

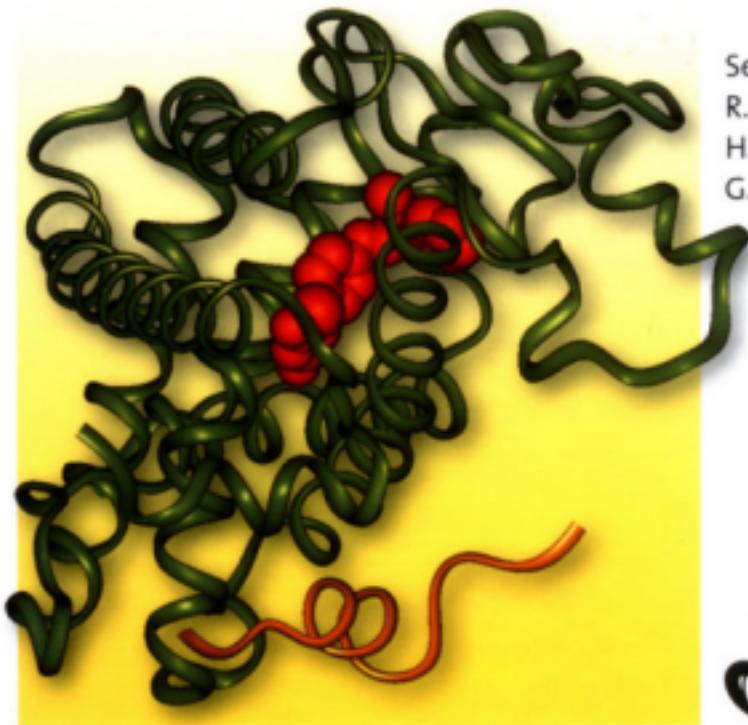
Edited by
R. E. Babine and S. S. Abdel-Meguid

WILEY-VCH

Protein Crystallography in Drug Discovery

Volume 20

Series Editors:
R. Mannhold,
H. Kubinyi,
G. Folkers



**Protein Crystallography
in Drug Discovery**
Edited by
R. E. Babine and S. S. Abdel-Meguid

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

Recently published Volumes:

Vol. 12 Drug Targeting

G. Molema, D. K. F. Meijer (eds.)

2001, ISBN 3-527-29989-0

Vol. 13 Pharmacokinetics and Metabolism in Drug Design

D. Smith, D. Walker, H. van de Waterbeemd

2001, ISBN 3-527-30197-6

Vol. 14 Bioinformatics – From Genomes to Drugs

T. Lengauer (ed.)

2001, ISBN 3-527-29988-2

Vol. 15 Drug-Membrane Interactions

J. K. Seydel, M. Wiese

2002, ISBN 3-527-30427-4

Vol. 16 BioNMR in Drug Research

O. Zerbe (ed.)

2002, ISBN 3-527-30465-7

Vol. 17 Quantum Medicinal Chemistry

P. Carloni, F. Alber (eds.)

2003, ISBN 3-527-30456-8

Vol. 18 Drug Bioavailability

H. van de Waterbeemd, H. Lennernäs, P. Artursson (eds.)

2003, ISBN 3-527-30438-X

Vol. 19 Protein-Ligand Interactions

H.-J. Böhm, G. Schneider (eds.)

2003, ISBN 3-527-30521-1

Protein Crystallography in Drug Discovery

Edited by

R. E. Babine and S. S. Abdel-Meguid



WILEY-
VCH

WILEY-VCH Verlag GmbH & Co. KGaA

Series Editors:

Prof. Dr. Raimund Mannhold

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

Prof. Dr. Hugo Kubinyi

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

Prof. Dr. Gerd Folkers

Department of Applied Biosciences
ETH Zürich
Winterthurerstr. 190
8057 Zürich
Switzerland
folkers@pharma.anbi.ethz.ch

Volume Editors:

Dr. Robert E. Babine

Suntory Pharmaceutical Research Laboratories
One Kendall Square
Cambridge, MA 02139
USA
robert.babine@sprlus.com

Dr. Sherin S. Abdel-Meguid

Suntory Pharmaceutical Research Laboratories
One Kendall Square
Cambridge, MA 02139
USA
sherin.s.abdel-meduid@sprlus.com

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.

The cover illustration shows the antidiabetic compound rosiglitazone bound to the nuclear receptor PPAR- γ (see chapter 1).

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>>

© 2004 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

Composition K+V Fotosatz GmbH, Beerfelden

Printing betz-druck gmbh, Darmstadt

Bookbinding Litges & Dopf Buchbinderei GmbH, Heppenheim

ISBN 3-527-30678-1

Contents

Preface XI

A Personal Foreword XIII

List of Contributors XV

| | | |
|---------|---|----|
| 1 | Molecular Recognition of Nuclear Hormone Receptor-Ligand Complexes | 1 |
| | <i>Robert E. Babine</i> | |
| 1.1 | Introduction | 1 |
| 1.1.1 | Nuclear Hormone Receptors: Ligand Binding Domains | 1 |
| 1.1.2 | Dimerization and Interactions with Co-activators and Co-repressors | 2 |
| 1.2 | Steroid Receptors | 5 |
| 1.2.1 | The Role of the Ligand | 5 |
| 1.2.1.1 | Estradiol Estrogen Receptor Complex | 5 |
| 1.2.1.2 | Other Estrogen Receptor Agonists Complexes | 8 |
| 1.2.1.3 | Estrogen Receptor Antagonists Complexes | 9 |
| 1.2.1.4 | Genistein – An ER- β Partial Agonist | 12 |
| 1.2.1.5 | R,R-5,11-cis-Diethyl-5,6,11,12-tetrahydrochrysene-2,8-diol: An ER- α Agonist and ER- β Antagonist | 14 |
| 1.2.2 | Structural Basis for Agonism and Antagonism and Partial Agonism: The Role of the Ligand | 15 |
| 1.2.3 | Progesterone/Progesterone Receptor Complex | 16 |
| 1.2.4 | Androgen Receptor Complexes | 17 |
| 1.2.5 | Glucocorticoid Receptor | 18 |
| 1.2.6 | Steroid Ligand Selectivity | 19 |
| 1.3 | The Vitamin D Receptor-Ligand Complexes | 22 |
| 1.4 | The Retinoic Acid Receptors RAR and RXR | 23 |
| 1.4.1 | Introduction | 23 |
| 1.4.2 | RAR- γ and RXR- α Retinoid Complexes | 25 |
| 1.4.3 | Selectivity of RAR Ligands and RAR Isotypes | 27 |
| 1.4.4 | RXR Complexes with Unnatural Ligands | 33 |
| 1.5 | PPAR: Isotype-Selective Ligands | 34 |
| 1.6 | Summary | 41 |
| 1.7 | Acknowledgements | 43 |

| | | |
|----------|---|----|
| 1.8 | References | 43 |
| 2 | Kinases | 47 |
| | <i>Jerry L. Adams, James Veal, and Lisa Shewchuk</i> | |
| 2.1 | Introduction | 47 |
| 2.2 | Structure and Function | 48 |
| 2.2.1 | Tertiary Structure | 48 |
| 2.2.2 | Catalysis and Substrate Binding | 49 |
| 2.2.3 | Regulation and Conformational Flexibility | 50 |
| 2.2.3.1 | Activation Loop Conformation | 50 |
| 2.2.3.2 | Glycine Rich Loop | 51 |
| 2.2.3.3 | C-Helix Orientation | 51 |
| 2.2.3.4 | Lobe Orientation | 53 |
| 2.2.3.5 | Solvent Channel | 53 |
| 2.3 | Crystallization | 54 |
| 2.3.1 | Defining the Construct | 54 |
| 2.3.2 | Mutagenesis | 55 |
| 2.3.3 | Phosphorylation | 55 |
| 2.4 | Inhibitor Design | 56 |
| 2.4.1 | Binding in ATP Cleft | 56 |
| 2.4.1.1 | ATP Binding Sites | 57 |
| 2.4.1.2 | Gatekeeper-Dependent Binding Pocket | 60 |
| 2.4.1.3 | Lipophilic Plug | 61 |
| 2.4.1.4 | Polar Surface Site | 61 |
| 2.4.2 | Conformational Considerations | 63 |
| 2.4.2.1 | Inhibitor-Induced Binding | 63 |
| 2.4.2.2 | What is the Most Appropriate Enzyme Form for Crystallography? | 64 |
| 2.4.2.3 | Homology Models and Surrogate Kinases | 66 |
| 2.4.3 | Paradigms for Kinase Drug Discovery | 68 |
| 2.4.3.1 | High Throughput Screening | 68 |
| 2.4.3.2 | Structure-Based Design | 69 |
| 2.4.3.3 | Mechanism-Based and Ligand Mimetic Design | 72 |
| 2.4.3.4 | Computational Chemistry and Virtual Screening | 72 |
| 2.4.4 | Selectivity | 73 |
| 2.5 | Conclusion | 75 |
| 2.6 | References | 76 |
| 3 | The Proteasome as a Drug Target | 83 |
| | <i>Tsunehiro Mizushima and Tomitake Tsukihara</i> | |
| 3.1 | Introduction | 83 |
| 3.2 | The Ubiquitin-Proteasome System | 83 |
| 3.2.1 | Role of the Ubiquitin-Proteasome System | 84 |
| 3.2.2 | 26S Proteasome | 85 |
| 3.2.3 | 20S Proteasome | 85 |
| 3.3 | Structure of the 20S Proteasome | 86 |

| | | |
|----------|--|-----------|
| 3.3.1 | Active Sites of Eukaryotic 20S Proteasomes | 88 |
| 3.3.2 | Novel Ntn-Hydrolase Active Site of the β 7 Subunit | 88 |
| 3.3.3 | Predicted Structure of Immunoproteasome and Substrate Specificities | 89 |
| 3.4 | Proteasome Inhibitors | 93 |
| 3.4.1 | Structure of Proteasome Inhibitor Complexes | 95 |
| 3.5 | Conclusions | 96 |
| 3.6 | Acknowledgements | 96 |
| 3.7 | References | 96 |
| 4 | Antibiotics and the Ribosome | 99 |
| | <i>Jeffrey L. Hansen</i> | |
| 4.1 | Introduction | 99 |
| 4.2 | The Ribosome | 99 |
| 4.2.1 | Introduction | 99 |
| 4.2.2 | Binding of tRNA | 100 |
| 4.2.3 | Peptidyl Transferase Activity | 101 |
| 4.2.4 | Structure of the Ribosome | 102 |
| 4.3 | Antibiotics | 103 |
| 4.3.1 | Introduction | 103 |
| 4.3.2 | Antibiotics that Bind to the 50S Subunit | 104 |
| 4.3.3 | MLS _B Antibiotics | 105 |
| 4.3.4 | Macrolides | 107 |
| 4.3.4.1 | Macrolides, 15- and 16-Membered | 107 |
| 4.3.4.2 | Binding Interactions Between the Lactone Ring and the Ribosome | 108 |
| 4.3.4.3 | Sugar Interactions with the Ribosome | 110 |
| 4.3.4.4 | A Covalent Bond | 111 |
| 4.3.4.5 | Macrolides, 14-Membered | 112 |
| 4.3.4.6 | Rational Drug Design of Macrolides | 112 |
| 4.3.5 | Lincosamides | 114 |
| 4.3.6 | Streptogramins | 114 |
| 4.3.7 | Chloramphenicol | 115 |
| 4.3.8 | Nucleoside Analogue Antibiotics | 117 |
| 4.3.8.1 | Puromycin | 117 |
| 4.3.8.2 | Aminoacyl-4-aminoxylosyl-cytosine Antibiotics | 118 |
| 4.3.9 | Other Antibiotics that Bind to the 50S Subunit | 119 |
| 4.3.9.1 | Sparsomycin | 119 |
| 4.3.9.2 | Anisomycin | 120 |
| 4.4 | Prospects for Rational Drug Design of Antibiotics that Bind to the Ribosome | 120 |
| 4.5 | Acknowledgements | 122 |
| 4.6 | References | 122 |

| | | |
|----------|--|------------|
| 5 | Structure-Based Design of Cathepsin K Inhibitors | 127 |
| | <i>Daniel F. Veber and Maxwell D. Cummings</i> | |
| 5.1 | Introduction | 127 |
| 5.2 | Background and Issues to be Addressed Using Protein Structure | 128 |
| 5.3 | Cysteine Protease Inhibitors: Historical Perspective | 131 |
| 5.4 | Diaminoketone-Based Inhibitors | 133 |
| 5.5 | Cyclic Diaminoketone Inhibitors | 136 |
| 5.6 | Alkoxyethyl and Thiomethyl Dipeptidyl Ketone-Based Inhibitors | 140 |
| 5.7 | Diacylcarbohydrazides | 142 |
| 5.8 | Conclusions | 143 |
| 5.9 | References | 145 |
| 6 | Structure-Based Design of Potent and Selective Cdk4 Inhibitors | 147 |
| | <i>Teruki Honma</i> | |
| 6.1 | Introduction | 147 |
| 6.2 | Homology Modeling of Cdk4 | 148 |
| 6.3 | Analysis of the ATP Binding Pocket | 151 |
| 6.4 | Strategies for Structure-Based Lead Identification: Virtual Screening of Known Compounds and de Novo Design | 152 |
| 6.4.1 | New de Novo Design Strategies | 153 |
| 6.4.2 | Evaluation of Chemical Availability by SEEDS | 155 |
| 6.5 | Structure-Based Generation of a New Class of Potent Cdk4 Inhibitors | 157 |
| 6.5.1 | Identification of New Scaffold Candidates Using LEGEND and SEEDS | 157 |
| 6.5.2 | From the Identified Scaffold to Lead Compounds | 159 |
| 6.5.3 | Confirmation of the Binding Mode | 159 |
| 6.6 | Structure-Based Design of Cdk4 Selective Inhibitors | 160 |
| 6.6.1 | Identification of Cdk4-Specific Amino Acid Residues | 161 |
| 6.6.2 | Library Design Based on the Locations of Cdk4-Specific Amino Acid Residues | 163 |
| 6.7 | Conclusion | 167 |
| 6.8 | Acknowledgements | 167 |
| 6.9 | References | 168 |
| 7 | Crystallization and Analysis of Serine Proteases with Ecotin | 171 |
| | <i>Sandra M. Waugh and Robert J. Fletterick</i> | |
| 7.1 | What is Ecotin? | 171 |
| 7.2 | Methods | 173 |
| 7.2.1 | Expression of Wild Type Ecotin | 173 |
| 7.2.2 | Purification of Wild Type Ecotin | 174 |
| 7.2.3 | Crystallization of Ecotin and Protease Complexes | 175 |
| 7.3 | Representative Examples of Ecotin and Protease Structures | 175 |
| 7.3.1 | Ecotin Defines the S7 Through S4' Subsites of Collagenase | 176 |

| | | |
|-----------|--|--|
| 7.3.2 | Ecotin as a Tight Binding Substrate | 177 |
| 7.3.3 | Ecotin Defines Regions Distal to the Factor Xa Protease Domain | 178 |
| 7.4 | Crystallization and Structure Determination of E ₂ P ₂ Complexes | 180 |
| 7.5 | Conclusion | 183 |
| 7.6 | References | 184 |
| 8 | X-ray Crystallography in the Development of Orthogonal Ligand-Receptor Pairs | <i>Donald F. Doyle and Lauren J. Schwimmer</i> |
| 8.1 | Introduction | 187 |
| 8.2 | Applications of OLRPs | 188 |
| 8.2.1 | Applications in Basic Research | 188 |
| 8.2.2 | Applications in Applied Research | 189 |
| 8.3 | Early Work | 189 |
| 8.4 | Structure-Guided Mutagenesis and OLRPs | 190 |
| 8.4.1 | Binding Pocket Mutations | 191 |
| 8.4.1.1 | Nuclear Receptors | 191 |
| 8.4.1.2 | Estrogen Receptor OLRPs | 193 |
| 8.4.1.3 | Retinoic Acid Receptor OLRPs | 195 |
| 8.4.1.4 | Retinoid X Receptor OLRPs | 195 |
| 8.4.2 | Distant Mutations | 197 |
| 8.5 | Other Examples of OLRPs | 200 |
| 8.5.1 | Chemical Inducers of Dimerization to Control Transcription | 200 |
| 8.5.2 | OLRPs and ATP Analogues | 202 |
| 8.6 | Summary | 204 |
| 8.7 | References | 206 |
| 9 | Engineering Proteins to Promote Crystallization | <i>Lei Jin and Robert E. Babine</i> |
| 9.1 | Introduction | 209 |
| 9.2 | Removing Protein Heterogeneity by Truncation | 210 |
| 9.3 | Removing Protein Heterogeneity by Point Mutation | 211 |
| 9.4 | Improving Crystal Packing by Point Mutation | 212 |
| 9.5 | Acknowledgements | 214 |
| 9.6 | References | 214 |
| 10 | High-throughput Crystallography | <i>Harren Jhoti</i> |
| 10.1 | Introduction | 217 |
| 10.2 | Technological Advances | 218 |
| 10.2.1 | Clone to Crystal | 218 |
| 10.2.1.1 | Protein Production | 218 |
| 10.2.1.2 | Crystallization | 219 |
| 10.2.2 | Crystal to Structure | 220 |
| 10.2.3 | Progress in Structural Genomics | 221 |

| | | |
|--------------------------|--|-----|
| 10.3 | High-throughput Crystallography in Lead Discovery | 223 |
| 10.3.1 | Protein-Ligand Crystal Structures | 224 |
| 10.4 | Fragment-Based Lead Discovery | 225 |
| 10.4.1 | Fragment-Based Lead Discovery Using X-ray Crystallography | 226 |
| 10.4.2 | Structure-Based Optimization of Fragment Hits | 229 |
| 10.5 | Conclusions | 230 |
| 10.6 | Acknowledgements | 231 |
| 10.7 | References | 231 |
| 11 | Micro-Crystallization | 235 |
| | <i>Carl L. Hansen, Morten Sommer, Kyle Self, James M. Berger, and Stephen R. Quake</i> | |
| 11.1 | Introduction | 235 |
| 11.2 | Microfluidics – Method and Design | 237 |
| 11.3 | Utility of Microfluidics for Crystallization | 242 |
| 11.4 | References | 253 |
| Subject Index 257 | | |

Preface

Among all new technologies in drug research, structure-based ligand design is one of the most powerful approaches. The drugs Captopril, Dorzolamide and Zanamivir, to mention only some prominent examples, resulted from rational design, based on the knowledge and analysis of protein 3D structures. The discovery of some other drugs, e.g. the more recent HIV protease inhibitors Nelfinavir and Amprenavir, was at least supported by protein crystallography studies. Many other drug candidates that resulted from structure-based design are in clinical development.

Whereas some early attempts of structure-based design failed due to inappropriate physicochemical and pharmacokinetic properties of the ligands, modellers are nowadays aware of the pitfalls in ligand design. Large, greasy ligands are avoided, as well as too polar compounds. According to favorable lead and drug properties, defined e.g. by the Lipinski rule of five, ligand design focuses on compounds with relatively low molecular weight, an intermediate lipophilicity range, and a limited number of hydrogen bond donors and acceptors.

In 1997, Klaus Gubernator and Hans-Joachim Böhm edited a volume on structure-based ligand design in this book series. A comprehensive review by Robert Babine, one of the editors of the current book, and Steven Bender, also published in 1997, discussed the design of aspartic, serine, cysteine, and metalloprotease inhibitors, and of immunosuppressants. The present book deals with some other families of important biological targets, e.g. nuclear receptors and kinases. In addition, several other attractive drug targets are reviewed. A special topic, the design of orthogonal protein-ligand pairs, will become important in personalized medicine. The book closes with chapters on recent progress in technologies: protein engineering to promote crystallization, micro-crystallization, and high-throughput crystallography. With its broad perspective, the book provides a state-of-the-art overview on important results and techniques that are relevant for protein 3D structure-based drug design.

The series editors are grateful to Robert Babine and Sherin Abdel-Meguid for their engaged work and to Frank Weinreich, Wiley-VCH, for ongoing support during the preparation of the book. We expect that volume 20 of "Methods and Principles in Medicinal Chemistry" will be another highlight of this successful series, which started only ten years ago.

September 2003

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

A Personal Foreword

When approached to initiate this book project the initial thoughts were to remember a review published in 1997 [1]. The first memories of that endeavor were quite painful, so the first instinct was to politely turn down the kind offer to initiate this project. On second reflection, researching and preparing the 1997 review was an educational and rewarding experience. Therefore, with the help of a co-editor, the invitation to initiate and complete this endeavor was accepted.

The objective of this book is to provide a forward looking overview of the use of protein crystallography in drug discovery. It has been organized so that the early chapters review and describe some mature and emerging topics that would fall under the ‘traditional structure-based design’ umbrella, the middle chapters provide focused accounts of specific works, and the final chapters delve into new and fertile areas of research. This book does not attempt to be comprehensive; thus the final lineup of chapters was a compromise between the interests of the editors and the willingness of the authors to contribute a chapter.

The first two chapters review nuclear hormone receptors and protein kinases. Both of these chapters provide an overview of the topic and shed some insight into how small molecule ligands can achieve selectivity between related protein targets. The next two chapters review topics that begin to ‘push the limit’ on the size of the complexes that can be used in drug design. Both the proteosome and the ribosome are very large biological macromolecules that are the targets for drug discovery. In the ribosome chapter, an important point is made regarding what conclusions are warranted, or not warranted, based upon the resolution of the x-ray diffraction data. The ribosome chapter also introduces the topic and challenges of antibiotic resistance in drug discovery. The next two chapters provide accounts of detailed structure-based design studies aimed at obtaining inhibitors of both cathepsin K and Cdk4. The cathepsin K chapter provides a nice account of the iterative structure-based design process. This work is especially notable for the use of an unexpected crystallographic result to move a project in a novel direction. The Cdk4 chapter also does an excellent job of introducing many computational methods that are used in drug discovery. For those readers interested in protein-based virtual screening of chemical databases, we also recommend the work by Bissantz *et al.* [2] in which they evaluate different docking/scoring combinations. Chapter 7 describes applications of the protease inhibitor ecotin, particularly its

use as a tool to obtain crystals of serine proteases and to study the interactions between serine proteases and their substrates. Chapter 8 reviews work on ‘orthogonal ligand-receptor pairs’ and the impact of crystallography on this area of research. While ‘traditional structure-based design’ uses the structure of a protein-ligand complex as a tool to design modified ligands, the work on ‘orthogonal ligand-receptor pairs’ uses the structure of the complex as a tool to design both modified proteins and modified ligands. This work has applications for deconvoluting some cellular processes and also provides useful tools in the area of chemical genetics. Protein/ligand pairs may also have future applications in gene therapy. Chapters 7 and 8 present examples that use crystallography to design mutant proteins for additional structural studies. Chapter 9 discusses the use of mutant proteins as an entry into obtaining high-resolution crystal structures. In numerous past examples, when a human protein proved difficult to crystallize, a protein from another species such as mouse, rat, or chicken was often used as a surrogate protein. A potential problem with this approach is that the active sites of the human and the surrogate enzyme might be different. The chapter on “Engineering Proteins to Promote Crystallization” reviews examples where the surface of a human protein is modified in a location remote from the active site to produce a mutant human protein that can give useful crystals. The next chapter discusses “High-throughput crystallography” which is an area of intense interest [3, 4]. This chapter provides a clear overview of the field, and has an informative section on crystallography in lead discovery. The final chapter describes miniaturization of crystallization utilizing the emergent technologies of microfluidics. This technology allows crystallization experiments to be performed on the nanoliter scale, thus conserving a very valuable resource in crystallography, the protein. Many of the technologies described in the final three chapters should impact crystallization and structure determination of the many new proteins identified in the human and other genomes.

We expect that this book will be a useful reference to practitioners of structure-based design. In addition, we hope that the technologies discussed in the later chapters will help researchers solve new problems in the next generation of structure-based design problems.

Cambridge, MA, April 2003

Robert E. Babine
Sherin S. Abdel-Meguid

- 1 BABINE, R. E. and S. L. BENDER, Molecular Recognition of Protein-Ligand Complexes: Applications to Drug Design. *Chem. Rev.* **1997**, *97*(5), 1359–1472.
- 2 BISSANTZ, C., G. FOLKERS, and D. RODNAN, Protein-based virtual screening of chemical databases. 1. Evaluation of different docking/scoring combinations. *J. Med. Chem.* **2000**, *43*(25), 4759–4767.
- 3 MOUNTAIN, V., Innovation – Astex, Structural Genomix, and Syrrx. *Chem. Biol.* **2003**, *10*, 95–98.
- 4 BERTINI, I., Structural genomics. *Acc. Chem. Res.* **2003**, *36*(3), 155.

List of Contributors

JERRY L. ADAMS

GlaxoSmithKline Pharmaceuticals
Medicinal Chemistry
1250 South Collegeville Road
Collegeville, PA 19426
USA

ROBERT E. BABINE

Suntory Pharmaceutical Research
Laboratories LLC
Structural and Computational
Chemistry
One Kendall Square, Bldg 700
Cambridge MA 02139
USA

JAMES M. BERGER

University of California at Berkeley
Department of Molecular
and Cell Biology
229 Stanley Hall
Berkeley, CA 94720-3206
USA

MAXWELL D. CUMMINGS

Computational Chemistry
3-Dimensional Pharmaceuticals
665 Stockton Drive, Suite 104
Exton, PA 19341
USA

DONALD F. DOYLE

Georgia Institute of Technology
School of Chemistry and Biochemistry
770 State Street
Atlanta, GA 30332-0400
USA

ROBERT J. FLETTERICK

Departments of Biochemistry
and Biophysics and Pharmaceutical
Chemistry
University of California
at San Francisco
600 16th Street, Box 2240
San Francisco, CA 94143-2240
USA

CARL L. HANSEN

California Institute of Technology
Department of Applied Physics
MS 128-95
Pasadena, CA 91125
USA

JEFFREY L. HANSEN

Department of Molecular Biophysics
and Biochemistry
Yale University
266 Whitney Avenue, 415 Bass Center
New Haven, CT 06520
USA

TERUKI HONMA

Banyu Tsukuba Research Institute
in collaboration with Merck Research
Laboratories
Okubo-3
Tsukuba 300-2611
Ibaraki
Japan

HARREN JHOTI

Astex Technology Ltd.
250 Cambridge Science Park
Milton Road
Cambridge, CB4 0WE
United Kingdom

LEI JIN

Suntory Pharmaceutical Research
Laboratories LLC
One Kendall Square, Bldg 700
Cambridge MA 02139
USA

TSUNEHIRO MIZUSHIMA

Japan Science and Technology
Corporation
Structure and Function
of Biomolecules
3-18-22 Honkomagome, Bunkyo-ku
Tokyo, 113-0021
Japan

STEPHEN R. QUAKE

California Institute of Technology
Department of Applied Physics
MS 128-95
Pasadena, CA 91125
USA

LAUREN J. SCHWIMMER

Georgia Institute of Technology
School of Chemistry and Biochemistry
770 State Street
Atlanta, GA 30332-0400
USA

KYLE SELF

Fluidigm Corporation
7100 Shoreline Court South
San Francisco, CA 94080
USA

LISA M. SHEWCHUK

GlaxoSmithKline Pharmaceuticals
Discovery Research
5 Moore Drive
Research Triangle Park, NC 27709
USA

MORTEN SOMMER

California Institute of Technology
Department of Applied Physics
MS 128-95
Pasadena, CA 91125
USA

TOMITAKE TSUKIHARA

Institute for Protein Research
Osaka University
3-2, Yamadaoka, Suita
Osaka 565-0871
Japan

JAMES VEAL

Serenex, Inc.
Informatics and Computational
Chemistry
323 Foster St.
Durham, NC 27701
USA

DANIEL F. VEBER

290 Batteson Rd.
Ambler, PA 19002
USA

SANDRA M. WAUGH

University of California
at San Francisco
Graduate Group in Biophysics
San Francisco, CA 94143-2240
USA

1

Molecular Recognition of Nuclear Hormone Receptor-Ligand Complexes

ROBERT E. BABINE

1.1

Introduction

1.1.1

Nuclear Hormone Receptors: Ligand Binding Domains

Nuclear hormone receptors (NHR) are multidomain proteins that function as transcription factors. They contain a central DNA binding domain (DBD) responsible for targeting the receptor to highly specific DNA sequences comprising a response element. The DBD is surrounded by two activation domains: the activation function 1 (AF-1) domain that resides at the N-terminus and the activation function 2 (AF-2) domain that resides at the C-terminal ligand-binding domain (LBD) [1, 2]. NHRs for which no natural ligand is known are referred to as orphan nuclear hormone receptors. Regulation of gene transcription by nuclear receptors requires the recruitment of proteins characterized as co-regulators, with ligand-dependent exchange of co-repressors for co-activators serving as the basic mechanism for switching gene repression to activation [2]. Upon activation by a small molecule ligand, or hormone, NHRs dimerize with another ligated NHR and recruit co-activators to turn on target genes. The binding of a ligand thus can act as a switch between gene activation and gene repression. This biological property has made the ligand binding domains of nuclear hormone receptors important drug targets.

Our understanding at the molecular level of how nuclear receptor ligands exert their effects has been dramatically enhanced by the elucidation of the crystal structures of the apo- and/or ligand-bound LBDs of several nuclear receptors (Tab. 1.1). These structures have revealed a common fold among LBDs, consisting of an antiparallel α -helical sandwich of 11–13 helices. Conventional nomenclature refers to these helices as helix-1, and helices-3 through 12. The region between helices 1 and 3 is variable and may contain zero, one or two helices. The helices fold to form a hydrophobic cavity into which the ligand can bind. The position of helix-12 relative to the other helices is dependent upon the presence or absence of a ligand and the nature of that ligand. The binding site that comprises the AF-2 domain is determined by the position of the C-terminal α -helix (helix-12).

Tab. 1.1 PDB codes of all nuclear hormone receptor-ligand-binding domains in the protein data bank as of December 2002.

| Protein | Apo structure | Ligated structures | Complex dimer structure |
|----------------|----------------------|--|--------------------------------|
| ER- α | | 1A52, 1ERE, 1ERR, 1QKT, 1QKU, 3ERD, 3ERT, 1L2I, 1GWQ, 1GWR | |
| ER- β | | 1QKM, 1QKN, 1HJ1, 1L2J | |
| AR | | 1I37, 1I38, 1E3G | |
| GR | | 1M2Z, 1NHZ | |
| TR- β | | 1BSX | |
| RAR- α | | | 1DKF |
| RAR- γ | | 1EXA, 1EXX, 1FCX, 1FCY, 1FCZ, 2LBD, 3LBD, 4LBD, 1FD0 | |
| RXR- α | 1G1U, 1LBD | 1FBY, 1G5Y, 1MV9, 1MVC, 1MZN | 1DKF, 1FM6, 1FM9, 1K74 |
| RXR- β | | 1H9U | |
| PR | | 1A28, 1E3K | |
| VDR | | 1DB1, 1IE8, 1IE9 | |
| PPAR- α | | 1I7G, 1K7L, 1KKQ | |
| PPAR- γ | 1PRG, 3PRG | 1I7I, 2PRG, 4PRG, 1KNU | 1FM6, 1FM9, 1K74 |
| PPAR- δ | 2GWX | 1GWX, 3GWX | |
| PXR | 1ILG | 1ILH | |
| ROR- β | | 1K4W | |
| ROR- α | | 1N83 | |
| HNF4- γ | | 1LV2 | |
| ERR3 | 1KV6 | | |

ER Estrogen Receptor, AR Androgen Receptor, GR Glucocorticoid Receptor, TR Thyroid Receptor, RAR Retinoic Acid Receptor, RXR Retinoid X Receptor, PR Progesterone Receptor, VDR Vitamin D Receptor, PPAR Peroxisome Proliferator-Activated Receptor, PXR Pregnan X Receptor, ROR Retinoic Acid-Related Orphan Receptor, HNF-4 Hepatocyte Nuclear Factor 4, ERR Estrogen-Related Receptor

1.1.2

Dimerization and Interactions with Co-activators and Co-repressors

The Glaxo group was the first to report the structure of a heterodimeric complex between two activated NHRs (peroxisome proliferator-activated receptor- γ) PPAR- γ and (retinoid X receptor- α) RXR- α . The structure (PDB entry: 1FM6) of this heterodimer complex contains six components: the two receptor LBDs, their two respective ligands, and two peptides derived from the steroid receptor co-activator-1 (SRC-1). These peptides contain a conserved LxxLL motif that is present in this class of co-activators. The complex is butterfly shaped, with both LBDs adopting the conserved “helical sandwich” fold previously reported for other ligand-bound nuclear receptors. PPAR- γ contains 13 α -helices and four short β -strands, while RXR- α is composed of 11 α -helices and two short β -strands [3]. This complex is illustrated in Fig. 1.1.

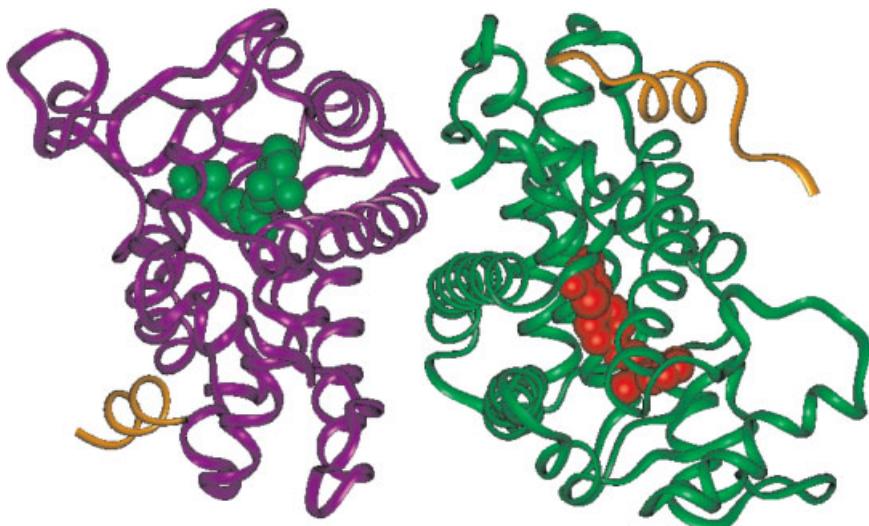


Fig. 1.1 Heterodimeric complex between PPAR- γ (green) bound to agonist rosiglitazone (red spheres), RXR- α (purple) bound to agonist *cis*-retinoic acid (green spheres). Pep-

tides (orange) derived from the SRC-1 co-activator protein are also shown bound to both NHRs. From PDB entry 1FM6.

For each NHR, the LxxLL motif of the SRC-1 peptides binds in a helical conformation to a groove on the protein defined by helices-3, -4 and -12 (Fig. 1.2). This groove is the binding site that comprises the AF-2 domain and it is determined by the position of helix-12. This “agonist” position of helix-12 is stabilized by the presence of an agonist ligand. Thus, a role of an activating ligand (agonist) is to stabilize a conformation of the LBD that allows binding of a co-activator protein.

The Glaxo group has also reported the crystal structure of a ternary complex containing the PPAR- α ligand-binding domain bound to the antagonist GW6471 and a co-repressor motif derived from the SMRT protein (PDB entry: 1KKQ) [4]. In this antagonist complex, helix-12 is found in a different position relative to the other helices than it is in the agonist structure. In this structure, the co-repressor motif adopts a three-turn α -helix and docks into a hydrophobic groove formed by helices-3, -4, and -12 (Fig. 1.3). Superposition of an agonist- and antagonist-bound PPAR- α reveals that the co-repressor-binding site partly overlaps with the co-activator-binding site. The additional helical turn in the co-repressor motif extends into the space that is left by the repositioning of helix-12, and prevents this helix from folding back into its active conformation. The binding of the co-repressor peptide is further reinforced by the antagonist, which blocks helix-12 from adopting an active conformation [4].

These structural studies have led to a proposal that nuclear receptors distinguish co-repressors from co-activators by the length of their helical interaction motifs [4]. The presence of an agonist ligand stabilizes a conformation of helix-12 that produces an AF-2 conformation that does not allow binding of co-repressors

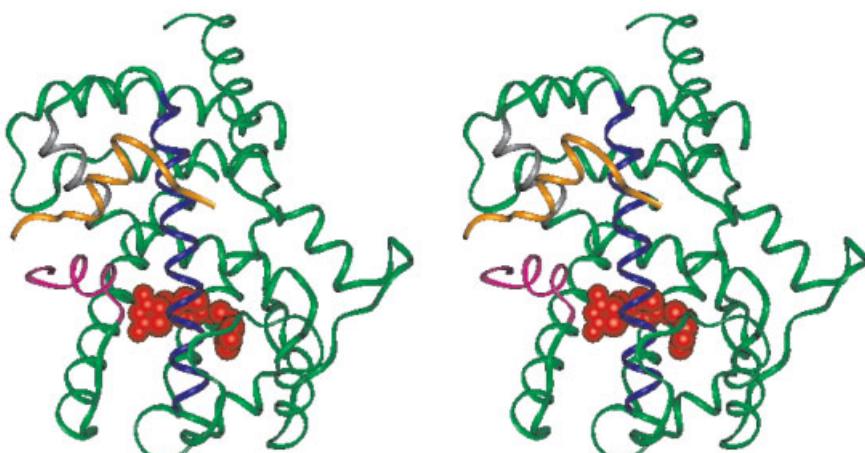


Fig. 1.2 A closer look at the PPAR- γ rosiglitazone SRC-1 peptide complex (wall eyed stereo). The helical nature of the NHR LBD is apparent. The agonist ligand (red spheres) induces the agonist conformation of helix-12

(magenta) of the LDB. This protein conformation has a groove defined by helices-3 (blue), -4 (gray) and -12. The LxxLL motif of the SRC-1 co-activator peptide (orange) binds to that groove. From PDB entry 1FM6.

