule Screening on Chemical Microarrays

G. METZ, H. OTTLEBEN, D. VETTER

9.1

Introduction

Since their introduction about a decade ago, combinatorial chemistry and high-throughput screening (HTS) have become indispensable tools in the drug-discovery process. The possibility to synthesize ever-increasing numbers of molecules through novel chemistries and automation is stimulating the development of higher screening capabilities through miniaturization and robotics. Robust biochemical assay development is providing the basis for large-scale screening of biological targets. While in the early days much effort and hope were directed towards managing a numbers game, the focus is shifting from quantity towards quality. For instance, the screening of compound mixtures is being replaced by screening of individual substances in a one-well-one-compound fashion. The design of general-purpose screening libraries as well as corresponding follow-up strategies has become a key aspect in small molecule discovery and optimization. Hits from high-throughput screening enter a selection process to become the subject of medicinal chemistry approaches in lead optimization. Strategies are employed to improve potency, selectivity, and physicochemical profile. If possible, several compound series are generated to allow for alternative routes in case of failure of one. Syntheses of analogues for further exploration are guided by a combination of medicinal chemistry knowledge and intuition, as well as quantitative structure-activity relationships, if available. Such studies help to define particular pharmacophoric features within the hit or lead molecule that constitute the underlying molecular recognition motifs between the ligand and its target and provide a hypothesis for its mode of action.

The focus in screening for biological activity assays is on detecting hits with activities in the low micromolar range. Compounds exhibiting this level of activity tend to be of molecular weight in the range of 300–600 Da and of substantial functional complexity. A conceptually different strategy can be envisioned that – instead of trying to identify and keep relevant features in a rather complex hit compound – aims at a stepwise discovery starting from molecular fragments. Such fragments need to be screened with techniques suitable for the detection of presumably weak interactions. Guided by early stage structure-affinity informa-

tion, more potent compounds can then be assembled either from a combination of low affinity binders or by chemical modification of the initial fragments.

This chapter describes the underlying philosophy of fragment-based discovery as well as the experimental approaches suitable for this promising discovery concept. While fragment-based ligand discovery was first adapted in computational methods, several experimental techniques have been introduced recently. Biophysical methods such as nuclear magnetic resonance (NMR) and X-ray crystallography have been successfully applied to fragment discovery. A novel screening technique based on chemical microarrays in combination with label-free affinity detection has emerged and will be discussed in detail.

9.2

Fragment Approaches

9.2.1

Conceptual Ideas

Primary screening efforts in drug discovery aim at the identification of hit molecules with the necessary characteristics to be developed into a promising lead molecule. The definition of favorable properties of the starting screening compounds has gained much attention. The design of libraries with drug-like characteristics generally follows the so-called “rule of five” which has been established by retrospective analysis of known drugs and allows a quick assessment based on simple properties, namely, preferred ranges for molecular weight and $\log P$ as well as the number of hydrogen bond donors and acceptors [1]. However, it has been proposed that the ideal profile for hit or lead compounds is different from that of the final drug molecule [2], in particular because hit or lead compounds must be amenable for further optimization. Three categories of lead compounds have been defined based on their physicochemical properties and typical affinities. First, it has been pointed out that hits from drug-like libraries rarely show high (<100 nM) affinities. Typical optimization schemes tend to increase both molecular weight and lipophilicity. Therefore, if applied to already large compounds, they would fall out of the preferred drug-like ranges. Second, non-drug-like hits with very high affinity at an early stage, e.g., natural compounds, usually are rather complex, and optimization towards drug-likeness is difficult. The third category, namely, small (100–350 Da) molecules with low $\log P$ (1–3) is being described as a favorable type of lead compound, as it still allows exploitation of additional (hydrophobic) interactions during combinatorial optimization schemes. Overall, the value of active screening compounds is judged considering affinity relative to units of molecular weight and lipophilicity. Teague et al. suggest that the quality of hits emerging from screening could be improved by tailoring screening libraries towards such lead-like characteristics [2].

In the case of hit identification, various techniques of library design and compound selection are applied to maximize the likelihood of discovery. The relation-

ship between the probability of discovering molecules in a screening assay and their structural complexity was studied using a model of receptor-ligand interactions [3]. The ligand and the active site are represented by a linear string of binary (“+” and “-”) features, and a good fit is indicated by complementarity of the aligned patterns. While molecular recognition is determined by a delicate interplay of physicochemical and steric complementarity, the apparent simplification still allows one to address some key aspects. Varying degrees of relative complexity between the active site and the ligand are represented by different lengths of the binary patterns. By calculating matching probabilities for varying ligand complexities, it is observed that the likelihood of finding any fit exponentially decreases with ligand complexity and that the curve describing the chance of finding just one unique match between ligand and active site features reaches a distinct maximum. The theoretical complexity of such single-mode binding was found to peak at a binary pattern length of three when compared to a site complexity of 12 features. For less complex ligands, multiple binding modes start to dominate the probability curve. While this finding cannot be directly translated into simple molecular features, it points towards a higher overall chance of hit discovery for relatively small molecules. However, smaller molecules will probably exhibit weaker binding, and therefore chances to experimentally detect binding will depend on the screening technology. Taking this into account, the peak of preferred ligand complexity would shift towards somewhat larger compounds. In summary, the existence of an optimal ligand complexity and its relation to the detection probability defines a “range of useful events” (Fig. 9.1) and leads to the postulation that screening simpler molecules is advantageous from a probabilistic point of view [3].

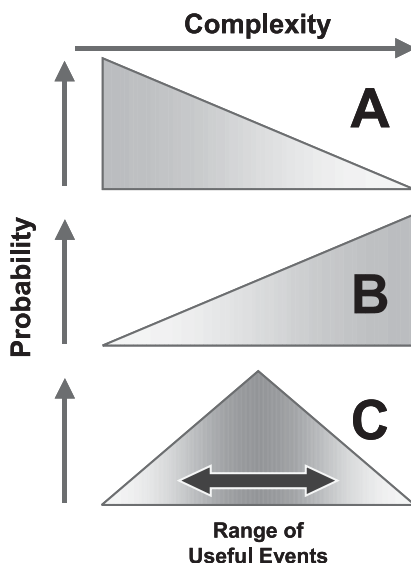


Fig. 9.1 The theoretical probability of molecular recognition based on a simple interaction model indicates that the likelihood of a unique binding mode decreases with increasing ligand complexity (A). The probability to experimentally detect a binding event is estimated to increase with complexity (B). The product probability for a so-called useful event, namely, the detection of a ligand with a unique binding mode, reaches a maximum at a medium ligand complexity (C). (See citation in text for details and discussion).

Molecular recognition addresses the aspect of not only affinity but also specificity, and the question of unique binding modes becomes fundamental for smaller molecules. In general, the multitude of energetically similar but structurally different binding modes increases with fewer interaction features of the ligand. The corresponding free energy of binding landscape is termed “frustrated,” and both native and non-native binding modes of small molecules must be considered [4]. Ideally, the native binding mode is separated from alternate positions in the binding site by an energy “stability gap,” which ensures that a rather specific recognition motif is present (Fig. 9.2). Small fragment-like compounds exhibiting such a preferred binding mode, so-called molecular anchors, may be more suitable to combinatorial optimization than stronger binders with iso-energetic multiple binding modes [5]. With anchors serving as a receptor-specific recognition motif, a native binding mode would imply that addition of structural features to a binding fragment will not dramatically alter its orientation. Such modular approaches rely on the observation that additivity if not synergy is obtained when preselected fragments are combined. A point in case is a detailed structural comparison of enzyme-ligand binding which showed that binding modes for individual fragments of the thymidylate synthase substrate correspond well with the whole substrate in the active site [6].

The relationship between the potential affinity of a given molecule and its actually measured binding strength to a given target has been addressed by analyzing individual functional group contributions to drug-receptor interactions [7]. Average binding energies of 10 common functionalities have been estimated based on a dataset of 200 drugs and enzyme inhibitors with known free energies of binding. The average binding energy for a given ligand, often also referred to as Andrews energy, can then be calculated by adding up the average contribution of the functional groups present in the molecule and taking into consideration an entropic correction term. The comparison of experimental and theoretical binding energy may then be used to classify the binding as better or worse than average, in other words as an indicator of a good or bad ligand-receptor fit. Poor binders were found to be large, flexible, and rich in polar groups. A similar finding was reported for 3000 screening compounds, where a large proportion exhibited lower observed affinities when compared to the average

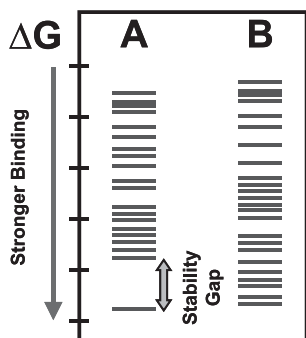


Fig. 9.2 Concept of a stability gap between different binding modes of a fragment within an active site. The assumed binding energy for two different fragments is shown as a simple scheme. Fragment A qualifies as a molecular anchor, as there is a difference in free energy of binding of the best placement to alternative placements. Such a stability gap is not observed for the second fragment B. Originally, these considerations were made based on docked placements of fragments and free energies as estimated from a scoring function (see citation in text).

Andrews energy based on the number of functionalities present. The general non-linearity of binding affinity and ligand size has been pointed out in comparing binding affinities relative to the number of non-hydrogen atoms in known protein-ligand complexes [8]. While a correlation of binding energy with size was found for smaller ligands, there also is evidence for an upper limit of observed affinities for larger ligands that therefore under-perform according to their Andrews energy. The authors note that as few as 7–10 atoms allow for nanomolar binding constants and also suggest to assess relative affinity during a drug-discovery process using expansion approaches. In a more pictorial phrase, it is the “bang per dalton” that has to be kept in mind when evaluating the outcome of hit and lead finding efforts. This ratio tends to be better for smaller-sized compounds.

9.2.2

Choice of Screening Fragments

The quality and quantity of fragment collections are critical issues for all fragment-based discovery methods. Some considerations are related to deconvolution strategies based on unique NMR shifts or unique mass or shape, while others are sensitive to the detection limits and noise level of the respective experiment. Besides these technical aspects, the goal is to identify those fragments that best suit the concept of molecular anchors or recognition motifs. On the one hand, a certain size and complexity make a defined binding mode more likely. On the other hand, in order to qualify as a fragment, screening compounds would be limited to a molecular weight range between 100 and 300 Da. The total number of fragments in a screening collection is related to the experimental throughput and the necessity to cover the “accessible fragment space” in order to increase the chances of finding a hit in the initial screen. The sufficient presence of functional groups in the screening fragments is of importance for subsequent hit expansion strategies such as combining or decorating the initial hits.

