

# Protein-Ligand Interactions From Molecular Recognition to Drug Design

*Edited by*

*H.-J. Böhm and G. Schneider*



WILEY-  
VCH

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

*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

*H.-J. Böhm*

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 *De Novo* Design 14

1.5.1 Successful Identification of Novel Leads Through Virtual Screening 14

1.5.2 *De novo* Ligand Design with LUDI 15

1.6 Outlook 16

1.7 Acknowledgments 17

1.8 References 17

|          |   |           |
|----------|---|-----------|
| <b>2</b> | <b>Introduction to Molecular Recognition Models</b>   | <b>21</b> |
|          | <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        |
| <b>3</b> | <b>Experimental Approaches to Determine the Thermodynamics of Protein-Ligand Interactions</b> | <b>51</b> |
|          | <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        |
| <b>4</b> | <b>The Biophore Concept</b>   | <b>73</b> |
|          | <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        |
| <b>5</b> | <b>Receptor-Ligand Interaction</b>                                       | <b>107</b> |
|          | <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        |

|          |  |            |
|----------|--|------------|
| <b>6</b> | <b>Hydrogen Bonds in Protein-Ligand Complexes</b>          | <b>137</b> |
|          | <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        |
| <br>     |  |            |
| <b>7</b> | <b>Principles of Enzyme-Inhibitor Design</b>               | <b>163</b> |
|          | <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        |

|          |   |            |
|----------|---|------------|
| <b>8</b> | <b>Tailoring Protein Scaffolds for Ligand Recognition</b>         | <b>187</b> |
|          | <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        |
| <br>     |   |            |
| <b>9</b> | <b>Small Molecule Screening of Chemical Microarrays</b>           | <b>213</b> |
|          | <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        |
| <br>     |   |            |
|          | <b>Subject Index</b>  | <b>237</b> |



## 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 & Molecu-  
lar 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-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 $\beta$ - $\gamma$ -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                                  |
| $\Delta G$ | Change in free energy                                |

|            |   |
|------------|---|
| $\Delta H$ | Change in enthalpy  |
| $\Delta S$ | Change in entropy   |
| $\Delta X$ | 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            | $\text{mol L}^{-1}$  |
| 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 ( $1.99 \text{ cal mol}^{-1} \text{ deg}^{-1}$ ; $8.31 \text{ J mol}^{-1} \text{ deg}^{-1}$ ) |
| 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  $\Delta G$  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-