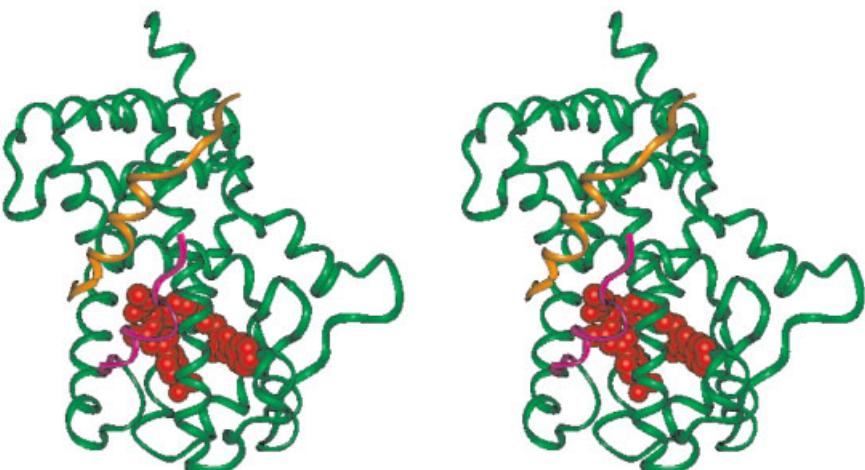


Fig. 1.3 Structure of the complex between PPAR- α (green), an antagonist ligand (red spheres) and a peptide (orange) derived from the SMRT co-repressor protein (wall eyed stereo). The orientation is close to that of

Fig. 1.2. The SMRT peptide binds as an extended helix and interacts with helix-12 (magenta) of PPAR- α , which is in a much different conformation than it is in an agonist structure. From PDB entry 1KKQ.

and allows the binding of co-activators. In the presence of an antagonist ligand, helix-12 is blocked from assuming the active AF-2 conformation, resulting in a larger pocket that can accommodate the three-turn α -helix of the co-repressor motif. The high degree of conservation of the co-repressor and co-activator interaction interfaces suggests that this is a model that applies to many of the nuclear receptors [4].

While not comprehensive, the remainder of this chapter will discuss in detail how ligands are capable of binding to NHRs and influencing the conformation of helix-12.

1.2

Steroid Receptors

While the chemical structures and biological properties of steroids have fascinated chemists and biologists alike for many decades, it is only recently that their receptors have been studied [1,5]. This section will give an overview of the structural biology of steroid receptor ligand complexes focusing on the role of the ligand.

1.2.1

The Role of the Ligand

1.2.1.1 Estradiol Estrogen Receptor Complex

Estradiol is the natural agonist ligand of the estrogen receptors (ER). There are two isoforms of the estrogen receptor, ER- α and ER- β . There are four structures of the complex between the LBD of ER- α and estradiol in the PDB. Expression, purification and crystallization of LBDs can often be problematic. Thus, the first structure solved (PDB entry: 1ERE) [6] used protein in which the free cysteines were carboxymethylated. The refined structure showed at least one of the cysteines was modified. Another structure (PDB entry: 1A52) [7] used protein that was refolded. The structure of this protein showed an artifact where two LBDs were connected by an intermolecular disulfide bond. Through much experimentation, conditions were found to express, purify and crystallize ER- α LBD without modification or refolding. The protein structure was deposited as PDB entry 1QKU [8] and it is this protein structure that will be discussed.

The mostly hydrophobic ligand estradiol is completely enclosed in the hydrophobic core of the ER- α LBD. This protein conformation with helix-12 folded over, and completely enclosing, the ligand-binding site is the agonist conformation of ER- α . While the ligand is completely enclosed within the ER- α protein, the fit is less than optimal; that is, there are several gaps between the surface of the ligand and the surface of the protein. It has been reported that the volume of the ligand-binding cavity (450 \AA^3) is nearly twice that of estradiol's molecular volume (245 \AA^3). The length and breadth of the ligand is well matched by the receptor, but there are large unoccupied cavities opposite the α face of the B-ring and the β face of the C-ring [6]. The estradiol ligand makes direct van der Waals (VDW) con-

tacts with residues from helix-3, helix-5, helix-7 and helix-11. While the presence of the agonist ligand stabilizes the conformation of helix-12, the ligand makes no direct VDW contacts with this helix.

The A-ring phenol makes hydrogen-bonding interactions with the side chains of Glu-353, located in the middle of helix-3, and Arg-394, located at the C-terminal end of helix-5. These two protein residues also interact with a common water molecule via hydrogen bonds. This arrangement is electrostatically complementary to the phenol and provides the basis for the known preference of an A-ring phenol group in ER ligands. The D-ring hydroxyl group makes a hydrogen bond with the side chain of His-524 that is located on helix-11. Thus, the two polar groups of the estradiol ligand make complementary polar interactions with the ER protein (Fig. 1.4). In the complex there are no unsatisfied polar interactions in the interior ligand-binding core of the protein. Thus, two types of contacts orient the ligand:

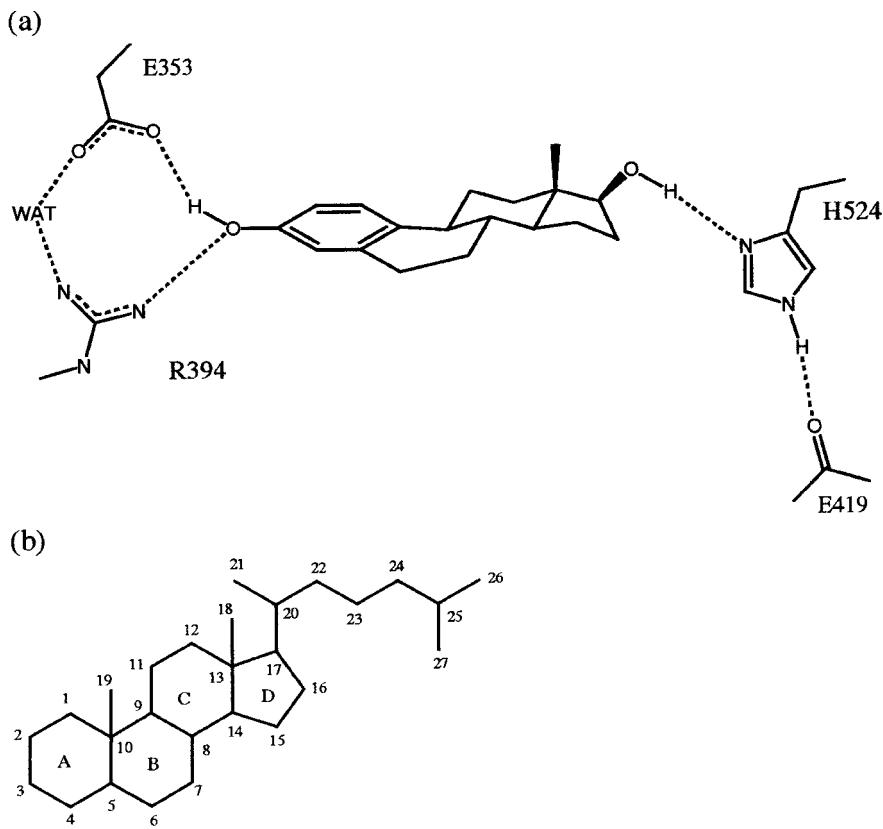


Fig. 1.4 (a) Schematic of estradiol binding to ER- α . There are polar interactions at both ends of the ligand; the A-ring phenol group interacts with the side chain of Glu-353 of he-

lix-3 and the D-ring hydroxyl group interacts with His-524. (b) Numbering scheme for steroid ring systems.

hydrogen bonds at the two ends and hydrophobic VDW contacts along the body of the ligand [7]. The combination of non-polar interactions and the specific polar interactions account for the ability of ER to bind estradiol selectively with high affinity and to exclude other steroids.

Estradiol is a pure agonist of estrogen receptor- α and the structure of this complex (PDB entry: 1QKU) [8] shows how an agonist ligand precisely and productively folds ER- α and stabilizes the unique agonist position of helix-12 (Figs. 1.4 and 1.5). When estradiol binds to ER- α , it interacts through the 17β -hydroxyl group with His-524, which in turn forms a hydrogen bond with the backbone carbonyl group of Glu-419 in the loop connecting helix-6 to helix-7. This glutamic acid contacts Glu-339 of helix-3 and Lys-531 of helix-11, forming a hydrogen bond network that favors the helix-12 agonist position. The loop between helix-1 and helix-3 accompanies the movement of helix-3. The precise positioning of helix-3 is an important feature for the constitution of the ligand-binding cavity. These interactions are essential for stabilizing the agonist conformation of helix-12. A triple mutant (C381S, C417S, C530S) of ER- α (PDB entry: 1QKT) displays limited transcriptional activity upon estradiol stimulation. The structure of this triple mutant, in complex with estradiol, shows helix-12 in an antagonist conformation. This structure helped to elucidate some of the structural features required for agonist activity [8].



Fig. 1.5 Structure of the complex between ER- α and estradiol. The agonist ligand is held by polar interactions on opposite ends of the ligand and it stabilizes the agonist conformation of helix-12 (black). From PDB entry 1QKU.

1.2.1.2 Other Estrogen Receptor Agonists Complexes

There are numerous examples of nonsteroidal estrogen receptor agonists. Diethyl stilbene (DES; Fig. 1.6a) is one example of a nonsteroidal agonist ligand for ER- α ; its structure has been determined (PDB entry: 3ERD) [9]. The structure of the ligand-binding core of the protein in this complex is nearly superimposable on the structure of the protein in the estradiol/ER- α complex. The A-ring of DES superimposes exactly with the A-ring of estradiol and results in interactions with Glu-353, Arg-394 and a structural water molecule. The D-ring phenol makes a hydrogen bond with His-524 but does so with a different geometry than does the hydroxyl group of estradiol (Fig. 1.6). Recall that in the estradiol/ER- α complex there are large unoccupied cavities opposite the α -face of the B-ring and the β -face of the C-ring. In the DES complex, the olefin group is out-of-plane with both aromatic rings. This results in the ethyl groups occupying the unoccupied cavities that are present in the estradiol complex and creating new unoccupied cavities in the positions filled by the B-, C- and D-rings of estradiol. Thus, like the estradiol complex, DES binds to the receptor by hydrogen bonding interactions at the two ends of the ligand and hydrophobic interactions in the middle. In addition, like the estradiol complex, there are several gaps between the surface of the ligand and the surface of the protein.

The structure of another nonsteroidal ligand, raloxifene-core (RalC; Fig. 1.6b), has been determined in complex with ER- α (PDB entry: 1GWQ) [43]. In this complex, the structure of the ligand-binding core of the protein in this complex is nearly superimposable on the structure of the protein in the estradiol/ER- α and DES/ER- α complexes and the A-ring phenol binds in the same manner as that of estradiol and DES. The D-ring phenol hydrogen bonds with His-524 but does so differently from the way estradiol or DES does. The central hydrophobic core of RalC makes hydrophobic interactions and leaves several gaps between the surface of the ligand and the surface of the protein. The position occupied by the B-ring of estradiol is unoccupied in this complex.

Upon examination of the structures of three complexes between ER- α and agonist ligands, some common features become apparent. The structure of the ligand-binding core of each protein is nearly identical. In each of the three examples discussed, the details of the protein-ligand interactions are different yet they bind to, and presumably stabilize, the same form of the receptor. Each complex is characterized by a less than optimal fit between the hydrophobic cores of the ligand and the protein. In each case the ligand is completely enclosed by the protein. There are no buried hydrophilic groups on either the protein or the ligand that are not involved in a hydrogen bond interaction with their ligand or protein partner, respectively. The role of the ligand is to precisely and productively fold ER- α and stabilize the unique agonist conformation of the protein.

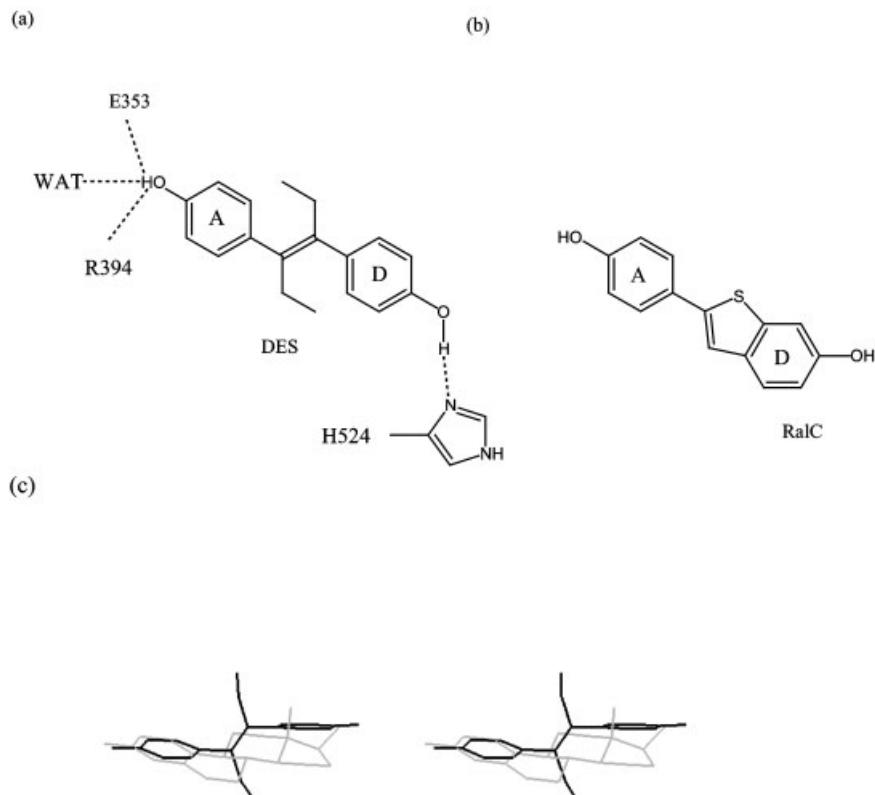


Fig. 1.6 (a) Illustration of the polar interactions of DES with ER- α . (b) Structure of raloxifene core (RalC). (c) Wall eyed stereo view of

1.2.1.3 Estrogen Receptor Antagonists Complexes

Raloxifene (RAL) is an antagonist ligand for the estrogen receptors. The structures of the complex between RAL and both ER- α (PDB entry: 1ERR) [6] and ER- β (PDB entry: 1QKM) [11] have been reported. In contrast to the ER- α agonist structures discussed above, RAL is not completely enclosed by the protein. In these complexes, solved in the absence of a co-repressor or co-repressor peptide, helix-12 does not fold over the ligand but resides on the protein surface that is usually occupied by co-activator peptides in the agonist co-activator complexes (Fig. 1.7). For the RAL/ER- α complex, the A-ring phenol of RAL interacts in the normal fashion with Glu-353, Arg-394 and a structural water molecule. The D-ring phenol interacts with the side chain of His-524 but does it differently than estradiol, DES or RalC (Fig. 1.8). In the agonist structures the side chain of His-524 forms hydrogen bonds to both a D-ring substituent and the backbone carbonyl of Glu-419. In the RAL complex His-524 interacts only with the ligand (a similar situation exists in the ER- β /RAL complex). While the B-ring of RAL overlaps with B-ring of estradiol the D-rings do not overlap



Fig. 1.7 Structure of the complex between ER- α (gray) and the selective antagonist raloxifene (solid spheres). In this structure, solved in the absence of a co-repressor peptide, helix-12 (bold) occupies the groove normally occupied by the LxxLL motif of co-activator peptides. From PDB entry 1ERR.

and the phenol hydroxyl is displaced 5.1 Å from the position of the 17 β -OH of estradiol. The side chain of RAL extends out of the ligand binding core making extensive hydrophobic contacts with helix-3, helix-5, helix-11 and the loop between helix-11 and helix-12. In this structure, the piperazine (F) ring occupies some of the same space that helix-12 occupies in the agonist structures. In addition, a salt bridge between Asp-351 and the piperazine ring nitrogen is observed. Also of note is that for the agonists structure between ER- α and RalC and the antagonist structure between ER- α and RAL the A- and D-rings are reversed.

For the protein ER- α , the structure of helix-3, helix-5 and the β -structure between helix-5 and helix-6, in contact with the ligand, is nearly identical in both the agonist and antagonist complex structure. However, the structure of helix-6 and the C-terminus of helix-11 that contact the ligand are different in the agonists and antagonist structures. As mentioned previously, helix-12 does not fold over the ligand and is in a completely different position in the antagonist structure.

The structure of ER- α in complex with the selective antagonist 4-hydroxytamoxifen (OHT; Fig. 1.9) has been reported (PDB entry: 3ERT) [9]. This structure is quite similar to the raloxifene/ER- α complex structure with the E-ring protruding out of the ligand-binding pocket. The A-ring phenol interacts with Arg-394 and Glu-353 in a manner similar to other ER- α ligand complexes (Fig. 1.9). The protein structure resembles that of the RAL complex in that helix-12 does not fold over the ligand but resides on the protein surface; helix-12 and the dimethylamino group makes a salt bridge with Asp-351. The conformation of the Ar-C=C-Ar group of OHT is different from the Ar-C=C-Ar conformation of DES.

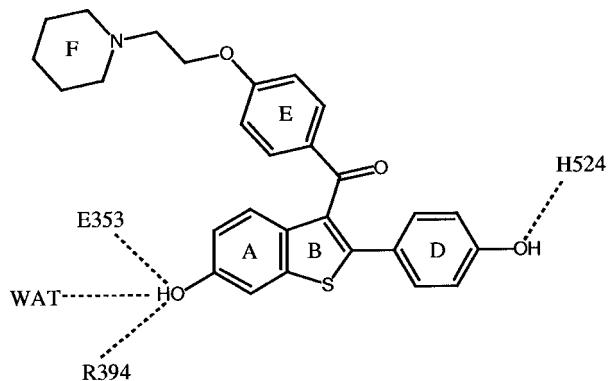


Fig. 1.8 Key polar interactions of the antagonist raloxifene with ER- α .

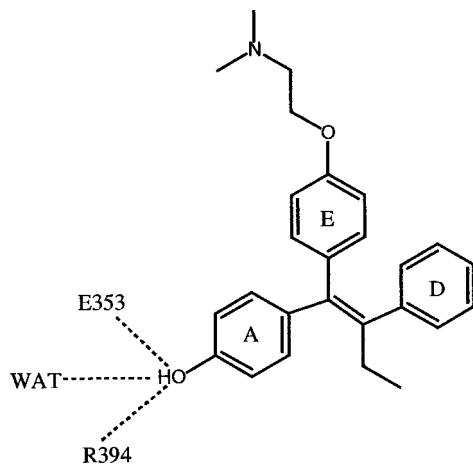


Fig. 1.9 Key polar interactions of the antagonist 4-hydroxytamoxifene with ER- α .

The structure of the pure antagonist ICI-164,384 (ICI; Fig. 1.10) in complex with rat ER- β has been reported (PDB entry: 1HJ1) [10]. ICI is a substituted estradiol with a 7- α -alkylamide group and it is this group that protrudes out of the ligand-binding pocket. To place this substituent out of the pocket requires that the ligand rotate approximately 180° (about the long hydroxyl to hydroxyl axis) in the binding site relative to estradiol. The A-ring phenol still interacts with the usual arginine and glutamine side chains; however, the D-ring alcohol interacts with His-430 and this histidine interacts with the backbone carbonyl of Glu-326 (Fig. 1.10). The primary consequence of the flipped binding mode of the ICI, compared with that of estradiol (normal mode), is an alteration in the nonpolar contacts made by

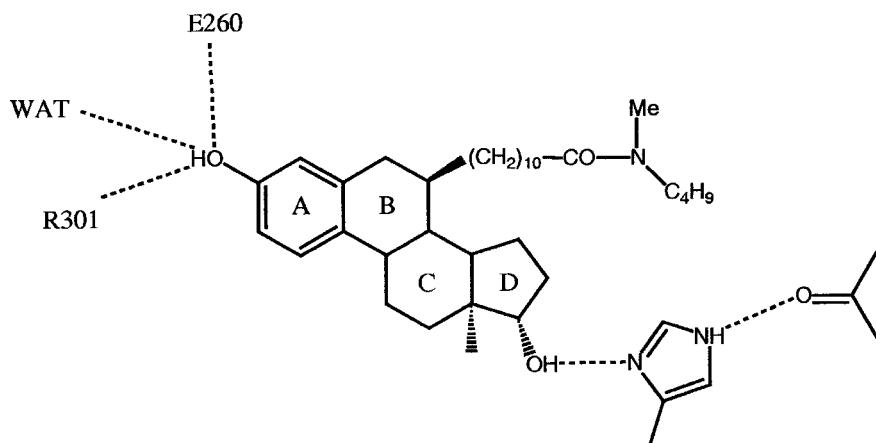


Fig. 1.10 Key polar interactions of the antagonist ICI-164,384 with ER- β .

the ligand's steroidal core. The position adopted by the central framework of ICI readjusts to occupy the cavity maximally and positions the ligand's two hydroxyl groups so that they can optimally interact with the protein. Consequently, the A-ring end of the steroid is shifted laterally away from helix-3 and toward helix-6 by about 1 Å. In contrast, the D-ring end of the molecule adopts a very similar spatial position in the flipped and "normal" orientations. The ER- β cavity accommodates these changes with only minimal alterations in the positioning of the cavity-lining residues [10].

The *N*-(*n*-butyl)-*N*-methylundecanamide side chain of ICI exits the binding pocket in an identical manner to that observed previously for RAL. However, once free from the confines of the ligand-binding pocket, the flexible side chain of the ICI antagonist adopts a conformation that is distinct from that observed with the corresponding regions of RAL and OHT. The ICI side chain, unlike the corresponding regions of RAL and OHT, is not tethered to the LBD through a salt bridge with Asp-258. Helix-12 is invisible in the experimental electron density maps, suggesting that it is highly mobile. The lack of a stable orientation of helix-12 can be directly attributed to the binding mode of ICI. The protruding 7*a*-substituent of ICI sterically prevents alignment of helix-12 over the cavity, as with other antagonists, but, in addition, the positioning of the terminal amide portion of the side chain precludes helix-12 from adopting an alternate orientation along the co-activator-binding cleft [10].

1.2.1.4 Genistein – An ER- β Partial Agonist

Genistein (GEN; Fig. 1.11a) is an isoflavonoid natural product that is an ER- β selective partial agonist. The structure of the complex between human ER- β and genistein has been reported (PDB entry: 1GKM) [11]. GEN binds across the ligand-binding cavity in a manner similar to estradiol with ER- α . The A-ring phenol of GEN inter-

acts with Arg-346, Glu-305 and a structural water molecule in a manner the same as most ligands. The D-ring phenol forms a hydrogen bond with His-475, which in turn makes a hydrogen bond with the backbone carbonyl of Glu-371. The other D-ring phenol forms an intramolecular hydrogen bond with the keto group of the C-ring; these polar groups occupy the α -position off the C- and D-rings of estradiol (Fig. 1.11). Unlike estradiol and other agonist ligands, helix-12 does fold over and enclose the partial agonist genistein. In this structure, helix-12 occupies a position on the surface of the protein between helix-3 and helix-5.

The primary determinant of helix-12 positioning appears to be the burial of its hydrophobic face against the protein. The positions of helix-12 in the presence of pure ER- α agonists, such as estradiol and diethylstilbestrol [6, 9], and that seen in the ER- β /GEN complex both fulfill this objective. In the archetypal “agonist” orientation, typified by the ER- α /estradiol complex, the N-terminal end of helix-12 appears to be stabilized by two hydrophilic interactions. The helix is capped by the side chain of a residue on helix-3, Asp-351, which interacts with the main chain amide of Leu-491. In addition, Tyr-537, located between helix-11 and -12, makes a hydrogen bond with helix-3 residue Asn-348. These two interactions effectively anchor the N-terminal end of helix-12 and position it so that Leu-540 can seal the ligand-binding cavity. This “agonist” conformation of helix-12 completely encloses the ligand within the

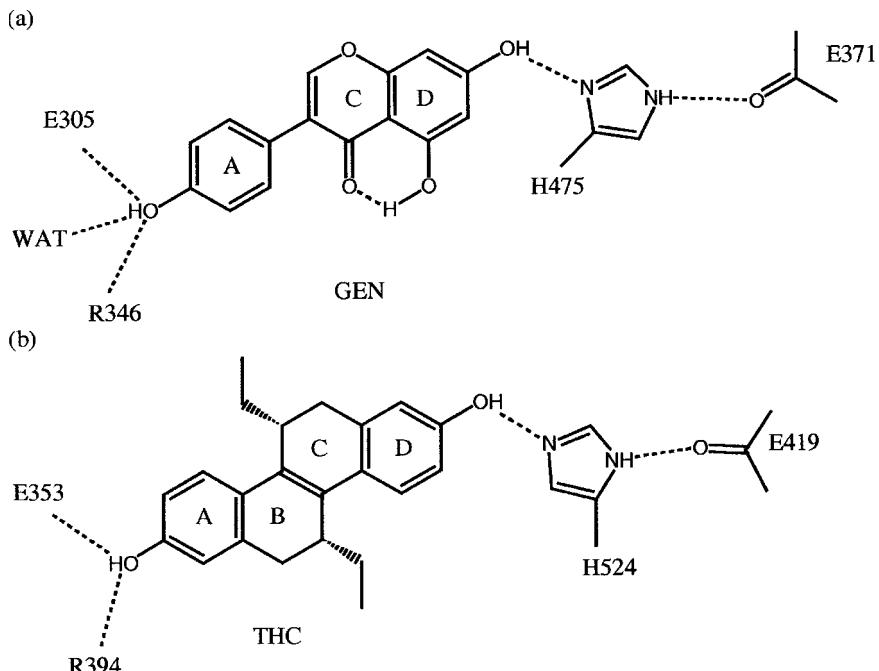


Fig. 1.11 (a) Key polar interactions of the partial agonist genistein with ER- β . (b) Structure of THC and its key interactions with

ER- α . THC is an ER- α agonist and an ER- β antagonist (see text for details).

binding cavity. The asparagine present at position 348 in ER- α , which interacts with Tyr-537, is replaced by lysine-300 in ER- β . Consequently, the “agonist” orientation of helix-12 may be less favored in ER- β due to loss of the hydrogen bonding interaction made by this residue to Tyr-398. However, the origin of GEN’s “destabilizing” influence on helix-12 remains unclear [11].

1.2.1.5 *R,R*-5,11-cis-Diethyl-5,6,11,12-tetrahydrochrysene-2,8-diol: An ER- α Agonist and ER- β Antagonist

The *R,R*-enantiomer of 5,11-cis-diethyl-5,6,11,12-tetrahydrochrysene-2,8-diol (THC; Fig. 1.11) exerts opposite effects on the transcriptional activity of the two estrogen receptor (ER) subtypes, ER- α and ER- β . THC acts as an ER- α agonist and as an ER- β antagonist. The crystal structures of THC bound to both ER- α (PDB entry: 1L2I) and ER- β (PDB entry: 1L2J) have been reported [12]. These two structures show how THC exerts different functional effects on the two isotypes of ER by stabilizing distinct conformations of the two receptors. The ER- α conformation favors co-activator binding, while the ER- β conformation precludes co-activator binding.

The complex between the C-2 symmetric THC and ER- α shows the A-ring phenol interacting with Arg-394 and Glu-353 and the D-ring phenol interacting with His-524 (Fig. 1.11b). Like other agonist structures, His-524 interacts with the backbone carbonyl of Glu-419. The two α -face ethyl substituents fill space in the middle of the ligand-binding core of the protein. The structure of the ER- α protein in the THC complex is nearly identical to its structure in the estradiol complex. For agonist structures, ligand binding stabilizes the conformations of several protein side chains that form part of the binding site for helix-12 or the loop between helix-11 and helix-12 [12].

THC binds to ER- β in a similar, but distinctly different manner than it does to ER- α . While the A-ring phenol makes very similar interactions in the ER- α and ER- β complexes, the D-ring fails to interact intimately with His-524 in the ER- β complex and it does not indirectly stabilize helix-12 in an agonist conformation. The B-ring ethyl group adopts a different conformation in the ER- β complex; this causes the core ring system to be tilted away from the floor of the binding pocket compared with the ER- α complex. As a result, the D-ring of THC fails to stabilize many of the ER- β side chains in a manner that stabilizes the agonist conformation of helix-12. Specifically, the side chain of His-475 fails to pack against that of Met-479 (equivalent to Met-528 in ER- α), causing the Met-479 side chain to be disordered. The backbone atoms of Met-479 and the residues flanking it adopt a random coil conformation (as compared with the helical conformation adopted by their counterparts in ER- α). As a result, both Leu-476 and Met-479 are not positioned appropriately to form interactions with relevant residues from helix-12 and the preceding loop that would stabilize the active conformation of helix-12. Thus, by positioning certain binding pocket residues in nonproductive conformations, the binding of THC to ER- β disfavors the agonist-bound conformation of helix-12 and shifts the equilibrium towards the inactive conformation of helix-12 [12].

1.2.2

Structural Basis for Agonism and Antagonism and Partial Agonism: The Role of the Ligand

The proposed role of agonists is to stabilize a precise conformation of the estrogen receptor that allows for productive dimerization and co-activator binding. The structures of agonist-bound NHRs reveal that helix-12 folds over the ligand into a position that stabilizes recruitment of a co-activator.

For the antagonist ligands OHT, RAL and ICI, the positioning of the ligand side chains precludes the agonist-bound conformation of helix-12 by steric hindrance. This mechanism of antagonist action has been referred to as “active antagonism”. The ER- β antagonist THC lacks a bulky side chain, and in its complex with ER- β , helix-12 is not sterically precluded from adopting the agonist-bound conformation. Instead, THC antagonizes ER- β by stabilizing nonproductive conformations of key residues in the ligand-binding pocket, thereby disfavoring the equilibrium to the agonist-bound conformation of helix-12 and leading to stabilization of an inactive conformation of helix-12. This mechanism of antagonism has been referred to as “passive antagonism”. There are many other examples of NHR ligands that act as antagonists even though they are smaller than the endogenous agonists of these NHRs; thus “passive antagonism” appears to be a common mechanism for antagonists [12].

The position of helix-12 for the ER- β antagonist THC is similar to the position of helix-12 for the ER- β partial agonist GEN. Based on the position of helix-12, ER- β in the conformation stabilized by THC and GEN should be incapable of interacting with co-activators and should not have agonist activity. However, GEN is a partial agonist and THC is a pure antagonist. The simplest model that explains these data is based on two hypotheses [12]. First, helix-12 in the unliganded ER- β LBD is in equilibrium between the inactive conformations observed in the THC and GEN ER- β structures and the active agonist-bound conformation. Second, ligands affect transcriptional activity by shifting this equilibrium rather than inducing a single static conformation of helix-12. Partial agonists, such as GEN, are incapable of shifting the equilibrium in favor of the active conformation to the same extent as full agonists, they are able only to increase the affinity of the receptor for co-activator incrementally. Partial agonist binding should permit the NHR to sample both inactive and active conformations of helix-12 [12] (see also [8]). On the other hand, passive antagonists may prevent the sampling of the agonist conformation or they may stabilize a protein conformation that allows binding to co-repressors [4].

There also exist a class of ligands referred to as SERMs (selective estrogen receptor modulators) that display tissue-selective pharmacology [13]. Raloxifene and tamoxifen are two clinically used SERMs for which structures are available. Crystal structures of the estrogen receptor bound to different ligands (estradiol, tamoxifen, or raloxifene) reveal that ligands of different sizes and shapes induce a spectrum of receptor conformational states. These states can be “interpreted” by the cellular complexion of co-regulators and the environment of the local promoter of

the target gene [13]. While the role of agonist's ligands is presumably understood, the structural basis of antagonism and partial agonism are less well understood. The structural basis of SERM activity is not understood at this time.

1.2.3

Progesterone/Progesterone Receptor Complex

Similar to the estrogen/estrogen receptor complex, the 3-keto steroid progesterone binds to the progesterone receptor (PR) by hydrogen bonding interactions at the two ends of the ligand and hydrophobic interactions in the middle of the ligand (Fig. 1.12). The two axial methyl substituents on the steroid core of progesterone contribute significantly to the overall shape complementarities between the steroid and PR. The 3-keto group of the progesterone A-ring forms hydrogen bonds to the side chains of Arg-766, Gln-725 and a water molecule that interacts with both Arg-766 and Gln-725. The side chain carbonyl (D-ring 17 β acetyl group) of progesterone makes a weak hydrogen bond with the side chain of Thr-894 (Fig. 1.12). There are no strong hydrogen bond interactions between the D-ring carbonyl oxygen atom of progesterone and the protein in PR, suggesting that the recognition of this group is made mainly through hydrophobic and steric interactions [14].

PR Arg-766 is located at the C-terminal end of helix-5 and aligns with ER- α Arg-394. PR Gln-725 is on helix-3 and aligns with ER- α Glu-353. The difference between a glutamine (PR) and a glutamic acid (ER) at the same position of helix-3 plays a role in the preference of PR for a 3-keto steroid and the preference of ER for an A-ring phenol. Sequence and structural alignments [15] show that androgen receptor (AR), glucocorticoid receptor (GR) and mineralocorticoid receptor (MR) also have a glutamine at this position in helix-3. These NHRs have 3-keto steroids as their natural ligands. However, some other NHRs that do not recog-

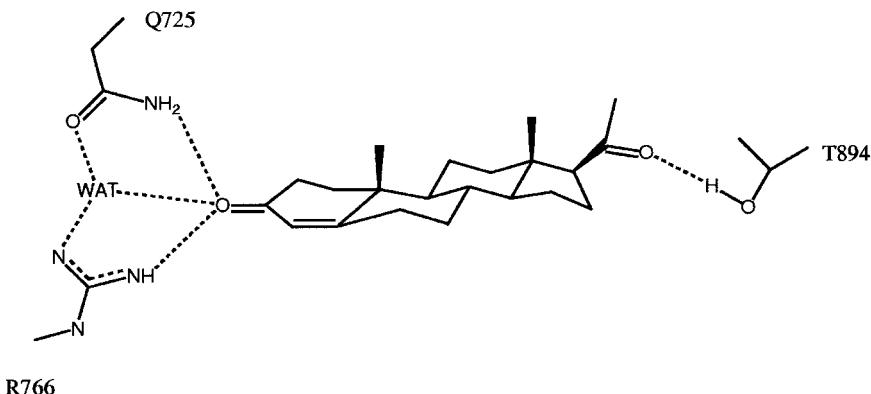


Fig. 1.12 Interactions of the 3-keto steroid progesterone with PR. Gln-725 from helix-3 forms a hydrogen bond with the 3-keto group of the ligand. In ER, the corresponding resi-

due that interacts with the A-ring phenol is a glutamic acid; thus, it appears that this interaction plays a key role in the selection of ketones by PR and phenols by ER.

nize 3-keto steroids, such as RXR, hepatocyte nuclear factor-4 (HNF-4) and *Drosophila melanogaster* nuclear receptor Ultraspirel (dUSP), also have a glutamine at this position. The PR Thr-894 residue is located on helix-11 and is in the *i+4* position relative to ER His-524.

1.2.4

Androgen Receptor Complexes

The structure of the androgen receptor (AR) complex with the natural agonist dihydroxytestosterone (DHT; Fig. 1.13) has been reported (PDB entry: 1I37) [16]. It binds to AR by hydrogen bonding interactions at the two ends of the ligand and hydrophobic interactions in the middle of the ligand. The two axial methyl substituents on the steroid core of DHT contribute significantly to the overall shape complementarities between the steroid and AR. The A-ring ketone group forms hydrogen bond interactions with Gln-711, Arg-752 and a structural water molecule that bridges those two residues. Met-745 makes VDW contacts with the A-ring and the C-19 β -methyl group. The D-ring hydroxyl group makes hydrogen bond interactions with the side chains of Asn-705, from helix-3, and the side chain of Thr-877 from helix-11 (Fig. 1.13).

Although there is only 55% sequence identity between the LBDs of AR and PR, there is a 77% sequence similarity, and as expected, the three-dimensional structures of these two LBDs are very similar. The ligand binding cores of the two proteins are quite similar in size and shape. DHT binds to the AR LBD in an almost identical fashion to the way progesterone binds to the PR LBD. Both agonists interact with helices 3, 5 and 11 of their respective LBDs. The interactions of ring A of each ligand with its receptor are similar, specifically those with the side chains of Gln-711, Met-745 and Arg-752 in the AR LBD, corresponding to those with Gln-725, Met-759 and Arg-766 in the PR LBD, and a conserved water molecule. The interactions of AR and PR with ring C are also similar, with close contacts to the main-chain of Leu-704 (Leu-718 in the PR LBD) and side chain of Asn-705 (Asn-719 in the PR LBD). The contact between C18 of DHT and the O γ 1 of Thr-877 is unique to the wild-type AR LBD, as the corresponding cysteinyl side chain is pointed away from the steroid in the PR LBD structure. These D-ring polar interactions are presumably responsible for the ability of AR to select the proper ligands correctly. The AR point mutant T877A shows promiscuity for other steroid ligands such as progestins, estrogens and cortisols that differ from DHT in substitution at position 17 on the D-ring. The structure of this mutant, complexed with DHT, has been reported (PDB entry: 1I38) and is nearly identical to the wild type structure. The replacement of T887 by alanine leaves additional space off the D-ring and may explain its promiscuity [16].