Systematic procedures for the identification of suitable fragments or substructures have been implemented that aim to analyze active molecules in order to identify biologically relevant motifs. Because the field is expanding, a diverse nomenclature is evolving; synonyms for chemical fragments are needles, molecular anchors, biophores, molecular frameworks, MULBITs, or base fragments. In a procedure called RECAP [9], fragmentation is limited to predefined bond types leading to virtual fragments with chemical functionalities that allow them to be used as building blocks for combinatorial chemistry. The method was applied to a number of databases, including the World Drug Index (WDI). By examination of fragment distribution across different therapeutic areas, specific motifs were identified for use in target-biased screening libraries. Typically, the molecular weight of most fragments was around 200 Da and the cleavage rules kept ring systems intact. While the approach mainly addressed the question of building block selection for combinatorial libraries, the authors also pointed out its usefulness for generating a fragment database for (computational) ligand buildup approaches. Different variations of the basic approach have been implemented [10–12].

In a related work, databases of known drugs were the basis for breaking down molecules into rings, linkers, and side chains [13, 14]. In this way both common molecular frameworks and side chains were identified. Interestingly, it was found that rather few frameworks and side chains represent the majority of compounds in drug-related compound collections such as the Comprehensive Medicinal Chemistry (CMC) database. Incorporating this finding into a general-purpose screening library resulted in the so-called SHAPES library for NMR-based screening [15]. Natural product databases were also used for the search of interesting molecular scaffolds [11].

It has long been pointed out [16] that certain substructural motifs, named privileged structures, are capable of serving as a starting point for ligands for more than one protein. Systematic modifications of such structures with substitution patterns have been successfully applied in medicinal chemistry [17]. Often these attractive scaffolds are rigid polycyclic heteroatomic systems that allow us to present binding elements in various directions within the binding site. The concept has been adopted as a design strategy for combinatorial libraries [18], and corresponding building blocks are being marketed under the name “optimers” [19]. Experimental screening of compounds enriched with motifs based on such privileged fragments is thought to increase the chances of identifying promising hits for medicinal chemistry programs. Even more, fragment-based screening would allow us to directly use known privileged fragments or close analogues thereof. On the other hand, screening fragments that are not necessarily related to known privileged structures might help us to discover novel chemical motifs displaying properties that qualify them to be called “privileged” [20].

In summary, fragment collections for the experimental screening techniques that will be outlined below either fall into the category of “diversity-oriented” collections or are selected with some bias. The focus is on “drug-relevant” substructures or structures that are targeted against a certain class of proteins where prior knowledge exists. In addition to finding novel leads based on fragment screening, the technique can also be applied in the optimization of a known binder where a particular substructure needs to be replaced by a bio-isoster. A prime example is the quest for alternatives to charged basic groups occupying the thrombin S1 binding pocket in order to enhance bioavailability [21].

9.2.3

Experimental Approaches

Fragment-based discovery methods have long been implemented in computational *de novo* drug design [22, 23]. In the multiple fragment-positioning methods, various functional groups are first placed within the active site, and, after preferred placements have been identified, linking of the fragments provides molecules that can be ranked according a chosen scoring function. Alternatively, sequential build-up strategies start from one placed fragment and successively add functional groups in order to “grow” the ligand, guided by the target structure and a suitable energetic scoring function. A similar strategy is also used in docking programs

using an incremental construction algorithm that first performs a fragmentation of the ligand and then, starting with a base fragment placement, continuously builds up the ligand in the binding site. Clearly, these techniques are very much related to the molecular anchor concept discussed above. Besides problems associated with *de novo* computational approaches such as the synthesizability of proposed virtual compounds, the most critical issue is the fragment ranking, which is calculated by either energy-based methods or rule-based scoring functions. These predictive limitations in computationally derived binding conformation make experimental methods that are guided by either observed activity or binding affinity very attractive.

In a hybrid approach, termed “biased needle screening,” *in silico* prescreening of molecular fragments is followed by a high-concentration bioassay, biophysical hit validation, and structure-based optimization [24]. Virtual pharmacophore screening of 350,000 compounds resulted in the identification of a 3000-member subset with molecular weights below 300 Da. These needle compounds were tested in an activity assay customized to pick up even weak binders. Several structural series could be confirmed, and a 3-D structure-guided optimization based on NMR and X-ray data gave novel, potent inhibitors of DNA gyrase. The authors state that initial HTS on this target did not deliver suitable lead structures. This finding supports the concepts outlined above and demonstrates the usefulness of needle screening as a new entry point to explore the chemical space.

The so-called “target-guided ligand assembly strategy” starts from a library of possible binding fragments where each member possesses a common chemical linkage group [25]. Monomer screening against the target is then performed in a bioassay at high concentrations in order to detect even weak binders that then can be connected with a set of flexible linkers for a second round of screening. The utility of the method was demonstrated by using the tyrosine kinase c-Src and a microtiter-based ELISA assay. After screening at two concentrations (1 mM and 500 μ M) and inspection of the hits, 37 reagents were selected out of 305 O-methyl oximes in the primary screening library. Using five different linkers, homo- and heterodimers of these reagents were synthesized in single wells, resulting in mixtures of compounds with different linker length. After identification of wells showing inhibition in a second screening assay at higher concentration, deconvolution of the mixtures by single-compound re-synthesis allowed to identify individual active substances and highlighted the importance of the linker length. The most potent compound identified in this manner exhibits an IC_{50} of 64 nM, a very large increase compared with IC_{50} values around 40 μ M for the individual reagents alone. As with other methods described above, the synergy achieved and the lower number of compounds needed for screening demonstrate the advantage of modular approaches in general.

In order to discover weakly bound, low-molecular-weight (approx. 250 Da) ligands, a “tethering” strategy has been suggested that relies on the formation of a disulfide bond between the ligand and a cysteine residue on the protein [26]. This cysteine would either be present in the wild type or be genetically engineered in order to target a specific site in the protein. A library of disulfide-containing mole-

cules was prepared and a mixture of a few substances (8–15) was incubated with the target protein under conditions that allow reversible reactions. It is expected that the formation of disulfide bonds to the protein is entropically stabilized for those compounds with an inherent affinity to the protein in proximity to the cysteine. Tethered complexes could be identified using mass spectrometry, provided that the library had been designed to contain molecules with unique molecular mass. The method was applied to thymidylate synthase containing a cysteine in the active site. Several library members showed binding corresponding to millimolar inhibition constants, as determined in an enzymatic assay. From a series of related ligands displaying different binding behavior, qualitative early SAR information was obtained. The exact binding mode was determined with the same ligand attached at different nearby sites through cysteine mutations; interestingly, the location of the tethered molecule was conserved, indicating little influence of the tether on the binding mode. Structure-based modifications in analogy to known substrates improved the initial hit to an inhibitor in the sub-micromolar range.

Techniques such as X-ray crystallography and NMR, which combine low affinity screening capability with structural information, are very powerful tools. NMR spectroscopy plays an increasingly prominent role among the biophysical screening methods, and the experimental schemes are continuously being improved to widen the scope of applications [27–32]. Different detection strategies based either on ligand or target resonance signals have evolved. The approaches share the ability to detect weak yet specific binders. Design principles for NMR screening compounds have been reviewed [33].

An early and frequently cited fragment-based experimental screening approach was termed “SAR by NMR” [34, 35]. Here, mixtures of small organic molecules together with ^{15}N -labeled protein were subjected to 2-D ^1H - ^{15}N NMR measurements. Protein chemical shift variations relative to spectra of the protein alone indicated a binding event. Based on the predetermined assignment of chemical shifts, both the location of the binding site in the protein and the binding molecule were identified. This initial binder was then used to saturate the protein in order to find a second small molecule binding to a proximal site in a new round of screening. After neighboring small molecules were identified and optimized through an analogue approach, combinations with various linkers were synthesized and assayed. The linker design was supported by the 3-D structure of the protein-ligand complex. As a result of a successful linkage, binders were obtained with high affinities even exceeding the product of the binding constants of the individual fragments due to linker-mediated entropic enhancement. The method has been successfully applied in a number of studies [36–40] elegantly combining the fragment-based strategy with use of structural information. However, the original method can be applied only if sufficient (~ 200 mg) ^{15}N -labeled, soluble (at 2 mM) protein of limited size (< 40 kDa) is available, and it requires chemical shift assignment before the actual screening.

The necessity for labeled proteins and the size limitation can be overcome by techniques that monitor not protein NMR signals but ligand resonances, either through line-broadening experiments, transferred NOE measurements, or relaxa-

tion- and diffusion-edited methods using pulsed field gradients. The NMR SHAPES strategy utilizes 1-D line broadening and 2-D transferred NOE measurements to identify binders in a mixture of compounds [15]. Potential weakly binding scaffolds (μM to mM) are selected based on an analysis of known drugs and represent molecular frameworks recurrently found in active therapeutic molecules. In contrast to the SAR by NMR method, the strategy does not aim to find highly potent ligands by NMR screening but instead empirically provides a basis for compound selection. In this way, libraries aimed at HTS screening may be biased by filtering against the target-specific chemical motifs identified by the SHAPES approach.

The use of X-ray crystallography for primary screening is conceptually related to the SAR by NMR technique. Again, the sensitivity of the method makes it well positioned for fragment-based discovery. Because organic solvent molecules contain functional groups representative of those found in screening compounds, it has been suggested to co-crystallize proteins with different solvents and to experimentally determine preferred locations of small organic molecules [41]. The position of several organic molecules was thought to provide initial templates for selection of screening compounds or rational ligand design efforts. In yet another fragment strategy termed CrystaLead [42], the electron density map of the protein was determined to identify protein and solvent densities in the unbound state. Next, the crystal was exposed to a mixture of small organic molecules in a soaking experiment, and binders were identified by their appearing electron densities. It was crucial for the identification of the binding compounds in the screening mixture that they were of diverse and unique molecular shapes. Weak binding compounds (up to high μM) were detected and then optimized in a structure-directed process. The CrystaLead technology was demonstrated in a urokinase screen of 6–8 compound mixtures against 9 crystals. In this way, 61 fragments were exposed to the crystal soaking experiment, and 5 binders with affinities in the low and high μM range were discovered. One compound was further optimized using additional structural knowledge from a known inhibitor and resulted in an optimized lead of 370 nM affinity to urokinase.

