Protein-Ligand Interactions
From Molecular Recognition to Drug Design

Edited by
H.-J. Böhm and G. Schneider

WILEY-VCH GmbH & Co. KGaA

Protein-Ligand Interactions
From Molecular Recognition to Drug Design

Edited by
H.-J. Böhm and G. Schneider

Methods and Principles in Medicinal Chemistry

Protein-Ligand Interactions
From Molecular Recognition to Drug Design

Edited by
H.-J. Böhm and G. Schneider

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.

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

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

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
Meyerhofstraße 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 Saatzucht 5
D-85350 Freising-Weihenstephan
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 *ß-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	*p*-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; °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 PRopErtY 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 cal mol^{-1} deg^{-1}; 8.31 J mol^{-1} 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)	*p*-(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. Böнм

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):

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

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). Unfortu-

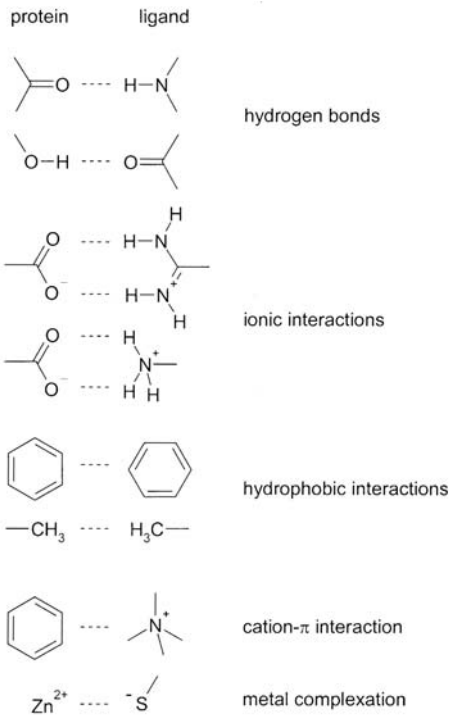

Fig. 1.1 Typical non-bonded interactions found in protein-ligand complexes. Usually, the lipophilic part of the ligand is in contact with the lipophilic parts of the protein (side chains of the amino acids Ile, Val, Leu, Phe, and Trp, perpendicular contact to amide bonds). In addition, several hydrogen 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-