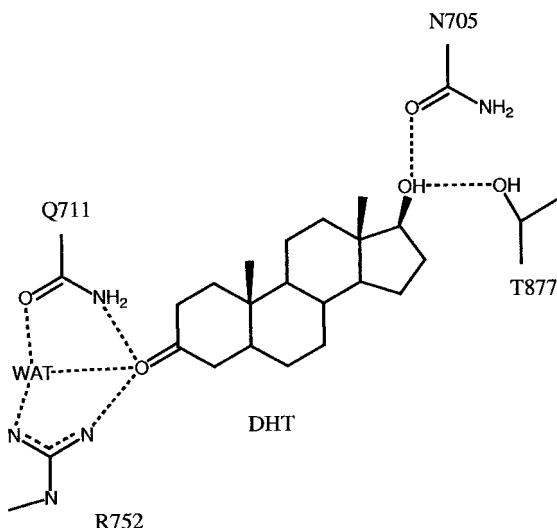


Fig. 1.13 Interactions of DHT with AR. As in other steroid NHR complexes, the ligand is oriented on both ends by polar interactions. Mutagenesis experiments have implicated Thr-877 as a key residue that determines ligand selectivity.

1.2.5

Glucocorticoid Receptor

The structure of the glucocorticoid receptor (GR) in complex with the agonist ligand dexamethasone (Fig. 1.14) has been reported (PDB entry: 1M2Z) [17]. As with the 3-keto steroid ligands for PR and AR, the A-ring carbonyl of dexamethasone forms direct hydrogen bonds to Arg-611 and Gln-570 (Fig. 1.14). Like other agonist-steroid receptor complexes, dexamethasone does not have a precise fit in the ligand-binding pocket and occupies 65% of the binding pocket volume. The high affinity binding of dexamethasone to GR is readily explained by the extensive hydrophobic and hydrophilic interactions between the ligand and the protein. One or more hydrophobic residues within the GR protein contact nearly every atom of the steroid core of dexamethasone. In addition, all of the hydrophilic groups of dexamethasone form hydrogen bonds with the protein. Compared with the ligand-binding pocket found in the PR, AR or ER structures, the GR pocket has an additional branch extending from the center in the side of the steroid pocket. This additional side pocket in GR is formed by the structural rearrangement of helices-6 and -7. The side chain of Asn-564 is oriented in a way to allow it to make hydrogen bonds to the C-ring 11-hydroxyl and the D-ring 24-hydroxyl. Furthermore, the 21-hydroxyl (off the C17 position) and the 22-carbonyl form hydrogen bonds with residues Gln-642 and Thr-739, respectively. The extensive hydrogen bond network between GR and the ligand observed here are likely to contribute to the high affinity binding of dexamethasone [17].

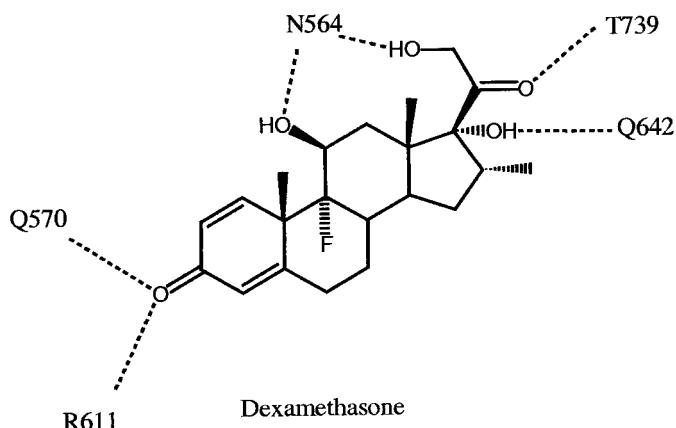


Fig. 1.14 Key polar interactions of dexamethasone with GR.

1.2.6

Steroid Ligand Selectivity

Steroid hormones such as cortisol, corticosterone, testosterone, progesterone and estrogen share a similar core chemical structure but mediate distinct biological responses. Structural comparisons of GR, AR, PR, and ER have now provided some insight into how specificity is achieved by the steroid receptors. In the structures of steroid receptors complexed with steroid agonists, the core steroid template assumes a common orientation with the A-ring oriented toward a conserved arginine from helix-5 and the D-ring toward helix-12. However, many subtle differences in the secondary structure and the topology of the ligand-binding pockets exist in different steroid receptors. In particular, helices-6 and -7 of GR deviate significantly from ER, AR and PR and produce a unique side pocket in GR. This pocket may account for the GR selectivity of glucocorticoids, which have larger substituents at the C17 α position compared with estrogen, progesterone and testosterone. Interestingly, the mineralocorticoids that selectively bind MR have similar substituents at the C17 α position. To date no structure for MR has been reported; however, MR may also have a similar pocket for these large C17 α substitutes, since the residues that form the GR side pocket are also conserved in MR (Fig. 1.15) [17].

Besides the shape differences, the polar atoms are also distributed differently in the steroid pockets with respect to the specific protein-ligand hydrogen bonds. For example, the polar substituents in steroid hormones are often located at positions C3 or C17. As discussed above, in the PR structure, the C3 ketone accepts hydrogen bonds from Arg-766 and Gln-725. These residues are also present in AR, GR and MR; however, in ER, the glutamine is replaced by glutamate, which prefers to accept a hydrogen bond from the phenolic group in the C3 position of the A-ring of estradiol [7]. These differences in hydrogen bond formation may explain why

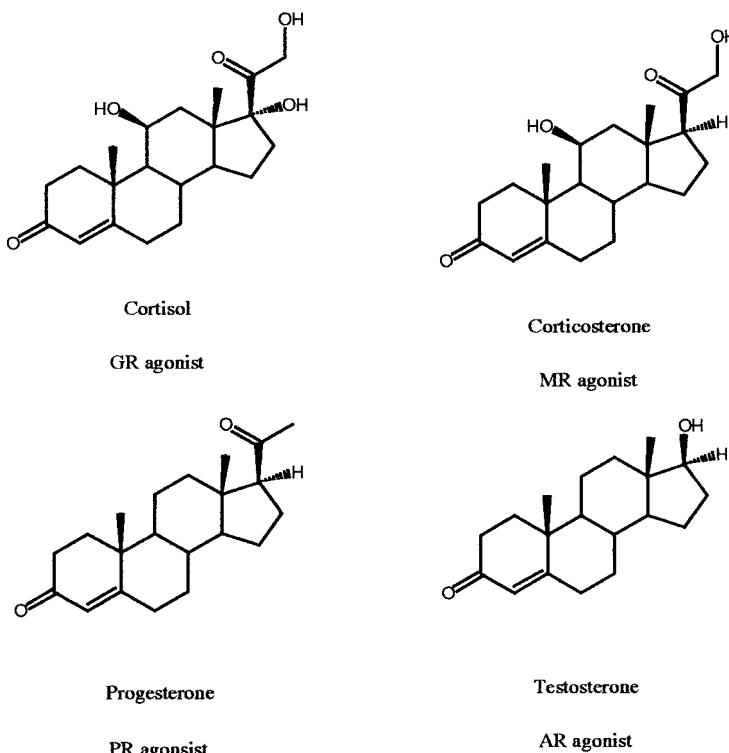


Fig. 1.15 Structure of some steroid ligands.

PR, as well as AR, GR and MR prefer the steroid hormones with a ketone at the C3 position, whereas ER prefers a phenol group. The larger GR side pocket can best explain the selectivity of GR ligands with larger substituents at the C17 position. While MR presumably has a similar pocket to GR the selectivity of MR for mineralocorticoids can be attributed to the differences in hydrogen bonding patterns between the receptor and the ligands. The MR-selective steroids, corticosterone, aldosterone and 11-deoxycorticosterone, all lack the 17 α -hydroxyl group, which forms a specific hydrogen bond with Gln-642 in the GR structure. At this position, MR has a hydrophobic leucine (Leu-848) that should disfavor the presence of a polar hydroxyl in this region [17]. Many of the structures between steroid receptors and agonist ligands show a less than optimal steric fit between the ligand and the protein. However, it appears that the subtle differences in the shape and electronic properties of the ligand binding pockets of steroid receptors are sufficient to either select the correct ligand or to deselect the incorrect one.

Structures exist for a common ligand, R1881, bound to both PR (PDB entry: 1E3K) and AR (PDB entry: 1E3G) [18]. This ligand shows good binding affinities to both proteins and the affinities are comparable to the affinity of the natural ligand for each protein (Fig. 1.16). PR can bind R1881 as well as progesterone, in

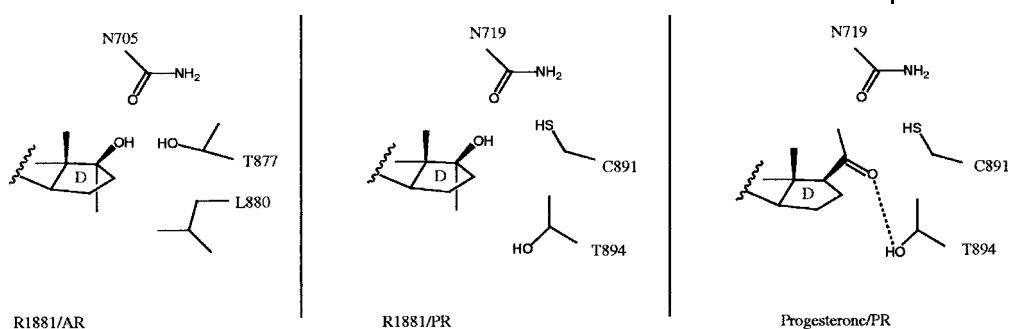


Fig. 1.16 Examples of interactions in the D-rings of steroids that account for ligand selectivity.

both the AR and PR complexes R1881 binds similarly with the A-ring ketone interacting with Arg-766 in PR (Arg-752 in AR) and Gln-725 in PR (Gln-711 in AR) and a structural water molecule.

The overall shape of the ligand-binding pocket is similar in both PR and AR and many of the nonpolar side chains contacting the ligand are the same in both proteins. There are a total of 18 amino acid residues in AR and PR that interact with the bound ligands (either R1881 or progesterone). Most of these residues are hydrophobic and interact mainly with the steroid scaffold, whereas a few are polar and may form hydrogen bonds to the polar atoms in the ligand. The 17β hydroxyl group of R1881 forms different hydrogen bonds, when bound to AR or PR. In AR, the 17β hydroxyl group is hydrogen-bonded to Asn-705 and Thr-877. The same pattern is observed in the PR-R1881 complex where the 17β hydroxyl group of R1881 also forms a hydrogen bond to Asn-719. In contrast to the AR complex, Cys-891 (Thr-877 in AR) shows a weak interaction with the 17β hydroxyl group of R1881. Thr-894 in PR is replaced by Leu-880 in AR, and a methyl group of this leucine makes a van der Waals contact with both the 17α methyl group and 17β hydroxyl group of R1881. This bulkier side chain in AR is very likely responsible for the failure of AR to recognize the 17β acetyl group of progesterone. While there is an extra polar residue (Thr-877 besides Asn-705 which is also present in PR) that can form an additional hydrogen bond to the 17β hydroxyl oxygen, it is likely that the decrease in pocket volume caused by the change of Thr-894 to Leu-880 inhibits the binding of bulkier ligands such as progesterone [18].

While there are few published examples that directly address issues of steroid selectivity, some trends are apparent. The steroid receptor-agonist-ligand complexes do not involve precise steric fits between the protein and the ligand. In each case, the ligand is completely enclosed within the ligand-binding core of the protein. While the steric match is not precise, there are few, or no, polar groups present that are not involved in a protein–ligand hydrogen bond. Thus, one element of selectivity involves having a polar protein group to complement a ligand polar group. The presence of a glutamate on helix-3 (E353 on ER- α) is responsible for the preference of ER for ligands containing an A-ring phenol. In PR, AR, GR and MR the corresponding position on helix-3 has a glutamine and is responsible

for the exclusion of A-ring phenol containing ligands and the preference of A-ring ketone containing ligands. Subtle changes elsewhere in the ligand-binding pocket appear to be responsible for the B-, C- and D-ring preference. It is not clear whether ligand selectivity of steroid receptors is due to the exclusion of the wrong ligand(s) or to the selection of correct ligand.

1.3

The Vitamin D Receptor-Ligand Complexes

The vitamin D receptor (VDR) binds the vitamin D metabolite $1\alpha,25$ -dihydroxyvitamin D3 (VDX; Fig. 1.17) as its natural agonist ligand. VDX is a seco-steroid hormone, that is, it is a steroid with an opened B-ring. The structure of the VDX/VDR complex has been reported (PDB entry: 1DB1) using a deletion mutant of VDR [19, 20]. The structure of the protein in the complex resembles a typical agonist conformation with helix-12 folded over and enclosing the ligand. The hydroxylated A-ring makes polar interactions with several groups on the protein. The pseudo-equatorial hydroxyl group forms hydrogen bonds with the side chains of Ser-237 and Arg-274. Arg-274 is located on helix-5 while Ser-237 is located on helix-3, this is the equivalent position of Glu-353 of ER- α and Gln-725 of PR. The axial hydroxyl group forms hydrogen bonds with the phenol group of Tyr-143 and the hydroxyl group of Ser-278. Tyr-143 is located at the C-terminus of helix-1 and Ser-278 is located at the near the C-terminus of helix-5, this is the equivalent position of Arg-394 of ER- α and Arg-766 of PR. The tertiary hydroxyl group off the D-ring interacts with two different histidine side chains, His-305, which is located between helix-6 and helix-7, and His-397, which is located on helix-11. His-305 also hydrogen bonds to the side chain of Gln-400 from helix-11 (Fig. 1.17).

Comparison of the agonist-bound structure of ER- α and VDR provides insight into how different NHRs can select different ligands. The gross structures of these two proteins differ in the relative positions of helix-7 and helix-11 relative to helix-3 and in the conformation of the protein between the end of helix-5 and the beginning of helix-7. These changes alter the size and shape of the binding pocket to better accommodate estradiol (ER- α) or VDX (VDR). Key side chain differences, such as VDR Ser-237 (Glu-353 in ER- α) and VDR Ser-278 (Arg-394 in ER- α), affect the electronic complementarity of the NHR for the appropriate ligand.

The crystal structures of the ligand-binding domain of the vitamin D receptor complexed to VDX (Fig. 1.18) and the 20-epi analogues, MC1288 (PDB entry: 1IE9) and KH1060 (PDB entry: 1IE8), show that the protein conformation is identical, further suggesting that for a given LBD the agonist conformation is unique. In all complexes, the A- to D-ring moieties of the ligands adopt the same conformation and form identical contacts with the protein. The explanation of the superagonist effect of the 20-epi analogs is to be found in higher stability and longer half-life of the active complex, thereby excluding different conformations of the ligand-binding domain [21].

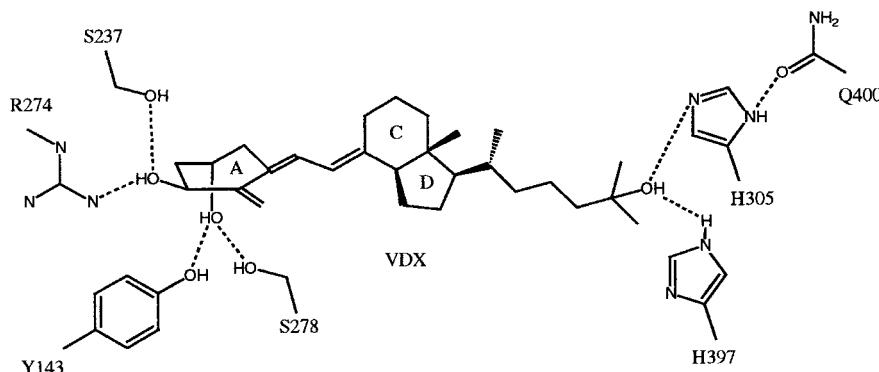


Fig. 1.17 Polar interactions of VDX with VDR.

Vitamin D₃-resistant rickets is associated with mutations to VDR, some of which effect ligand binding and ligand-dependent transactivation. The missense mutation R274L causes a >1000-fold reduction in VDX responsiveness and is no longer regulated by physiological concentrations of the hormone. There have been two reports of the use of the VDR structure to design selective ligands that bind the mutant receptor in preference to the wild type VDR [22, 23].

In one study, it was proposed that the R274L mutation removed a polar interaction between the protein and the natural ligand VDX and created a hydrophobic hole in its place. SS-III, a derivative of VDX, was designed to fill that hole. It was prepared and found to be 286 times more potent than VDX against the mutant protein [22] (Fig. 1.18).

In another study, computer-aided molecular design was used to generate a focused library of nonsteroidal analogues of the VDR agonist LG190155 (Fig. 1.19) that were uniquely designed to complement the R274L protein associated with Vitamin D₃-resistant rickets. Half of the designed analogues exhibit substantial activity in the R274L mutant. The seven most active designed analogues (such as A-11; Fig. 1.19) were more than 16 to 526 times more potent than VDX in the mutant receptor. Significantly, the analogues are selective for the nuclear VDR and did not stimulate cellular calcium influx, which is associated with activation of the membrane-associated vitamin D receptor (Fig. 1.19) [23].

1.4

The Retinoic Acid Receptors RAR and RXR

1.4.1

Introduction

Retinoid acids (RAs), the active retinoid derivatives of vitamin A, regulate complex gene networks. The pleiotropic effects of active retinoids are transduced by their cognate nuclear receptors, retinoid X receptors (RXRs) and retinoic acid receptors

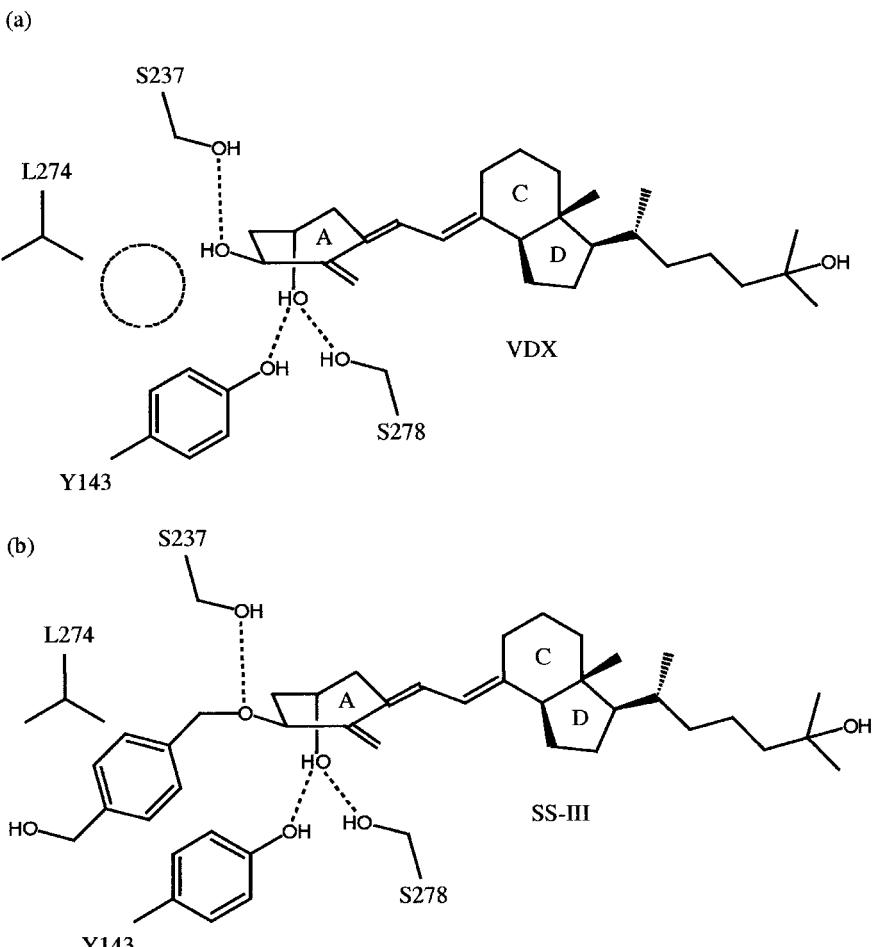


Fig. 1.18 (a) The R274L VDR mutation fails to bind the natural ligand VDX. (b) The li-

gand SS-III was designed and found to bind to this mutant protein.

(RARs), which act as transcriptional regulators activated by two stereoisomers of retinoic acid (RA): 9-cis RA (9cRA) and all-trans RA (atRA). Among the nuclear receptors, RXR occupies a central position and plays a crucial role in many intracellular signaling pathways as a ubiquitous heterodimerization partner with numerous other members of this superfamily. Whereas RARs bind both retinoic acid isomers, RXRs exclusively bind 9cRA. The various RAR (α , β and γ) and RXR (α , β and γ) isoforms are encoded by different genes, while their multiple isoforms, which differ in their N-terminal region, are the results of differential promoter usage and alternative splicing. It has been shown that RXRs play a unique role among NHRs since they are able to heterodimerize with a number of members of the NHR superfamily [24]. This section will give an overview of the

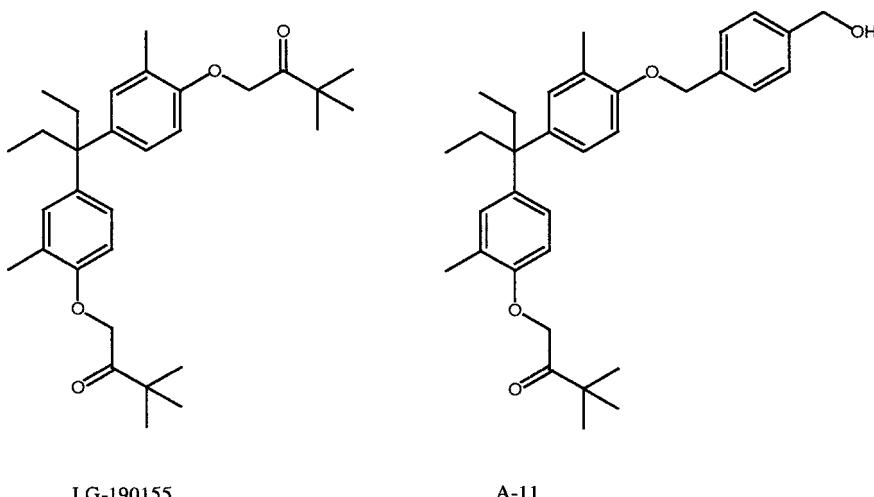


Fig. 1.19 Structures of ligands for VDR R274L mutant.

structural biology of retinoid receptor-ligand complexes focusing on the role of the ligand.

1.4.2

RAR- γ and RXR- α Retinoid Complexes

The structure of atRA bound to RAR- γ was one of the first reported structures of the ligand-binding domain of an NHR bound to a ligand (PDB entry: 2LBD) [25, 26]. Like the steroid receptors the hydrophobic atRA ligand is buried, but fits loosely, within the ligand binding hydrophobic core of the protein. The protein adopts a typical agonist-bound conformation with helix-12 folding over, and enclosing, the ligand. The carboxylate group of atRA makes hydrogen bonds with both the backbone NH and side chain OH of Ser-289. Ser-289 makes two hydrogen bonds to the guanidine side chain of Arg-278. The carboxylate group of atRA also forms a hydrogen bond with a structural water molecule which is also hydrogen bonded to the backbone carbonyl of Leu-233. Arg-278 is located at the end of helix-5 and Ser-289 is located on a turn between helix-5 and helix-6 (Fig. 1.20).

Comparing the agonist-bound structures of RAR- γ and ER- α shows some significant differences. There is a significant change in the positions of helices-1, -3, -6, -7 and -9, and a small change in the positions of helices-11 and -12 relative to helices-4 and -5. In addition, the protein conformation between the end of helix-5 and the beginning of helix-7 is very different between ER- α and RAR- γ . As with ER- α , there is an arginine at the end of helix-5 (Arg-278 in RAR- γ , Arg-394 in ER- α). Glu-353 on helix-3 of ER- α is replaced by Cys-237 in RAR- γ . Unlike ER and the other steroid receptors, this residue in RAR- γ (Cys-237) does not appear to play a key role in ligand recognition. As was seen for VDR, changes in the gross

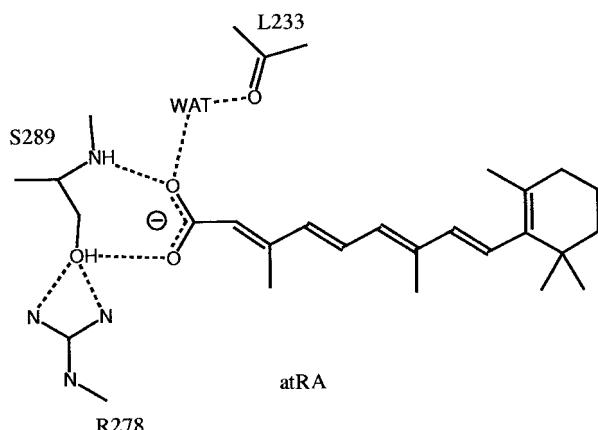


Fig. 1.20 Schematic of interactions between atRA and RAR- γ .

structure of the protein result in changes in the relative positions of the helices. These changes alter the size and shape of the binding pockets to better accommodate the appropriate ligands. Clearly different NHRs have evolved structures that are complementary to their respective ligands.

RAR- γ is a permissive receptor and will also bind 9cRA as an agonist; the structure of the protein in this complex (PDB entry: 3LBD) [27] is nearly identical to that in the atRA complex. The polar carboxylate group of 9cRA interacts with RAR- γ in almost the same manner that atRA does and the rings of 9cRA and atRA fill the same hydrophobic binding site. However, the less than perfect fit of these two ligands to RAR- γ allows the tetraene chain to adopt different conformations (20-methyl group on opposite sides of the pocket) in the two complexes. In each complex the 19-methyl group occupies approximately the same space. The difference in chain conformations, coupled with a loose steric fit, allows for either a *cis* or *trans* olefin to be accommodated at the 9–10 position (Fig. 1.21).

In contrast to RAR, RXR can select 9cRA in preference to atRA. The structure of the RXR- α /9cRA complex has been reported (PDB entry: 1FBY) [24]. This structure exhibits the typical agonist protein conformation with helix-12 folded over the ligand. The carboxylate of 9cRA forms hydrogen bonds to the backbone NH of Ala-237 and the guanidine group of Arg-316. In addition, the carboxylate makes a hydrogen bond to a water molecule; that water forms hydrogen bonds to the backbone carbonyl of Leu-309 and to another water that is hydrogen bonded to the side chain of Gln-275. Arg-316 is located at the end of helix-5 and is in the equivalent position to Arg-278 in RAR- α and Ala-237 is in the equivalent position to Ser-289 of RAR- α (Fig. 1.22). Of interest, Gln-275 is located on the same position of helix-3 as Gln-725 in PR. As discussed previously this glutamine is conserved in several other steroid receptors that bind 3-keto steroids and is equivalent to Glu-353 in ER- α . Recall that in the steroid receptors this residue plays a key role in A-ring recognition. However, in RXR- α this glutamine residue is solvent exposed and makes an indirect polar

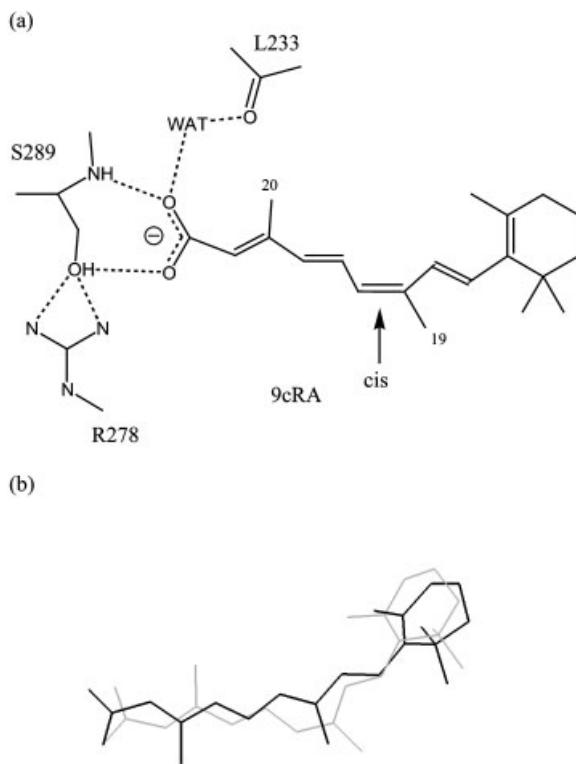


Fig. 1.21 (a) Schematic of the interactions between 9cRA and RAR- γ . (b) Superposition of atRA (black) and 9cRA (gray) bound to RAR- γ .

interaction with the ligand. The hydrophobic ligand-binding pocket has a distinct L-shape to it. The *cis*-olefin of 9cRA allows the ligand to bend and accommodate this binding site. While the fit of 9cRA to RXR- α is not precise, the shape of the binding site precludes the binding of incorrect ligands such as atRA. This readily explains why RXR- α is selective for 9cRA as opposed to atRA.

1.4.3

Selectivity of RAR Ligands and RAR Isotypes

A very elegant experiment has been reported where enantiomers of an unnatural RAR- γ -selective ligand have been determined [28]. Of special significance is the observation that one of the enantiomers is effectively inactive in the biochemical assays. While acquiring co-crystal structures of many active ligands with the same protein is now commonplace, obtaining structures of inactive (i.e., weakly active) compounds is more challenging. This work provides a high-resolution structure (PDB entry: 1EXX) of the complex between an inactive compound (BMS270395,

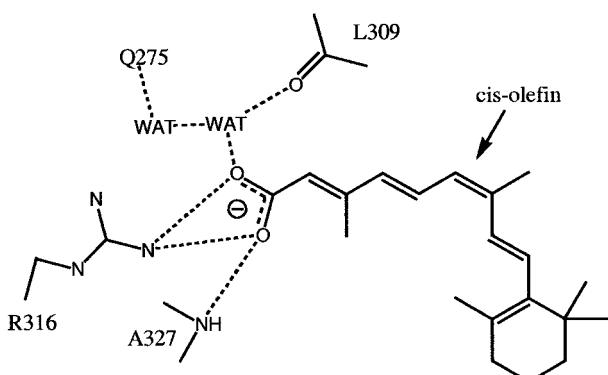


Fig. 1.22 Schematic of the polar interactions between 9cRA and RXR- α . The shape of the RXR- α protein is complementary to 9cRA and explains its selectivity.

$K_d \approx 500 \mu\text{M}$) and RAR- γ . The structure (PDB entry: 1EXA) of the more potent enantiomer (BMS270394, $K_d=500 \text{ nM}$) and RAR- γ was also determined for direct comparison (Fig. 1.23).

The conformational profiles and energies of the two enantiomers are, by definition, mirror images of each other. Thus, the ligand conformational arguments made by Klaholz et al. [28] are not correct. If one superimposes the mirror image of the inactive enantiomer on the active isomer it is apparent that the bound conformations of the two ligands are similar. The A-rings (except for the fluorine), amide groups and alcohol moiety superimpose almost exactly. The major difference in the conformations is the torsion angle ($\text{C}=\text{O}-\text{CH}-\text{C}(\text{Ar})-\text{C}(\text{Ar})$; it is 53° and -102° for the active and inactive enantiomer, respectively. A quick MM2 calculation predicts a 119° torsion angle is a low energy local conformational minimum [29]. These calculations suggest that it is unlikely that the conformation energy difference between the active and inactive enantiomers is close to 4 Kcal mol^{-1} . It would appear that ligand conformational strain does not play a dominant role in the large difference in affinity between the two enantiomers. Thus, the major component of the large difference in affinity between these two isomers must come from protein-ligand interactions.

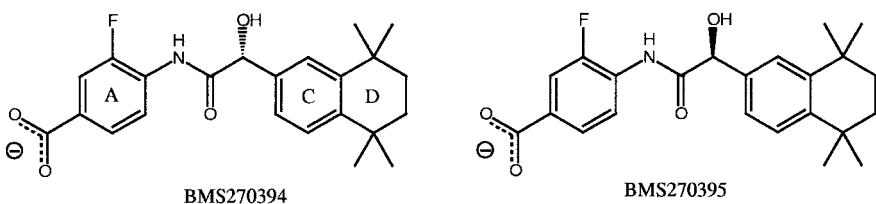


Fig. 1.23 Structures of enantiomeric ligands, BMS270394 is a 500 nM RAR- γ ligand while its enantiomer BMS270395 is a very poor RAR- γ ligand.

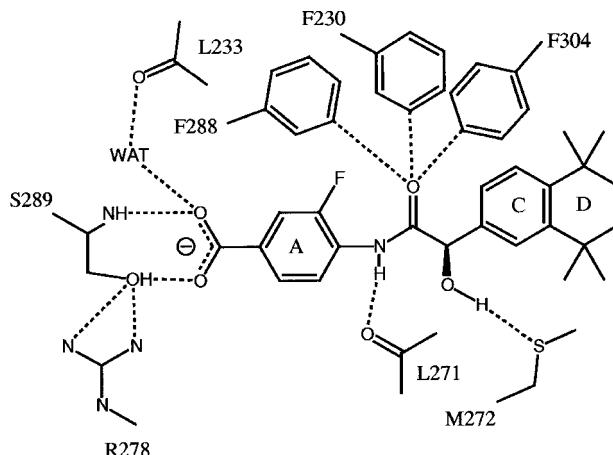


Fig. 1.24 Key interactions between the active enantiomer BMS270394 and RAR- γ . A key interaction involves a hydrogen bond between the ligand hydroxyl group and the sulfur of Met-272. Notice that the linking amide group makes complementary interactions with the enzyme.

The structure of the more active enantiomer (BMS270394) shows that the ligand binds to RAR- γ in a typical agonist conformation (Fig. 1.24). The carboxylate group of the ligand forms hydrogen bonds to Ser-289 and a water molecule in the same fashion that atRA does. The hydrophobic CD ring system makes VDW contacts in the hydrophobic ligand-binding core. The hydroxyl group of the ligand makes a hydrogen bond (3.2 Å) with the thioether of Met-272. The amide group also interacts well with the protein, the NH hydrogen bonds to the backbone carbonyl of Leu-271 and the amide carbonyl interacts with three aromatic H atoms; two of these interactions are much shorter than the third. The shortest of these interactions (3.17 Å) is with Phe-230. The interactions of carbonyl oxygens with aromatic hydrogens are well preceded, favorable interactions in protein-ligand complexes. These are weakly polar interactions between the positive dipole of an aromatic C–H bond and the electrons of a carbonyl oxygen. The fluorine atom exhibits short VDW contacts with the Ca and $\text{C}\beta$ atoms of Ala-234. For the more active isomer, all of the ligand's polar groups make complementary polar interactions with the protein, the fluorine atom comes into close VDW contact with the protein, and the hydrophobic groups are buried in hydrophobic parts of the protein.

The structure of the weakly active enantiomer (BMS270395) shows that the ligand also binds to RAR- γ in a typical agonist conformation (Fig. 1.25). The carboxylate group of the ligand makes hydrogen bonds to Ser-289 and a water molecule in the same fashion that BMS270394 does. The hydrophobic CD ring system creates VDW contacts in the hydrophobic ligand-binding core in much the same manner as BMS270394. In addition, the hydroxyl group of the inactive ligand also

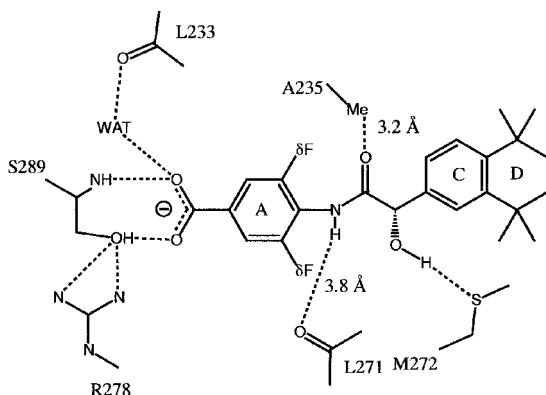


Fig. 1.25 Key interactions between the “inactive” enantiomer BMS270395 and RAR- γ . Like the active enantiomer, a key interaction involves a hydrogen bond between the ligand hydroxyl group and the sulfur of Met-272. In this isomer the amide group does not make favorable interactions with the protein and the A-ring shows some disorder in the positions of the fluorine atoms.

forms a strong hydrogen bond (3.2 Å) with the thioether of Met-272. If the “inactive” and the “active” enantiomer adopted the same bound conformation, the hydroxyl group would still have had room to fit, indicating that the interaction between the hydroxyl group and the methionine is quite significant in determining the structure of the complex. Thus, the “inactive” enantiomer makes many of the same interactions with the protein that the „active“ enantiomer does. The major differences between the two are the position of the linking amide group and the position and occupancy of the fluorine atom and the conformation of the ligand. In the “inactive” enantiomer complex, the amide group does not make complementary interactions with the protein, and unlike the “active” enantiomer, the polar amide group is buried in the protein without making favorable compensatory interactions. Also, unlike the “active” enantiomer, BMS270395 shows two different orientations for the fluorine atom, in both orientations there are sub-optimal interactions between the fluorine and the protein. While the aromatic A-rings are in nearly the same position in both ligands, subtle differences in the position of the A-ring appear to be responsible for the disorder in the fluorine atoms. Furthermore, while the bound conformations of the two ligands are different, as discussed above, the conformational differences are not likely to be too severe. It appears that the burial of the polar amide and the less than optimal fit of the fluorine atom are the key determinant of enantiomer selectivity.