Crystallographic screening has gained significant interest in the commercial setting where high-throughput crystallography laboratories and technological advances in both hardware and software have increased the rate of protein-ligand crystal structure determination. The advent of structural genomics initiatives has spurred the development of robotics and automated data interpretation [43, 44]. The latest developments have recently been reviewed [45–47]. These technological advances are now being paired with fragment-based approaches. Results have been reported [46] where defined binding could be elucidated for very small fragments (<200 Da), supporting the fact that despite small size and presumably weak interactions, fragments are able to make key interactions in order to act as molecular anchor [4]. In a study with a protein kinase, such a weak initial fragment was optimized into a nanomolar inhibitor by the aid of structure-based design [45].

9.3

Chemical Microarrays

9.3.1

Background

Chemical microarrays can be defined as collections of chemical compounds covalently immobilized on a carrier in a 2-D pattern. After the widespread adoption of DNA-based microarrays in basic and commercial research, there is a growing interest to extend the array concept beyond genetics applications. Exploratory work is underway to lay out proteins and cells in array formats [48–51]. In parallel, various routes are being taken to realize arrays of small molecules such as synthetic chemicals, peptides, or natural products. The principle of a regular 2-D arrangement of chemical diversity is well known from high-throughput screening, which currently serves as the paradigm in *de novo* small molecule discovery. In this highly industrialized undertaking, compounds are solubilized and deposited in wells of microtiter plates as either single compounds or mixtures. A recent modification of HTS is the deposition of chemical-containing droplets on flat surfaces [52] to reduce sample consumption. In the latter, the target protein is captured in a hydrogel that is subsequently added to cover the surface and re-solubilize the screening compounds. It is an inherent problem of HTS that small molecule solubility can vary greatly and can hardly be accounted for. Therefore, the concentration of any given compound in the screening mixture is difficult to predict and essentially unknown at the primary screening stage. This limitation is especially apparent in a fragment-based screening regime. Because higher compound concentrations are employed to facilitate the detection of binders below the usual micromolar cutoff, DMSO tolerance of the target protein or insolubility of the small molecules restricts this approach to robust assay systems and more hydrophilic fragment diversity.

When we set out to develop a platform for chemical microarrays, compatibility with fragment-oriented diversity was a major design principle. We saw great promise in also solubilizing hydrophobic compounds by hydrophilic or amphiphilic spacer moieties. Such spacer groups are required for covalent tethering of the ligands to the array surface and for access to protein-binding pockets. The amphiphilic nature of the spacer aids hydration of small molecules associated with high $\log P$ values. The flexibility and the length of the spacer chain allow for the ligands to reach even so-called deep protein-binding sites.

Covalent tethering of small molecules is generally seen as an obstacle for deriving structure-activity relationships. The compounds are not free to orient themselves for optimal fit with the protein receptor, the spacer might sterically clash with the protein surface, and the spacer attachment itself might affect the electronic properties of the ligands. Consequently, in order to generate covalent modifications of known drugs or other compounds known to be bioactive, the decision for regioselective spacer attachment should be based on a receptor-ligand complex structure if possible. Nevertheless, for *de novo* discovery of small molecule frag-

ments, the spacer-mediated, structure-independent compound solubilization outweighs the above-mentioned potential disadvantage of restricted orientation. Furthermore, fragments can be conjugated to the spacer moiety by more than one type of coupling chemistry and through various functional groups present on the fragment itself. A third aspect of the small molecule “display” format is the additional information generated through knowledge of the tether site. The fact that possible orientations of a tethered compound within a binding site are more limited than in a homogeneous assay can facilitate building a hypothesis for a mode of action. The directionality can be used either in ligand-alignment procedures for SAR-type studies or as a bias in docking algorithms in order to evaluate possible binding modes.

Other advantages of using chemical microarrays of peptides and organic compounds are in the minimization of biological sample consumption and the resulting increase in screening throughput. Array approaches hold the promise of mass production and industrialization. A simple format for reliably storing and reading chemical diversity is attractive compared to the current cumbersome fashion of operating refrigerated warehouse-type storage and dispensing systems.

Chemical microarrays promise to enable function-blind screening of large numbers of novel targets. In general, chemical microarray approaches open up the opportunity to map interactions and discover small molecule binders for a given protein of interest even before understanding the protein’s function. Once the arrays are produced, the only prerequisite for array-based screening is the preparation of purified, homogeneous, and soluble protein. Advances in protein-production and -purification methods, such as expression systems optimization, protein folding, and affinity-tag development, complement array-based screening. It can be safely estimated that several thousands of novel, non-membrane-bound, putative target proteins will be derived from the knowledge of the human genome [53]. Chemical microarrays provide a potentially powerful alternative to high-throughput screening for small molecule *de novo* discovery on this wealth of novel targets.

9.3.2

On-array Synthesis

Chemical diversity for chemical microarrays can be accessed in two ways: importing compounds onto the chip surface or synthetically creating molecules directly on the support. In the pioneering work by Fodor and coworkers, a combination of solid-phase organic synthesis and photolithographic techniques was applied to the *in situ* creation of oligonucleotide and peptide microarrays [54, 55].

Technological challenges associated with this approach have limited it to the generation of oligonucleotide arrays. DNA consists of only four different nucleotides, and differences in reactivities among the four nucleotides on glass surfaces are well understood, and the solid-phase synthesis is highly optimized. Hence, highly miniaturized synthesis of a vast diversity of different single-strand probes became feasible, and corresponding commercial products derived from photolithography are in widespread use. In contrast, the application of on-array synthesis

techniques is more demanding for the generation of peptidic and even more challenging for combinatorial chemical libraries. In these cases, a much larger number of building blocks with greater ranges of reactivities is required. The greatest hurdle to high-density, *in situ* array synthesis on glass is quality control of the numerous products generated. Techniques to infer physicochemical characteristics of monomolecular films in micrometer-sized areas are just beginning to evolve and are still far away from becoming a routine application such as high-pressure liquid chromatography-coupled mass spectrometry (LC-MS), which is used for conventional combinatorial compound library analysis.

A different way to produce chemical microarrays *in situ* is spot synthesis of combinatorial libraries on cellulose sheets [56]. Spot synthesis is configured as an open system to be operated at room temperature. Despite attempts to replace cellulose with polypropylene as a synthesis support [57], cellulose is still the support of choice for spot synthesis, and reaction conditions have to be compatible with the restricted chemical stability of cellulose. Due to the acid lability of such membranes, the diversity content of these arrays was initially restricted to the synthesis of peptides. Recently, a method was described that could widen the scope of spot synthesis arrays. Germeroth and coworkers [57] succeeded in the assembly of a library of 8000 cellulose-bound 1,3,5-triazines under mild reaction conditions. They employed a strategy that took advantage of a temperature-dependent, successive displacement of cyanuric chlorides by different nucleophiles in a first report of the synthesis of small organic compounds on cellulose sheets.

In addition to the limited range of cellulose-compatible synthesis protocols, two further drawbacks remain that are inherent to on-array approaches. Because synthesis takes place directly on the surface that is subsequently used for screening, quality control of the synthesis products is restricted to the surface-bound molecules. Standard cleave-and-characterize procedures involving LC-MS analytical techniques are not practical. This is a severe problem even for the synthesis of peptides where well-established protocols are available and becomes more pronounced when novel chemistries have to be employed. Secondly, each array produced is unique, which renders the production rather costly and prevents the generation of numerous copies of the same array for high-throughput applications.

9.3.3

Off-array Synthesis and Spotting

Some of the problems associated with on-array *in situ* synthesis can be overcome by a technology recently published by Stuart Schreiber's group. In this work, the compound diversity was generated in solution or solid-phase formats different from the array layout, and chemical microarrays were subsequently produced by spotting pre-synthesized molecules [58–60]. In a proof-of-principle experiment, three different organic compounds were immobilized on a chip surface [58]. For this purpose a silanated glass slide was derivatized to give a surface densely functionalized with thiol-reactive maleimide groups. Onto these surfaces, a high-precision robot delivered approximately 1 nl of a solution of the three different organic

compounds, which were pre-synthesized to contain a spacer with a thiol group. In this way, an array of 10,800 spots each 200–250 μm in diameter was produced on a 2.5×7.5 cm glass slide. The chips were afterwards probed with respective fluorescence-labeled proteins for selectively binding their ligands. The fluorescence intensity recorded on the spot reflected the binding affinity of the respective ligand. This finding demonstrates the feasibility of semi-quantitative measurements of ligands with binding affinities in the low micro- to nanomolar range. In a second communication a similar technical setup was combined with a different chemical immobilization strategy [59]. The same set of organic probe molecules was pre-synthesized, this time containing primary alcohol groups, and spotted directly onto thionyl-chloride-activated glass slides. A small combinatorial library comprising 78 compounds was prepared in analogous fashion and probed with a target protein, and a new “hit” was identified in addition to the positive controls. Although the initial results of this approach were very encouraging, discovering new ligands for broader protein diversity would require a larger number of organic compounds. Moreover, the surface chemistries employed in both papers did not allow a tight control of the number and density of the compounds on each spot of every chip. This might become limiting when a more quantitative analysis of the results is required or when lower affinity interactions are to be analyzed.

In the approach cited above, both the nanoliter droplet deposition technique and the surface design of the array support were adopted from DNA microarray fabrication [61]. The presentation of oligonucleotides or DNA on surfaces as well as the readout of such chips in hybridization experiments are facilitated by the fact that both ligand and receptor molecules are relatively similar with respect to solubility, charge density, and pI. The physicochemical properties of nucleotides make it possible to design surfaces that show relatively little background binding under various conditions. However, the situation is quite different if proteins are the target receptors and protein binding rather than hybridization is of interest. Proteins differ dramatically in stability, solubility, and hydrophobicity and can be basic or acidic. The challenge here is to design a surface that resists the unspecific binding of a wide range of different proteins. In addition, probing a DNA microarray by hybridization is relatively simple, as DNA is rather stable and the binding constant for a double-strand formation is relatively high, allowing one to subject the molecules to stringent hybridization conditions. The interactions of small organic molecules or peptides with proteins are often much weaker and range mainly from milli- to micromolar binding constants. This requires high-performance surfaces to minimize unspecific binding.