Both of these structures show a hydrogen bond to the thioether of Met-272. Clearly, this is an important recognition element for binding to RAR- γ . In both RAR- α and RAR- β , the residue corresponding to Met-272 is an isoleucine. BMS270394 is much less potent against both RAR- α and RAR- β . While the structure of BMS270394 with either RAR- α or RAR- β has not been reported, one can speculate that the burial of the hydroxyl group near the isoleucine is responsible for the weaker affinity of these ligands for these isotypes.

A paper has been published that discusses the structural basis for isotype selectivity of the RARs [30]. This paper reported structures of three RAR ligands bound to RAR- γ , BMS184394 (PDB entry: 1FCX), an RAR- γ -selective agonist, CD564 (PDB entry: 1FCY), an RAR- β/γ co-agonist and BMS181156 (PDB entry:

1FCZ), a panagonist (Fig. 1.26). Residues that directly contact the ligand are mostly conserved in all three isoforms. The only differences are that Met-272 in RAR- γ is replaced by an isoleucine in both RAR- α and - β , and that Ala-234 in RAR- γ is replaced by a serine in RAR- α .

BMS181156 has approximately the same activity against all three RAR isotypes and it is the most potent of these three ligands against RAR- γ . The structure of its complex bound to RAR- γ shows the bicyclic ring system bound in a hydrophobic site similar to BMS270394 and the carboxylate interacting with Ser-289 in the same manner as other RAR- γ ligands. This ligand binds to RAR- γ in a manner similar to 9cRA with the carbonyl group of BMS181156 occupying the same space as the C19 methyl group of 9cRA. In the BMS181156 complex, the carbonyl group makes close contacts with the C α atom of Phe-304 and the C γ atom of Met-272. These close contacts have been described as C-H \cdots O=C hydrogen bonds [30]. The conformation of the Met-272 side chain is different in this complex to that in the BMS270394 and BMS270395 complexes, in the BMS181156 complex the sulfur atom does not interact directly with the ligand.

CD564 shows comparable affinity towards RAR- γ and RAR- β but is 40-fold less potent against RAR- α . It is 5-fold less potent against RAR- γ than BMS181156. The structure of its complex bound to RAR- γ shows the bicyclic ring system bound in a hydrophobic site similar to BMS181156 and the carboxylate interacting with Ser-289 in the same manner as other RAR- γ ligands. The carbonyl group of CD564 makes a close contact with C ζ of Phe-304, which is in a different conformation than it is in the BMS181156 complex. The loss in potency against RAR- α is attributed to a steric clash between the naphthalene ring of the ligand and a serine side chain. In RAR- γ (and RAR- β), this residue is Ala-234 and, in the complex structure, this residue is in close contact with the naphthalene ring. For

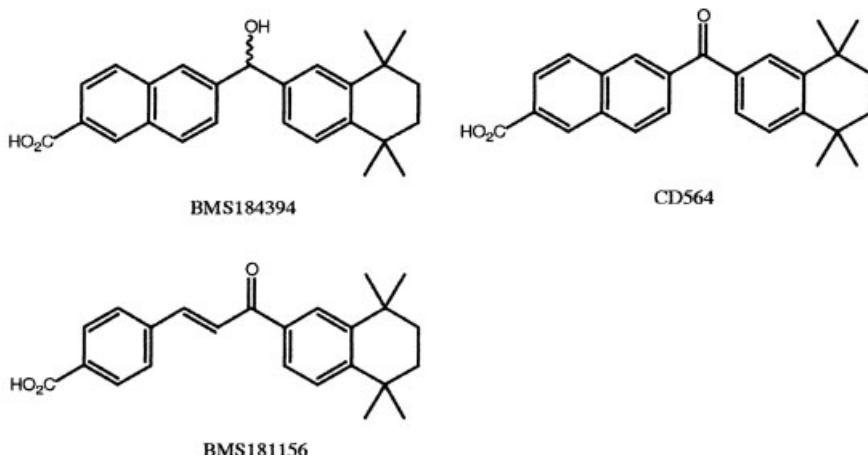


Fig. 1.26 Structures of RAR ligands.
BMS184394 is an RAR- γ -selective ligand,

CD564 is an RAR- β/γ co-agonist, and BMS181156 is an RAR panagonist.

BMS181156 the acyclic linker avoids this steric clash allowing potent binding to RAR-*a*.

BMS184394, a mixture of enantiomers, is 100-fold less potent than BMS181156 against RAR-*γ*; however, it shows 10-fold selectivity over RAR-*β* and 100-fold selectivity over RAR-*a*. This ligand binds to RAR-*γ*, as a single enantiomer, in a manner nearly identical to CD564. The major difference in the ligand is the difference between a ketone and an alcohol. As with BMS270394 and BMS270395, the alcohol group of BMS184394 forms a hydrogen bond to the thioether of Met-272. The side chain of Met-272 is in the same conformation as it is in the BMS270394 and BMS270395 structures. Thus, this side chain adopts different conformations depending upon the presence or absence of a hydrogen bond between the ligand and the thioether. The data suggests that the interaction of a ligand alcohol group with the thioether of Met-272 plays a key role in RAR-*γ* selectivity. The significant loss of affinity upon going from a ketone (CD564) to an alcohol (BMS184394) suggests that the observed C–H···O=C hydrogen bonds in the ketones play an important role in stabilizing the complex and the alcohols cannot take advantage of this type of interaction. For RAR-*γ*, the alcohol-containing ligands can gain sufficient affinity by forming a hydrogen bond with Met-272 and thus becomes a selective ligand.

SR11254, the oxime of CD564, is a potent (4 nM) RAR-*γ*-selective agonist [31]. Its structure in complex with RAR-*γ* (PDB entry: 1FD0) shows that it binds in a manner nearly identical to CD564 (Fig. 1.27) [32]. The ligand binds as a ~1:1 mixture of *E* and *Z* oximes. In both isomers, the hydroxyl group forms a hydrogen bond to the Met-272 thioether. The side chain of Met-272 resembles the conformation found in other complexes in which the ligand forms a hydrogen bond to the Met-272 side chain. In this structure, close contacts between the oxime oxygens and protein methyl groups are observed. The *Z*-isomer interacts with the CE methyl group of Met-272 and the *E*-isomer interacts with the C_γ2 methyl group of Ile-275. These have been characterized as C–H···O hydrogen bonds and are proposed to help stabilize the complex. The Klaholz and Moras paper is also noteworthy for its discussion of C–H···O hydrogen bonds [32].

The above studies provide a structural explanation for how very similar NHRs (i.e., the isotypes of RAR) can selectively bind different ligands. In each case a key polar interaction (ligand carboxylate) and the burial of a substantial amount of hydrophobic surface allows binding to occur. Tight binding occurs when there are no bad steric or electrostatic interactions. The case of enantiomer selectivity (BMS270394 vs. BMS270395) occurred when only one of the enantiomers could bury an amide group and make complementary interactions. RAR-*γ* and RAR-*β* differ by only one amino acid group in the ligand-binding pocket (RAR-*γ* Met-272 is an Ile in RAR-*β*). Ligands that can form a hydrogen bond to the thioether of Met-272 tend to be selective for RAR-*γ*. RAR-*β* and RAR-*a* differ by only one amino acid group in the ligand-binding pocket (RAR-*γ* Ala-234 is an Ala in RAR-*β* and a Ser in RAR-*a*). Selectivity for RAR-*β* over RAR-*a* can be achieved with ligands that make close VDW contacts with the side chain of Ala-234 in RAR-*γ* (or RAR-*β*). These ligands presumably bind poorly to RAR-*a* due to a bad steric con-

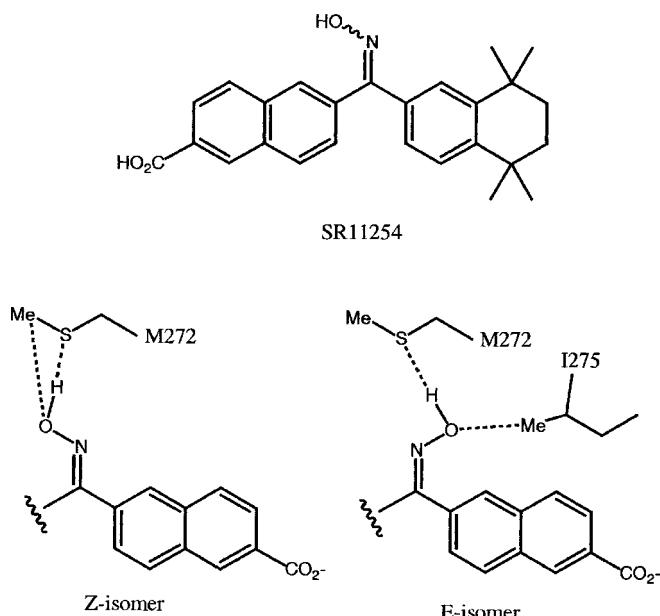


Fig. 1.27 Key interactions of the RAR- γ -selective agonist with RAR- γ .

tact. Thus, it is possible to achieve selectivity against very similar isoforms of the same NHR and it is possible to rationalize these selectivities based on structure. It is likely that in the future subtle selectivity will be rationally designed into ligands.

1.4.4

RXR Complexes with Unnatural Ligands

The structures of two structurally related RXR-selective ligands have been reported. The complex between BMS-649, RXR- α and a co-activator peptide (PDB entry: 1MVC) [33] and between LG-268 and RXR- β (PDB entry: 1H9U) [34] are similar and will be discussed briefly below. The carboxylate group of both ligands interacts with the backbone NH of Ala-327 (RXR- α , Ala-398 RXR- β) and Arg-316 (RXR- α , Arg-387 RXR- β) in the same manner that 9cRA does. The D-ring binds in the same place that the ring of 9cRA does (Fig. 1.28). The shape of the ligand is complementary to the L-shaped RXR ligand-binding pocket. This shape complementarity explains why these ligands are selective for RXR over RAR.

One significant difference between these two structures is the position of helix-12; in the complex between RXR- α , BMS-649 and a co-activator peptide, helix-12 is in the standard agonist position. In the complex between LG-268 and RXR- β helix-12 is in a novel position not seen in other NHR structures. In this position LG-268 does not fold over and cap the ligand. Additional data indicate that LG-

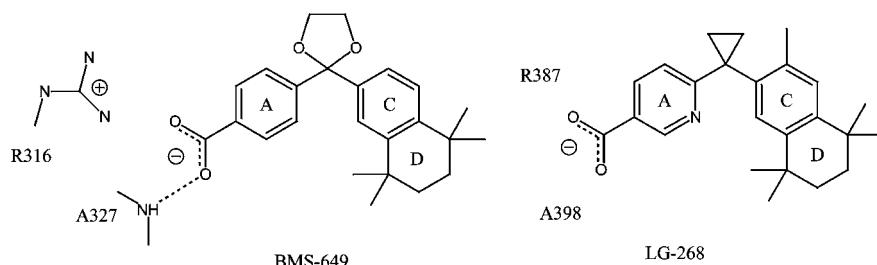


Fig. 1.28 Structures of RXR ligands.

268 is unable to release co-repressors from RXR unless co-activators are also present; this suggests that certain RXR ligands alone may be inefficient at repositioning helix-12 [34].

1.5

PPAR: Isotype-Selective Ligands

Peroxisome proliferator-activated receptors (PPARs), are members of the steroid/retinoid nuclear receptor family of ligand-activated transcription factors. There are three isotypes of the PPAR family, PPAR- α , PPAR- δ and PPAR- γ . Various natural fatty acids (FAs) and eicosanoids serve as ligands for the PPARs. Interestingly, several unsaturated FAs that activate the PPARs *in vitro* have pharmacological effects similar to those reported for the synthetic PPAR ligands. Based upon these observations and the promiscuous ligand-binding properties of the PPARs, it has been suggested that these receptors serve as physiological sensors of lipid levels, linking FA concentrations to glucose and lipid homeostasis [35]. The biological profiles of each have led to a number of ligated structures having been reported. This section will focus upon how ligands can select between these three isotypes. Unlike the RARs, there are several amino acid differences in the subtypes that directly contact the ligand. Thus, there are more options for obtaining isotype selective ligands and it may be more challenging to obtain ligands that show high affinity for all isotypes.

The structure of the FA eicosapentaenoic acid (EPA) bound to PPAR- δ has been determined (PDB entry: 3GWX) [35]. The ligand-binding pocket of PPAR- δ assumes roughly a “Y” shape with each of the three arms approximately 12 Å in length. One arm of the Y is composed of a mix of hydrophobic residues and polar residues and is capped by two residues from helix-12. This is the only arm of the Y that is substantially polar in character. Another arm of the Y is composed of hydrophobic residues and is sealed by the last residue of helix-1, and the loop between helices-1 and -2. The third arm of the Y-shaped pocket is composed mainly of hydrophobic residues with a few polar residues. The interior of the ligand-binding pocket is accessible via a channel that is exposed to solvent [35].

The EPA complex crystal has two molecules in the asymmetric unit; each independent molecule has two distinct conformations for the EPA ligand (Fig. 1.29). For EPA in each complex molecule, the acid group and the first eight carbons adopt very similar conformations. The hydrophobic tails of EPA diverge in the different bound conformations with each occupying a different arm of the Y. The carboxylate head group of EPA interacts with several polar residues in one arm of the Y. Surprisingly, the details of these interactions are different in each of the four copies of the EPA in the crystal. There are three polar residues that interact with the carboxylates, His-323 from helix-5, His-449 from helix-11 and Tyr-473 from helix-12. His-449 forms a buried hydrogen bond to the side chain of Lys-367. PPAR- δ , and the PPARs in general, use a helix-12 residue to make a specific hydrogen bond to agonist ligands, unlike many of the other NHRs.

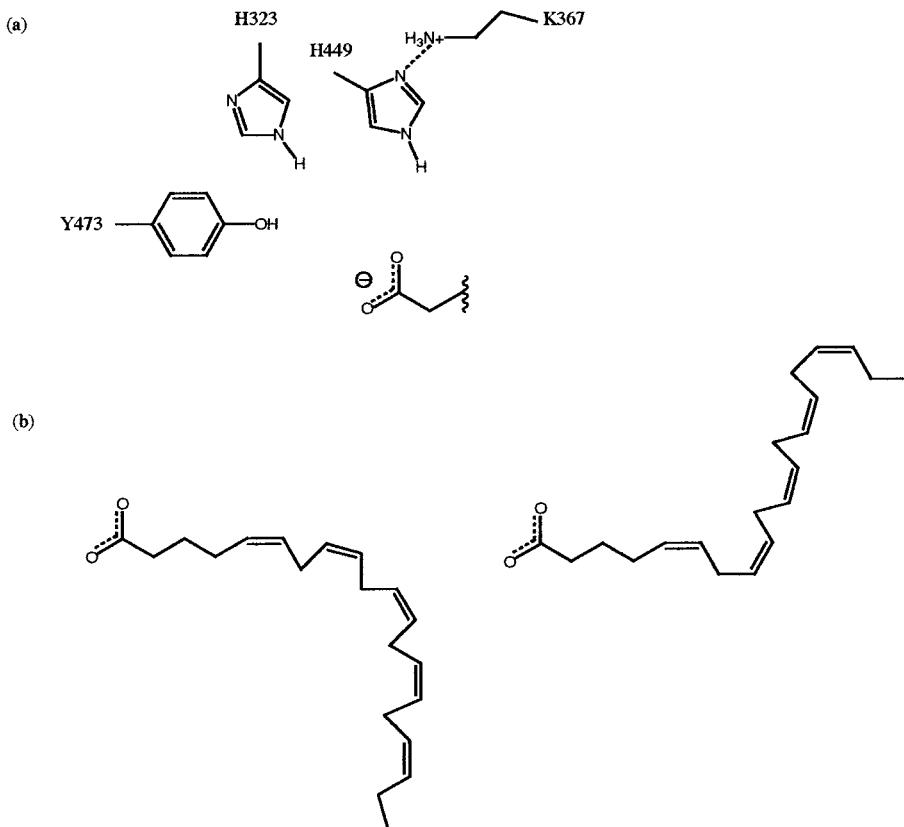


Fig. 1.29 Interactions of EPA with PPAR- δ . (a) The carboxylate head group of the ligand interacts with three protein residues, His-323, His-449 and Tyr-473. (b) Two distinct conformations of the ligand.

mations of the hydrophobic tail of the ligand were observed. The ligand-binding site of PPAR- δ is Y-shaped and each conformation of EPA fills a different branch of the Y.

The structure of the fibrate ligand GW2433, a PPAR- α/δ co-agonist, has also been reported bound to PPAR- δ (PDB entry: 1GWX; Fig. 1.30) [35]. In the structure, the carboxylate forms hydrogen bonds to two histidines (His-323 and His-449) and to Tyr-473 from helix-12. The rest of the molecule makes substantial VDW contacts with all three branches of the Y-shaped ligand-binding pocket. These are mostly hydrophobic interactions between ligand and protein; however, a hydrogen bond between the ligand urea NH and the side chain of Cys-285 is observed. While GW2433 is a PPAR- α/δ co-agonist, the structure of its complex with PPAR- α has not been disclosed. As will be discussed below, the residue corresponding to His-323 in PPAR- α is a tyrosine (Tyr-314).

AZ-242 is a PPAR- α/γ co-agonist and its structure has been determined in complex with both PPAR- α (PDB entry: 1I7G) and PPAR- γ (PDB entry: 1I7I) [36]. Like PPAR- δ , both PPAR- α and PPAR- γ have Y-shaped ligand binding sites; however, AZ-242 only fills two of the three sites. PPAR- α has four polar residues that interact with the carboxylates, Tyr-314 from helix-5, His-440 from helix-11, Tyr-464 from helix-12 and Ser-280 from helix-3. His-449 forms a buried hydrogen bond to the side chain of Lys-358. Tyr-314, His-440 and Tyr-464 are in the equivalent positions as His-323, His-449 and Tyr-473 in PPAR- δ . Thus, a major difference between PPAR- α and PPAR- δ is the change of a histidine in PPAR- δ to a tyrosine in PPAR- α . The dihydrocinnamate group of AZ-242 has a different local conformation than that of the phenoxy acid group of GW2433, as a result the carboxy

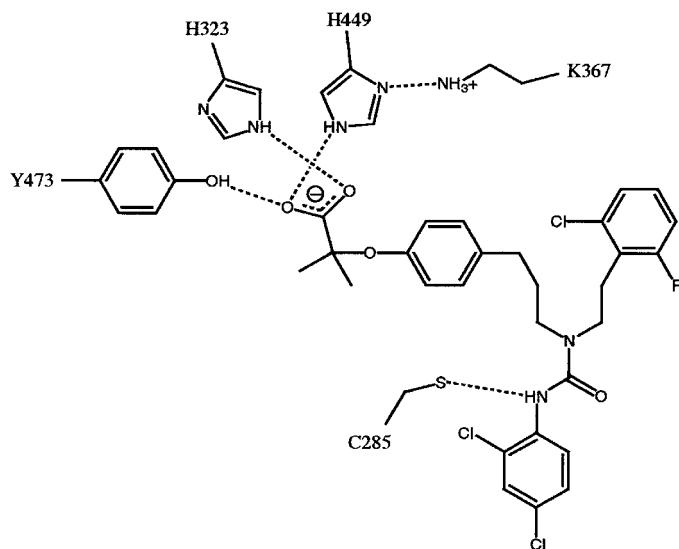


Fig. 1.30 Interactions of the PPAR- α/δ co-agonist with PPAR- δ . The Y-shape of the ligand is complementary to the Y-shape of the ligand binding pocket. One of the three arms of the

Y makes polar interactions with the carboxylate head group of the ligand, the interactions in the other two arms are mostly hydrophobic.

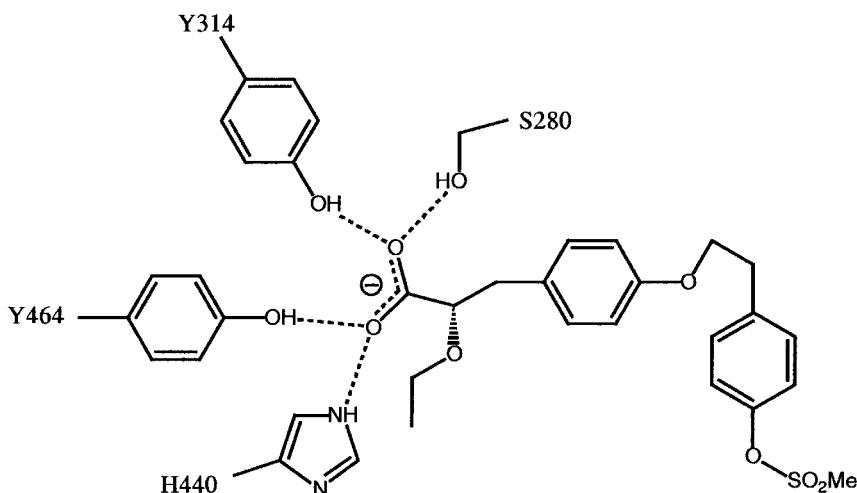


Fig. 1.31 Key polar interactions of PPAR- α/γ co-agonist AZ-242 with PPAR- α .

groups of both ligands interact with the protein in distinct manners. This may explain why AZ-242 is a PPAR- α agonist and not a PPAR- δ agonist (Fig. 1.31).

The structure of AZ-242 bound to PPAR- γ has also been reported. The conformation of the dihydrocinnamate group resembles the conformation found in the PPAR- α complex. In PPAR- γ , Tyr-314 in PPAR- α is replaced by His-323, thus the carboxylate interacts with two histidines (His-323 and His-449) and one tyrosine (Tyr-473). This arrangement of polar groups is the same as found in PPAR- δ ; however, there are subtle differences in the position of each residue relative to each other in the two isotypes. In addition, one of the carboxylate oxygens also forms a hydrogen bond with Ser-289. One difference between the two AZ-242 complexes is the conformation of the phenoxy “tail” of the ligand; this results in slight differences in the final position of the phenyl sulfonate group. One difference between PPAR- γ and PPAR- α is the presence of Cys-275 in PPAR- α , which is Gly-284 in PPAR- γ . The different tail conformations of the phenoxy linker allow the sulfonate group of AZ-242 to avoid the Cys-275 side chain in PPAR- α (Fig. 1.32).

The thiazolidinediones (TZDs) are a class of antidiabetic agents that were subsequently shown to be potent PPAR- γ agonists. TZDs are usually selective for PPAR- γ although exceptions are known [37, 38]. The structure of the TZD rosiglitazone in complex with PPAR- γ has been reported both in the presence (PDB entry: 1FM6) [3] and absence (PDB entry: 2PRG) [39] of RXR- α ligated with 9-cis-retinoic acid. Both complexes are quite similar and the discussion will center around the heterodimeric complex. Like AZ-242, rosiglitazone occupies two arms of the Y-shaped ligand-binding site. The polar TZD head forms hydrogen bonds with His-323 and Tyr-473 and also makes longer polar interactions with His-449 and Ser-289 (Fig. 1.33). Thus, the TZD group mimics a carboxylate group. The nonpolar part of rosiglitazone makes VDW contacts within the ligand-binding site. The

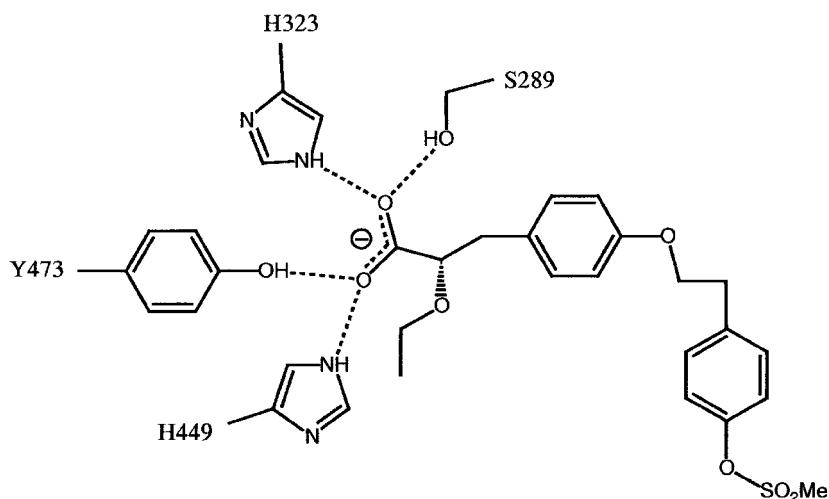


Fig. 1.32 Key polar interactions of PPAR- α/γ co-agonist AZ-242 with PPAR- γ . Notice that

the equivalent residue of PPAR- α Tyr-314 is His-323 in PPAR- γ .

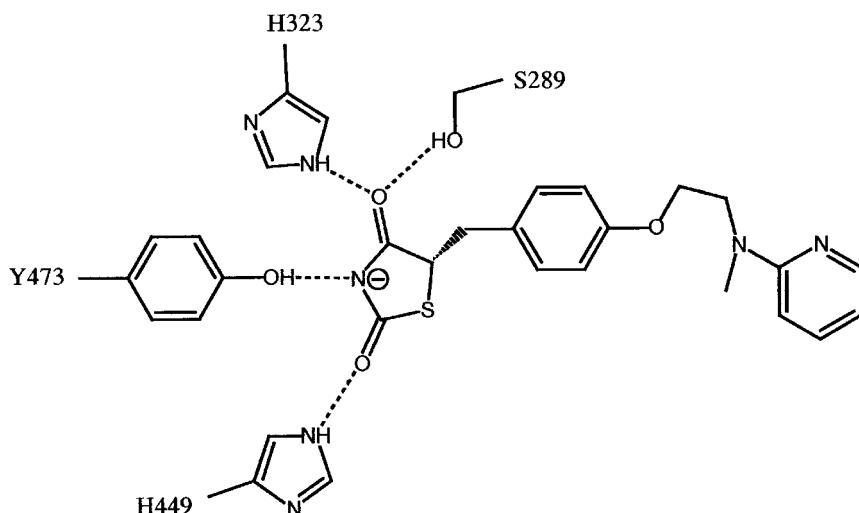


Fig. 1.33 Key polar interactions of PPAR- γ -selective agonist rosiglitazone with PPAR- γ . The TZD group acts as a carboxylate mimetic.

larger size of a TZD group compared with a carboxylate group usually allows it to be selective for PPAR- γ over the other isotypes. For PPAR- δ , the carboxylate binding site is narrower than in the other isotypes; thus, the larger TZD head group cannot interact with the polar groups on the PPAR- δ protein. The presence of the Tyr-314 precludes the TZD head group from making the proper polar contacts with the PPAR- α protein. Consistent with this proposal, the PPAR- γ H323Y mu-

tant binds rosiglitazone 50-fold weaker than wild type PPAR- γ . In addition, the PPAR- α Y314H mutant binds rosiglitazone weakly while the wild type PPAR- α does not bind it to any measurable extent [40]. Clearly, the steric demands of the carboxylate-binding pocket play a major role in determining isotype selectivity. It has been speculated that the shift in the TZD head group caused by Tyr-314 in PPAR- α results in an unfavorable conformation in the remainder of the TZD ligand. The TZD KRP-297 has been reported to show dual PPAR- α/γ activity [41]. Unlike rosiglitazone, KRP-297 has a *meta*-substituted side chain across the central phenyl rings. The *meta*-substituted side chain may allow an improved fit in the PPAR- α protein [40].

Farglitazar (GI262570) is a PPAR- γ/α co-agonist that shows 1000-fold selectivity for PPAR- γ over the α -isotype. The structure of its complex with PPAR- γ has been reported in the heterodimeric complex with ligated RXR- α (PDB entry: 1FM9) [3]. Farglitazar binds to PPAR- γ in a similar manner that AZ242 does; the dihydrocinnamate moieties bind nearly identically and, like AZ242, the carboxylate interacts with His-449, Tyr-473, His-323 and Ser-289 (Fig. 1.34). The tails of the two ligands bind by VDW contacts in the same pocket although the details of the binding are different. Thus, only two of the three branches of the Y-shaped pocket are occupied. The most dramatic difference between these ligands is the size of the substituent adjacent to the carboxylate. For AZ242 it is an O-ethyl group, and for farglitazar it is the much larger amino benzophenone group. The side chain of Phe-363 moves in the farglitazar structure relative to the AZ242 structure to open a pocket that accommodates the second ring of the benzophenone group.

A close analogue of farglitazar, GW409544, is also a co-agonist that binds only ten times poorer towards PPAR- α (Fig. 1.35). Neither farglitazar nor GW409544 bind to, or activate, PPAR- δ to any measurable extent. The structures of

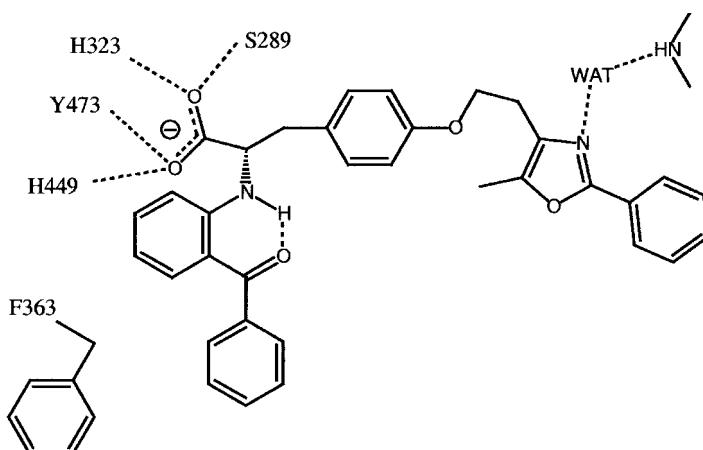


Fig. 1.34 Key polar interactions of PPAR- γ -selective agonist farglitazar with PPAR- γ . The larger benzophenone group fits into a pocket

created by the movement of Phe-363 relative to its position in the rosiglitazone complex.

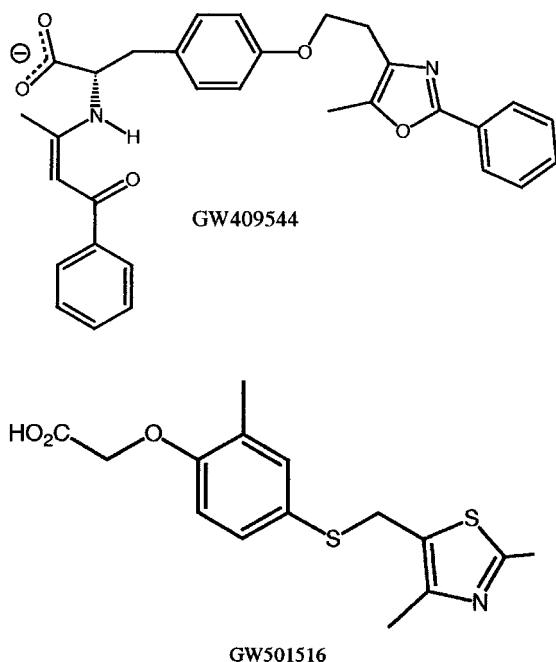


Fig. 1.35 Structure of the PPAR- α/γ co-agonist GW409544 and the PPAR- δ -selective agonist GW501516.

GW409544 bound to both PPAR- γ (PDB entry: 1K74) and to PPAR- α (PDB entry: 1K7L) have been reported [40]. GW409544 binds to PPAR- γ in a manner very similar to, but not identical to, the way farglitazar binds to PPAR- γ . The α -amino side chains bind differently in the two structures. GW409544 binds to PPAR- α and PPAR- γ in nearly identical manners despite slight changes in key protein residues contacting the ligand. The equivalent residue to PPAR- γ Phe-363 is Ile-354 in PPAR- α , and as previously discussed PPAR- γ His-323 is substituted by the larger Tyr-314 in PPAR- α . A modeling exercise proposed that for farglitazar to interact with the larger Tyr-314 in PPAR- α a shift in the ligand position would result in a steric clash between a benzophenone ring and Phe-273. The structure of the complex between GW409544 and PPAR- α suggests that adding three atoms to the vinyllogous amide of GW409544 to generate farglitazar would result in a steric clash with Phe-273. Consistent with this proposal, the PPAR- α Y314H mutant potently binds farglitazar [40].

A very interesting discussion about the key factors that result in PPAR subtype specificity has been published by the Glaxo group [40]. This discussion is now summarized below. The PPAR- α and PPAR- γ ligand-binding pockets are significantly larger than the PPAR- δ pocket because of the narrowing of the pocket adjacent to helix-12. It is notable that only a handful of potent PPAR- δ ligands have been described. Ligands such as TZDs and L-tyrosine-based agonists like farglitazar show little or no binding to PPAR- δ . In both cases, their acidic head groups seem to be too large to fit within the narrow PPAR- δ pocket. In contrast, the po-

tent PPAR- δ agonist GW501516 (Fig. 1.36) contains an unsubstituted phenoxy-acetic acid head group that complements the narrow PPAR- δ ligand-binding pocket. Fibrate ligands, which generally bind to PPAR- δ only at high micromolar concentrations, contain small alkyl substituents adjacent to the carboxylate group. A PPAR- δ mutant M417V, which allows fibrate ligands to bind to PPAR- δ , has been reported [42]. This mutation is likely to increase the size of the PPAR- δ pocket to facilitate the binding of the small alkyl substituents adjacent to the carboxylate. Thus, the reduced size of the PPAR- δ pocket is a major determinant of ligand binding to this subtype. In comparison with PPAR- δ , the PPAR- α and PPAR- γ ligand-binding pockets are closer in size and shape to each other. The Glaxo group has found that a major determinant of selectivity between these two subtypes is the substitution of Tyr-314 in PPAR- α for His-323 in PPAR- γ . These amino acids form part of the network of hydrogen-bonding residues that are involved in the activation of the receptor by its acidic ligands. Overlay of the PPAR- α and PPAR- γ crystal structures reveals that the larger volume of the Tyr-314 side chain in PPAR- α forces a 1.5 Å shift in the position of the high-affinity ligand GW409544. The structurally related ligand farglitazar is unable to accommodate this shift because of a steric interaction with PPAR- α Phe-273. As a result, farglitazar shows 1000-fold selectivity for PPAR- γ over PPAR- α . The point mutations of Y314H in PPAR- α and H323Y in PPAR- γ demonstrate that these single amino acids are, in large part, responsible for determining the subtype selectivity of farglitazar. In each case, a 10–100-fold shift in the potency of the ligand was observed. Compared with farglitazar, GW409544 has three atoms removed to allow it to shift within the PPAR- α pocket without clashing with Phe-273. The potent dual PPAR- α/γ agonist activity of GW409544 results from a complementary match of the re-engineered ligand with both the PPAR- α and PPAR- γ ligand-binding pockets. TZD ligands also respond to the point mutation of Y314H in PPAR- α and H323Y in PPAR- γ with a corresponding increase in PPAR- α and decrease in PPAR- γ activity, respectively. These data suggest that the TZDs, which do not contain the large N-substituents present in the L-tyrosine-based ligands, also have difficulty accommodating the 1.5 Å shift required to bind to PPAR- α . It was speculated that the shift in the TZD head group results in an unfavorable conformation in the remainder of the molecule. It is interesting to note that a single amino acid difference in PPARs has such a dramatic impact on ligand selectivity, given that the PPAR pocket is composed of more than 25 amino acids [40].

1.6 Summary

NHRs function as either activators or repressors of gene expression. Small molecule ligands (hormones) are the regulators of these proteins. Structural studies have characterized at least two distinct forms of ligated NHRs, agonist-bound NHR with a peptide derived from a co-activator and antagonist-bound NHR with a peptide derived from a co-repressor. These define two states for the NHR, the

first one is the “on” state where genes are expressed and the second is the “off” state where gene expression is repressed. There are other structures that may, or may not, represent intermediate states between “on” and “off”.

The discussions above, while not comprehensive, have focused on how ligands can bind to and alter the conformational state of the NHR. More importantly, the discussion has focused on the important issue of ligand selectivity for a specific NHR. A system that is regulated by ligands is only effective if the proteins that are being regulated respond only to a specific ligand, or to just a few key ligands. Thus, the ability of estradiol to regulate ER- α and ER- β and not to regulate other related proteins such as GR and AR is the key to this regulatory system. The key trends about how ligands selectively bind to and agonize their NHR target protein will now be summarized.

A striking observation for many of these agonist NHR complexes is the lack of high steric complementarity between the mostly hydrophobic ligand and the ligand-binding core of the protein. The complex between estradiol and ER- α shows that there are large unoccupied cavities between the ligand and the protein. In spite of this, estradiol is a potent and selective ligand for ER- α and the closely related ER- β .

Another good example is found in the complexes between PPAR- γ and rosiglitazone (Fig. 1.33) and the larger ligand farglitazar (Fig. 1.34). Both are potent selective ligands; however, the larger ligand is 90 times more potent. While the polar thiazolidinedione moiety of rosiglitazone plays a key role for its isotype selectivity for PPAR- γ , the larger benzophenone moiety of farglitazar is required for PPAR- γ selectivity. Thus, for farglitazar a precise relative fit of the polar carboxylate functionality and one of the apolar groups work together to achieve PPAR- γ selectivity. As discussed previously, the Glaxo group has shown that changes to the benzophenone moiety of farglitazar produce ligands that agonize both PPAR- γ and PPAR- α . They have also determined that a single polar amino acid change between PPAR- γ (His-323) and PPAR- α (Tyr-314, equivalent to PPAR- γ His-323) is the primary determinant of PPAR- γ selectivity for both rosiglitazone and farglitazar. The conclusion reached is that the PPAR isotype selectivity is determined by the polar ligand group making a precise interaction with the protein, with the remainder of the ligand being well tolerated by the protein.