A common drawback of glass slides as supports for chemical microarrays, as well as of cellulose or polypropylene sheets, is the protein compatibility of their respective surface chemistries. It is inherently difficult to render polypropylene or silanated glass resistant to unspecific protein adsorption and is even more challenging to control a critical parameter such as ligand density on these polymers. For instance, engineering for biocompatibility of resins used in organic synthesis has become a major challenge for “bead-binding assay” screens [62].

In contrast, self-assembling monolayers (SAMs) of thiols on gold are not only among the best-characterized and best-defined synthetic surfaces but also can be designed to exhibit very low background protein binding. SAMs give excellent control over surface properties at the molecular level. The basic principles of SAM formation and applications in protein binding to immobilized ligands have been reviewed [63, 64]. In general, an alkane thiol is chemisorbed onto a gold surface, and the packing of the hydrocarbon chains creates a dense monolayer on the gold. Attaching an oligo(ethylene oxide) to the hydrocarbon chain confers excellent resistance to nonspecific protein and DNA binding, biocompatibility, and non-fouling properties to the surface [65–67].

Reactive SAMs form uniform layers that contain reactive groups such as amines or carboxylic acids at their surface [68]. The ratio of alkane thiols in a mixture of molecules with and without a reactive group allows one to control the presentation of ligands in a defined surface density (Fig. 9.3). Carboxylic acid-containing surfaces can be activated by pentafluorophenyl or *N*-hydroxy succinimide esters. Reactive SAM technology has recently been used for peptide [69] and protein arrays [70] as well. Using the immobilized ligand benzene sulfonamide together with carboanhydase, it was demonstrated that the unspecific binding can be minimized and that there was no irreversible adsorption. The measurements also were not complicated by mass transport [71, 72]. However, when rate constants for association and dissociation were analyzed, some lateral steric effects affected the binding kinetics. Ligands displayed on mixed self-assembled monolayers ligands in well-defined sur-

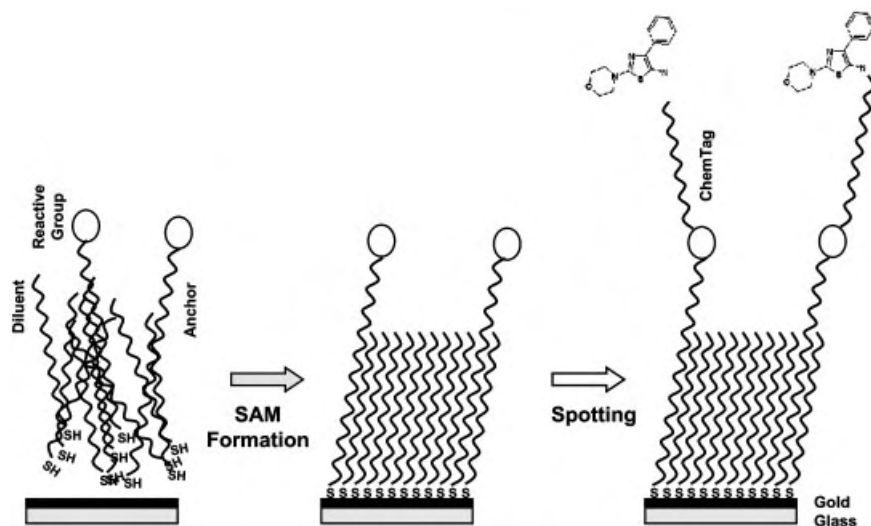


Fig. 9.3 Creating of a chemical microarray occurs in two steps. A self-assembled monolayer is formed on gold-coated glass from a diluent and an anchor molecule carrying a reactive group. Spotting of ligands attached

to a ChemTag linker results in covalent immobilization of organic molecules. The uniform density of the ligands at each spot is defined by the ratio of anchor and diluent molecules.

face densities offer control over critical parameters to minimize unspecific or irreversible binding or deviation of the binding kinetics from those observed in solution.

In our laboratories, we developed a proprietary SAM technology comprising a functionalized anchor molecule diluted with unfunctionalized spacer molecules. The chemical functionality is used to capture spotted ligand molecules, which themselves are carrying a linear spacer and functionalization molecule, named ChemTag. The system was tested against numerous well-known pairs of interacting molecules such as receptor:ligand, enzyme:co-factor/inhibitor, and antigen:antibody for detection of specific interactions. Using a variety of proteins differing in hydrophobicity and molecular weight, the results showed that – irrespective of the protein chosen and even at high concentration – unspecific binding was minimized. These experiments proved that interactions between proteins and ligands of low molecular weight (<200 Da) as well as weak affinities up to the millimolar range can be specifically detected, thus fulfilling a major requirement for the use of chemical microarrays for fragment-oriented screening applications.

Building on the advantages provided by reactive mixed self-assembled monolayers, we designed both a compatible spacer chemistry and a miniaturized format for highly parallel solid-phase synthesis. Compounds are synthesized through combinatorial chemistry on a solid phase that is preloaded with the ChemTag spacer molecule. This linear spacer carries two functional groups, located at alpha and omega sites. The alpha group is used for transient conjugation to the synthesis support through a standard linker and remains protected in this fashion during the synthesis cycles. The omega group is available for ligand synthesis and can be functionalized in various ways to allow for different coupling chemistries. Cleavage from the solid phase by splitting the linker-alpha group bond yields free, spacer-modified compounds in solution, with the alpha group free to be conjugated to the reactive sites on the microarray SAM surface. The common spacer group acts as a chemical shuttle to transport substances from one solid phase to another (Fig. 9.4). After solid-phase assembly, products are separated from the synthesis resin and stored in mother microtiter plates. At this stage, aliquots are subjected to quality control using LC-MS. After quality assessment, nanoliter amounts are transferred from the mother plates onto the surfaces of microarrays coated with the reactive mixed SAM. High-precision spotting of the substances in a custom-built automated environment ensures reproducible immobilization of the compounds onto the microarrays.

A great variety of organic molecules can be attached through one of the different omega-site chemistries of the ChemTag spacer. The array content of our chemical microarrays ranges from the immobilization of single fragments to combinatorial libraries. As an example for a fragment array, up to 1536 individual monomers have been attached to the ChemTag and have been immobilized on a single array. Using different chemistries, multifunctional fragments can be immobilized in more than one way. In addition, binary libraries with a batch size of 10,000 compounds have been synthesized on a nanoscale level (40 nmoles/compound) and then spotted on a single array. This rather low synthesis scale of chemical products is sufficient for the production of hundreds of ready-to-screen arrays (Fig. 9.5).

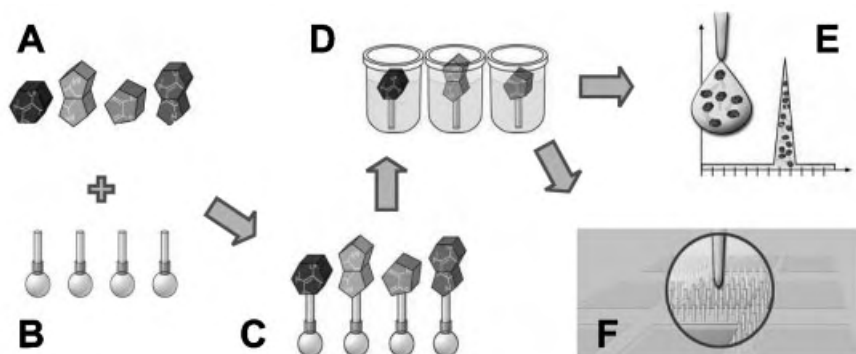


Fig. 9.4 The chemical microarray production starts by assembling combinatorial libraries or single fragments (A) to the ChemTag, which itself is attached to a solid-phase resin (B) resulting in compounds attached to the resin (C). After cleavage from the resin, the

tagged compounds are stored in mother plates (D) and LC/MS quality control can be performed (E). The final step is spotting of the compounds onto a prefabricated microarray carrying a reactive self-assembled monolayer (F).



Fig. 9.5 Chemical microarrays as they are used at Graffinity. The array is based on a microtiter plate footprint and carries up to 10,000 individual compounds spotted in rows and columns onto an optical microstructure.

9.4

Screening on Microarrays

9.4.1

Detection Technology

Chemical microarrays that are based on glass or cellulose or other polymer surfaces were typically interrogated for protein binding by fluorescence or chemiluminescence, respectively. Fluorophore-labeled protein samples were incubated with chemical microarrays, washed, and subsequently scanned by commercial microarray readers. The use of orthogonally labeled proteins was also described [58, 59]. Compound libraries on cellulose sheets are not readily subjected to fluorescence imaging, but a multi-step process of target protein and target-directed antibody together with enzyme-conjugated secondary antibody binding serves to obtain high-sensitivity images of protein-binding patterns.

In our approach towards chemical microarray screening, we wanted to exploit the fact that self-assembled monolayers readily lend themselves to label-free detection based on surface plasmon resonance (SPR). The key to this approach is a thin gold metal film on a glass support that allows for both the physical effect of plasmon formation and the formation of high-quality self-assembled monolayers. Label-free imaging for chemical microarray readout is attractive because it reduces unspecific binding compared to chemiluminescence, as antibody and enzyme labels are not brought into contact with the array. The approach also allows us to observe the direct binding pattern of the unmodified target protein, as covalent modifications such as fluorophore conjugations are obviated.

Molecular recognition of an immobilized ligand and its solubilized binding partner can be detected by a physical phenomenon called surface plasmon resonance [73]. In general, the sensor chip comprises a glass prism covered with gold and coated with a SAM presenting the potential ligands (Fig. 9.6). Protein solution is added to a reservoir on top. Upon binding, an affinity-dependent mass change occurs at the interface of the detector surface and the liquid above. This influences the dielectric properties of the gold layer. The readout is achieved by measuring the exact resonance condition for an energy transfer from the photons of a light beam to the electrons in the gold layer placed on top of a glass prism. An incident beam from the bottom is reflected at the gold surface and three factors, namely, the angle of incidence, the wavelength, and the refractive index at the interface, determine the resonance condition. For a fixed angle, the wavelength-dependent minimum of the reflected light is quantitatively related to the mass change due to binding. Likewise, an angular dependence at a defined wavelength can be recorded.

Optical biosensors and in particular those based on SPR have gained importance in many areas. The technique itself, biological applications, and the availability of commercial instruments have been reviewed [74]. The most common application in the area of protein-ligand interactions is the detailed study of binding kinetics [75–79]. It relies on immobilization of one of the binding partners on a

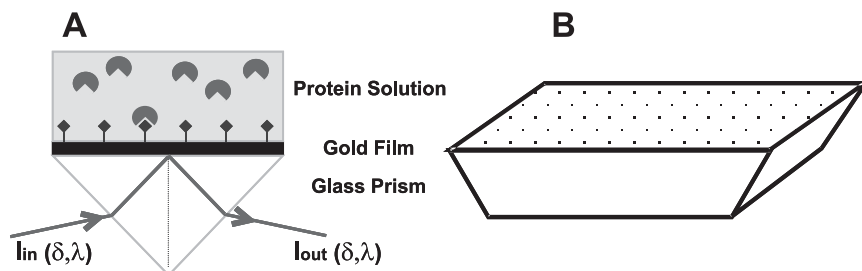


Fig. 9.6 Basic setup (A) for surface plasmon resonance-based detection of molecular recognition comprises a gold-covered glass prism and the observation of the reflected light intensity in dependence of the angle and wavelength. Binding of a soluble protein to an immobilized ligand influences the resonance

condition at the gold interface. In the case of high-density chemical microarrays (B), the actual sensor fields are arranged in rows and columns to allow the parallel detection of up to 10,000 individual binding events using imaging technologies.

surface, and complex formation is optically detected in real time by adding the solubilized partner, e.g., by means of a microfluidic flow chamber. The sensitivity of the technique allowed the detection of small molecules binding to immobilized protein receptors in a number of cases. However, sensitivity is at the limits and throughput is low if the protein is immobilized and the surface is exposed to individual ligands in a sequential fashion. These restrictions do not apply for the chemical microarray format, where a high number of small molecules are immobilized and binding to the target protein is detected in parallel (Fig. 9.6). Because SPR is sensitive to mass change, the binding of large proteins increases sensitivity as compared to the setup where the protein is immobilized.

The need for higher throughput has led to the development of instruments that are capable of working either sequentially or parallel on several sensor fields [80]. However, parallel detection comes into play only when the sensor technology is combined with 2-D arrays opening the technique for screening applications. We designed and built a number of Plasmon Imager devices to fulfill this need and to enable us to perform label-free, simultaneous binding detection for up to 10,000 immobilized small organic molecules against a macromolecular receptor in solution. The instruments are based on imaging technology where the reflected light from the entire array is captured by a CCD camera. The resonance condition for surface plasmons is determined by stepwise variation of the wavelength of the incident light. For each sensor field on the array, the light intensity (in percent of 100) of the reflected light is recorded against the wavelength (in nm), and for each spot a nm value for minimal light intensity is obtained. In order to generate an output that corresponds to protein-binding pattern recognition, two measurements are necessary. The difference between a measurement with buffer alone and after adding the target protein defines the SPR signal, i.e., the nanometer shift of the resonance condition.

9.4.2

Protein Affinity Fingerprints

Chemical microarray screening data are conveniently displayed in a 2-D format of squares arranged in rows and columns. This representation reflects the actual spatially encoded ligand positions on the grid layout of the microarray. Therefore, each square represents a single spot on the physical array and thus also a single array compound. The SPR shift that is obtained after exposing the chemical microarray to a protein can be shown as a grayscale or in color codes. We developed a proprietary visualization tool for point-and-click interrogation of each data point for both the chemical structure and the measured affinity associated with it. The visualized dataset is called an affinity fingerprint for the protein of interest (Fig. 9.7). In the case of binary libraries, constructed from two sets of monomers, visualization of the affinity fingerprints easily deconvolutes the presence of building blocks in hits. By arranging the compounds in rows and columns corresponding to the presence of a single monomer, the occurrence of many strong signals in one column indicates the significant contribution or even dominance of a monomer in the combinatorial hit compound. Besides such prominent fragments, indicated by “strong” rows or columns, hits also can be found for certain binary combinations where each monomer has little or even no “tolerance” for the second building block in order to show affinity. Typically, both types of hits are found. Overall, affinity fingerprints are the large-scale documentation of patterns of molecular recognition between small organic molecules and proteins.

Immobilization of a collection of commercially available monomers on chemical microarrays represents the equivalent of the single fragment screening methods described above. Fig. 9.8 shows the affinity fingerprint of a chemical microarray containing 1200 individual fragments. The immobilized compounds all qualify as needle compounds, with molecular weights from 80 to 350 Da. Fragments can be selected that display diversity in pharmacophoric motifs and/or are enriched in drug-related chemical motifs as outlined above. The relationship between molecular weight and obtained SPR signal shows that the smallest fragments with detectable signals have a molecular weight of around 150 Da. Binding constants of chemical microarray hit fragments are typically observed to be in the micromolar range. For example, a well-known needle-type molecule, namely, amino-benzamidine with a K_i of 80 μM against thrombin, is detectable within a fragment affinity fingerprint.

After identification of binders from the array screening, the ChemTag can be replaced by a series of small substructures, leading to soluble compounds based around the original array hit structure. This so-called tag-replacement strategy makes use of the tether site to expand the array hits. The functional group originally connected to the ChemTag provides both a readily available linking chemistry and a bias towards an attachment point, with a higher chance of pointing towards additional interaction sites or subpockets within the protein. With tag replacement, areas are explored that are not yet occupied by the remainder of the ligand, and attachment sites are likely avoided that might result in clashes with the protein. Case

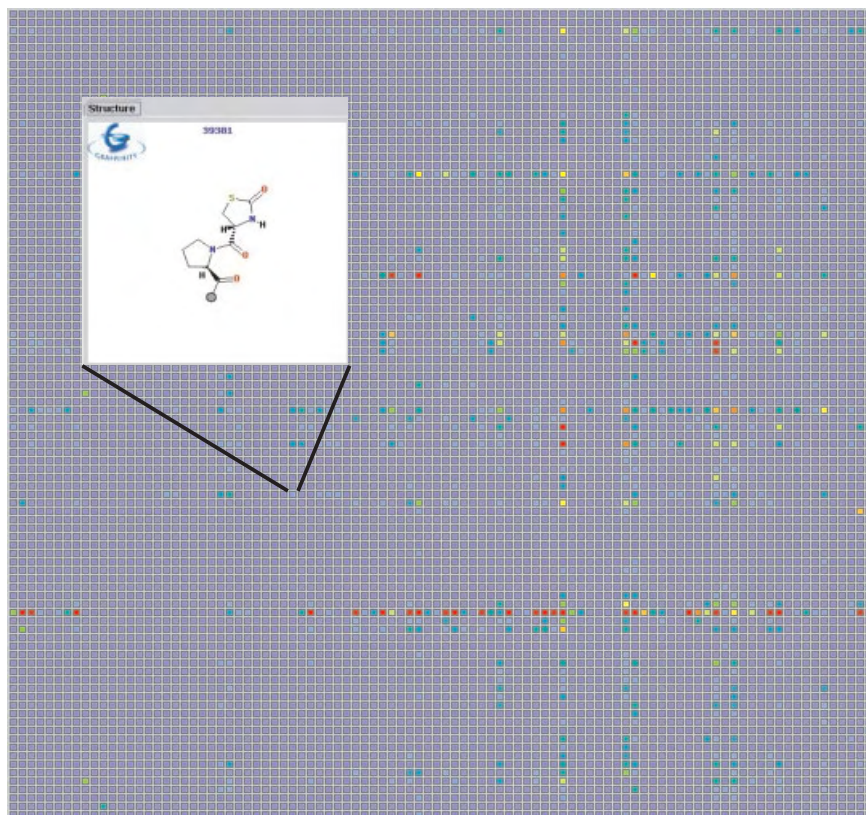


Fig. 9.7 Affinity fingerprint of a target protein probed against a chemical microarray presenting 9216 immobilized binary compounds. The color range goes from red to orange to yellow to green to blue to code for decreasing SPR signal. Rows and columns each represent one of two monomers of the binary library

and clearly a pattern of prominent rows and columns points to the presence of certain monomer fragments being highly populated among the hits. The pop-up structure displays an array compound example, with the Chem-Tag attachment site indicated by a gray-filled circle.

studies showed that the affinity of micromolar array hit compounds could be increased by one or two orders of magnitude. Such tag-replacement series link chemical microarray screening with follow-up medicinal chemistry efforts.

9.5 Conclusion

A number of technological advances in the area of surface chemistry and biophysical detection technologies have made fragment-based screening an attractive and promising tool for the early stages in drug discovery. The approach provides a

new entry point for lead discovery, as it aims at finding drug fragments in a first round of screening and using this information for an iterative buildup of chemical complexity. Such a procedure promises to address a number of issues encountered in the discovery and development of larger, more-complex screening compounds. Fragment-based screening allows us to investigate the relative contribution of the monomers in a reagent-based SAR evaluation, helping to avoid combinatorial explosion. Despite combinatorial chemistry, the number of accessible products by far exceeds the screening capabilities, and library diversity analysis tries to help reduce redundancy in compound collections to achieve a better coverage of “chemicals space.” Prescreening of monomers helps to better cover diversity by focusing combinatorial design around fragments that fared well in an early stage. Then, either smaller screening libraries for optimization can be synthesized based on these preferred monomers, or next rounds of testing can be based on fragment analogues.

The experimental techniques that allow the detection of small and weak binders either are based on high-concentration bioassays or fall into the category of affinity screening. Biological assays require careful setup and might not be available in general. NMR and crystallography have been used for some time now for fragment-screening purposes, and elegant techniques have been developed to discover small fragments binding to target proteins. These techniques open the possibility of detailed insight into the binding mode of ligands, and their attraction lies in the seamless integration of structure-based optimization. Nevertheless, the experimental procedures are technically demanding and are limited either by crystallization conditions or NMR-related protein requirements.