Many steroid hormones are mostly hydrophobic molecules that have polar groups at opposite ends of the molecule (A- and D-rings). For these steroid receptors a key polar residue on a common position of helix-3 helps the estrogen receptors (Glu-353 in ER- α) select for A-ring phenols, while PR, AR, GR and MR, which have a glutamine at this position, are selective for 3-keto A-rings (Figs. 1.4 and 1.12). Further ligand discrimination occurs primarily by additional polar interactions near the D-ring (Figs. 1.12, 1.13, 1.14 and 1.15). For the steroid receptors, agonist selectivity appears to be determined by matching of polar groups on the ends of the ligand with complementary protein polar groups. The hydrophobicity of the ligands and the apolar nature of the ligand binding site provides a mechanism for obtaining high affinity, provided there are no polar mismatches, in spite of less than optimal steric complementarity.

Another good example of this is the case of two enantiomers (BMS270394 and BMS270394, Fig. 1.23) of the same ligand binding to RAR- γ . Both ligands orient the carboxylate, hydroxyl group, the A-, C- and D-rings in approximately the same manner. The linking amide group of the more potent isomer makes complementary interactions with the protein (Fig. 1.24) while the less potent isomer does not (Fig. 1.25). Thus we see a vivid example of where effective agonist binding occurs when a hydrophobic ligand makes a less than optimal steric fit in the core of the NHR and has the matching of polar ligand groups with complementary protein polar groups. The “inactive” isomer must bury the polar amide group in the protein interior and thus bind very weakly.

Comparison of RAR and RXR shows distinctly different shapes of their ligand-binding pockets. Thus, RAR can effectively bind either all-*trans*-retinoic acid or 9-*cis*-retinoic acid as agonists, while RXR can only accommodate the bent ligand 9-*cis*-retinoic acid. These data make it tempting to speculate that selectivity is determined not by a precise match of the “right” ligand with the “right” NHR, but rather by the exclusion of “wrong” ligands by a mismatch of polar groups or an impossible steric fit.

To conclude this chapter, the molecular recognition of ligands by NHRs is determined by a number of subtle factors. Agonism or antagonism is due to the ability of a given ligand to bind to the appropriate form of the protein. It is likely that nonpolar interactions provide binding affinity even in the absence of good shape complementarity. Selectivity appears to be guided by the avoidance of interactions, either polar or apolar, that are unfavorable.

1.7

Acknowledgements

I would like to thank Dr. Michael J. Rynkiewicz for careful review of this document and for several suggestions that helped shape the final version.

1.8

References

- 1** OLEFSKY, J. M. Nuclear receptor minireview series. *J. Biol. Chem.* **2001**, *276*, 36863–36864.
- 2** ROSENFIELD, M. G., GLASS, C. K. Coregulator codes of transcriptional regulation by nuclear receptors. *J. Biol. Chem.* **2001**, *276*, 36865–36868.
- 3** GAMPE, R. T., JR., MONTANA, V. G., LAMBERT, M. H., MILLER, A. B., BLEDSOE, R. K., MILBURN, M. V., KLIEWER, S. A., WILLSON, T. M., XU, H. E. Asymmetry in the PPARgamma/RXRalpha crystal structure reveals the molecular basis of heterodimerization among nuclear receptors. *Mol. Cell* **2000**, *5*, 545–555.
- 4** XU, H. E., STANLEY, T. B., MONTANA, V. G., LAMBERT, M. H., SHEARER, B. G., COBB, J. E., MCKEE, D. D., GALARDI, C. M., PLUNKET, K. D., NOLTE, R. T., PARKS, D. J., MOORE, J. T., KLIEWER, S. A., WILLSON, T. M., STIMMEL, J. B. Structural basis for antagonist-mediated recruitment of nu-

- clear co-repressors by PPARalpha. *Nature* **2002**, *415*, 813–817.
- 5 EVANS, R. M. The Steroid and Thyroid Hormone Receptor Superfamily. *Science* **1988**, *240*, 889–895.
- 6 BRZOZOWSKI, A., PIKE, A., DAUTER, Z., HUBBARD, R., BONN, T., ENGSTROM, O., OHMAN, L., GREENE, G., GUSTAFSSON, J., CARLQUIST, M. Molecular basis of agonism and antagonism in the oestrogen receptor. *Nature* **1997**, *389*, 753–758.
- 7 TANENBAUM, D. M., WANG, Y., WILLIAMS, S. P., SIGLER, P. B. Crystallographic comparison of the estrogen and progesterone receptor's ligand binding domains. *PNAS* **1998**, *95*, 5998–6003.
- 8 GANGLOFF, M., RUFF, M., EILER, S., DUCLAUD, S., WURTZ, J. M., MORAS, D. Crystal structure of a mutant hERalpha ligand-binding domain reveals key structural features for the mechanism of partial agonism. *J. Biol. Chem.* **2001**, *276*, 15059–15065.
- 9 SHIAU, A. K., BARSTAD, D., LORIA, P. M., CHENG, L., KUSHNER, P. J., AGARD, D. A., GREENE, G. L. The structural basis of estrogen receptor/coactivator recognition and the antagonism of this interaction by tamoxifen. *Cell* **1998**, *95*, 927–937.
- 10 PIKE, A. C., BRZOZOWSKI, A. M., WALTON, J., HUBBARD, R. E., THORSELL, A. G., LI, Y. L., GUSTAFSSON, J. A., CARLQUIST, M. Structural insights into the mode of action of a pure antiestrogen. *Structure (Camb.)* **2001**, *9*, 145–153.
- 11 PIKE, A. C. W., BRZOZOWSKI, A. M., HUBBARD, R. E., BONN, T., THORSELL, A.-G., ENGSTROM, O., LJUNGGREN, J., GUSTAFSSON, J.-A., CARLQUIST, M. Structure of the ligand-binding domain of oestrogen receptor beta in the presence of a partial agonist and a full antagonist. *EMBO J.* **1999**, *18*, 4608–4618.
- 12 SHIAU, A. K., BARSTAD, D., RADEK, J. T., MEYERS, M. J., NETTLES, K. W., KATZENELLENBOGEN, B. S., KATZENELLENBOGEN, J. A., AGARD, D. A., GREENE, G. L. Structural characterization of a subtype-selective ligand reveals a novel mode of estrogen receptor antagonism. *Nat. Struct. Biol.* **2002**, *9*, 359–364.
- 13 KATZENELLENBOGEN, B. S., KATZENELLENBOGEN, J. A. Biomedicine: Enhanced: Defining the “S” in SERMs. *Science* **2002**, *295*, 2380–2381.
- 14 WILLIAMS, S., SIGLER, P. Atomic structure of progesterone complexed with its receptor. *Nature* **1998**, *393*, 392–396.
- 15 WURTZ, J. M., BOURGUET, W., RENAUD, J. P., VIVAT, V., CHAMBON, P., MORAS, D., GRONEMAYER, H. A canonical structure for the ligand-binding domain of nuclear receptors. *Nat. Struct. Biol.* **1996**, *3*, 87–94.
- 16 SACK, J. S., KISH, K. F., WANG, C., ATTAR, R. M., KIEFER, S. E., AN, Y., WU, G. Y., SCHEFFLER, J. E., SALVATI, M. E., KRYSSTEK, S. R., JR., WEINMANN, R., EINSPAHR, H. M. Crystallographic structures of the ligand-binding domains of the androgen receptor and its T877A mutant complexed with the natural agonist dihydrotestosterone. *PNAS* **2001**, *98*, 4904–4909.
- 17 BLEDSOE, R. K., MONTANA, V. G., STANLEY, T. B., DELVES, C. J., APOLITO, C. J., MCKEE, D. D., CONSLER, T. G., PARKS, D. J., STEWART, E. L., WILLSON, T. M., LAMBERT, M. H., MOORE, J. T., PEARCE, K. H., XU, H. E. Crystal structure of the glucocorticoid receptor ligand binding domain reveals a novel mode of receptor dimerization and coactivator recognition. *Cell* **2002**, *110*, 93–105.
- 18 MATIAS, P. M., DONNER, P., COELHO, R., THOMAZ, M., PEIXOTO, C., MACEDO, S., OTTO, N., JOSCHKO, S., SCHOLZ, P., WEGG, A., BASLER, S., SCHAFER, M., EGNER, U., CARRONDO, M. A. Structural evidence for ligand specificity in the binding domain of the human androgen receptor. Implications for pathogenic gene mutations. *J. Biol. Chem.* **2000**, *275*, 26164–26171.
- 19 ROCHEL, N., WURTZ, J. M., MITSCHLER, A., KLAHOZ, B., MORAS, D. The crystal structure of the nuclear receptor for vitamin D bound to its natural ligand. *Mol. Cell* **2000**, *5*, 173–179.
- 20 ROCHEL, N., TOCCINI-VALENTINI, G., EGEA, P. F., JUNTUNEN, K., GARNIER, J.-M., VIJKO, P., MORAS, D. Functional and structural characterization of the insertion region in the ligand binding domain of the vitamin D nuclear receptor. *Eur. J. Biochem.* **2001**, *268*, 971–979.

- 21** TOCCHINI-VALENTINI, G., ROCHEL, N., WURTZ, J. M., MITSCHLER, A., MORAS, D. Crystal structures of the vitamin D receptor complexed to superagonist 20-epi ligands. *PNAS* **2001**, *98*, 5491–5496.
- 22** SWANN, S. L., BERGH, J. J., FARACH-CARSON, M. C., KOH, J. T. Rational design of vitamin D3 analogues which selectively restore activity to a vitamin D receptor mutant associated with rickets. *Org. Lett* **2002**, *4*, 3863–3866.
- 23** SWANN, S. L., BERGH, J., FARACH-CARSON, M. C., OCASIO, C. A., KOH, J. T. Structure-based design of selective agonists for a rickets-associated mutant of the vitamin D receptor. *J. Am. Chem. Soc.* **2002**, *124*, 13795–13805.
- 24** EGEA, P. F., MITSCHLER, A., ROCHEL, N., RUFF, M., CHAMBON, P., MORAS, D. Crystal structure of the human RXR{alpha} ligand-binding domain bound to its natural ligand: 9-cis retinoic acid. *EMBO J.* **2000**, *19*, 2592–2601.
- 25** RENAUD, J. P., ROCHEL, N., RUFF, M., VIVAT, V., CHAMBON, P., GRONEMEYER, H., MORAS, D. Crystal structure of the RAR-gamma ligand-binding domain bound to all-trans retinoic acid. *Nature* **1995**, *378*, 681–689.
- 26** WAGNER, R. L., APRILETTI, J. W., MCGRATH, M. E., WEST, B. L., BAXTER, J. D., FLETTERICK, R. J. A structural role for hormone in the thyroid hormone receptor. *Nature* **1995**, *378*, 690–697.
- 27** KLAHOZ, B. P., RENAUD, J. P., MITSCHLER, A., ZUSI, C., CHAMBON, P., GRONEMEYER, H., MORAS, D. Conformational adaptation of agonists to the human nuclear receptor RAR gamma. *Nat. Struct. Biol.* **1998**, *5*, 199–202.
- 28** KLAHOZ, B. P., MITSCHLER, A., BELEMA, M., ZUSI, C., MORAS, D. Enantiomer discrimination illustrated by high-resolution crystal structures of the human nuclear receptor hRARgamma. *PNAS* **2000**, *97*, 6322–6327.
- 29** BABINE, R. E. Unpublished, 2002.
- 30** KLAHOZ, B. P., MITSCHLER, A., MORAS, D. Structural basis for isotype selectivity of the human retinoic acid nuclear receptor. *J. Mol. Biol.* **2000**, *302*, 155–170.
- 31** YU, K. L., SPINAZZE, P., OSTROWSKI, J., CURRIER, S. J., PACK, E. J., HAMMER, L., ROALSVIG, T., HONEYMAN, J. A., TORTOLANI, D. R., RECKEK, P. R., MANSURI, M. M., STARRETT, J. E., Jr. Retinoic acid receptor beta,gamma-selective ligands: synthesis and biological activity of 6-substituted 2-naphthoic acid retinoids. *J. Med. Chem.* **1996**, *39*, 2411–2421.
- 32** KLAHOZ, B., MORAS, D. C.-H.-O Hydrogen bonds in the nuclear receptor RAR-gamma – a potential tool for drug selectivity. *Structure (Camb.)* **2002**, *10*, 1197.
- 33** EGEA, P. F., MITSCHLER, A., MORAS, D. Molecular recognition of agonist ligands by RXRs. *Mol. Endocrinol.* **2002**, *16*, 987–997.
- 34** LOVE, J. D., GOOCH, J. T., BENKO, S., LI, C., NAGY, L., CHATTERJEE, V. K., EVANS, R. M., SCHWABE, J. W. The structural basis for the specificity of retinoid-X receptor-selective agonists: new insights into the role of helix H12. *J. Biol. Chem.* **2002**, *277*, 11385–11391.
- 35** XU, H. E., LAMBERT, M. H., MONTANA, V. G., PARKS, D. J., BLANCHARD, S. G., BROWN, P. J., STERNBACH, D. D., LEHMANN, J. M., WISELY, G. B., WILLSON, T. M., KIEWER, S. A., MILBURN, M. V. Molecular recognition of fatty acids by peroxisome proliferator-activated receptors. *Mol. Cell* **1999**, *3*, 397–403.
- 36** CRONET, P., PETERSEN, J. F., FOLMER, R., BLOMBERG, N., SJÖBLOM, K., KARLSSON, U., LINDSTEDT, E. L., BAMBERG, K. Structure of the PPARalpha and -gamma ligand binding domain in complex with AZ 242; ligand selectivity and agonist activation in the PPAR family. *Structure (Camb.)* **2001**, *9*, 699–706.
- 37** WILLSON, T. M., BROWN, P. J., STERNBACH, D. D., HENKE, B. R. The PPARs: from orphan receptors to drug discovery. *J. Med. Chem.* **2000**, *43*, 527–550.
- 38** BROOKS, D. A., ETGEN, G. J., RITO, C. J., SHUKER, A. J., DOMINIANNI, S. J., WARSHAWSKY, A. M., ARDECKY, R., PATERNITI, J. R., TYHONAS, J., KARANEWSKY, D. S., KAUFFMAN, R. F., BRODERICK, C. L., OLDFHAM, B. A., MONTROSE-RAFIZADEH, C., WINNERSKSI, L. L., FAUL, M. M., McCARTHY, J. R. Design and synthesis of 2-methyl-2-[4-(2-[5-methyl-2-aryloxazol-4-yl]ethoxy)phenoxy]propionic acids: a new

- class of dual PPARalpha/gamma agonists. *J. Med. Chem.* **2001**, *44*, 2061–2064.
- 39** NOLTE, R.T., WISELY, G.B., WESTIN, S., COBB, J.E., LAMBERT, M.H., KUROKAWA, R., ROSENFELD, M.G., WILLSON, T.M., GLASS, C.K., MILBURN, M.V. Ligand binding and co-activator assembly of the peroxisome proliferator-activated receptor-gamma. *Nature* **1998**, *395*, 137–143.
- 40** XU, H.E., LAMBERT, M.H., MONTANA, V.G., PLUNKET, K.D., MOORE, L.B., COLLINS, J.L., OPLINGER, J.A., KLEWER, S.A., GAMPE, R.T., JR., MCKEE, D.D., MOORE, J.T., WILLSON, T.M. Structural determinants of ligand binding selectivity between the peroxisome proliferator-activated receptors. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 13919–13924.
- 41** MURAKAMI, K., TOBE, K., IDE, T., MOCHIZUKI, T., OHASHI, M., AKANUMA, Y., YAZAKI, Y., KADOWAKI, T. A novel insulin sensitizer acts as a coligand for peroxisome proliferator-activated receptor-alpha (PPAR-alpha) and PPAR-gamma: effect of PPAR-alpha activation on abnormal lipid metabolism in liver of Zucker fatty rats. *Diabetes* **1998**, *47*, 1841–1847.
- 42** TAKADA, I., YU, R.T., XU, H.E., LAMBERT, M.H., MONTANA, V.G., KLEWER, S.A., EVANS, R.M., UMESONO, K. Alteration of a single amino acid in peroxisome proliferator-activated receptor- $\{\alpha\}$ (PPAR $\{\alpha\}$) generates a PPAR $\{\delta\}$ phenotype. *Mol. Endocrinol.* **2000**, *14*, 733–740.

2

Kinases

JERRY L. ADAMS, JAMES VEAL, and LISA SHEWCHUK

2.1

Introduction

The regulation of glycogen phosphorylase function by reversible phosphorylation was first described in the mid-1950s. However, the potential importance of protein kinases was not fully appreciated until the advent of DNA cloning and sequencing in the mid-1970s. The initial estimate of 1000–2000 human protein kinases made in 1987 has now been refined using the full human genome to 518 putative protein kinase genes or approximately 1.7% of all human genes [1]. With nearly a third of all proteins containing at least one covalently bound phosphate, it is easy to envisage an important regulatory role for the enzymes that attach (kinases) and remove (phosphatases) inorganic phosphate from proteins. Because kinases are involved in the regulation of all aspects of cellular function, protein kinase inhibition may be a widely applicable strategy for the treatment of disease. Protein kinases now figure prominently among the molecular targets currently being pursued by the pharmaceutical industry to treat a wide spectrum of diseases, these include cancer, diabetes and its complications, rheumatoid arthritis, Alzheimer's, hypertension and stroke.

The realization of the therapeutic promise of protein kinase inhibitors has proven to be a daunting task. Two major obstacles to be overcome were achieving access to these intracellular targets and selectivity of inhibition. With over 500 kinases in the human genome, it is not surprising that selectivity has proven to be the more difficult of the two problems. However, as will be discussed in this review, X-ray crystallography has proven to be the central technology driving our understanding of the structural features governing inhibitor binding and providing direction in the search for selective kinase inhibitors.

2.2

Structure and Function

2.2.1

Tertiary Structure

A great level of structural knowledge has accumulated for protein kinases over the past decade (for various reviews see [2–9]). The minimal kinase domain consists of approximately 300 amino acids: for example, Cyclin-Dependent Kinase 2 (CDK2) has 298 residues (Fig. 2.1) [10, 11]. The kinase domain can be further divided into two lobes or sub-domains, N-terminal and C-terminal, with the lobes connected via a five to six residue “hinge” or “linker” region. The N-terminal lobe is the smaller subunit and consists of approximately 80 amino acids. It is primarily beta sheet in nature but does contain a significant alpha helix segment, termed the C-helix (nomenclature derives largely from the initial Cyclic AMP-Dependent Kinase, PKA, structure [9, 12]). By contrast the C-terminal lobe, in addition to being substantially larger, is largely α -helical in nature.

A binding cleft is formed between the two lobes, and ATP docks into this region. Hydrogen bonds are formed between the N6 exocyclic amino group of ATP and a backbone carbonyl oxygen of the hinge region noted above, as well as between the N1 of ATP and a backbone NH, again from the hinge region. Note

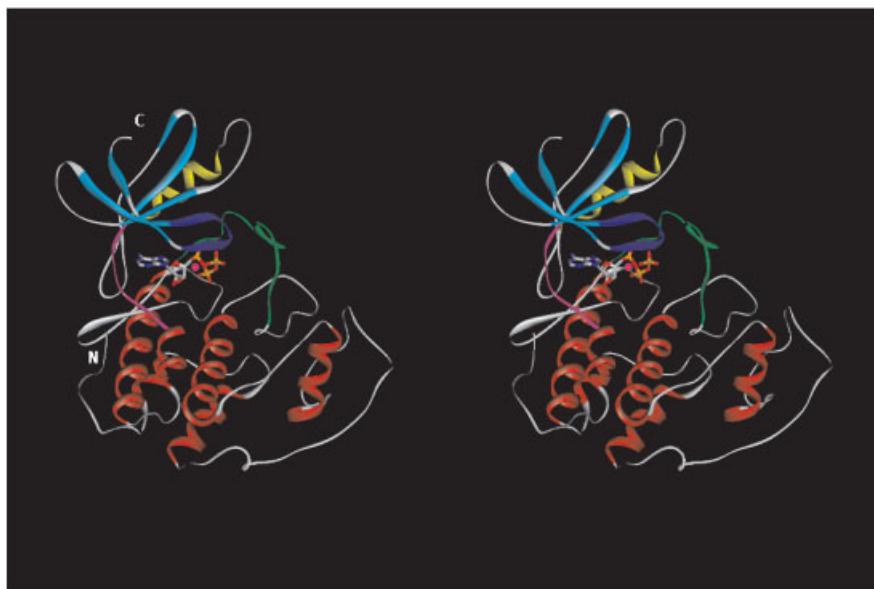


Fig. 2.1 Stereo ribbon diagram of human CDK2 with bound ATP (1HCK.PDB). Structural elements are colored as follows, glycine-rich loop in dark blue, N-terminal beta sheet

in cyan, the C-helix in yellow, linker or hinge region in magenta, activation loop in green and C-terminal lobe helices in red.

that, in addition to these two groups, the hinge region also projects a second backbone carbonyl oxygen into the binding site in the region of adenine H2. This functionality is significant for synthetic ligand-hydrogen bonding interactions as discussed later. Beyond hydrogen bonds to the adenine ring, the ribose hydroxyls of ATP are often observed to show hydrogen bonds to a polar residue (Asp 86 of CDK2 [10, 11], Asp 1083 of insulin receptor kinase [13], and equivalent residues) that resides at the beginning of the C-terminal lobe as defined. The adenine and ribose rings are also sandwiched by a series of hydrophobic residues from both the N- and C-terminal lobes. In particular for CDK2, residues Val-18, Ala-31, Val-64, and Leu-134 have van der Waals contacts with ATP as do equivalent residues of other kinases. An additional residue, Phe-80 for CDK2, is also significant and contacts ATP in the vicinity of C5 and the N6 exocyclic amino group and has been termed the gatekeeper. This residue shows high variability across kinases. As brief examples, it is Met for insulin receptor kinase (IRK), Thr for p38, Gln for ERK2, and Val for VEGFR2 [14–19]. The gatekeeper nomenclature derives from the fact that the residue straddles both the ATP binding site and an additional region deep within the kinase that small molecule inhibitors often utilize (see Sections 2.2.3.1 and 2.4.2.2). The final component of ATP, the triphosphate group, projects as would be expected into a highly hydrophilic and open region framed by a set of strictly conserved residues involved in catalysis. It should be noted at this point that although the small molecule inhibitors designed for protein kinases are predominantly ATP competitive, they commonly occupy regions of substantial volume in addition to where ATP would normally bind. As a result, the “ligand” binding site, when referring to inhibitors, is not synonymous with the ATP binding site.

2.2.2

Catalysis and Substrate Binding

Crystallographic studies, in particular with non-hydrolyzable ATP and transition state analogues as well as substrate analogues, have provided a detailed understanding of the likely nature and mechanism of substrate binding and catalysis [13, 20–24]. As noted, the triphosphate group interacts with conserved residues of the protein kinase. More specifically, the α and β phosphates are oriented via a salt bridge with a lysine (Lys-33 of CDK2) and the lysine itself is further fixed in an appropriate location for catalysis by a conserved glutamic acid residue (Glu-51) residing on the C-helix. Asp-145 and Asn-132 of CDK2 and equivalent residues complete a coordination sphere between the phosphate groups and accompanying magnesium counterions.

Peptide substrate then docks onto the protein kinase, in general presumably occupying a cleft along the C-terminal lobe, as exemplified by the peptide inhibitor in the PKA-AMPPNP-PKI and IRK-ATP structures. Catalysis appears to be via a dissociative transition state mechanism and a planar phosphate intermediate [20, 22]. The incoming peptide hydroxyl is oriented via Asp-127, which is in turn further stabilized via a hydrogen bond to Asn-132. These latter two residues that

are part of the signature-catalytic loop DxxxxN motif that designates proteins as potential protein kinases.

2.2.3

Regulation and Conformational Flexibility

Protein kinases are highly regulated owing to their central role in signal transduction [3, 5, 25]. While a full description of regulatory mechanisms is outside the remit of this review, it is nonetheless important to consider in the context of how it affects ligand binding. More specifically, regulatory control of protein kinases often involves significant conformational changes in structure that alter the shape of the binding site afforded to a small molecule inhibitor. Four mechanisms of regulation and their conformational effects will be mentioned: activation loop conformation, glycine-rich loop conformation, C-helix position and lobe orientation.

2.2.3.1 Activation Loop Conformation

The original unliganded IRK structure [14] revealed a protein in which the ATP binding site was occluded by residues from the so called “activation loop” or “T-loop”. This C-terminal extended region is present in all protein kinase structures to date and, as defined here, begins with a well conserved Asp-Phe-Gly sequence motif (DFG starting with Asp-145 of CDK2 and Asp-1150 of IRK) and extends until just upstream of a consensus Ala/Thr-Pro-Asp/Glu triplet. The loop contains one or more phosphorylation sites, i.e., it is a substrate for another protein kinase or autophosphorylation via protein dimerization. It is useful to consider the activation loop as having a range of conformations, in equilibrium, with two relative extremes. One end of the equilibrium is as seen in the IRK-AMPPNP structure [13] and represents the catalytically active conformation. Specifically, the aspartic acid of the DFG motif is strictly conserved and involved in catalysis (Section 2.2.2), and so must be properly positioned to coordinate magnesium ions. The adjacent phenylalanine occupies a position adjacent to the C-helix and packs against a series of hydrophobic residues from other regions of the protein. The ATP site is unhindered by protein such that ATP can freely dock and form suitable hydrogen bonds to the linker strand with the γ phosphate fully accessible. By contrast, at the other end of the equilibrium, a conformation is observed as represented by the initial IRK structure. The phenylalanine of the DFG motif is translocated by approximately 10 Å to roughly a position where, in an ATP -iganded structure, the ribose would reside. Additionally, the aspartic acid residue conformation is altered via rotation of the peptide backbone ζ angle.

Phosphorylation of the activation loop appears to drive the equilibrium towards a catalytically active conformation due to the ability of the phosphorylated residue to form stabilizing salt bridges with a series of lysine or arginine residues. However, it should be noted that, based on the evidence to date, the activation loop can occupy a range of conformations including one close to an active conforma-

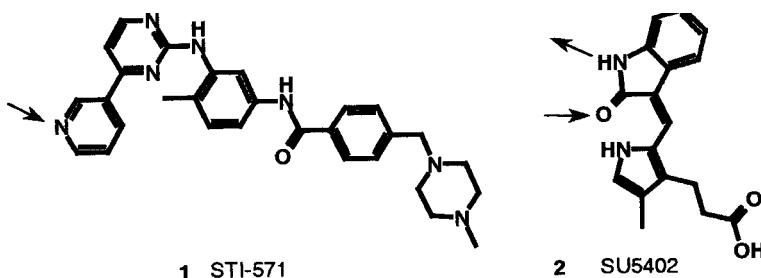
tion in the absence of phosphorylation and an inactive one despite phosphorylation. In short, it is a true equilibrium with energetics and relative populations of different forms varying from kinase to kinase depending on a range of factors including phosphorylation, additional protein and substrate binding, and internal energy of the conformation. The significance for ligand binding, particularly for small molecule inhibitors, is that the nature of the ATP pocket and adjacent regions is significantly altered. For example, STI-571 **1** (Gleevec, Scheme 2.1), as described in more detail later, binds solely to an inactive type conformation [26, 27] with a portion of the ligand residing where the phenylalanine of the Abl DFG triad would reside in an active conformation. Inactive conformations have now been observed for a range of both serine/threonine and tyrosine kinases including Abl, IRK, p38, Tie2, and Twitchin kinase [14, 26–29] indicating that the phenomena is likely widespread across kinases, and so again emphasizing the importance to ligand design. One final point to note is that the region immediately upstream of the DFG motif is relatively constant to date across X-ray structures. In particular, the carbonyl oxygen of the residue preceding the aspartic acid is typically involved in a buried hydrogen bond to either a tyrosine phenol or histidine epsilon nitrogen. This hydrogen bond holds the carbonyl group fixed, and consequently, the backbone NH of the aspartic acid is also fixed via the rigid peptide bond. The major movement of the loop is, as a result, initiated via rotation around the ζ torsion of the aspartic acid.

2.2.3.2 Glycine Rich Loop

The glycine-rich loop (G-rich loop) is defined by the GxGxxG motif that is highly prevalent in protein kinases and occurs in the first 10–20 residues of the kinase domain. Other residues within this region, e.g., Thr-14 and Tyr-15 of CDK2, may be targets for phosphorylation and consequential inhibition of kinase activity. With respect to inhibitor ligand design, the important feature of this region is its inherent flexibility. In particular, ligand structures such as FGFR-SU5402 **2** [30] and Abl-Gleevec **1** (Scheme 2.1) [26, 27] demonstrate the ability of the loop to mold or collapse onto the ligand to form favorable contacts, particularly with the aromatic residue that typically precedes the final glycine. It is difficult to determine based on structural evidence to date, but one can speculate that there is less likelihood of seeing a recurrent pattern for how the loop adjusts in conformation, as opposed to the activation loop conformational changes where patterns are clearly present.

2.2.3.3 C-Helix Orientation

Recall that the C-helix contains a strictly conserved glutamic acid residue. It is apparent from multiple structures of different protein kinases that modulation of the conformation of this helix is a common regulatory theme. Additionally, there are multiple ways of regulating C-helix conformation, and as with other modes of regulation, ligand binding is affected. For CDK2 [10, 11], the C-helix is properly



Scheme 2.1 Arrows indicate the hydrogen bonding pattern of the inhibitor to the kinase backbone determined by X-ray crystallography.

positioned by the binding of cyclin proteins to give a tertiary complex, upon ATP binding, that is catalytically active. In the absence of these proteins, the C-helix rotates on its axis, projecting the glutamic acid more towards the solvent, and translates away from the catalytic lysine. This conformational change creates additional space in a region adjacent to the ATP binding site that a small molecule inhibitor can occupy without steric interference. Similar patterns of C-helix displacement are also observed in Hck and Src Kinase X-ray structures [6, 31–35], but the degree of change is distinct. For Src and Hck, C-helix conformation is not modulated via interactions with cyclins but instead through phosphorylation and/or interaction with other protein domains.

Another mechanism of modulating the C-helix conformation is employed in common for the so called AGC family of kinases: PKA, PKB, PRK, SGK1 and MSK1 are among the members of this family. These kinases utilize a hydrophobic motif (FxxF) beyond the kinase C-terminus to pack adjacent to the C-helix and lock it into an active conformation [12, 36–40]. For proper packing to occur, the evidence to date is that an acid functionality must be immediately adjacent and downstream of the second phenylalanine of the hydrophobic motif. This acidic group can be in the form of either a phosphorylated residue (e.g., PKB/SGK1), an aspartic acid (PRK2), or a C-terminus-free carboxylate (PKA). Crystallographic studies [38–40] for PKA and mutated PKB (S474D) show the acidic functionality to form a salt bridge or hydrogen bonds with residues from the N-terminal lobe [41]. This interaction stabilizes the C-helix location. There are two key features for ligand design considerations. First, as with the other mechanisms described above, the nature of the ligand binding site in the region of the C-helix is affected by the regulatory state. However, there is also a second feature, unique to AGC family members. In addition to the hydrophobic motif, they also have other conserved sequence motifs, which based on PKA and recent PKB structures, impact the shape of the front of the ATP binding site in the vicinity of C2 of ATP. Namely, a phenylalanine (PKA 327, PKB 439), that is upstream of the hydrophobic motif, projects into the front of the ATP site, stacking over a glycine–glycine step that is at the end of the hinge region, and forms an aromatic herringbone type interaction with the C2/H2 of the adenine ring conformation [36, 38–41]. In

other kinases this region is fully accessible, and commonly occupied by inhibitors, so a substantially different binding profile is thus possible for AGC family members, relative to other protein kinases.

2.2.3.4 Lobe Orientation

An additional feature of protein kinase structure and function that impacts ligand binding is the orientation of the N-terminal lobe relative to the C-terminal lobe. Catalysis is, as noted, dependent on residues from both lobes, and as such, the proper spatial orientation between the domains is required for effective catalysis. In the absence of ATP, substrate, or other needed co-factors and adapter proteins, the domains are observed to orient differently via conformational changes in the hinge region as well as a few other confined sequence elements (e.g., an extended region just downstream of the C-helix). The net effect is to alter the shape of the ATP binding cleft and, consequently, the shape of the synthetic ligand needed for inhibition. This effect is perhaps observed to be most pronounced to date for ERK2 and p38 MAPK kinases [16–19, 42]. For p38, the binding site is substantially more open, relative to, e.g., CDK2 or IRK, and this feature as a result alters the profile for the type of ligand needed to inhibit p38 [43].

2.2.3.5 Solvent Channel

The lysine–glutamic acid salt bridge has been described above, but one aspect that has not been emphasized is that this interaction occurs relatively deep in the binding site, in the vicinity of the N7/C8 positions of ATP. Additionally, the backbone NH of the aspartic acid that is part of the DFG motif also projects into the sterically accessible space adjacent to N7 of ATP. These two consistent features across protein kinases (Lys–Glu salt bridge and Asp NH) create a polar region buried within the ATP binding site. The hydrophilicity of this site can be further enhanced by the presence of polar amino acids in this region, for example at the gatekeeper residue or the residue immediately preceding the DFG motif. The net result is to favor and stabilize the presence of a solvent channel of two to three water molecules deep within the ATP binding site. The precise positioning of the water molecules is modulated by the types of amino acids in this region, and the number of water molecules can be expanded significantly for catalytically inactive conformations. Additionally, sterically bulky residues such as Phe-80 of CDK2 may block the presence of some of the water molecules. Inhibitors may either interact favorably with the solvent located in this region or displace it in part or in whole. An increasingly common motif is to target the Asp NH as will be discussed. Regardless of whether the solvent is to be displaced or retained, it is a key aspect of inhibitor design.

2.3

Crystallization

2.3.1

Defining the Construct

One of the most challenging aspects of obtaining a kinase crystal structure is designing the initial constructs. Recent advances in high-throughput cloning techniques, affinity purification tags and crystallization robots have however allowed for multiple constructs to be tested rapidly in parallel.

Initial constructs are often designed based on previously solved crystal structures and usually contain only the catalytic kinase core. However, many kinases are multi-domain proteins and inclusion of additional domains can increase both protein stability and solubility. The minimal kinase domain typically contains five β -strands and one α -helix in the N-terminal lobe and 7–8 α -helices in the C-terminal lobe. In tyrosine kinases, the N-terminal lobe usually starts near a highly conserved tryptophan, 14 residues upstream of the GxGxxG motif. The start of the N-terminus is less well conserved in serine/threonine kinases.

Limited proteolysis has proven useful for identifying the boundaries of the N- and/or C-terminal lobes as well as identifying additional segments or domains that may provide stability. Typically the full length protein is first expressed and purified. The protein is then digested with a panel of proteases, both in the presence and absence of known ligands or inhibitors, and the digests are analyzed by SDS-PAGE. The N- and C-terminus of the stable fragments are identified by N-terminal sequencing and mass spectrometry. In the case of EphB2 and T β R-1, an N-terminal juxtamembrane or regulatory domain was included in the crystallography construct in order to obtain data quality crystals [44, 45]. Regions C-terminal to the kinase domain were included in the CaM and Twitchin kinase constructs and were found to be pseudosubstrates in the crystal structure [46, 47].

The majority of kinases solved to date have been expressed using a baculovirus expression system in insect cells. Only a few kinases, including p38, ERK2, PAK and EphB2, have been successfully expressed in *E. coli* [17, 19, 45, 48]. Constructs typically contain an affinity tag to aid in the purification. Both His₆ and cleavable GST tags [45, 48, 49] have been used successfully. Newer expression systems, including the Invitrogen Echo® system, Gibco/Life Technologies Gateway® system or the Novagen pTriEx-1 cloning system, allow one to test multiple affinity tags in both *E. coli* rapidly and insect cells with minimal subcloning [50].

To improve the stability and/or solubility of the protein during expression and purification, peptide substrates, binding partners and ligands have been added. The structures of many cyclin-dependent kinases (CDKs) have been solved by co-expression and purification with protein activators and inhibitors. The structure of CDK5 was solved by co-expression with the protein activator p25 [51]. The structure of CDK6 was solved by reconstitution of a CDK6/cyclinD/p18^{INK4c} complex *in vitro* during the purification [52]. The solubility and stability of GSK3, during purification and crystallization, was significantly improved by addition of a

39 amino acid peptide derived from the C-terminus of its binding partner FRAT1 [53].

2.3.2

Mutagenesis

In some cases, mutation of surface or active site residues has been required to improve stability and solubility. Mutation of surface cysteines prevented the formation of disulfide-linked aggregates in FGFR1 [54]. Mutation of highly conserved, active site residues, to create a kinase-dead mutant, was required for a number of kinases whose over-expression proved to be toxic to cells. The conserved aspartic acid in the DFG motif was mutated to an asparagine in CDK5 while the conserved lysine in PAK was mutated to an arginine [48, 51].

Removal of flexible loops within the kinase domain has also been required for some kinases. Many receptor tyrosine kinases contain an additional domain within their C-terminal lobe, referred to as the kinase-insert domain. Removal of this highly charged domain was required to obtain data quality crystals of VEGFR2 [15].

Some proteins have not crystallized despite considerable effort. In such cases, mutation of a closely related kinase to resemble the kinase of interest has provided valuable structural information. Ikuta and co-workers mutated three residues within the ATP binding pocket of CDK2 to the corresponding residues in CDK4. The resulting mutant protein was crystallized and used to design CDK4-selective inhibitors [55].

2.3.3

Phosphorylation

Heterogeneous phosphorylation is often a problem when kinases are expressed in insect cells. Multiple approaches have been used to solve this problem. Proteins have been completely dephosphorylated by incubation with λ protein phosphatase or alkaline phosphatase [38, 39, 56]. Ion exchange and isoelectric focusing chromatography have been used to separate proteins with multiple phosphorylation states. An γ -aminophenyl ATP-sepharose column was used to separate different phosphorylated states of human c-Src [34]. Alternatively, serine/threonine or tyrosine phosphorylation sites can be mutated to alanine or phenylalanine, respectively [42]. For tyrosine kinases with multiple autophosphorylation sites, the active site aspartic acid can be mutated to an asparagine, creating a kinase dead mutant [57].