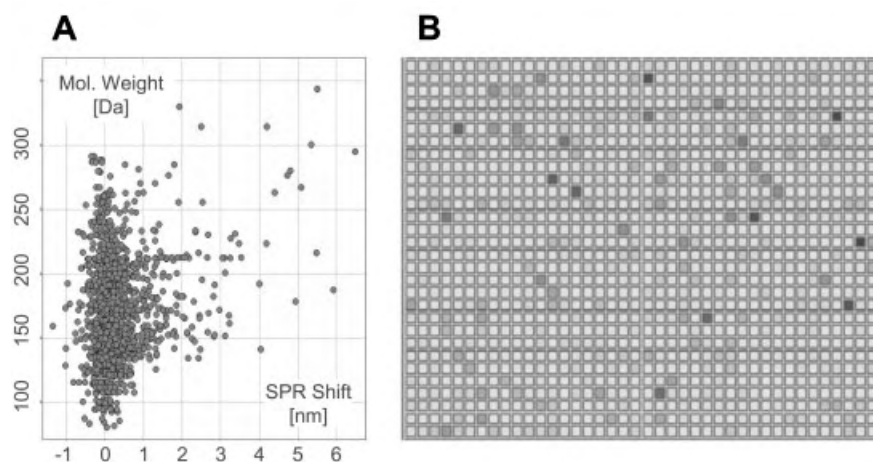


Fig. 9.8 SPR signal of a chemical microarray presenting 1200 immobilized fragments shown as a function of molecular weight (A). Even small structures around 150 Da show detectable signal. The corresponding affinity

fingerprint of the target protein probed against the chemical fragment microarray is shown as a grayscale, with stronger binders being darker (B).

A novel emerging technology, label-free screening on chemical microarrays, brings together SPR, a well-known technique for studying molecular interactions, with self-assembled monolayer surface chemistry to display small organic molecules in a miniaturized format. Progress in the area of label-free imaging now enables the parallel label-free affinity fingerprinting of 10,000 immobilized organic molecules against a target protein. The sensitivity of the method allows for the detection of weak interactions and makes routine, empirical fragment-based discovery a reality. The high-throughput, standardized screening format opens the possibility for chemical genomics applications aiming to move chemistry upstream in the discovery process. The goal is the early integration of chemical information either for target validation or for rapid identification of novel leads on a genomic scale.

9.6

Acknowledgement

The authors are grateful for discussions with the colleagues of Graffinity Pharmaceuticals and want to emphasize the team effort that made SPR based screening on chemical microarrays possible.

9.7

References

- 1 LIPINSKI, C.A., LOMBARDO, F., DOMINY, B.W., and FEENEY, P.J., *Adv. Drug Deliv. Rev.* **2001**, 46, 3–26.
- 2 TEAGUE, S.J., DAVIS, A.M., LEESON, P.D., and OPREA, T., *Angew. Chem. Int. Ed.* **2001**, 38, 3743–3748.
- 3 HANN, M.M., LEACH, A.R., and HARPER, G. J., *Chem. Inf. Comput. Sci.* **2001**, 41, 856–864.
- 4 REJTO, P. A. and VERKHIVKER, G. M., *Proc. Natl. Acad. Sci. U.S.A* **1996**, 93, 8945–50.
- 5 REJTO, P. A. and VERKHIVKER, G. M., *Pac. Symp. Biocomput.* **1998**, 362–373.
- 6 STOUT, T.J., SAGE, C. R., and STROUD, R. M., *Structure.* **1998**, 6, 839–848.
- 7 ANDREWS, P. R., CRAIK, D. J., and MARTIN, J. L., *J. Med. Chem.* **1984**, 27, 1648–1657.
- 8 KUNTZ, I. D., CHEN, K., SHARP, K. A., and KOLLMAN, P. A., *Proc. Natl. Acad. Sci. U.S.A* **1999**, 96, 9997–10002.
- 9 LEWELL, X. Q., JUDD, D. B., WATSON, S. P., and HANN, M. M., *J. Chem. Inf. Comput. Sci.* **1998**, 38, 511–522.
- 10 MAKINO, S., KAYAHARA, T., TASHIRO, K., TAKAHASHI, M., TSUJI, T., and SHOJI, M. J., *Comput. Aided Mol. Des* **2001**, 15, 553–559.
- 11 SCHNEIDER, G., LEE, M. L., STAHL, M., and SCHNEIDER, P. J., *Comput. Aided Mol. Des* **2000**, 14, 487–494.
- 12 LEWELL, X. Q. and SMITH, R. J., *Mol. Graph. Model.* **1997**, 15, 43–46.
- 13 BEMIS, G. W. and MURCKO, M. A., *J. Med. Chem.* **1996**, 39, 2887–2893.
- 14 BEMIS, G. W. and MURCKO, M. A., *J. Med. Chem.* **1999**, 42, 5095–5099.
- 15 FEJZO, J., LEPRE, C. A., PENG, J. W., BEMIS, G. W., AJAY, MURCKO, M. A., and MOORE, J. M., *Chem. Biol.* **1999**, 6, 755–769.
- 16 EVANS, B. E., RITTLE, K. E., BOCK, M. G., DIPARDO, R. M., FREIDINGER, R. M., WHITTER, W. L., LUNDELL, G. F., VEBER, D. F., ANDERSON, P. S., CHANG, R. S., *J. Med. Chem.* **1988**, 31, 2235–2246.
- 17 PATCHETT, A. and NARGUND, R., *Annual Reports in Medicinal Chemistry* **2000**, 35, 289–297.

- 18 NICOLAOU, K., PFEFFERKORN, J., ROECKER, A., CAO, G., BARLUENGA, S., and MITCHELL, H. J., *Am. Chem. Soc.* **2000**, 122, 9939–9953.
- 19 www.arraybiopharma.com. 2002.
- 20 HAJDUK, P. J., BURES, M., PRAESTGAARD, J., and FESIK, S. W., *J. Med. Chem.* **2000**, 43, 3443–3447.
- 21 LUMMA, W. C., WITHERUP, K. M., TUCKER, T. J., BRADY, S. F., SISCO, J. T., NAYLER-OLSEN, A. M., LEWIS, S. D., LUCAS, B. J., and VACCA, J. P., *J. Med. Chem.* **1998**, 41, 1011–1013.
- 22 APOSTOLAKIS, J. and CAFLISCH, A., *Comb. Chem. High Throughput. Screen.* **1999**, 2, 91–104.
- 23 MURCKO, M. A. In: CHARIFSON, P. S., EDITOR, *Practical Application of Computer-Aided Drug Design*, Dekker, **1997**.
- 24 BOEHM, H. J., BOEHRINGER, M., BUR, D., GMEINDER, H., HUBER, W., KLAUS, W., KOSTREWA, D., KUEHNE, H., LUEBBERS, T., MEUNIER-KELLER, N., and MUELLER, F., *J. Med. Chem.* **2000**, 43, 2664–2674.
- 25 MALY, D. J., CHOONG, I. C., and ELLMAN, J. A. *Proc. Natl. Acad. Sci. U.S.A* **2000**, 97, 2419–2424.
- 26 ERLANSON, D. A., BRAISTED, A. C., RAPHAEL, D. R., RANDAL, M., STROUD, R. M., GORDON, E. M., and WELLS, J. A., *Proc. Natl. Acad. Sci. U.S.A* **2000**, 97, 9367–9372.
- 27 PELLECCIA, M., SEM, D., and WUETHRICH, K., *Nat. Rev. Drug Discov.* **2002**, 1, 211–218.
- 28 VAN DONGEN, M., WEIGELT, J., UPPENBERG, J., SCHULTZ, J., and WIKSTROM, M., *Drug Discov. Today* **2002**, 7, 471–478.
- 29 DIERCKX, T., COLES, M., KESSLER, H., *Curr. Opin. Chem. Biol.* **2001**, 5, 285–291.
- 30 BRADLEY, D., *Modern Drug Discovery* **2001**, 29–34.
- 31 ROBERTS, G. C., *Drug Discov. Today* **2000**, 5, 230–240.
- 32 MOORE, J. M., *Curr. Opin. Biotechnol.* **1999**, 10, 54–58.
- 33 LEPRE, C. A., *Drug Discov. Today* **2001**, 6, 133–140.
- 34 SHUKER, S. B., HAJDUK, P. J., MEADOWS, R. P., and FESIK, S. W., *Science* **1996**, 274, 1531–1534.
- 35 HAJDUK, P. J., MEADOWS, R. P., and FESIK, S. W., *Science* **1997**, 278, 497–499.
- 36 HAJDUK, P. J., DINGES, J., MIKNIS, G. F., MERLOCK, M., MIDDLETON, T., KEMPF, D. J., EGAN, D. A., WALTER, K. A., ROBINS, T. S., SHUKER, S. B., HOLZMAN, T. F., and FESIK, S. W., *J. Med. Chem.* **1997**, 40, 3144–3150.
- 37 HAJDUK, P. J., ZHOU, M. M., and FESIK, S. W., *Bioorg. Med. Chem. Lett.* **1999**, 9, 2403–2406.
- 38 HAJDUK, P. J., DINGES, J., SCHKERYANTZ, J. M., JANOWICK, D., KAMINSKI, M., TUFANO, M., AUGERI, D. J., PETROS, A., NIENABER, V., ZHONG, P., HAMMOND, R., COEN, M., BEUTEL, B., KATZ, L., and FESIK, S. W., *J. Med. Chem.* **1999**, 42, 3852–3859.
- 39 HAJDUK, P. J., BOYD, S., NETTESHEIM, D., NIENABER, V., SEVERIN, J., SMITH, R., DAVIDSON, D., ROCKWAY, T., and FESIK, S. W., *J. Med. Chem.* **2000**, 43, 3862–3866.
- 40 HAJDUK, P. J., GOMTSYAN, A., DIDOMENICO, S., COWART, M., BAYBURT, E. K., SOLOMON, L., SEVERIN, J., SMITH, R., WALTER, K., HOLZMAN, T. F., STEWART, A., MCGARAUGHTY, S., JARVIS, M. F., KOWALUK, E. A., and FESIK, S. W., *J. Med. Chem.* **2000**, 43, 4781–4786.
- 41 ALLEN, K., BELLAMAINA, C., DING, X., JEFFERY, C., MATTOS, C., PETSKE, G., and RINGE, D., *J. Phys. Chem.* **1996**, 100, 2605–2611.
- 42 NIENABER, V. L., RICHARDSON, P. L., KLIGHOFER, V., BOUSKA, J. J., GIRANDA, V. L., and GREER, J., *Nat. Biotechnol.* **2000**, 18, 1105–1108.
- 43 ANDERSON, S. and CHIPLIN, J., *Drug Discov. Today* **2002**, 7, 105–107.
- 44 GOODWILL, K., TENNANT, M., and STEVENS, R., *Drug Discov. Today* **2001**, 6, S113–S118.
- 45 CARR, R. and JHOTI, H., *Drug Discov. Today* **2002**, 7, 522–527.
- 46 BLUNDELL, T., JHOTI, H., and ABELL, C., *Nat. Rev. Drug Discov.* **2002**, 1, 45–54.
- 47 STEWART, L., CLARK, R., and BEHNKE, C., *Drug Discov. Today* **2002**, 7, 187–196.
- 48 BUSSOW, K., CAHILL, D., NIETFIELD, W., BANCROFT, D., SCHERZINGER, E., LEHRACH, H., and WALTER, G., *Nucleic Acids Res.* **1998**, 26, 5007–5008.
- 49 ARENKOV, P., KUKHTIN, A., GEMMELL, A., VOLOSHCHUK, S., CHUPEEVA, V., and

- MIRZABEKOV, A., *Anal. Biochem.* **2000**, 278, 123–131.
- 50 ZHU, H., BILGIN, M., BANGHAM, R., HALL, D., CASAMAYOR, A., BERTONE, P., LAN, N., JANSEN, R., BIDLINGMAIER, S., HOUEK, T., MITCHELL, T., MILLER, P., DEAN, R. A., GERSTEIN, M., and SNYDER, M., *Science* **2001**, 293, 2101–2105.
- 51 MACBEATH, G. and SCHREIBER, S. L., *Science* **2000**, 289, 1760–1763.
- 52 WARRIOR, U., BURNS, D., and KOFRON, J., Arrayed Compound Screening (ARCS): A Well-less Screening Platform with Advantages of Miniaturization. 2001. Eighth Annual HTT EXPO Advancing Drug Development **2001**.
- 53 VENTER, J. C. et al., *Science* **2001**, 291, 1304–1351.
- 54 FODOR, S. P. A., LEIGHTON, R. J., PIRUNG, M. C., STRYER, L., LU, A. T., and SOLAS, D., *Science* **1991**, 251, 767–773.
- 55 JACOBS, J. W. and FODOR, S. P., *Trends Biotechnol.* **1994**, 12, 19–26.
- 56 FRANK, R., *Tetrahedron* **1992**, 48, 9217–9232.
- 57 SCHARN, D., WENSCHUH, H., REINEKE, U., SCHNEIDER-MERGENER, J., and GERMEROOTH, L., *J. Comb. Chem.* **2000**, 2, 361–369.
- 58 MACBEATH, G., KOEHLER, A., and SCHREIBER, S. L., *J. Am. Chem. Soc.* **1999**, 121, 7967–7968.
- 59 HERGENROTHER, P. J., DEPEW, K., and SCHREIBER, S. L., *J. Am. Chem. Soc.* **2000**, 122, 7849–7850.
- 60 KURUVILLA, F. G., SHAMJI, A. F., STERNSON, S. M., HERGENROTHER, P. J., and SCHREIBER, S. L., *Nature* **2002**, 416, 653–657.
- 61 SCHENA, M., SHALON, D., DAVIS, R. W., and BROWN, P. O., *Science* **1995**, 270, 467–470.
- 62 LAM, K. S., LEBL, M., and KRCHNAK, V., *Chem. Rev.* **1997**, 97, 411–448.
- 63 ULMAN, A., *Chem. Rev.* **1996**, 96, 1533–1554.
- 64 MRKSICH, M. and WHITESIDES, G. M., *Trends Biotechnol.* **1995**, 13, 228–235.
- 65 SIGAL, G. B., BAMDAD, C., BARBERIS, A., STROMINGER, J., and WHITESIDES, G. M., *Anal. Chem.* **1996**, 68, 490–497.
- 66 MRKSICH, M., *Journal of American Chemical Society* **1995**, 117, 12009–12010.
- 67 JUNG, L. et al., *Langmuir* **2000**, 16, 9421–9432.
- 68 TSANG, S. K., CHEH, J., ISAACS, L., JOSEPH-MCCARTHY, D., CHOI, S. K., PEVEAR, D. C., WHITESIDES, G. M., and HOGLE, J. M., *Chem. Biol.* **2001**, 8, 33–45.
- 69 HOUSEMAN, B. T., HUH, J. H., KRON, S. J., and MRKSICH, M., *Nat. Biotechnol.* **2002**, 20, 270–274.
- 70 HODNELAND, C. D., LEE, Y. S., MIN, D. H., and MRKSICH, M., *Proc. Natl. Acad. Sci. U.S.A* **2002**, 99, 5048–5052.
- 71 LAHIRI, J., ISAACS, L., GRZYBOWSKI, B., CARBECK, J., and WHITESIDES, G. M., *Langmuir* **1999**, 15, 7186–7198.
- 72 LAHIRI, J., ISAACS, L., TIEN, J., and WHITESIDES, G. M., *Anal. Chem.* **1999**, 71, 777–790.
- 73 McDONNELL, J. M., *Curr. Opin. Chem. Biol.* **2001**, 5, 572–577.
- 74 BAIRD, C. L. and MYSZKA, D. G., *J. Mol. Recognit.* **2001**, 14, 261–268.
- 75 ABERY, J., *Modern Drug Discovery* **2001**, 35–40.
- 76 RICH, R. L. and MYSZKA, D. G., *J. Mol. Recognit.* **2001**, 14, 223–228.
- 77 RICH, R. L. and MYSZKA, D. G., *J. Mol. Recognit.* **2001**, 14, 273–294.
- 78 MYSZKA, D. G., *J. Mol. Recognit.* **1999**, 12, 390–408.
- 79 MORTON, T. A. and MYSZKA, D. G., *Methods Enzymol.* **1998**, 295, 268–294.
- 80 COOPER M. A., *Nat. Rev. Drug Disc.* **2002**, 1, 515–528.

Subject Index

a

ab initio calculations 10, 144
 acceptor-proton-donor system 141
 π -acceptor systems 143
 accessible surface 203
 accuracy of scoring functions 12
 activation energy 54
 active analogue approach 78
 active conformation 110
 active site 165
 additive models 7, 24
 ADME properties 74, 98
 β -adrenoceptor 64
 affibodies 190
 affinity chromatography 60
 affinity fingerprint 231
 agonists 64, 66, 108
 alanine scan 128
 allosteric effects 36
 ALMOND 83
 alternative binding mode 7
 AMBER force field 14
 amide hydrogen bond 149
 Andrews analysis 9, 216
 angiotensin-converting enzyme (ACE) 78
 antagonists 64, 66, 108
 antibodies 120, 187
 anticalins 195, 199
 antigen 120, 188
 antigen column 12
 anti-receptor antibodies 124
 antisense strategies 36
 anti-thrombotic drugs 182
 aphrodisin 191
 apolar surfaces 150
 apolipoprotein D (ApoD) 193
 APPA 167
 aprotic solvents 22

aptamers 125
 aqueous solutions 32
 Argatroban 172
 aromatic ion pairs 43
 array approaches 223
 artificial ligand-binding proteins 190
 artificial receptor proteins 41, 207
 association constant 24, 113
 association rate 113
 atomic properties 84
 atom-type definitions 86
 azacrown ether 22

b

bacteriorhodopsin 108
 BCUT descriptors 94
 bifurcated hydrogen bond 142
 bilin-binding protein (BBP) 193
 bimolecular interaction 113
 binding affinity 4, 10
 binding assays 112
 binding constant 4
 binding kinetics 112
 binding mode 12, 87, 173, 182, 216
 binding of solvent molecules 180
 binding patterns 174
 binding pocket 13
 biophore 74
 biotin 22, 190
 bit-string 86
 Bjerrum theory 39
 Boltzmann distribution 11, 147
 bond lengths 146
 bulk water 6, 150
 buried polar groups 4
 buried water 150, 178

c

calcium channel blocker 89
 calorimetry 55
 Cambridge Crystallographic Database 84
 carbonic anhydrase 15
 catalytic mechanism of RNA
 cleavage 61
 catalytic triad 163
 cation- π interactions 43
 CATS descriptor 89
 CDR-3 188
 chalice-like shape 191
 charge transfer 138, 142
 charge-assisted hydrogen bonds 5
 charge-charge interaction 38
 chelate effect 22
 chemical microarrays 222, 227
 chemical shift 146
 chemokines 107
 chemotypes 74
 Chem-X 90
 chimeric receptors 126
 chip surface 224
 chiral resolution machine 35
 circular dichroism 59
 clinical trials 182
 clique-detection algorithm 79, 84
 closed systems 51
 combinatorial methods 124
 – biotechnology 194
 – chemistry 95, 97
 – design 198, 202
 – docking 168
 CoMFA method 82
 competition analysis 115
 complexation strength 26
 compound mixtures 213
 Concord 86
 conformational
 – analysis 78
 – coupling 36
 – freedom 32
 – induction 111
 – selection 110, 129
 – space 78, 87
 constitutive receptor activity 110
 cooperativity 8, 36
 Corina 86
 cost function 81
 Coulombic interactions 22, 39
 coupling experiment 146
 coverage 79
 cross-linking 130

crystal structures 204
 crystallographic analysis 202, 221
 cyclodextrin 22
 CYP3A4 98
 cytochrome P450 98

d

database searching 74, 85
 Daylight fingerprint 88
de novo design 14f., 95, 157, 218
 Debye-Hückel equation 38
 degrees of freedom 67
 delta effect 181
 design rules 166
 diafiltration 59
 dielectric conditions 13
 differential scanning calorimetry 55
 digitoxigenin 204
 digoxigenin 199, 203
 dipole-dipole interaction 141
 directed-tweak approach 87
 DISCO 78
 dispersive interactions 22, 28, 43, 45
 dissociation rate 113
 distance map 78
 ditopic recognition 28
 DNA microarray 225
 DOCK 14, 86
 docking 8, 13, 85
 donor-acceptor interactions 10, 38, 39,
 140
 drug candidates 74
 drug-design 63, 152
 drug-like characteristics 17, 182, 214
 DrugScore function 11

e

edge-to-face interactions 43
 efficacy 64
 electron sharing 146
 electronic configuration 142
 electrophoresis 59
 electrostatic
 – attractions 51
 – fields 83, 142
 – interactions 77, 138
 empirical scoring functions 7ff.
 endothelin antagonists 91
 endothermic interaction 56
 energy
 – calculation 87
 – function 11
 – transfer 51

enrichment 79
 – factor 14
 – rates 14
 enthalpy change 55
 enthalpy-driven binding 5, 6
 enthalpy-entropy compensations 24, 31
 entropic
 – change 53, 54
 – penalty 152
 entropy-driven binding 5, 31, 66
 enzyme inhibition 163
 equilibrium 53
 – constant 54
 – dialysis 59
 Euclidian distance 89
 experimental artifacts 31
 exposed salt bridge 5

f

Factor Xa inhibitors 95
 fibronectin 189
 field-based methods 83
 First Law of thermodynamics 52
 FKBP inhibitors 14
 flexibility 34
 FlexX 12
 FLOG 13
 fluorescein 201
 – conjugates 196
 fluorescence anisotropy 59
 fluorometric assays 116, 117
 force field-based methods 8f., 157
 fragment-based ligand discovery 76, 214
 fragment database 217
 free energy correlations 45
 free energy of binding 8, 53, 149, 155
 free energy perturbation 9
 functional groups 73
 – contributions 9
 fusion proteins 112, 126

g

GASP 79
 gas-phase enthalpy of binding 9
 gauche interactions 33
 Gaussian approximation 83
 gel (exclusion) chromatography 59
 genetic algorithm 79, 93
 genetically modified receptors 112
 geometric fitting 26, 29
 geometry of hydrogen bonds 139
 Gibbs free energy 57
 – of binding 4