In many cases the active, phosphorylated form of the protein is preferred for structure-based drug design studies. Quantitative autophosphorylation has been achieved for many kinases including IRK, IGF-1 and VEGFR2 [13, 15, 49]. Residual ATP and ADP must be removed following phosphorylation to ensure homogeneity.

Many serine/threonine kinases require phosphorylation by an upstream kinase for activation. This requires cloning the upstream kinase as well as additional purification steps after phosphorylation [38, 39, 58]. In some cases, the active

form of the protein can be mimicked by replacing phosphorylated serine or threonine residues with aspartic or glutamic acid [38, 39].

2.4 Inhibitor Design

2.4.1 Binding in ATP Cleft

The region of the ATP binding site is the current focus of kinase inhibitor drug discovery research. At first blush the pursuit of selective kinase inhibitors competitive with a substrate common to all kinases, ATP, and which is present at intracellular concentrations (2–5 mM) that are 10–1000-fold in excess of the K_m , would appear to be a poor strategic approach. This belief was widely held into the mid-1990s at which time several reports emerged of kinase-selective ATP-competitive inhibitors. The structural basis for both binding in the ATP site and features governing kinase selectivity was rapidly developed in subsequent X-ray crystallographic studies and as discussed below provides a framework for current kinase discovery efforts. While many alternative binding sites exist which could provide the basis for non-ATP competitive approaches to kinase regulation, none are likely to prove as universally applicable as the compounds that bind into the ATP site. The primary attractiveness of the ATP site is that it affords a deep hydrophobic

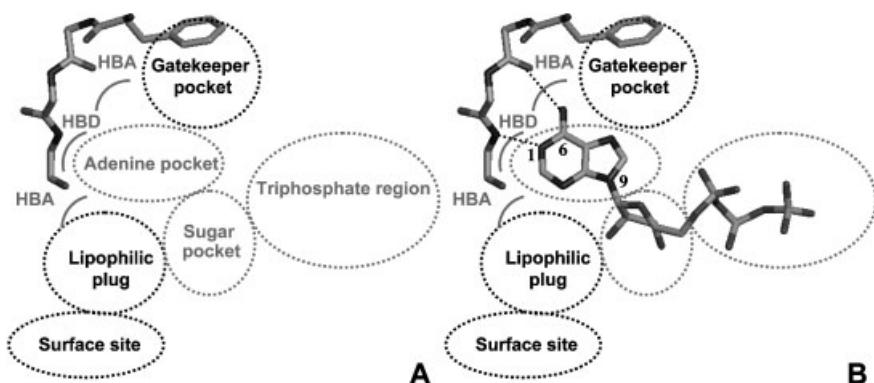
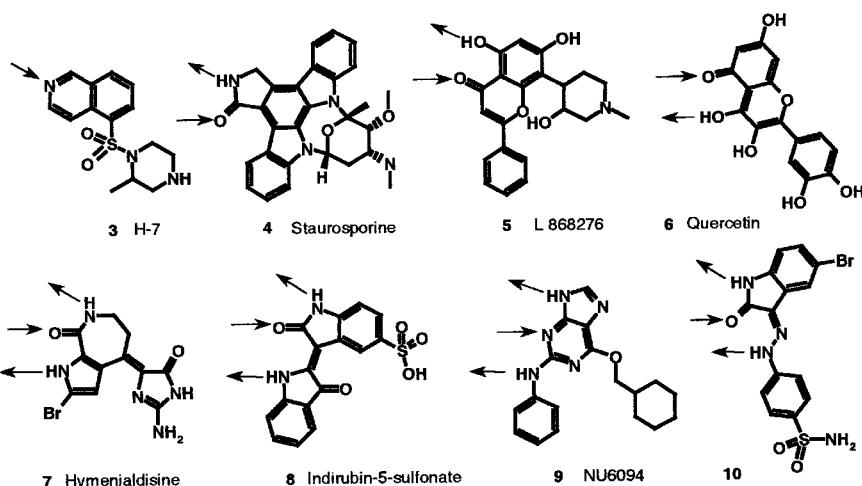


Fig. 2.2 (A) Traxler/Bower Kinase inhibitor pharmacophore. The perspective given is a top down view from the N-terminal to C-terminal domain in which the plane of the adenine ring matches that of the page. A short segment of the kinase linker region is shown which includes the three hydrogen bonding residues (minus side chains) plus the gatekeeper residue. In this orientation the top of

the graphic corresponds to the back of the pocket and the bottom of the figure to the front or solvent. HBA and HBD denote hydrogen-bond acceptor and donor, respectively. The graphics were generated using WebLab ViewerLite from MSI and the indicated PDB files. (B) Complex of ATP, cyclin A and human CDK2 (PDB.1FIN).



Scheme 2.2 Arrows indicate the hydrogen bonding pattern of the inhibitor to the kinase backbone determined by X-ray crystallography.

cleft for the adenine of ATP (or in some cases the guanine of GTP) that is common to all kinases and yet highly varied in shape. The commonality of binding elements provides inhibitor design features that can be applied to all kinases, whereas the exploitation of unique structural features provides the basis for achieving selectivity.

At present there are >100 protein kinase-inhibitor structures in publicly available structural databases which span 28 kinases and a variety of inhibitor structural classes. Based upon an analysis of these data a classification of ATP binding regions has been proposed by Traxler and Furet [59] and Bower (personal communication from Michael J. Bower, December 1999). In the subsequent discussion a slightly modified version of this classification will be used to organize the trends seen across kinases and inhibitor classes (see Fig. 2.2). In the Traxler model, five sites were proposed of which three (adenine, sugar and phosphate-binding sites) can be directly related to ATP and two additional lipophilic sites which lay outside of the region occupied by ATP. In the Bower model, an additional polar site on the surface of the protein was proposed.

2.4.1.1 ATP Binding Sites

As previously described, the purine binding site is well engineered to complement both the planar aromatic nature and hydrogen bonding pattern of adenine. That the great majority of known inhibitors mimic these features highlights the essential role the purine binding site plays in the binding of both ATP and inhibitors. The key features of a purine mimetic are hydrogen bonding functionality, which at a minimum mimics the hydrogen bond formed by N1 of adenine and a planar aromatic or heteroaromatic ring system (Fig. 2.2B). Many inhibitors also

contain a group which occupies the ribose-binding site and engages residues in the protein which either hydrogen bond to a ribose hydroxyl or to groups involved in the ligation of Mg and the α -phosphate of ATP. The isoquinoline sulfonamide **3** (Fig. 2.3 A) is good example of an inhibitor which incorporates these binding elements [36]. While inhibitors must minimally contain a single hydrogen bond acceptor, more commonly seen are inhibitors which mimic the dual hydrogen bond donor–acceptor properties of adenine. Examples of donor–acceptor inhibitors (Scheme 2.2) are staurosporine **4** (NHCO of lactam) [60] and des-chloroflavorpiridol **5** (carbonyl and phenolic hydroxyl) [61]. Also observed are alternate hydrogen bond donor–acceptor complexes in which the protein acceptor and donor atoms are located on the same amino acid residue (N+3 from the gatekeeper). Quercetin **6**, whose phenolic oxygens bind to the kinase amide backbone, is an example of this alternative pattern [62] (Fig. 2.3 B). Interestingly, in casein kinase 2 which utilizes GTP, as well as ATP, as the phosphate donor, this alternate hydrogen bond donor–acceptor pair has also been observed for the binding of GTP [63]. Finally, the structures of hymenialdisine **7** [64], indirbin-5-sulfonate **8** [65], NU6094 **9** [66] and **10** (Fig. 2.5 B) [67] in CDK2 provide examples of inhibitors that make all three possible hydrogen bonding contacts within the hinge region.

The triphosphate binding site of kinases presents a conserved set of polar side chains to bind the phosphate portion of Mg–ATP and to position the γ -phosphate for transfer to the acceptor residue (Section 2.2.2). In general, the triphosphate site is largely unoccupied in the kinase-inhibitor complexes thus far described. A notable exception to this trend is balanol **11a**, a potent polyphenolic natural product inhibitor of PKA and PKC. In the PKA structure, balanol occupies the entire ATP binding site, making extensive hydrogen bonding contacts to the glycine-rich loop and other polar phosphate-binding residues (Fig. 2.4 A) [68]. A detailed investigation of synthetic balanol analogues revealed that the removal of polar groups which bound in the triphosphate pocket (**11b–d**, OH and CO₂H) had no effect on PKA potency, although PKC inhibition was reduced by 10–1000-fold (Tab. 2.1)

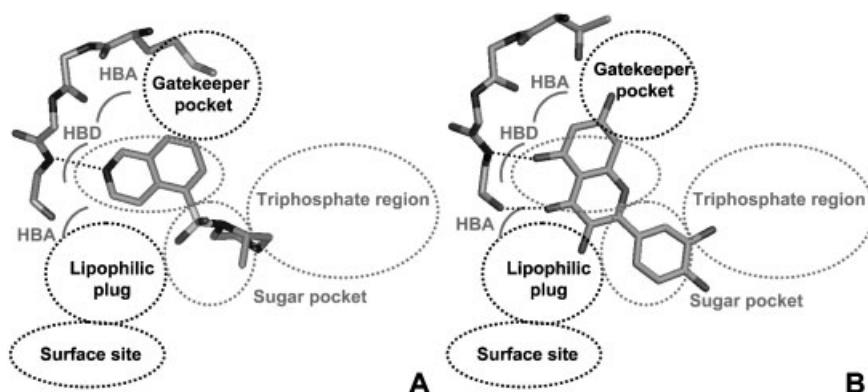


Fig. 2.3 (A) Complex of the catalytic subunit of bovine PKA with **3**, H-7, (PDB.1YRD).

(B) Complex of quercetin, **6**, with the Src family kinase, human HCK (PDB.2HCK).

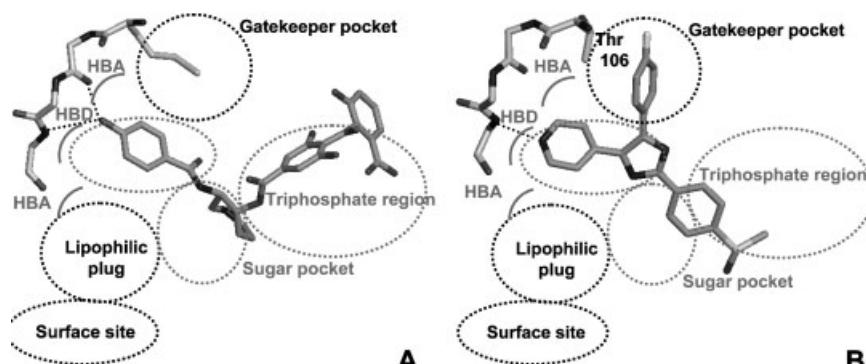


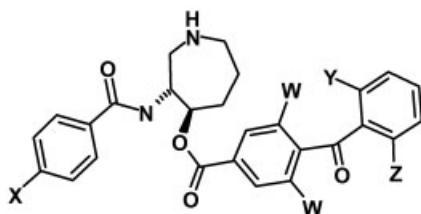
Fig. 2.4 (A) Complex of the catalytic subunit of activated murine PKA with balanol, **11a**,

(PDB:1BX6). (B) Complex of SB 203580, **12**, in human p38 (PDB:1AU9).

[69]. Also prepared was a 5'-deoxybalanol analogue, **11e**, which lacks the phenolic group that hydrogen bonds the amide backbone in the hinge region and again this change had no effect on PKA inhibition. Using these balanol analogues and a number of the isoquinoline sulfonamides, a theoretical model was developed to calculate ligand-binding affinities to PKA [70]. These calculations, which were able to reproduce most of the experimental data, found that the nonpolar binding interactions made the primary contribution to high-affinity binding. Hydrogen bonding interactions in general provided no net contribution to binding as the energetics of binding to the kinase were negated by the requirement to desolvate these polar groups. However, for the poorly solvated isoquinoline inhibitors, the single hydrogen bond made to the amide backbone in an otherwise lipophilic

Tab. 2.1 Potency of balanol analogues prepared by Koide et al. against PKA and PKC [69].

| Compound | PKA K_I (nM) | PKC K_I (nM) | X | Y | Z | W |
|------------|----------------|----------------|----|----|-------------------|----|
| 11a | 4.7 | 5.3 | OH | OH | CO ₂ H | OH |
| 11b | 3.9 | 604 | OH | H | CO ₂ H | OH |
| 11c | 11 | 5000 | OH | OH | H | OH |
| 11d | 3.4 | 80 | OH | OH | CO ₂ H | H |
| 11e | 3.5 | 69 | H | OH | CO ₂ H | OH |

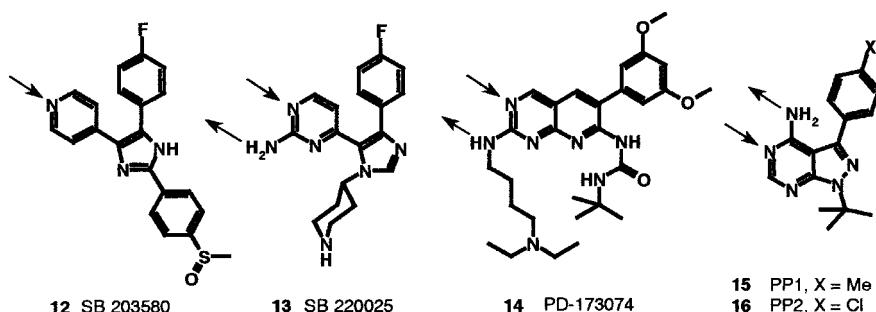


binding pocket does make a positive electrostatic contribution to binding. These conclusions, which are consistent with binding studies in other protein-small molecule systems, may provide an explanation for the paucity of kinase inhibitors which make extensive contact in the highly polar triphosphate-binding pocket.

2.4.1.2 Gatekeeper-Dependent Binding Pocket

The solution by Tong and co-workers of the co-crystal structure of SB 203580, **12**, (Scheme 2.3) in p38, a serine/threonine kinase, provided the first structural data to explain how kinase selectivity could be achieved with an ATP-competitive inhibitor [71]. This structure revealed the binding of the fluorophenyl group into a pocket behind and orthogonal to the adenine of ATP. Residues Thr-106 and Lys-53 provide the most extensive contacts with the fluorophenyl group by forming the walls of the pocket above and below the plane of the aromatic ring (Fig. 2.4 B). Recognizing that a limited number of kinases have side chains at position 106 the size of threonine or smaller, Tong proposed a potential role of Thr-106 in determining kinase selectivity. For example, the closely related ERK2, which has the larger glutamine side chain at position 106, is 1000-fold less sensitive to SB 220025, **13** ($IC_{50}=20\ \mu M$) relative to p38 α [43].

The publication by Tong et al. was quickly followed by similar findings for several tyrosine kinases. Specifically, the co-crystal structures of PD 173074, **14**, in FGFR1-14 (Fig. 2.5 A) [72], PP1, **15**, in Hck [32], PP2, **16**, in Lck [73] and **15** in c-Src [74] all highlighted the significance of this newly discovered binding pocket. Further compelling evidence that a single amino acid residue, which is often termed the gatekeeper, as it controls access to this binding pocket, was the primary factor governing access to this pocket was provided by mutagenesis studies with p38 γ [75], p38 α [76] and c-Src [74]. Two important conclusions from these studies were: (1) that the introduction of a small gatekeeper residue into any protein kinase was sufficient to confer sensitivity to inhibitors that occupied the gatekeeper pocket, and (2) that kinase sensitivity to these inhibitors was closely related to the size of the gatekeeper residue, being greatest for glycine, followed by alanine and then serine and valine (Tab. 2.2).



Scheme 2.3

Tab. 2.2 The effect of site-specific mutagenesis of gatekeeper residues Met-106 (p38 γ) and Thr-338 (Src) on inhibitors which occupy the gatekeeper specificity pocket [74, 75].

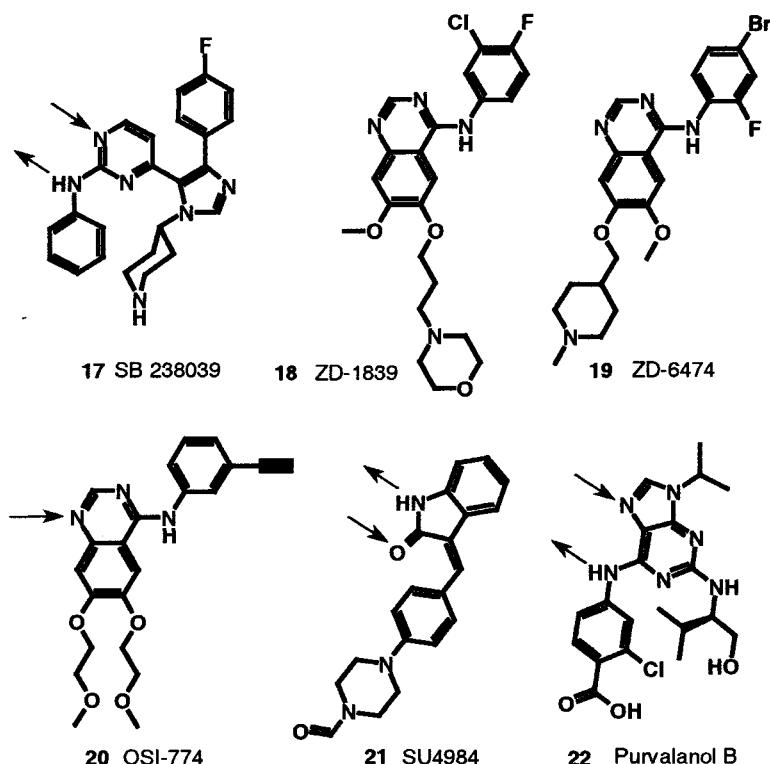
| <i>p38γ/SAPK3-12</i> | | <i>Src-15</i> | |
|--|-----------------------------|-------------------|-----------------------------|
| <i>Gatekeeper</i> | <i>IC₅₀ (nM)</i> | <i>Gatekeeper</i> | <i>IC₅₀ (nM)</i> |
| Gly | 30 | Gly | 5 |
| Ala | 10 | Ala | 5 |
| Ser | 50 | Ser (wildtype) | 400 |
| Thr | 300 | Thr | 100 |
| Leu | 45 000 | Val | 100 |
| Gln | 50 000 | Ile | 5000 |
| Met (wildtype) | >100 000 | Met | 8000 |
| Phe | >100 000 | Phe | 8000 |

2.4.1.3 Lipophilic Plug

The lipophilic environment created by residues that form the adenine-binding pocket extends sufficiently beyond the opening of the pocket to create an additional lipophilic binding area in front of the adenine. This pocket may not be accessible to inhibitors that bind into tightly closed adenine pockets. Inhibitor accessibility to this pocket may also be restricted for some kinase families. For example, the AGC family kinases (PKA, PKC, AKT) have a C-terminal tail that extends back into N-terminal domain and caps the adenine site with a phenylalanine (see Section 2.2.3.3). Inhibitor-kinase structures that bind an aryl ring in this site include **10** in CDK2 (Fig. 2.5B) [67], **9** in CDK2 [66], and **17** (Scheme 2.4) in p38 [77]. For these later two examples, the substitution of the inhibitor NH₂ with NHPh, which places the aryl group in this lipophilic pocket, affords a 10-fold increase in potency.

2.4.1.4 Polar Surface Site

The extension of the inhibitor beyond the lipophilic site that caps the adenine pocket has been successfully employed either to gain access to solvent (water) or to sites on the surface of the kinase. In general the inhibitor scaffolds that have been pursued as potential drug candidates are relatively flat lipophilic heteroaromatics, and the resulting compounds have very poor aqueous solubility. The attachment of a polar water-solubilizing group, typically a tertiary amine, to take advantage of this access to the solvent has been a key strategy for many kinase drug discovery efforts. Published examples for the anilinoquinazoline class of inhibitors are ZD1839 **18**, an EGFR inhibitor [78] and ZD6474 **19**, a VEGFR2 inhibitor [79]. While no crystal structures of these examples have been published, the recent publication of OSI-774, **20**, in complex with EGFR confirms the anticipated binding mode of these compounds [80]. The crystal structures of **14** (Fig. 2.5A) [72] and SU4984, **21**, [30] in FGFR1 confirm that the water-solubilizing tertiary amines introduced onto these scaffolds project into the aqueous environment.



Scheme 2.4

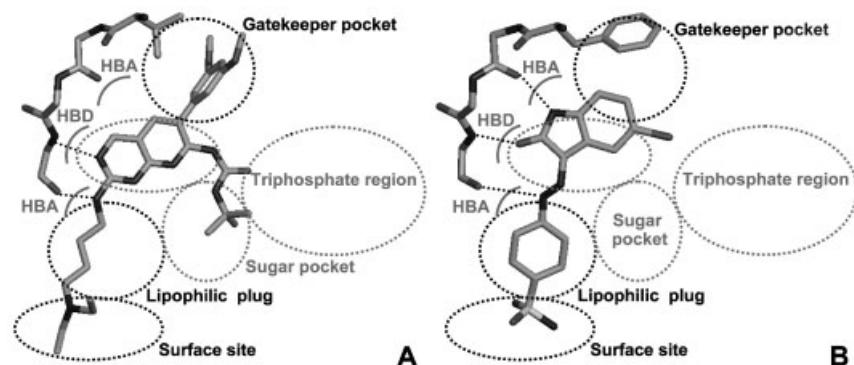


Fig. 2.5 (A) Structure of unphosphorylated FGFR1 soaked with a solution of 14

(PDB:2FG1). (B) Structure of 10 in complex with inactive human CDK2 (PDB:0.1 fvt).

As would be expected for functionality that remains water-solvated, the introduction of these amines often has little impact on inhibitor potency. Conversely, the observation of a significant effect on potency for functionality which extends be-

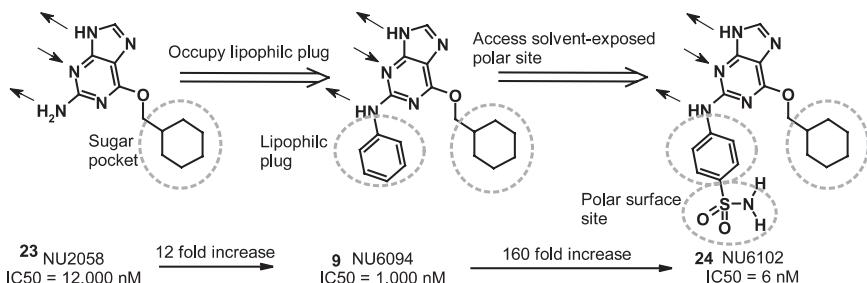


Fig. 2.6 Structure-based design guided by iterative crystallography using activated cyclin A-bound Thr-160-phosphorylated CDK2.

beyond the lipophilic plug raises the expectation of a specific binding site on the protein surface. The hydrogen bonding interaction of an aryl sulfonamide with a backbone amide NH of Asp-86 appears to be the main driver for the >100-fold enhancement in binding affinity seen for **10** (Fig. 2.5 B) [67] and **24** (Fig. 2.6) [66].

2.4.2

Conformational Considerations

The conservation of kinase structure and mechanism across the entire protein kinase family dictates a conservation of shape and electrostatics for the ATP binding site when ATP is bound to the catalytically active form of the kinase. However, no such commonality of features is required for the ATP binding site of kinases in their inactive state. In the last few years protein crystallography has provided many insights into the structural biology of protein kinase regulation. The picture that has emerged is complex. Multiple mechanisms have been described for regulation of kinase activity and the catalytic state of the protein is often the sum of competing mechanisms of activation and de-activation. The distortion of the kinase away from its active conformation is the common theme for these regulatory mechanisms. In many cases the conformational change required for activation is the direct result of a phosphorylation in the activation loop [81]. For the cyclin-dependent kinases an additional event, the binding of an accessory protein, cyclins, is required to obtain an active conformation [82]. An in depth account of these mechanisms can be found in recent reviews by Huse and Kuriyan [5], Johnson and Lewis [83] and Engh and Bossemeyer [2]. The following sections consider the impact of conformational flexibility on inhibitor design.

2.4.2.1 Inhibitor-Induced Binding

The structures of two potent PKA inhibitors, staurosporine **4** [84] and balanol **11a** [68], reveal a protein binding surface complementary to that of the inhibitor. Noting the large protein-inhibitor contact surfaces in these structures, Engh and Bossemeyer investigated the relationship of buried surface area to inhibitor potency.

The intuitive notion, that inhibitor potency should be related to binding area, can be theoretically derived as a logarithmic relationship between inhibitor potency and buried surface area [85]. When the data were plotted for PKA (log IC₅₀ versus contact area), all of the compounds fell close to the theoretical line. One remarkable finding is that this correlation held for compounds whose binding induced a significant conformational change in PKA, for example **4** and **11a**. Staurosporine is a particularly instructive example as it is a potent inhibitor of many kinases and in all cases induces a protein conformation complementary to itself. The most likely explanation is that the induced conformations are energetically easily accessible, an assumption well-supported by structural studies which have examined the activation mechanisms of kinases. Several authors noting the tight fit of staurosporine have made reference to the phenomena of hydrophobic collapse. This picture, that the kinase envelopes and molds to the shape of inhibitor, fits well with the conclusions of Hunenberger et al. in which they noted that hydrophobics and not hydrogen bonding was the chief driver of inhibitor affinity [70].

When Engh expanded the analysis to include all the available kinase-inhibitor structures, the fit was poorer. Outliers included STI-571 **1**, which has the greatest buried surface area, but is only moderately potent and a number of compounds that bound more potently than would be expected based upon their surface contacts. An explanation for the relatively poor binding affinity of **1** is suggested from its structure as determined in the un-activated form of Abl [27]. This structure reveals that binding of the inhibitor forces the activation loop into a catalytically inactive conformation which appears to be energetically unfavorable relative to other active conformations. This conclusion was supported by inhibition studies demonstrating **1** to be a potent inhibitor of the un-activated kinase ($K_I=37\text{ nM}$), but a weak inhibitor of the fully activated Tyr-393 phosphorylated Abl ($K_I=7000\text{ nM}$). Moreover, the insensitivity of the fully activated kinase was predicted by modeling which suggested that Tyr-393 phosphorylation would favor the catalytically active conformation of the activation loop. That inhibitor binding can force the kinase into higher energy conformations is a further consequence of conformational plasticity that promises to play an important role in the design of selective kinase inhibitors.

Finally, what about those inhibitors that bind better than predicted based upon an analysis of buried surface area? Examples given by Engh and Bossemeyer include CDK2-**22**, Hck-**15**, Lck-**16** and p38-**13** [2]. This tighter binding could reflect a higher degree of inhibitor–enzyme complementation. The structures of **13** and related pyridinylimidazoles in p38 offer support for this explanation. Unlike many of the structures discussed thus far, the binding of pyridinylimidazoles appears to induce very little change in p38 over that of the catalytically inactive apo-structure [43].

2.4.2.2 What is the Most Appropriate Enzyme Form for Crystallography?

The most pervasive mechanism for the regulation of kinases is that effected by other kinases and phosphatases that add and remove phosphate at multiple sites. Another common mechanism by which kinase activity and substrate binding (ATP and acceptor protein) are regulated is through the binding to proteins that

lie both upstream and downstream to the kinase transmitting the signal. In most of these examples, the regulation of kinase activity is probably effected by conformational changes that either favor or disfavor the attainment of the catalytically active ATP-bound conformation. In principle, consideration of these inactive conformations in designing an inhibitor is not required, if the sole mechanism of the inhibitor is to compete with ATP for the binding to the catalytically active form of the kinase. In practice, it is not that simple, as there are several examples known in which the inhibitors bind equally well to both the un-activated and activated kinase. Also, unfortunately in most cases it has proven much more difficult to purify and crystallize the active protein. Hence, many of the inhibitor-kinase structures have been solved with an inactive form of the kinase.

Recently Davies and colleagues have reported their studies on the design of CDK2 inhibitors. Whereas previous structure-based design efforts had used the structure of inactive monomeric CDK2, prior to initiating their design efforts, the Davies group undertook a comparative study of both the inactive and activated cyclin A-bound Thr-160-phosphorylated structures of CDK2 [86]. This comparison revealed two major changes which might affect inhibitor binding. They also solved the structures of an inhibitor, indirubin-5-sulfonate, bound to both kinase forms and performed calculations to quantify the effect of structural differences between these two forms on inhibitor binding. Having established an empirical basis for using the activated kinase as a tool in structure-based design, in a subsequent publication the authors presented a successful example of structure-based inhibitor design based upon the activated CDK2 structure [66]. Starting with a 12 μ M inhibitor, NU2058 23, a >1000-fold improvement in potency was achieved for the optimized compound, NU6102 24 (Fig. 2.6).

While the above example supports activated kinases as the most relevant form for structure-based design, the behavior of STI-571 1, which binds some 200-fold more potently to the un-activated form of Abl, suggests that inactive kinase structures may also have value [27]. The Abl-1 structure revealed that inhibitor binding locks the activation loop in an unusual inactive conformation opening up a new pocket deeper in the ATP site. Potential advantages of inhibitors that reform the ATP binding site are enhanced selectivity and the ability to lock the kinase into a non-activatable conformation. However, these advantages are not fully realized for 1, as it is not uniquely selective (Abl, c-Kit and PDGFR are all equally inhibited) [87] and there are no data demonstrating that 1 inhibits activation. Another potential advantage is that the inhibitor may not have to compete with ATP for binding to the un-activated kinase. Given the high intracellular concentration of ATP, the cellular potency of inhibitors that are strictly competitive with ATP should be greatly attenuated, and as such binding of the inhibitor to an un-activated enzyme form with low ATP affinity might circumvent the disadvantage inherent to ATP-competitive inhibitors [88].

Recent Abl structural studies comparing the binding of 1 to PD 173955, 25, suggest that for this example the unique binding conformation may come at a high cost [26]. PD 173955, which has a reduced contact surface with the enzyme, does not bind in the pocket formed by movement of the activation loop. However,

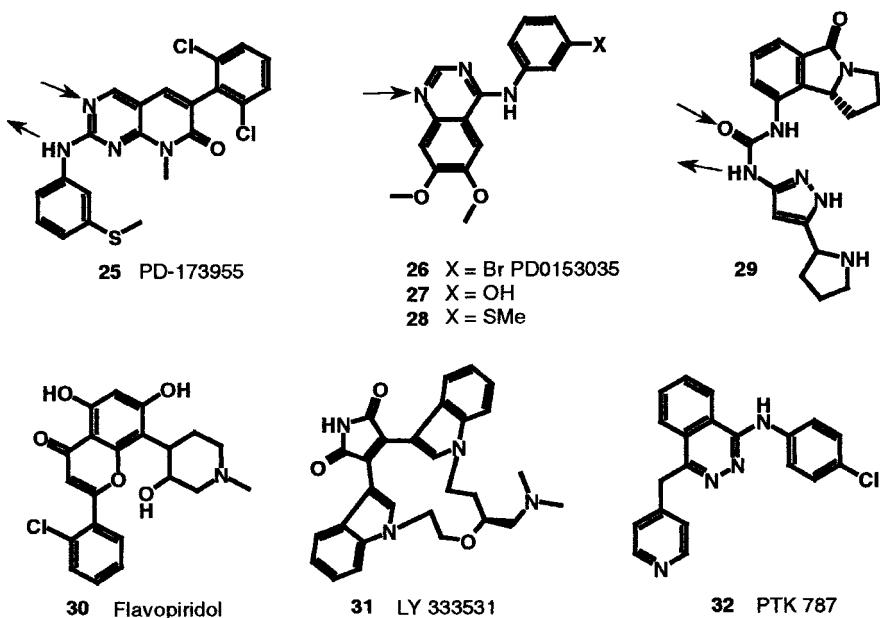
inhibition studies demonstrate that **25** inhibits both the un-activated and activated kinase equally well. The explanation proposed by the authors is that the larger binding surface of **1** is required to pay the penalty for binding to the novel inactive conformation. In contrast **25**, which does not bind to this novel activation loop conformer, is an equipotent inhibitor of both un-activated and activated Abl. The authors propose that the differences in cellular potency of these two inhibitors (**25** is some 10-fold more potent in a cellular assay) may be a reflection of their differing binding modes and conclude an advantage for the inhibitor which can bind to multiple enzyme forms (Scheme 2.5).

The rationale of using the activated CDK2 structure to design **24** are clear. However, these studies leave unanswered the question as to whether this structure provided a result superior to that based upon the inactive kinase. Indeed, many groups have used the inactive form of CDK2 to successfully aid their lead optimization efforts. The Abl example illustrates how inactive kinase structures can lead to unique design possibilities. To re-iterate, there are several mitigating factors that make any blanket statement on choice of enzyme form unwise. As will be outlined in the subsequent section, much progress has been made in kinase design irrespective of the activation state, a result which suggests that any structure is better than none.

2.4.2.3 Homology Models and Surrogate Kinases

The conservation of tertiary structure and mechanism, which complicates the discovery of highly selective kinase inhibitors, provides for a portability of learning from one kinase to the next. Furthermore, the more closely two kinases are related, the more likely the information can be applied to the second kinase. One way in which structural knowledge is extrapolated from one kinase to the next is the process of homology modeling [89]. To construct a homology model the backbone tertiary structure of a known kinase is used as the template upon which the side chain residues of the target kinase are attached. Using various minimization algorithms, the homology model is refined and thereby provides a starting point for inhibitor docking or design. As discussed elsewhere in this chapter, homology modeling is a routinely used tool that can provide valuable insights.

The application of inhibitor-binding modes from one kinase to the next is another example of the portability of structural information. The universal commonality of an adenine mimetic in inhibitor templates, and the corresponding engagement of common backbone-binding elements in the kinase provides a powerful rationale for this process. Structural proof of a common binding mode in PKA, c-Src, Lck, CDK2 and Chk1 has been demonstrated for staurosporine, which potently inhibits all of these kinases [31, 73, 84, 90, 91]. Structural confirmation of a common binding mode for **15** and **16** to several members of the Src family of cytosolic tyrosine kinases has also been obtained [32, 73, 74]. In the above examples it can be argued that the unique structure of staurosporine or the close similarity of Src family members predisposes these systems to exhibit a common binding mode. However, even in highly unfavorable cases a “conservation” of binding



Scheme 2.5

modes has been observed. The anilinoquinazoline **26**, which has a sub-nanomolar IC₅₀ against EGFR, weakly inhibits p38 (IC₅₀=6.3 μM) and is all but inactive against CDK2 (IC₅₀=250 μM). Nonetheless, the binding mode observed for **27** in CDK2 (IC₅₀=1 μM) and **28** in p38 (IC₅₀=5 μM) [42], is the same as that recently observed for the related anilinoquinazoline, **20**, in EGFR [80].

There are several ways in which one may want to exploit the conservation of binding mode to advance inhibitor design. Two very common situations are the lack of protein crystallography for the target kinase at the initial stages of a new effort and the failure to obtain protein suitable for crystallography. To illustrate the later problem, despite the continued efforts of several groups to purify and crystallize Raf and MEK, the kinases upstream of ERK, no structures have been forthcoming. On the other hand, multiple groups have succeeded in crystallizing CDK2, p38 and members of the Src family. In our own labs when the above situations arise we routinely use the binding mode obtained from the more structurally tractable kinase as a “surrogate” for the target kinase. Ikuta et al. have published a clever application of the surrogate kinase concept to assist in the design of CDK4-selective inhibitors. Their approach was to design and crystallize a CDK2 mutant containing a region of the CDK4 ATP binding site in place of that of CDK2. This CDK4 mimic, which retained activity and demonstrated CDK4-like inhibitor selectivity, was used to design an inhibitor **29** that demonstrated high selectivity for CDK4 versus CDK2. The crystal structure of the CDK4 mimic-**29** complex was consistent with their design which identified a single amino acid as the

primary determinate for inhibitor selectivity [55]. Ideally, as in the case above, one would like to have available a surrogate kinase closely related to the target of interest, as this will make it more likely that the inhibitor will bind to the chosen surrogate and that information obtained will be more readily applied to the target. However, this is not a requirement. When faced with a new structural class of unknown binding mode, the solution of the first structure in any kinase is of great value. The process is not foolproof and as with any other co-crystal structure the utility of the information is dependent upon its predictive value.

2.4.3

Paradigms for Kinase Drug Discovery

One gauge of the impact of protein crystallography on kinase inhibitor design is the increasing frequency of its use. The majority of recent publications employ molecular modeling and report the use of protein kinase-inhibitor structures in the design process. Further evidence of the growing importance of crystallography is the increasing frequency of new kinase structures. In 2002 the list of newly solved structures included EGFR [80], AKT2 [38, 39], MAPKAPK2 [92], Aurora A [93], EphA2, and FAK [94]. The following discussion illustrates how this information is being applied to inhibitor design.

2.4.3.1 High Throughput Screening

Where to start for the design of a kinase inhibitor? Does one follow the example set for proteases using a knowledge of mechanism and substrate to design inhibitors and then use iterative cycles of crystallography and design? For now and into the foreseeable future the answer is no. Instead, with rare exception, the starting point for kinase design has been a high throughput screening (HTS) lead. As has been true for other drug targets, natural products have proven to be a rich source of kinase inhibitor scaffolds. The solution of kinase crystal structures with natural products, such as staurosporine **4**, L868276 **5**, quercetin **6**, hymenalsene **7**, and balanol **11a** (Scheme 2.2 and Tab. 2.1) have provided an understanding of binding in the ATP site which under-pins much of the current design efforts. Furthermore, modifications of these natural products has led to the discovery of compounds which are being studied in clinical trials for the treatment of cancer (flavopyridol, **30**) [95] and diabetic retinopathy (LY33351, **31**) [96]. The screening of large compound libraries has also provided a rich source of new inhibitor scaffolds. An unusually successfully example is PTK 787 **32**, a KDR inhibitor screening lead which has advanced into clinical trials [97]. Major inhibitor classes found as kinase leads using HTS that have led to clinical candidates are the quinazolines [98], oxindoles [99], anilinopyrimidines [100] and the recently discovered diaryl ureas [29].

2.4.3.2 Structure-Based Design

In a 1995 publication Furet et al. at Ciba Geigy used the available PKA structures to predict the binding interactions of PKA with staurosporine [101]. This prediction, which was subsequently confirmed with a co-crystal structure, stands as an early illustration of how crystallography would guide the discovery of kinase inhibitors. With the increasing success of crystallographers, many inhibitor design efforts for a novel kinase can now begin with a similar level of knowledge. In the absence of such information, inhibitor-docking experiments using a homology model or alternatively, crystallography data from a surrogate kinase can be used to guide inhibitor design efforts. The following section details several examples illustrating the current status of structure-based design. For more comprehensive reviews on inhibitor design see Traxler and Furet [59], Toledo and Lydon [60], Scapin [102], and Garcia-Echeverria et al. [103].

The staurosporine 4 binding model developed by Furet et al. correctly identified the lactam amide as the donor–acceptor that bound to the linker backbone [101]. Furthermore, the authors hypothesized the indole rings to occupy the sugar and back hydrophobic pocket (referred to as the sugar and gatekeeper pockets in Fig. 2.2A). This model led the group to design less rigid analogues which retained a cyclic imide to maintain the essential backbone-binding element identified in staurosporine (Fig. 2.7). While the bisindolyl maleimides, 33, maintained the potent PKC inhibition seen with 4 [104], the dianilinophthalimides, 34, proved to favor inhibition of EGFR kinase [105]. The success of these new templates provided support for the original modeling and design hypothesis, which at this time had yet to be structurally verified, and furthermore demonstrated the possibility of achieving more selective kinase inhibition with an ATP-competitive inhibitor. Docking studies using the dianilinophthalimides in a homology model of the EGFR kinase led to the development of a model similar to that previously proposed for 4. Based upon this model, a sulfur-aromatic interaction of a meta-halo aryl group of the inhibitor with Cys-773 present in the ribose-binding pocket was proposed as a key-binding element [106]. Furthermore, since Cys-733 is uniquely present at this position in the EGFR family, this interaction could explain the enhanced selectivity of these compounds for EGFR kinase. Continuing their focus

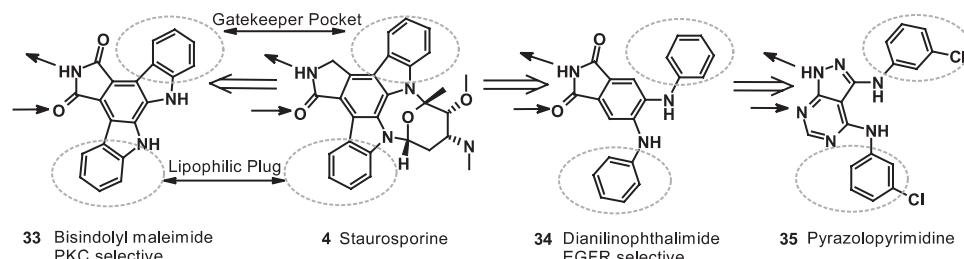


Fig. 2.7 The structure-based design of new inhibitor templates based upon a model of staurosporine bound to PKA.

on EGFR, which had been identified as a therapeutically attractive target, the Novartis (Ciba-Geigy) group pursued two more related series, the pyrazolopyrimidines, **35** [106] and the pyrrolopyrimidines, **36** [107]. In both series a preference for a meta-halo aryl group was demonstrated for the ring which was presumed by modeling to fit in the lipophilic pocket adjacent to Cys-773 (Fig. 2.8).

Independently, 4-anilinoquinazolines were being pursued as EGFR inhibitors at Parke-Davis [108] and Zeneca [109]. The obvious structural similarities between the Novartis series and the quinazolines extended to the SAR, which indicated in both series a preference for a meta-halo group on one of the aryls. These similarities led the Parke-Davis group to propose that a common binding mode for **26** and **36** would be expected (Fig. 2.8). However, docking studies on the EGFR homology model developed at Parke-Davis led this group to propose an alternative to the Novartis model, in which the meta-halo aryl group was positioned in an inner lipophilic pocket next to the gatekeeper. In a comparison of both docking models against all the Novartis data, the Parke-Davis model was better able to fit all the data [110]. A recent X-ray crystal structure of EGFR with anilinoquinazoline **20** matched the Parke-Davis model [80]. The successful design of CI-1033, **37**, a selective inhibitor which binds irreversibly to EGFR through reaction of the arylamide with Cys-773, provides additional evidence to indicate that the Parke-Davis model will hold for other anilinoquinazolines and closely related scaffolds [98, 111]. However, further extrapolation of this result is premature. Changes in template substitution and differing interactions between kinases have been shown to affect the binding orientation of kinase inhibitor templates [43, 112].

It is important to note that in the 3–5 year period prior to crystallographic confirmation of these models, they were used successfully by several research groups to design new templates and to advance lead optimization programs. The modeling predictions of Trumpp-Kallmeyer at Parke-Davis stand out in this regard [113]. In a publication outlining the development of a binding model for two series of 6-aryl pyridopyrimidines **38** (Fig. 2.9), the authors present in detail the methodology

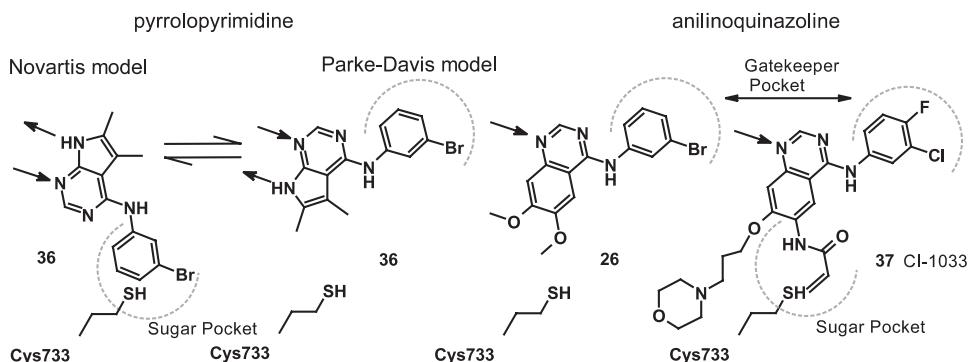


Fig. 2.8 The pyrrolopyrimidine- and anilinoquinazoline-EGFR binding models proposed by researchers at Novartis and Parke-Davis.

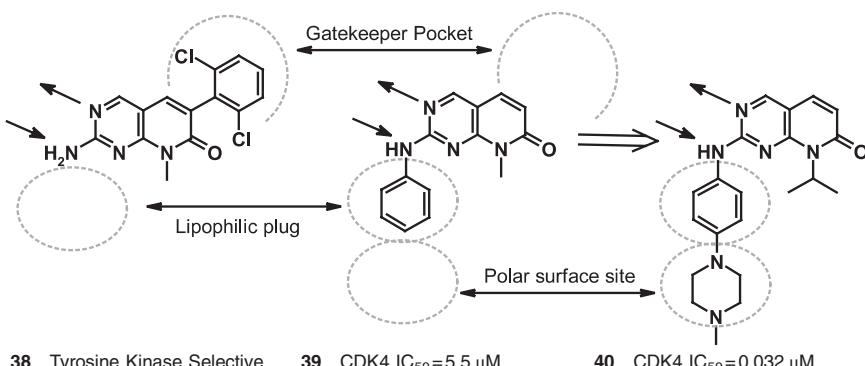


Fig. 2.9 The transformation of a tyrosine kinase-selective inhibitor **38** into a CDK-selective template (**39** and **40**) was achieved by adjusting the template substitution pattern.

and rationale used to develop the model. Because these new series potently inhibited FGFR and c-Src but also displayed significant activity against PDGFR and EGFR, homology models were built and docking studies performed for all four kinases. The analysis of docking studies were guided by the SAR and rationalize the selectivity for inhibition of the these four tyrosine kinases. Their conclusion of a 4-anilinoquinazoline-like binding mode for **38** (and related series) was verified in a crystal structure of **14** in FGFR1 [72]. Furthermore, the more general conclusion on the importance of the gatekeeper (Val-561 of FGFR1) as a key specificity residue for all kinase classes is now widely recognized.

An excellent example illustrating how structural knowledge guides design is the transformation of the 6-aryl-pyridopyrimidones **38** from a tyrosine kinase inhibitor-specific template to favor inhibition of the cyclin-dependent kinases (Fig. 2.9) [114]. Docking of the **38** into CDK2, using the same binding mode developed for the tyrosine kinases, was not possible as access to the gatekeeper-dependent binding pocket was blocked by the large phenylalanine residue which is present in all the cyclin kinases. The authors concluded this to be the primary reason these compounds were not cyclin-dependent kinase inhibitors, and if true, then pyridopyrimidones lacking this 6-aryl group should gain affinity for the cyclin kinases while losing affinity for tyrosine kinases. Pyridopyrimidines, **39** and **40**, having hydrogen in place of aryl at the 6-position were synthesized and proved to be potent cyclin kinase inhibitors.

Concurrent with these modeling efforts, several X-ray crystallographic studies were published for inhibitor-kinase complexes containing selective inhibitors. These included p38 [18, 43, 71], CDK2 [61, 82, 90, 115], several of the Src kinases [31, 32, 73, 74] and FGFR [30, 72]. The consistent theme of inhibitor-protein interactions (outlined in Sections 2.4.1/2.4.2) elucidated by these studies have provided the foundation for subsequent inhibitor design. Much of this design has been the reworking and optimization of templates discovered in screening, such as quinazolines, purines, diaryl imidazoles. Exemplary of this work are a structure-guided

iterative design cycle described by Davies et al. for the optimization of **24**, a purinyl inhibitor of CDK2 [116], and the design of CI-1033 **37**, an anilinoquinazoline containing a warhead acrylamide, which confers irreversible inhibition to the EGFR family of kinases [98].

2.4.3.3 Mechanism-Based and Ligand Mimetic Design

The skeptics that predicted the impossibility of a highly kinase-selective ATP-competitive inhibitor have now been silenced. However, the reasoning from which this skepticism arose remains to frustrate attempts to employ the rationale substrate-based design algorithms that have proven fruitful for other protein classes. Structure-based design efforts which rely on close analogues of either ATP or a peptide or a combination of both in a multi-substrate analogue have met with limited success, and there pursuit till now has been largely as an academic exercise to probe mechanistic details of the kinase reaction. Exemplary of this work is an ATP-peptide bi-substrate analogue inhibitor for which the kinetics and structure have been determined with the insulin receptor kinase [117]. For further information on the status of these approaches see the reviews by Lawrence and Niu [88] and Parang and Cole [118].

2.4.3.4 Computational Chemistry and Virtual Screening

Honma et al. from Banyu Pharmaceuticals have published one of the few examples of kinase inhibitor design which begins with a de novo ligand-building program to generate novel ligands [119]. Realizing the limitations of the de novo lead generation program (LEGEND), Honma chose to filter the structures generated by the de novo software by developing a program (SEEDS) to extract the essential information from the de novo structures and search it against a database of commercially available and/or synthetically accessible ligands. The protein residues chosen as starting points for design of a CDK4-specific inhibitor were those that make the backbone donor–acceptor hydrogen bonds to adenine and the conversed lysine (Lys-33) and glutamic acid (Glu-51) which were known to be involved in the binding of several CDK2 inhibitors. The coordinates used were from a CDK4 homology model based upon CDK2. Following further visual filtering to eliminate undesired skeletons and functional groups, a set of 350 compounds were purchased and screened. Hits were ranked based upon their structural tractability as leads. Optimization of this lead was performed in an iterative manner, making small arrays followed by evaluation of SAR using docking back into the homology model to guide the selection of reagents for subsequent arrays. The end result was **41**, a 42 nM inhibitor of CDK4, which was shown in an X-ray with CDK2 to mimic the main features of the proposed binding mode [119].

With the exception of the de novo ligand building, the design process used by Banyu to discover the CDK4 inhibitor provides a good introduction to the types of computational approaches that are routinely being employed with increasing success. Given the success of the de novo ligand design, it is worth considering why

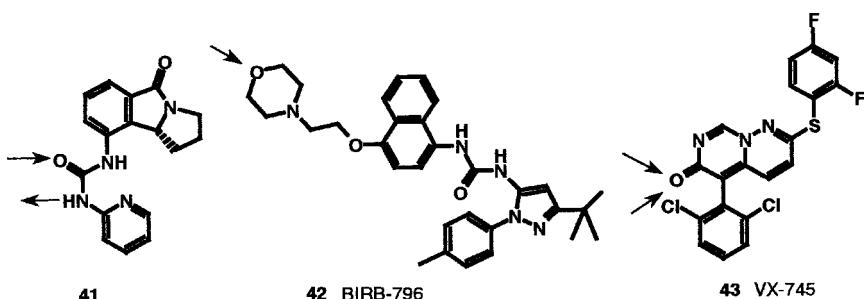
this has not been more widely used. Based upon our own experience, the most likely answer is that it has been tried but has yet to offer any advantage over the other available methods. In the Banyu implementation of de novo design, no “wet” chemistry was performed prior to screening of commercially available ligands. Given this restriction, one could have directly employed an alternate strategy of 3D pharmacophore searching of commercial databases. Such a process of pharmacophore generation and virtual screening has been described by Kinetix Pharmaceuticals [60]. Also Furet et al. from Novartis have reported on two examples where structure-based design was combined with database searching to discover leads for CDK1 and CDK2 [120, 121]. The development of kinase inhibitor libraries and the assembly of kinase screening sets is an another way the generic pharmacophore elements described above and in the Traxler/Bower model have been applied to advance lead discovery. To run a complete HTS screen of one million compounds is time consuming and expensive. As such there is considerable value in developing strategies to assemble smaller kinase inhibitor-enriched screening sets. For example, purvalanol B 22, a potent CDK2 inhibitor, was discovered in a purine-based combinatorial library [122]. Finally, it should be noted that the chances of getting a lead from HTS are increasing as the screening collections are populated with compounds prepared for other kinase inhibitor efforts. For example, compounds from the pyridinylimidazole class of p38 inhibitors have served as leads for c-Raf [123] and Alk5 [124]. For further details on the role of computational chemistry in kinase inhibitor structure-based design strategies and the range of computational tools being applied in this area see a recent overview by Woolfrey and Weston [89].

2.4.4

Selectivity

A realistic goal for inhibitors binding in the ATP site would be to realize that in most cases very high selectivity is not obtainable, is unnecessary and undeterminable. That absolute selectivity is not required is demonstrated by 1, the first kinase inhibitor approved for use in humans, which in addition to inhibiting the Bcr-Abl target, also blocks c-Kit and PDGFR [87]. Such selectivity being undeterminable relates to the technical challenge of developing the methodology to assay all 518 putative human kinases. Setting aside further discussion of what level of selectivity is required for a specific target, the fact remains that if inhibitor selectivity is insufficient the undesired side-effects are expected to be intolerable. Hence, developing strategies to achieve the required level of selectivity for the given target will be a top priority for a kinase drug discovery effort.

As noted earlier, unlike the phosphate-binding region, where the majority of residues are identical across all kinases, the adenine and ribose-binding pockets are lined by a less rigidly conserved set of residues. We have already covered examples illustrating the effect the gatekeeper has on the selectivity of inhibitors that access the gatekeeper-dependent binding pocket. The exploitation of Cys-773 in the EGFR receptor to develop irreversible inhibitors (37) is another example



Scheme 2.6

where a specific residue has been exploited to enhance potency and selectivity. The interpretation of X-ray structural data in light of inhibitor SAR led to the identification of these examples and promises to identify more differences that can be exploited to enhance selectivity (for additional examples see [30, 91]). However, in other cases where selective inhibition has been observed, elucidating the features responsible for selectivity has proven more difficult [64]. Looking further beyond the ATP binding site, several groups have discovered inhibitor binding sites at the solvent interface that impart selectivity. Two examples are the interaction of the sulfonamide of inhibitors **10** [67] and **24** [116] with D86 of CDK2 and the exploitation of the change from lysine (CDK2) to threonine (CDK4) at position of **89** to prepare a CDK4-selective inhibitor, **29** [55].

The conformational mobility of kinases, which is key for controlling the catalytic activity and substrate selectivity of kinases, may also prove to be key for obtaining inhibitor selectivity. Gleevec **1** [27], which inhibits Abl, and BIRB-796 **42** [29], which inhibits p38 α , both bind to catalytically inactive kinase conformations in which the phenylalanine of the conserved DFG motif in the activation loop is displaced by the binding of the inhibitor to the aspartate NH (Scheme 2.6). As noted in Section 2.2.3.1, this behavior has been observed for additional kinases and is likely to have applicability for the design of selective inhibitors of many kinases. Structural studies on the p38-VX-745, **43**, complex have revealed an inhibitor-induced flip of the Gly-110 peptide geometry which allows the inhibitor carbonyl oxygen to form two hydrogen bonds to amide NH groups in the linker region [125]. Researchers at Merck, who first reported this observation, speculate that the Gly-110 of p38 is uniquely suited for this conformational flip relative to kinases which do not have a glycine at this position and that this could explain the enhanced selectivity of **43** for p38. Several new classes of p38 inhibitors have been designed to take advantage of this feature and demonstrate improved kinase selectivity [125, 126].

2.5 Conclusion

Structural insights gained from crystallographic studies with individual members of the kinase family have resulted in the elucidation of a single, universally applicable kinase motif. Similarly, crystallographic studies of kinase-inhibitor complexes are providing an understanding of inhibitor affinity and selectivity that can be broadly applied. As reviewed in this chapter, when coupled with computational techniques, researchers have been able to design new inhibitor scaffolds and improve inhibitor selectivity. Recognition of the high strategic value of having crystallographic information at the beginning of a lead optimization effort is a driving force for improving the techniques to obtain crystalline kinases and the rapid solution of their structure, and if this fails, for the use of a surrogate kinase. This recognition and the elucidation of the general principles of kinase inhibitor binding is also driving the synthesis of kinase inhibitor libraries and the assembly of focus screening sets comprised of compounds which fit these kinase inhibitor descriptors.

It is important to note that current methodologies for the generation of homology models are inadequate for the task of generating the novel conformations that have been revealed by protein crystallography. Consequently, the best starting point for a kinase structure-based lead optimization effort based upon a novel inhibitor scaffold is to solve the structure of the ligand-kinase complex. NMR offers an alternative experimental technique for studying the solution dynamics of protein-ligand interaction and this technique has been successfully used in the design of novel inhibitors of several protein classes [127]. At present, the utility of NMR for the study of kinases is limited by the low resolution of structural information which may be gained and the general difficulty of the experiments. Another promising approach is the application of novel computational techniques to predict kinase-inhibitor structures. However, for this to be realized new approaches must be devised to account for the surprising conformational flexibility of kinases. Perhaps, by supplying a learning set of allowed motions, crystallography can provide the information required for the implementation of new conformational searching and inhibitor-docking algorithms.

Protein crystallography is anticipated to continue its dominant role as a spectroscopic technique to drive the optimization of ATP-competitive kinase inhibitors. Moreover, structural insights revealed by crystallography are likely to be essential for maximizing the potential of alternative non-ATP-competitive inhibitors. Examples of non-ATP-competitive inhibitors that have been explored are antagonists of SH2-binding domains [128] and the MEK inhibitors first discovered by scientists at Parke-Davis [129]. In the case of the SH2-binding domain, protein crystallography has provided a detailed understanding of the structural features required for high affinity binding. However, the identified binding motif presented significant challenges for the development of small molecular weight cell-permeable inhibitors, and thus far chemists have been unable to develop drug-like inhibitors of these protein-protein interaction domains. In contrast, while the binding site of the

MEK inhibitors remains unknown, these compounds have been developed into a compound appropriate for clinical development [130]. In their role as proteins that integrate and transduce cell-signalling stimuli, kinases often have multiple activators and multiple substrates plus additional interactions which modulate their activity. Thus there exist many potential sites for intervention beyond the substrates involved in catalysis (ATP and phosphoacceptor). Finding leads which bind at these sites will require new, non-ATP-centric screening methodologies. Once identified, protein crystallography will again be expected to lead the way towards fully exploiting these new opportunities.

2.6

References

- 1 MANNING, G., WHYTE, D. B., MARTINEZ, R., HUNTER, T., SUDARSANAM, S., *Science* **2002**, *298*, 1912–1934.
- 2 ENGH, R.A., BOSSEMAYER, D., *Pharmacol. Ther.* **2002**, *93*, 99–111.
- 3 HUBBARD, S.R., *Prog. Biophys. Mol. Biol.* **1999**, *71*, 343–358.
- 4 HUBBARD, S.R., *Front Biosci.* **2002**, *7*, d330–d340.
- 5 HUSE, M., KURIYAN, J., *Cell* **2002**, *109*, 275–282.
- 6 SICHERI, F., KURIYAN, J., *Curr. Opin. Struct. Biol.* **1997**, *7*, 777–785.
- 7 SOWADSKI, J.M., EPSTEIN, L.F., LANKIEWICZ, L., KARLSSON, R., *Pharmacol. Ther.* **1999**, *82*, 157–164.
- 8 TAYLOR S.S., KNIGHTON, D.R., ZHENG, J., TEN EYCK, L.F., SOWADSKI, J.M., *Annu. Rev Cell Biol.* **1992**, *8*, 429–462.
- 9 TAYLOR S.S., RADZIO-ANDZELM, E., *Structure* **1994**, *2*, 345–355.
- 10 DE BONDT, H.L., ROSENBLATT, J., JANCA-RIK, J., JONES, H.D., MORGAN, D.O., KIM, S.H., *Nature* **1993**, *363*, 595–602.
- 11 JEFFREY, P.D., RUSSO, A.A., POLYAK, K., GIBBS, E., HURWITZ, J., MASSAGUE, J., PAVLETICH, N.P., *Nature* **1995**, *376*, 313–320.
- 12 KNIGHTON, D.R., ZHENG, J.H., TEN EYCK, L.F., ASHFORD, V.A., XUONG, N.H., TAYLOR, S.S., SOWADSKI, J.M., *Science* **1991**, *253*, 407–414.
- 13 HUBBARD, S.R., *EMBO J.* **1997**, *16*, 5572–5581.
- 14 HUBBARD, S.R., WEI, L., ELLIS, L., HENDRICKSON, W.A., *Nature* **1994**, *372*, 746–754.
- 15 MCTIGUE, M.A., WICKERSHAM, J.A., PINKO, C., SHOWALTER, R.E., PARAST, C.V., TEMPICKY-RUSSELL, A., GEHRING, M.R., MROCKOWSKI, B., KAN, C.C., VILLA-FRANCA, J.E., APPELT, K., *Structure Fold. Des.* **1999**, *7*, 319–330.
- 16 WANG, Z., HARKINS, P.C., ULEVITCH, R.J., HAN, J., COBB, M.H., GOLDSMITH, E.J., *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 2327–2332.
- 17 WILSON, K.P., FITZGIBBON, M.J., CARON, P.R., GRIFFITH, J.P., CHEN, W., McCAFFREY, P.G., CHAMBERS, S.P., SU, M.S., *J. Biol. Chem.* **1996**, *271*, 27696–27700.
- 18 WILSON, K.P., McCAFFREY, P.G., HSIAO, K., PAZHANISAMY, S., GALULLO, V., BEMIS, G.W., FITZGIBBON, M.J., CARON, P.R., MURCKO, M.A., SU, M.S., *Chem. Biol.* **1997**, *4*, 423–431.
- 19 ZHANG, F., STRAND, A., ROBBINS, D., COBB, M.H., GOLDSMITH, E.J., *Nature* **1994**, *367*, 704–711.
- 20 ABLOOGLU, A.J., TILL, J.H., KIM, K., PARANG, K., COLE, P.A., HUBBARD, S.R., KOHANSKI, R.A., *J. Biol. Chem.* **2000**, *275*, 30394–30398.
- 21 BOSSEMAYER, D., ENGH, R.A., KINZEL, V., PONSTINGL, H., HUBER, R., *EMBO J.* **1993**, *12*, 849–859.
- 22 COOK, A., LOWE, E.D., CHRYSINA, E.D., SKAMNAKI, V.T., OIKONOMAKOS, N.G., JOHNSON, L.N., *Biochemistry* **2002**, *41*, 7301–7311.

- 23** MADHUSUDAN, TRAFNY, E.A., XUONG, N.H., ADAMS, J.A., TEN EYCK, L.F., TAYLOR, S.S., SOWADSKI, J.M., *Protein Sci.* **1994**, *3*, 176–187.
- 24** TILL, J.H., ABLOOGLU, A.J., FRANKEL, M., BISHOP, S.M., KOHANSKI, R.A., HUBBARD, S.R., *J. Biol. Chem.* **2001**, *276*, 10049–10055.
- 25** JOHNSON, L.N., NOBLE, M.E., OWEN, D.J., *Cell* **1996**, *85*, 149–158.
- 26** NAGAR, B., BORNMANN, W.G., PELLICENA, P., SCHINDLER, T., VEACH, D.R., MILLER, W.T., CLARKSON, B., KURIYAN, J., *Cancer Res.* **2002**, *62*, 4236–4243.
- 27** SCHINDLER, T., BORNMANN, W., PELLICENA, P., MILLER, W.T., CLARKSON, B., KURIYAN, J., *Science* **2000**, *289*, 1938–1942.
- 28** KOBE, B., HEIERHORST, J., FEIL, S.C., PARKER, M.W., BENIAN, G.M., WEISS, K.R., KEMP, B.E., *EMBO J.* **1996**, *15*, 6810–6821.
- 29** PARGELLIS, C., TONG, L., CHURCHILL, L., CIRILLO, P.F., GILMORE, T., GRAHAM, A.G., GROB, P.M., HICKEY, E.R., MOSS, N., PAV, S., REGAN, J., *Nat. Struct. Biol.* **2002**, *9*, 268–272.
- 30** MOHAMMADI, M., McMAHON, G., SUN, L., TANG, C., HIRTH, P., YEH, B.K., HUBBARD, S.R., SCHLESSINGER, J., *Science* **1997**, *276*, 955–960.
- 31** LAMERS, M.B., ANTSON, A.A., HUBBARD, R.E., SCOTT, R.K., WILLIAMS, D.H., *J. Mol. Biol.* **1999**, *285*, 713–725.
- 32** SCHINDLER, T., SICHERI, F., PICO, A., GAZIT, A., LEVITZKI, A., KURIYAN, J., *Mol. Cell* **1999**, *3*, 639–648.
- 33** SICHERI, F., MOAREFI, I., KURIYAN, J., *Nature* **1997**, *385*, 602–609.
- 34** XU, W., HARRISON, S.C., ECK, M.J., *Nature* **1997**, *385*, 595–602.
- 35** XU, W., DOSHI, A., LEI, M., ECK, M.J., HARRISON, S.C., *Mol. Cell* **1999**, *3*, 629–638.
- 36** ENGH, R.A., GIROD, A., KINZEL, V., HUBER, R., BOSSEMAYER, D., *J. Biol. Chem.* **1996**, *271*, 26157–26164.
- 37** FRODIN, M., ANTAL, T.L., DUMMLER, B.A., JENSEN, C.J., DEAK, M., GAMMELTOFT, S., BIONDI, R.M., *EMBO J.* **2002**, *21*, 5396–5407.
- 38** YANG, J., CRON, P., GOOD, V.M., THOMPSON, V., HEMMINGS, B.A., BARFORD, D., *Nat. Struct. Biol.* **2002**, *9*, 940–944.
- 39** YANG, J., CRON, P., THOMPSON, V., GOOD, V.M., HESS, D., HEMMINGS, B.A., BARFORD, D., *Mol. Cell* **2002**, *9*, 1227–1240.
- 40** ZHENG, J., KNIGHTON, D.R., TEN EYCK, L.F., KARLSSON, R., XUONG, N., TAYLOR, S.S., SOWADSKI, J.M., *Biochemistry* **1993**, *32*, 2154–2161.
- 41** KNIGHTON, D.R., ZHENG, J.H., TEN EYCK, L.F., XUONG, N.H., TAYLOR, S.S., SOWADSKI, J.M., *Science* **1991**, *253*, 414–420.
- 42** SHEWCHUK, L., HASSELL, A., WISELY, B., ROCQUE, W., HOLMES, W., VEAL, J., KUYPER, L.F., *J. Med. Chem.* **2000**, *43*, 133–138.
- 43** WANG, Z., CANAGARAJAH, B.J., BOEHM, J.C., KASSISA, S., COBB, M.H., YOUNG, P.R., ABDEL-MEGUID, S., ADAMS, J.L., GOLDSMITH, E.J., *Structure* **1998**, *6*, 1117–1128.
- 44** HUSE, M., CHEN, Y.G., MASSAGUE, J., KURIYAN, J., *Cell* **1999**, *96*, 425–436.
- 45** WYBENGA-GROOT, L.E., BASKIN, B., ONG, S.H., TONG, J., PAWSON, T., SICHERI, F., *Cell* **2001**, *106*, 745–757.
- 46** GOLDBERG, J., NAIRN, A.C., KURIYAN, J., *Cell* **1996**, *84*, 875–887.
- 47** HU, S.H., PARKER, M.W., LEI, J.Y., WILCE, M.C., BENIAN, G.M., KEMP, B.E., *Nature* **1994**, *369*, 581–584.
- 48** LEI, M., LU, W., MENG, W., PARRINI, M.C., ECK, M.J., MAYER, B.J., HARRISON, S.C., *Cell* **2000**, *102*, 387–397.
- 49** PAUTSCH, A., ZOEPHEL, A., AHORN, H., SPEVAK, W., HAUPTMANN, R., NAR, H., *Structure (Camb.)* **2001**, *9*, 955–965.
- 50** SUZENBACHER, G., GRUEZ, A., ROIG-ZAMBONI, V., SPINELLI, S., VALENCIA, C., PAGOT, F., VINCENTELLI, R., BIGNON, C., SALOMONI, A., GRISEL, S., MAURIN, D., HUYGHE, C., JOHANSSON, K., GRASSICK, A., ROUSSEL, A., BOURNE, Y., PERRIER, S., MIALLAU, L., CANTAU, P., BLANC, E., GENEVOIS, M., GROSSI, A., ZENATTI, A., CAMPANACCI, V., CAMBILLAU, C., *Acta Crystallogr. D. Biol. Crystallogr.* **2002**, *58*, 2109–2115.

- 51 TARRICONE, C., DHAVAN, R., PENG, J., ARECES, L.B., TSAI, L.H., MUSACCHIO, A., *Mol. Cell* **2001**, *8*, 657–669.
- 52 JEFFREY, P.D., TONG, L., PAVLETICH, N.P., *Genes Dev.* **2000**, *14*, 3115–3125.
- 53 BAX, B., CARTER, P.S., LEWIS, C., GUY, A.R., BRIDGES, A., TANNER, R., PETTMAN, G., MANNIX, C., CULBERT, A.A., BROWN, M.J., SMITH, D.G., REITH, A.D., *Structure (Camb.)* **2001**, *9*, 1143–1152.
- 54 MOHAMMADI, M., SCHLESSINGER, J., HUBBARD, S.R., *Cell* **1996**, *86*, 577–587.
- 55 IKUTA, M., KAMATA, K., FUKASAWA, K., HONMA, T., MACHIDA, T., HIRAI, H., SUZUKI-TAKAHASHI, I., HAYAMA, T., NISHIMURA, S., *J. Biol. Chem.* **2001**, *276*, 27548–27554.
- 56 BINNS, K.L., TAYLOR, P.P., SICHERI, F., PAWSON, T., HOLLAND, S.J., *Mol. Cell Biol.* **2000**, *20*, 4791–4805.
- 57 ARNOLD, L.D., DIXON, R., TALANIAN, R., GAZA-BULSECO, G., BUMP, N., SESHDARI, T., BELLAMACINA, C., ALLEN, K., HOFFKEN, W., RIEBEL, A., XU, Y., abstract for *Proceedings of the American Association for Cancer Research* **2002**, *43*, 848.
- 58 BELLON, S., FITZGIBBON, M.J., FOX, T., HSIAO, H.M., WILSON, K.P., *Structure Fold. Des.* **1999**, *7*, 1057–1065.
- 59 TRAXLER, P., FURET, P., *Pharmacol. Ther.* **1999**, *82*, 195–206.
- 60 TOLEDO, L.M., LYDON, N.B., *Structure* **1997**, *5*, 1551–1556.
- 61 DE AZEVEDO, W.F., LECLERC, S., MEIJER, L., HAVLICEK, L., STRNAD, M., KIM, S.H., *Eur. J. Biochem.* **1997**, *243*, 518–526.
- 62 TRAXLER, P., GREEN, J., METT, H., SE-QUIN, U., FURET, P., *J. Med. Chem.* **1999**, *42*, 1018–1026.
- 63 NIEFIND, K., PUTTER, M., GUERRA, B., IS-SINGER, O.G., SCHOMBURG, D., *Nat. Struct. Biol.* **1999**, *6*, 1100–1103.
- 64 MEIJER, L., THUNNISSEN, A.M., WHITE, A.W., GARNIER, M., NIKOLIC, M., TSAI, L.H., WALTER, J., CLEVERLEY, K.E., SALINAS, P.C., WU, Y.Z., BIERNAT, J., MANDELKOW, E.M., KIM, S.H., PETIT, G.R., *Chem. Biol.* **2000**, *7*, 51–63.
- 65 HOESSEL, R., LECLERC, S., ENDICOTT, J.A., NOBEL, M.E., LAWRIE, A., TUNNAH, P., LEOST, M., DAMIENS, E., MARIE, D., MARKO, D., NIEDERBERGER, E., TANG, W., EISENBRAND, G., MEIJER, L., *Nat. Cell Biol.* **1999**, *1*, 60–67.
- 66 DAVIES, T.G., BENTLEY, J., ARRIS, C.E., BOYLE, F.T., CURTIN, N.J., ENDICOTT, J.A., GIBSON, A.E., GOLDING, B.T., GRiffin, R.J., HARDCastle, I.R., JEWSBURY, P., JOHNSON, L.N., MESGUICHE, V., NEWELL, D.R., NOBLE, M.E., TUCKER, J.A., WANG, L., WHITFIELD, H.J., *Nat. Struct. Biol.* **2002**, *9*, 745–749.
- 67 BRAMSON, H.N., CORONA, J., DAVIS, S.T., DICKERSON, S.H., EDELSTEIN, M., FRYE, S.V., GAMPE, R.T., JR., HARRIS, P.A., HASSELL, A., HOLMES, W.D., HUNTER, R.N., LACKEY, K.E., LOVEJOY, B., LUZZIO, M.J., MONTANA, V., ROCQUE, W.J., RUSNAK, D., SHEWCHUK, L., VEAL, J.M., WALKER, D.H., KUYPER, L.F., *J. Med. Chem.* **2001**, *44*, 4339–4358.
- 68 NARAYANA, N., DILLER, T.C., KOIDE, K., BUNNAGE, M.E., NICOLAOU, K.C., BRUNTON, L.L., XUONG, N.H., TEN EYCK, L.F., TAYLOR, S.S., *Biochemistry* **1999**, *38*, 2367–2376.
- 69 KOIDE, K., BUNNAGE, M.E., GOMEZ, P.L., KANTER, J.R., TAYLOR, S.S., BRUNTON, L.L., NICOLAOU, K.C., *Chem. Biol.* **1995**, *2*, 601–608.
- 70 HUNENBERGER, P.H., HELMS, V., NARAYANA, N., TAYLOR, S.S., MCCAMMON, J.A., *Biochemistry* **1999**, *38*, 2358–2366.
- 71 TONG, L., PAV, S., WHITE, D.M., ROGERS, S., CRANE, K.M., CYWIN, C.L., BROWN, M.L., PARGELLIS, C.A., *Nat. Struct. Biol.* **1997**, *4*, 311–316.
- 72 MOHAMMADI, M., FROUM, S., HAMBY, J.M., SCHROEDER, M.C., PANEK, R.L., LU, G.H., ELISEENKOVA, A.V., GREEN, D., SCHLESSINGER, J., HUBBARD, S.R., *EMBO J.* **1998**, *17*, 5896–5904.
- 73 ZHU, X., KIM, J.L., NEWCOMB, J.R., ROSE, P.E., STOVER, D.R., TOLEDO, L.M., ZHAO, H., MORGENSTERN, K.A., *Structure Fold. Des.* **1999**, *7*, 651–661.
- 74 LIU, Y., BISHOP, A., WITUCKI, L., KRAYBILL, B., SHIMIZU, E., TSien, J., UBERSAX, J., BLETHROW, J., MORGAN, D.O., SHOKAT, K.M., *Chem. Biol.* **1999**, *6*, 671–678.
- 75 EYERS, P.A., CRAXTON, M., MORRICE, N., COHEN, P., GOEDERT, M., *Chem. Biol.* **1998**, *5*, 321–328.
- 76 GUM, R.J., McLAUGHLIN, M.M., KUMAR, S., WANG, Z., BOWER, M.J., LEE, J.C.,