GOLD 10
 Golpe 83
 goodness-of-fit 79
 GPCR ligands 90
 G-protein-coupled receptors 107ff., 127
 G-protein subfamilies 109
 GRID 82, 84
 gyrase inhibitor 15

h

Hammett equation 45
 hapten 188, 198
 HARPick 93
 hashed fingerprints 91
 heat capacity 55
 heat of reaction 54
 hemoglobin 36
 high-density lipoprotein (HDL) 193
 high-throughput screening 213, 222
 HIV reverse transcriptase inhibitors 12
 hole-size rule 30
 homology model 99
 Hoogsteen base pairing 36
 host-guest complexes 21
 hybridization 142
 hybridoma cells 122
 hydrogen-bond
 – acceptor / donor 76
 – proton 146
 – enthalpy 147
 – formation 147
 – capability 140
 – geometry 139
 – network 155, 179
 – potential 137, 142
 hydrogen bonds 16, 24, 39, 137, 180
 – in drug design 53
 hydrogen placement 156
 hydrophobic
 – collapse 172
 – core 191, 201
 – effects 51, 151
 – entropic contributions 31
 – forces 137
 – interactions 6, 10
 – interface 15, 76, 151, 188
 – ligands 193
 hypervariable loops 188

i

immobilized water 152
 immunization 188
 immunoblotting 124

in vitro selection 125, 188
 induced fit 36, 108
 informative design 96
 infrared spectroscopy 145
 inhibitor design process 164
 interaction sites 15, 84
 internal energy 52
 intramers 125
 intramolecular hydrogen bond 5
 intrinsic activity 64
 inverse agonists 109
 inverse Boltzmann technique 11
 ionophores 22, 26, 34
 irreversible antagonists 60
 isothermal titration calorimetry 4, 55

k

knowledge-based methods 8, 11

l

lead compounds 214
 lead identification 91
 Lewis-type complexes 39
 library 127
 – design 74, 92
 ligand complexity 215
 ligand pocket 204, 207
 linear free energy correlations 25
 linear response theory 8
 Lipinski's rules 154
 lipocalins 190
 lipophilic interactions, *see hydrophobic*
 local environment 149
 lock-and-key principle 4, 21
 London forces 53
 lone pairs 141, 157
 LUDI 15, 84, 96, 168

m

macrocytic effect 24
 macromolecular antigens 209
 Maximum Auto-Cross Correlation 83
 Melagatran 182
 membrane proteins 107ff.
 membrane-like environments 119
 Michaelis complex 167
 Michaelis constant 54
 microarrays 222
 microcalorimetry 55, 60
 MM2 force field 9
 molecular anchors 216
 molecular dynamics 156
 Molecular Dynamics simulations 8

molecular
 – evolution 169
 – fragments 15
 – recognition 21
 – shapes 221
 – size 13
 monoclonal antibodies 122
 monomer screening 219
 motional freedom 31
 multi-pharmacophore descriptors 94
 multiple 3-D conformations 88
 multiple linear regression 10
 multi-site interactions 22
 mutagenesis studies 62
 mutated residues 203

n

Napap 173
 Napsagatran 171
 needle 171
 – compounds 231
 – screening 96
 negative cooperativity 37
 negative entropy change 63
 negative ionizable (acid) 76
 neuraminidase inhibitors 84
 neuropeptides 107, 111
 neutron diffraction 139
 neutrophil gelatinase-associated lipocalin (hNGAL) 193
 NMR shifts 32
 NMR studies of hydrogen bonds 145
 non-bonded interactions 3, 5
 non-covalent interactions 22
 non-natural ligands 199
 non-specific binding 113
 nucleobase stacking 44

o

off-rates 166
 olfactory receptors 108
 opioid receptors 65
 optical biosensors 229
 optimizers 218
 optimization of selectivity and sensitivity 34
 orexin receptors 111
 organic complexes 22
 OWFEG (one window free energy grid)
 method 9

p

paratope 188
 partial agonists 110

partial charges 157
 partition equilibrium 59
 partition function 11
 partitioning method 92
 penalty functions 10, 78
 penicillins 166
 peptide backbone bond 140
 peptide ligands 127
 P-glycoprotein (P-gp) 99
 phage display 121f.
 phage M13 196
 phagemid-display technique 196
 pharmacophore 73, 126
 – features 76
 – fingerprints 88
 – hypothesis 129
 – model 87
 – searching 87
 – space 88
 pharmacophoric groups 73, 75
 photolithographic techniques 223
 pK_a shifts 13
 PLP scoring function 10
 PMF function 11
 polar functional groups 4, 16
 polarization effects 44
 polytopic interactions 22, 28
 porphyrin complexes 44
 positive cooperativity 37
 positive ionizable (base) 76
 posttranslational modifications 108, 124
 potential of mean force 11
 PPACK 167
 pregnane X receptor (PXR) 99
 privileged fragments 218
 progressive design 170
 property profile 94
 protein
 – antigens 189
 – engineering 209
 – flexibility 9, 14
 – folding 45
 – mutants 36
 – scaffolds 209
 – structures 141
 protonation state 13
 pseudo receptor 83

q

quantification of intermolecular
 forces 38
 quenching 197

r

radioligand binding 59
 – assay 115
 – techniques 64
 random library 198
 random mutagenesis 195
 rational protein design 187
 reaction coordinate 54
 reaction-energy diagram 54
 RECAP 217
 receptor-ligand-binding process 7
 receptor
 – mutagenesis 126
 – selectivity 111
 – site points 77
 recognition pocket 164
 recombinant antibody technology 188
 recombinant receptor proteins 190
 recursive partitioning 81
 refractive index 119
 reorganization of hydrogen bonds 150
 retinol-binding protein 190
 reversible antagonist 60
 reversible bimolecular reaction 57
 rhodopsin 108
 ribonuclease A 60
 ribonuclease inhibitor 60

s

salt bridges 24, 38
 SANDOCK 14
 SAR By NMR 220
 saturation analysis 114
 scaffold 189
 scaffold-hopping 88
 SCAMPI 81
 Scatchard plots 68, 113
 scoring functions 7f., 38, 158
 secondary interactions 25
 sedimentation equilibrium 59
 selective antagonists 126
 selectivity 167
 – in molecular recognition 34
 self-assembling monolayers 226
 semi-synthetic combinatorial
 library 196
 sensor chip 118
 separation assays 115
 serine proteases 96, 163
 – inhibitor 74
 shape indices 87
 SHAPES library 218
 shape-similarity searching 87

β -sheets 33
 signal transduction cascades 108
 similarity methods 91, 95
 simulated annealing 83, 93
 site-directed mutagenesis 126
 SMILES 86
 solvation effects 17
 solvent effects 30
 solvophobic interactions 22
 spacer groups 22
 specific binding 113
 spectroscopy 59
 spot synthesis 224
 SPROUT 95
 stability gap 216
 stacking 29, 35, 43
 standard free energy change 57
 state function 8
 steady-state conditions 67
 steady-state dialysis 59
 stereoselectivity 35
 steric
 – accessibility 141
 – constraint 85
 – fields 82
 – fit 4
 strain in host-guest complexes 30
 streptavidin 190
 structural plasticity 206
 structure-activity relationship (SAR) 181
 structure-based design 3, 16, 95, 97, 154
 substance P 129
 substituent constants 25
 SuperStar 85
 supramolecular complexes 21f.
 surface electrostatic properties 74
 surface plasmon resonance (SPR) 112, 118, 229
 surface polar groups 150
 surface water 150
 synthetic complexes 46
 synthetic peptides 120
 π -system 39, 43

t

target macromolecule 187
 tethered complexes 220
 thermal
 – conductivity 56
 – unfolding 208
 thermodynamic parameters 51, 54, 64, 148
 thermodynamics 51ff.
 thermolysin inhibitors 154
 thrombin 163
 – inhibitor 15
 total free energy of binding 22
 training data 10, 12
 transfection 112
 transition state analogues 164
 transmembrane helices 108, 126
 trypsin family 163

u

ultrafiltration 59
 unfavorable protein-ligand interactions 4, 17
 UNITY 76
 unspecific binding 227

v

valinomycin 34
 van der Waals charge-transfer 39
 van der Waals interactions 24, 43, 137
 van't Hoff equation 54, 58
 van't Hoff plot 58
 vancomycin 22
 vibrational frequency 145
 virtual compounds 93, 219
 virtual screening 8, 14, 98, 165
 vitamin A 191

w

water 31f., 149, 177f.
 – molecules 5, 141, 144
 – in hydrogen bonds 6
 – structure 13
 Watson-Crick base pairs 39
 weak binders 219
 weak hydrogen bonds 42
 weighting factor 11



WILEY-VCH

Protein-Ligand Interactions From Molecular Recognition to Drug Design

Edited by H.-J. Böhm
and G. Schneider

**Methods
and Principles
in Medicinal
Chemistry**

Volume 19

The lock-and-key principle formulated by Emil Fischer as early as the end of the 19th century has still not lost any of its significance for the life sciences. The basic aspects of ligand-protein interaction may be summarized under the term 'molecular recognition' and concern the specificity as well as stability of ligand binding.

Molecular recognition is thus a central topic in the development of active substances, since stability and specificity determine whether a substance can be used as a drug. Nowadays, computer-aided prediction and intelligent molecular design make a large contribution to the constant search for, e. g., improved enzyme inhibitors, and new concepts such as that of pharmacophores are being developed. An up-to-date presentation of an eternally young topic, this book is an indispensable information source for chemists, biochemists and pharmacologists dealing with the binding of ligands to proteins.

ISBN 3-527-30521-1



www.wiley-vch.de