- ADAMS, J. L., LIVI, G. P., GOLDSMITH, E. J., YOUNG, P. R., *J. Biol. Chem.* **1998**, *273*, 15605–15610.
- 77** ADAMS, J. L., BOEHM, J. C., GALLAGHER, T. F., KASSIS, S., WEBB, E. F., HALL, R., SORENSEN, M., GARIGIPATI, R., GRISWOLD, D. E., LEE, J. C., *Bioorg. Med. Chem. Lett.* **2001**, *11*, 2867–2870.
- 78** BARKER, A. J., GIBSON, K. H., GRUNDY, W., GODFREY, A. A., BARLOW, J. J., HEALY, M. P., WOODBURN, J. R., ASHTON, S. E., CURRY, B. J., SCARLETT, L., HENTHORN, L., RICHARDS, L., *Bioorg. Med. Chem. Lett.* **2001**, *11*, 1911–1914.
- 79** HENNEQUIN, L. F., STOKES, E. S., THOMAS, A. P., JOHNSTONE, C., PLE, P. A., OGILVIE, D. J., DUKES, M., WEDGE, S. R., KENDREW, J., CURWEN, J. O., *J. Med. Chem.* **2002**, *45*, 1300–1312.
- 80** STAMOS, J., SLIWKOWSKI, M. X., EIGENBROT, C., *J. Biol. Chem.* **2002**, *277*, 46265–46272.
- 81** HUBBARD, S. R., TILL, J. H., *Annu. Rev. Biochem.* **2000**, *69*, 373–398.
- 82** KIM, S. H., SCHULZE-GAHMEN, U., BRANDSEN, J., AZEVEDO JR., W. F., *Prog. Cell Cycle Res.* **1996**, *2*, 137–145.
- 83** JOHNSON, L. N., LEWIS, R. J., *Chem. Rev.* **2001**, *101*, 2209–2242.
- 84** PRADE, L., ENGH, R. A., GIROD, A., KINZEL, V., HUBER, R., BOSSEMAYER, D., *Structure* **1997**, *5*, 1627–1637.
- 85** KLEBE, G., BOHM, H. J., *J. Recept. Signal Transduct. Res.* **1997**, *17*, 459–473.
- 86** DAVIES, T. G., TUNNAH, P., MEIJER, L., MARKO, D., EISENBRAND, G., ENDICOTT, J. A., NOBLE, M. E., *Structure (Camb.)* **2001**, *9*, 389–397.
- 87** DRUKER, B. J., *Oncogene* **2002**, *21*, 8541–8546.
- 88** LAWRENCE, D. S., NIU, J., *Pharmacol. Ther.* **1998**, *77*, 81–114.
- 89** WOOLFREY, J. R., WESTON, G. S., *Curr. Pharm. Des* **2002**, *8*, 1527–1545.
- 90** LAWRIE, A. M., NOBLE, M. E., TUNNAH, P., BROWN, N. R., JOHNSON, L. N., ENDICOTT, J. A., *Nat. Struct. Biol.* **1997**, *4*, 796–801.
- 91** ZHAO, B., BOWER, M. J., McDEVITT, P. J., ZHAO, H., DAVIS, S. T., JOHANSON, K. O., GREEN, S. M., CONCHA, N. O., ZHOU, B. B., *J. Biol. Chem.* **2002**, *277*, 46609–46615.
- 92** MENG, W., SWENSON, L. L., FITZGIBBON, M. J., HAYAKAWA, K., TER HAAR, E., BEHRENS, A. E., FULGHUM, J. R., LIPPKE, J. A., *J. Biol. Chem.* **2002**, *277*, 37401–37405.
- 93** CHEETHAM, G. M., KNEGTEL, R. M., COLL, J. T., RENWICK, S. B., SWENSON, L., WEBER, P., LIPPKE, J. A., AUSTEN, D. A., *J. Biol. Chem.* **2002**, *277*, 42419–42422.
- 94** NOWAKOWSKI, J., CRONIN, C. N., MCREE, D. E., KNUTH, M. W., NELSON, C. G., PAVLETICH, N. P., ROGERS, J., SANG, B. C., SCHEIBE, D. N., SWANSON, R. V., THOMPSON, D. A., *Structure (Camb.)* **2002**, *10*, 1659–1667.
- 95** SENDEROWICZ, A. M., *Invest New Drugs* **1999**, *17*, 313–320.
- 96** JIROUSEK, M. R., GILLIG, J. R., GONZALEZ, C. M., HEATH, W. F., McDONALD, J. H., III, NEEL, D. A., RITO, C. J., SINGH, U., STRAMM, L. E., MELIKIAN-BADALIAN, A., BAEVSKY, M., BAILAS, L. M., HALL, S. E., WINNEROSKI, L. L., FAUL, M. M., *J. Med. Chem.* **1996**, *39*, 2664–2671.
- 97** BOLD, G., ALTMANN, K. H., FREI, J., LANG, M., MANLEY, P. W., TRAXLER, P., WIETFELD, B., BRUGGEN, J., BUCHDUNGER, E., COZENS, R., FERRARI, S., FURET, P., HOFMANN, F., MARTINY-BARON, G., MESTAN, J., ROSEL, J., SILLS, M., STOVER, D., ACEMOGLU, F., BOSS, E., EMMENEGGER, R., LASER, L., MASSO, E., ROTH, R., SCHIACHTER, C., VETTERLI, W., *J. Med. Chem.* **2000**, *43*, 2310–2323.
- 98** BRIDGES, A. J., *Chem. Rev.* **2001**, *101*, 2541–2572.
- 99** SUN, L., TRAN, N., LIANG, C., HUBBARD, S., TANG, F., LIPSON, K., SCHRECK, R., ZHOU, Y., McMAHON, G., TANG, C., *J. Med. Chem.* **2000**, *43*, 2655–2663.
- 100** ZIMMERMANN, J., CARAVATTI, G., METT, H., MEYER, T., MULLER, M., LYDON, N. B., FABBRO, D., *Arch. Pharm. (Weinheim)* **1996**, *329*, 371–376.
- 101** FURET, P., CARAVATTI, G., LYDON, N., PRIESTLE, J. P., SOWADSKI, J. M., TRINKS, U., TRAXLER, P., *J. Comput. Aided Mol. Des.* **1995**, *9*, 465–472.
- 102** SCAPIN, G., *Drug Discov. Today* **2002**, *7*, 601–611.
- 103** GARCIA-ECHEVERRIA, C., TRAXLER, P., EVANS, D. B., *Med. Res. Rev.* **2000**, *20*, 28–57.

- 104** BIT, R. A., DAVIS, P. D., ELLIOTT, L. H., HARRIS, W., HILL, C. H., KEECH, E., KUMAR, H., LAWTON, G., MAW, A., NIXON, J. S., *J. Med. Chem.* **1993**, *36*, 21–29.
- 105** TRINKS, U., BUCHDUNGER, E., FURET, P., KUMP, W., METT, H., MEYER, T., MULLER, M., REGENASS, U., RIHS, G., LYDON, N., *J. Med. Chem.* **1994**, *37*, 1015–1027.
- 106** TRAXLER, P., BOLD, G., FREI, J., LANG, M., LYDON, N., METT, H., BUCHDUNGER, E., MEYER, T., MUELLER, M., FURET, P., *J. Med. Chem.* **1997**, *40*, 3601–3616.
- 107** TRAXLER, P. M., FURET, P., METT, H., BUCHDUNGER, E., MEYER, T., LYDON, N., *J. Med. Chem.* **1996**, *39*, 2285–2292.
- 108** FRY, D. W., KRAKER, A. J., McMICHAEL, A., AMBROSO, L. A., NELSON, J. M., LEOPOLD, W. R., CONNORS, R. W., BRIDGES, A. J., *Science* **1994**, *265*, 1093–1095.
- 109** GIBSON, K. H., GRUNDY, W., GODFREY, A. A., WOODBURN, J. R., ASHTON, S. E., CURRY, B. J., SCARLETT, L., BARKER, A. J., BROWN, D. S., *Bioorg. Med. Chem. Lett.* **1997**, *7*, 2723–2728.
- 110** PALMER, B. D., TRUMPP-KALLMEYER, S., FRY, D. W., NELSON, J. M., SHOWALTER, H. D., DENNY, W. A., *J. Med. Chem.* **1997**, *40*, 1519–1529.
- 111** FRY, D. W., BRIDGES, A. J., DENNY, W. A., DOHERTY, A., GREIS, K. D., HICKS, J. L., HOOK, K. E., KELLER, P. R., LEOPOLD, W. R., LOO, J. A., McNAMARA, D. J., NELSON, J. M., SHERWOOD, V., SMAILL, J. B., TRUMPP-KALLMEYER, S., DOBRUSIN, E. M., *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 12022–12027.
- 112** SCHULZE-GAHMEN, U., BRANDSEN, J., JONES, H. D., MORGAN, D. O., MEIJER, L., VESELY, J., KIM, S. H., *Proteins* **1995**, *22*, 378–391.
- 113** TRUMPP-KALLMEYER, S., RUBIN, J. R., HUMBLET, C., HAMBY, J. M., HOLLIS SHOWALTER, H. D., *J. Med. Chem.* **1998**, *41*, 1752–1763.
- 114** BARVIAN, M., BOSCHELLI, D. H., COSSROW, J., DOBRUSIN, E., FATTAEY, A., FRITSCH, A., FRY, D., HARVEY, P., KELLER, P., GARRETT, M., LA, F., LEOPOLD, W., McNAMARA, D., QUIN, M., TRUMPP-KALLMEYER, S., TOOGOOD, P., WU, Z., ZHANG, E., *J. Med. Chem.* **2000**, *43*, 4606–4616.
- 115** DE, A. W., Jr., MUELLER-DIECKMANN, H. J., SCHULZE-GAHMEN, U., WORLAND, P. J., SAUSVILLE, E., KIM, S. H., *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 2735–2740.
- 116** DAVIES, T. G., PRATT, D. J., ENDICOTT, J. A., JOHNSON, L. N., NOBLE, M. E., *Pharmacol. Ther.* **2002**, *93*, 125–133.
- 117** PARANG, K., TILL, J. H., ABLOOGLU, A. J., KOHANSKI, R. A., HUBBARD, S. R., COLE, P. A., *Nat. Struct. Biol.* **2001**, *8*, 37–41.
- 118** PARANG, K., COLE, P. A., *Pharmacol. Ther.* **2002**, *93*, 145–157.
- 119** HONMA, T., HAYASHI, K., AOYAMA, T., HASHIMOTO, N., MACHIDA, T., FUKASAWA, K., IWAMA, T., IKEURA, C., IKUTA, M., SUZUKI-TAKAHASHI, I., IWASAWA, Y., HAYAMA, T., NISHIMURA, S., MORISHIMA, H., *J. Med. Chem.* **2001**, *44*, 4615–4627.
- 120** FURET, P., MEYER, T., STRAUSS, A., RACUGLIA, S., RONDEAU, J. M., *Bioorg. Med. Chem. Lett.* **2002**, *12*, 221–224.
- 121** FURET, P., MEYER, T., MITTL, P., FRETZ, H., *J. Comput. Aided Mol. Des.* **2001**, *15*, 489–495.
- 122** GRAY, N. S., WODICKA, L., THUNNISSEN, A. M., NORMAN, T. C., KWON, S., ESPINOZA, F. H., MORGAN, D. O., BARNES, G., LECLERC, S., MEIJER, L., KIM, S. H., LOCKHART, D. J., SCHULTZ, P. G., *Science* **1998**, *281*, 533–538.
- 123** ADAMS, J. L., LEE, D., *Curr. Opin. Drug Discov. Dev.* **1999**, *2*, 96–109.
- 124** CALLAHAN, J. F., BURGESS, J. L., FORNWALD, J. A., GASTER, L. M., HARLING, J. D., HARRINGTON, F. P., HEER, J., KWON, C., LEHR, R., MATHUR, A., OLSON, B. A., WEINSTOCK, J., LAPING, N. J., *J. Med. Chem.* **2002**, *45*, 999–1001.
- 125** COLLETTI, S. L., FRIE, J. L., DIXON, E. C., SINGH, S. B., CHOI, B. K., SCAPIN, G., FITZGERALD, C. E., KUMAR, S., NICHOLS, E. A., O'KEEFE, S. J., O'NEILL, E. A., PORTER, G., SAMUEL, K., SCHMATZ, D. M., SCHWARTZ, C. D., SHOOP, W. L., THOMPSON, C. M., THOMPSON, J. E., WANG, R., WOODS, A., ZALLER, D. M., DOHERTY, J. B., *J. Med. Chem.* **2003**, *46*, 349–352.
- 126** STELMACH, J. E., LIU, L., PATEL, S. B., PIVNICHNY, J. V., SCAPIN, G., SINGH, S., HOP, C. E., WANG, Z., STRAUSS, J. R., CAMERON, P. M., NICHOLS, E. A., O'KEEFE, S. J., O'NEILL, E. A., SCHMATZ, D. M., SCHWARTZ, C. D., THOMPSON, C. M., ZALLER, D. M., DOHERTY, J. B., *Bioorg. Med. Chem. Lett.* **2003**, *13*, 277–280.

- 127 SHUKER, S. B., HAJDUK, P. J., MEADOWS, R. P., FESIK, S. W., *Science* **1996**, *274*, 1531–1534.
- 128 CODY, W. L., LIN, Z., PANEK, R. L., ROSE, D. W., RUBIN, J. R., *Curr. Pharm. Des.* **2000**, *6*, 59–98.
- 129 PANG, L., SAWADA, T., DECKER, S. J., SALTIEL, A. R., *J. Biol. Chem.* **1995**, *270*, 13585–13588.
- 130 HERRERA, R., SEBOLT-LEOPOLD, J. S., *Trends Mol. Med.* **2002**, *8*, S27–S31.

3

The Proteasome as a Drug Target

TSUNEHIRO MIZUSHIMA and TOMITAKE TSUKIHARA

3.1

Introduction

The ubiquitin-proteasome pathway is the major proteolytic system in the cytosol and nucleus of all eukaryotic cells. The proteasome, a multisubunit proteolytic complex, is the central enzyme underlying nonlysosomal protein degradation. The ubiquitin-proteasome pathway also plays an important role in the regulation of many physiological processes as well as in the development of a number of major human diseases. For example, degradation of the tumor suppressor p53 [1], and p27^{Kip1}, an inhibitor of cyclin-dependent kinases [2], can promote tumorigenesis. Furthermore, the ubiquitin-proteasome pathway also plays an essential role in immune surveillance [3], muscle atrophy [4] and regulation of metabolic pathways [5–7]. Recent studies have shown that proteasome inhibitors potently induce apoptosis in many types of cancer cells, but exhibit reduced cytotoxicity in normal cells [8, 9]. Crystal structures of proteasome-inhibitor complexes revealed the specific inhibitor domains essential for proteasome binding.

3.2

The Ubiquitin-Proteasome System

The ubiquitin-proteasome pathway is an important mediator of intracellular proteolysis in eukaryotes (Fig. 3.1). Protein degradation in eukaryotic cells is a means of controlling the level of intracellular proteins rapidly, irreversibly and in a timely manner, allowing precise regulation of cellular function. The ubiquitin system acts by covalently attaching ubiquitin to selected substrate proteins using a cascade of three enzymes termed E1 [ubiquitin (Ub)-activating enzyme], E2 [Ub-conjugating enzyme] and E3 (ubiquitin ligase) [10–12]. Ubiquitin is bound reversibly to proteins by an isopeptide linkage between the carboxyl-terminus of ubiquitin and a lysine ε-amino group in the substrate protein. Protein ubiquitination is initiated by the ATP-dependent formation of a high-energy thioester bond between E1 and the C-terminus of ubiquitin. The activated ubiquitin is then transferred to one of several E2s, forming a thioester bond with E2 that catalyzes the formation

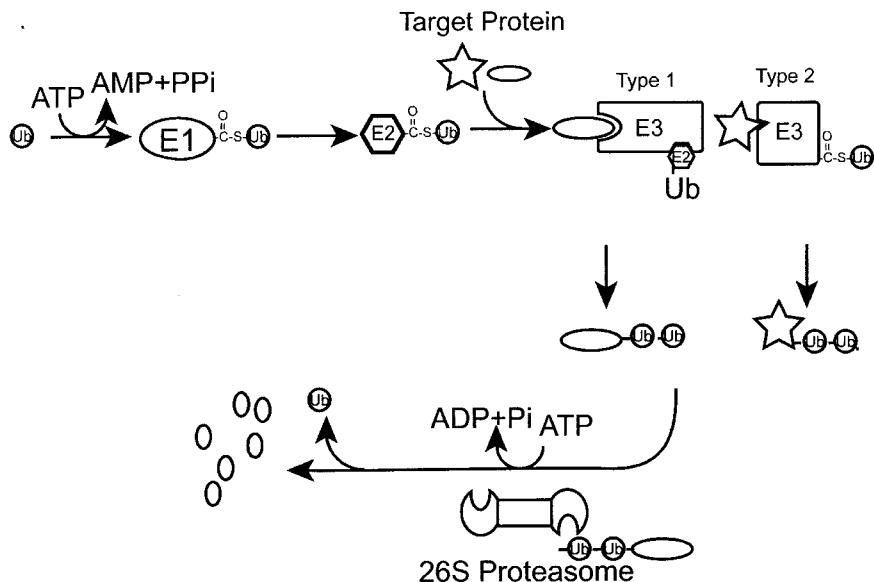


Fig. 3.1 The ubiquitin-proteasome system. Protein substrates are conjugated to multiple molecules of ubiquitin by a three-step process

mediated by E1, E2 and E3. The ubiquitinated substrate is rapidly degraded by the 26S proteasome.

of the isopeptide bond between ubiquitin and the substrate protein. In some cases, ubiquitin is transferred directly to target proteins by E2 but, frequently, the additional participation of E3 is required. Finally, ubiquitin conjugation to a lysine residue of a substrate protein-attached ubiquitin moiety usually results in the formation of poly-ubiquitin chains, which serves as a signal for degradation by the proteasome. In the ubiquitination pathway, E3 plays a decisive role in the selection of proteins to be degraded, because it specifically binds to protein substrates [1, 13]. The precise role played by E3 is still not clear; there seem to be multiple mechanisms by which E3 and its two subtypes, type 1 and type 2, recognize target proteins. Type-1 E3 is thought to ligate E2 by associating with the target protein and E2, and type-2 E3 is linked to ubiquitin via a thioester bond, which acts directly to form an isopeptide bond between ubiquitin and the substrate protein. Thus, the ubiquitin-proteasome system is a two-step process: firstly, the substrate is identified by the ubiquitination machinery, and secondly, the poly-ubiquitylated protein is recognized by the proteasome.

3.2.1

Role of the Ubiquitin-Proteasome System

Apoptosis is a major form of cell death and is important both in controlling cell number during development and in the removal of damaged cells. Defective apoptosis is involved in the pathogenesis of several diseases including certain cancers,

such as B-cell chronic lymphocytic leukaemia, where there is an accumulation of quiescent tumor cells. The ubiquitin-proteasome system is an important regulator of cell growth and apoptosis. It regulates the level of many proteins involved in the control of apoptosis, including some members of the Bcl-2 family, which comprises a number of inhibitors of apoptosis proteins [14], and some caspases. Several lines of evidence indicate abnormal accumulation of ubiquitin-conjugated proteins in a wide variety of neurodegenerative disorders that often lead to apoptotic cell death. Accordingly, alterations to the regulation of the ubiquitin pool in cells may play an important role in the pathogenesis of such diseases. Alternatively, functional disorders of the proteasome may also cause abnormal accumulation of ubiquitinated proteins in the cells, which could then lead to apoptosis.

Proteasome inhibition affects the stability of various cell-cycle regulatory proteins. Cyclins, cyclin-dependent kinase inhibitors and tumor suppressors (e.g., cycline B1, p27, p53) are all substrates for the ubiquitin-proteasome pathway, and inhibiting their requisite degradation has been shown clearly to sensitize cells to apoptosis. For examples, degradation of p53, a tumor suppressor, and p27, an inhibitor of cyclin-dependent kinases, can promote tumorigenesis. Many currently used anticancer agents exert their anticancer effects by inducing apoptosis thus conferring a resistance to tumor progression. Therefore, clinical manipulation of the proteasome could be useful in the treatment of cancer.

3.2.2

26S Proteasome

Proteasomes are unusually large multisubunit proteolytic complexes consisting of two primary components: (1) a central catalytic machine (equivalent to the 20S proteasome) and (2) two terminal regulatory subcomplexes termed PA700 that are attached to both ends of the central core in opposite orientations, forming the enzymatically active 26S proteasome [15, 16]. The 26S proteasome is an approximately 2.5 MDa complex made up of two copies each of at least 31 different subunits, which are highly conserved among all eukaryotes. PA700 itself can be further dissected into two multisubunit substructures, a lid and a base. The base most probably unfolds substrates and translocates them into the 20S proteasome [17, 18]. The role of the lid is still unclear.

3.2.3

20S Proteasome

The 20S proteasome is a protease complex comprising 28 subunits. It is a barrel-shaped structure formed by the axial stacking of four rings made up of two outer α rings and two inner β rings arranged in $\alpha\beta\beta\alpha$ order. Proteasomes are members of the N-terminal nucleophile- (Ntn-) hydrolase superfamily [19]. Their N-terminal threonine residues are exposed as the nucleophile in peptide bond hydrolysis [20, 21]. In eukaryotic cells, three of the β -type subunits have N-terminal threonine residues, are active and have specificities determined largely by the nature of their

S1 pockets [22]. The three major activities, the peptidylglutamyl-hydrolyzing (PGPH), trypsin-like and chymotrypsin-like activities, have been assigned to the three active subunits of $\beta 1$, $\beta 2$ and $\beta 5$, respectively, based on mutational and crystal structure analyses. Furthermore, it has been claimed, based on studies with model peptides and inactivation by inhibitors, that mammalian proteasomes also contain two other peptidase activities, referred to as “branched-chain amino acid-preferring” and “small neutral amino acid-preferring” (SNAAP).

Mammals have seven different α and ten different β proteasome subunit genes [23, 24]. In jawed vertebrates, three additional β subunits, $\beta 1i$, $\beta 2i$ and $\beta 5i$, induced by interferon- γ (IFN γ), a major immunomodulatory cytokine [25–27], are catalytically active and replace the constitutively active subunits $\beta 1$, $\beta 2$ and $\beta 5$, respectively, leading to the formation of the “immunoproteasome”, a structure that is responsible for immunological processing of intracellular antigens and that enhances the generation of ligands for class I molecules of the major histocompatibility (MHC) complex [3, 28–30]. $\beta 1i$ and $\beta 5i$ are encoded by genes in the MHC class II gene region adjacent to transporters associated with antigen-processing genes whose products are involved in the transport of antigenic peptides from the cytosol to the endoplasmic reticulum [31].

In mammalian cells, IFN γ also induces expression of another proteasome activator, consisting of several α and several β subunits, known alternately as PA28 or the 11S regulator (REG) [16, 32]. PA28, an activator of the 20S proteasome, associates with the 20S proteasome independent of ATP binding to form a football-like structure [33] and markedly stimulates the activities of various peptidases of the 20S proteasome *in vitro*. Unlike the 26S proteasome, this complex, however, fails to enhance the hydrolysis of large protein substrates with native or denatured structures, even when they are ubiquitinated. Recently, it was found that a “hybrid proteasome”, comprising the 20S proteasome flanked by PA28 on one side and PA700 on the other, functions as a unique ATP-dependent protease [34, 35]. Indeed, it has been shown that PA28 and PA700 can bind to the 20S proteasome simultaneously [36]. Another *in vitro* study showed that PA28 alone could enhance the generation of antigenic peptides by inducing dual substrate cleavage by the 20S proteasome [37]. The counterpart genes of PA28, however, are not present in budding yeast [24]. These properties indicate that the 20S proteasome in higher eukaryotes plays a special role in the immune system.

3.3

Structure of the 20S Proteasome

The overall shape of the bovine 20S proteasome is an elongated cylinder with large central cavities and narrow constrictions [38]. The approximate length and diameters of the bovine 20S proteasome are 150 and 115 Å, respectively (Fig. 3.2). The general architecture of the 20S proteasome was completely conserved among *T. acidophilum*, yeast and mammals. The arrangement of subunits in the bovine 20S proteasome is identical with that in the yeast 20S proteasome.

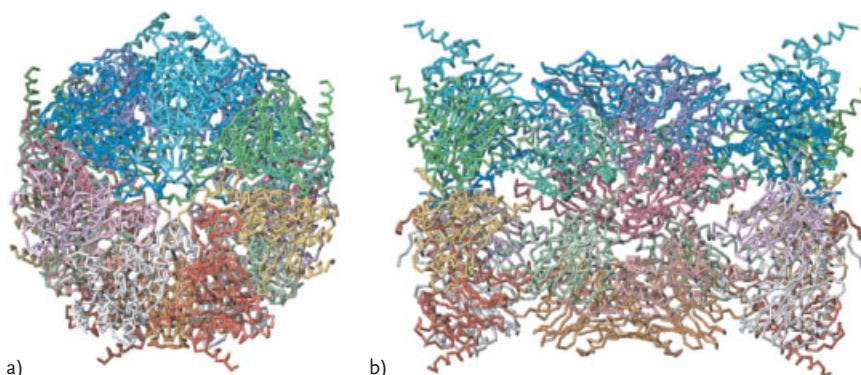


Fig. 3.2 (a) Top view of the bovine 20S proteasome showing Ca atoms only. (b) Side view of the bovine 20S proteasome.

All α subunits have a β -sandwich structure (Fig. 3.3 a) formed by two five-anti-parallel β -sheets with topology S8, S1, S2, S9 and S10 in the upper β -sheet and S7, S6, S5, S4 and S3 in the lower β -sheet. The sandwich structure is surrounded by helical layers comprising helices H1 and H2 on one side and H3, H4 and H5 on the other, as seen in *T. acidophilum* and yeast. In every α subunit, the H0 helix is found on the N-terminal side of S1. All β subunits of the bovine proteasome, like those in *T. acidophilum* and yeast, exhibit a β -sandwich structure (Fig. 3.3 b) similar to that of the α subunits, consisting of strands S1 to S10. The polypeptide chain of the β subunits starts with strand S1; the absence of helix H0 provides access to the interior of β -sandwich. The bovine β -type subunits exhibit structural

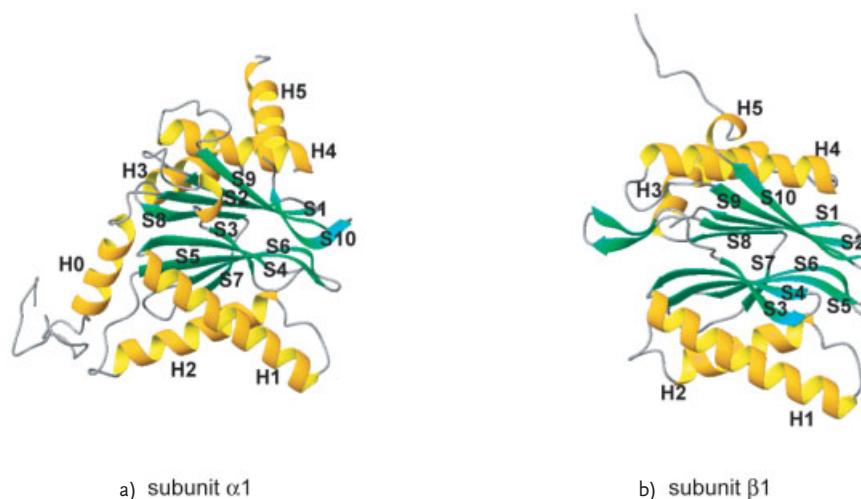


Fig. 3.3 (a) Ribbon drawing of the $\alpha1$ subunit. (b) Ribbon drawing of the $\beta1$ subunit. α Helices are labeled H, β strands S.

variations in the N-terminal residues preceding Thr-1 in the numbering system of the *T. acidophilum* proteasome and in the C-terminal residues following S10. Each β subunit structure of the bovine enzyme was almost identical with that of the corresponding subunit of the yeast enzyme. The post-translational processing of eukaryotic β subunits found in the yeast proteasome was confirmed in the structure of the bovine 20S proteasome. The three active subunits β 1, β 2 and β 5 had an N-terminal Thr residue, similar to all β subunits in *T. acidophilum* and active β subunits in yeast. The bovine subunits β 6 and β 7 had 10 and 8 extra N-terminal residues, respectively, compared with the active subunits. These were partly processed during the structural organization. The bovine β 4 subunits preserved the Met residue at the N-terminus.

3.3.1

Active Sites of Eukaryotic 20S Proteasomes

A catalytic system formed by Thr-1, Glu-17 and Lys-33 has been defined in the *T. acidophilum* proteasome by structural and mutational studies. Close to Thr-1 are residues Ser-129, Ser-169 and Asp-166, which seem to be required for the structural integrity of the site but may also be involved in catalysis. These residues are invariant in the active subunits. The structure of the active sites of the bovine proteasome were compared by superimposing the functional amino acid residues Thr-1, Glu-17, Arg-19, Lys-33, Ser-129, Asp-166, Ser-169 and Gly-170. Root mean square (R.m.s.) deviations for all atoms of these functional residues in the active bovine subunit pairs β 1– β 2, β 1– β 5, and β 2– β 5 were 0.4 Å, 0.3 Å and 0.4 Å, respectively, demonstrating that the functional core of each active subunit is well conserved.

The α ring of the bovine proteasome was superimposed onto the β ring of the yeast proteasome by a least-squares fitting of main chain atoms. R.m.s. deviations between all atoms of the functional amino acids in the bovine β 1, β 2 and β 5 subunits and the corresponding subunits in the yeast 20S proteasome were 0.3 Å, 0.3 Å and 0.2 Å, respectively, significantly smaller than the mean value of 0.8 Å for the main chain atoms of β 1, β 2 and β 5. This indicated that the catalytic sites including the substrate-binding regions of β 1, β 2 and β 5 were well conserved between the bovine and yeast proteasomes.

3.3.2

Novel Ntn-Hydrolase Active Site of the β 7 Subunit

The locations of the following functional groups around the N-terminal Thr-1 of the bovine β 7 subunit are consistent with the structure of an Ntn-hydrolase-active site (Fig. 3.4A). The N-terminal Thr-1 forms a hydrogen bond with Asn-104 O δ . Thr-1 O γ -H forms a hydrogen bond with Asp-59 O δ . Arg-91 of β 1 forms a salt bridge with Asp-56 of β 7. An oxyanion hole is formed by the Tyr-88 OH of β 1 or Arg-99 N η of β 7. A water molecule is found near Thr-1 and is replaced by substrate upon formation of the enzyme-substrate complex (Fig. 3.4B). Although the

locations of functional groups are different from those in $\beta 1$, $\beta 2$ and $\beta 5$, a reasonable reaction mechanism for substrate hydrolysis can be developed based on this arrangement (Fig. 3.5) [39]. These structural features suggest that the Thr-1 of bovine $\beta 7$ is part of a novel Ntn-hydrolase-active site. Furthermore, the S1 pocket of $\beta 7$ is much smaller than those in $\beta 1$, $\beta 2$ or $\beta 5$ (Fig. 3.4C). This active site should have SNAAP activity [40].

The Thr-1 residues of the bovine $\beta 7$ subunit and active subunits were far from one another when the main chain atoms of the subunits were superimposed. The conformation of the $\beta 7$ N-terminal main chain indicated that the new active site was not in the chamber formed by the two rings but was close to the gate formed by the α and β rings (Fig. 3.4D).

Thr-1, Asp-56, Arg-99 and Asn-104 in the $\beta 7$ subunit, and Tyr-88 of the $\beta 1$ subunit, are conserved between the bovine and yeast proteasomes. The three-dimensional locations of the functional groups of these residues are also similar in the two species, except for Thr-1. Structural comparison of the two structures suggested that the N-terminal chain in the yeast $\beta 7$ subunit can change its conformation to match that of the novel Ntnhydrolase-active site in the bovine proteasome. However, SNAAP activity has not been reported for the yeast enzyme.

3.3.3

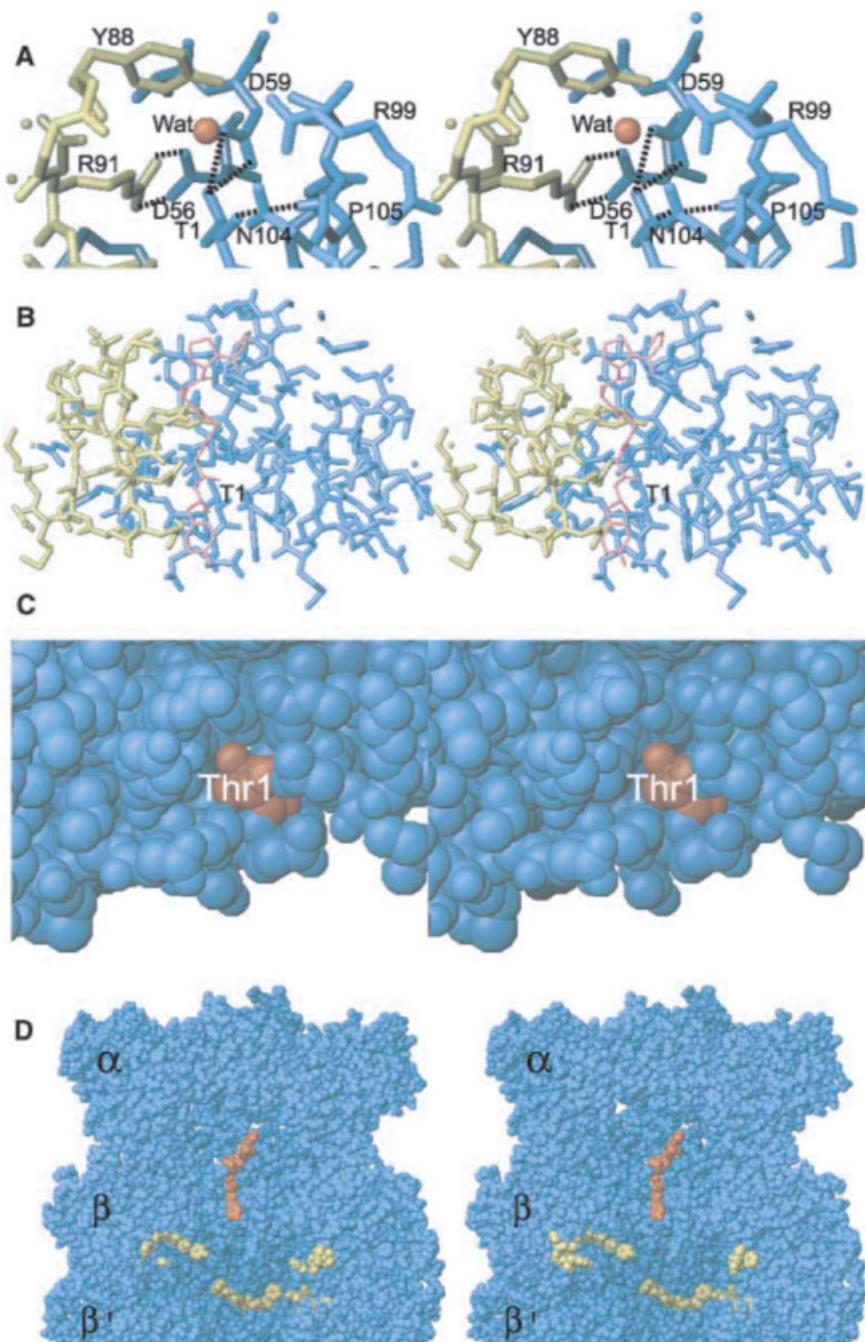
Predicted Structure of Immunoproteasome and Substrate Specificities

Human $\beta 1i$, $\beta 2i$ and $\beta 5i$ subunits are 59.2%, 57.7% and 68.6% identical, respectively, with their constitutively active counterparts $\beta 1$, $\beta 2$ and $\beta 5$, and 46.0%, 46.0% and 57.6% identical, respectively with the $\beta 1$, $\beta 2$ and $\beta 5$ subunits in yeast. Indeed, all induced subunits of the human 20S proteasome have higher sequence similarity to the corresponding subunits of the human constitutive proteasome than to their yeast counterparts. Both of the human proteasomes share the remaining four β subunits and all the α subunits. Therefore, the X-ray structure of the bovine 20S proteasome was used for more accurate prediction of the three-dimensional structure of the immunoproteasome (Fig. 3.6).

As in the yeast proteasome, each S1 pocket of the bovine 20S proteasome is formed by two subunits [22]. Val-20, Phe-31, Ser-35 and Leu-49 of the $\beta 1i$ subunit

Fig. 3.4 Stereo views of the new Ntn-hydrolase active site projected along the same directions. (A) Residues involved in the catalytic core are shown by single letter notation and the residue number in its respective bovine subunit. $\beta 1$ and $\beta 7$ subunits are colored yellow and blue, respectively. The red sphere is a water molecule. (B) Predicted structure of the model substrate Phe-Gly-Pro-Ala-Gly-Gly-Tyr fit into the $\beta 7$ active site. $\beta 1$, $\beta 7$ subunits and the model substrate are depicted in yellow, blue and red, respectively. (C) The active site

was drawn by a space-filling model to show its hollow surface. Thr-1, colored red, is located at the bottom of this hollow. (D) The predicted substrate structures in the 20S proteasome. The red model is a substrate of the novel $\beta 7$ active site; the three yellow models are substrates of the $\beta 1$, $\beta 2$ and $\beta 5$ active sites. The α ring, β ring and β' ring are depicted in part for convenience. The novel Ntn-catalytic site is in a different position and a different orientation (red) from the other three active sites (yellow).



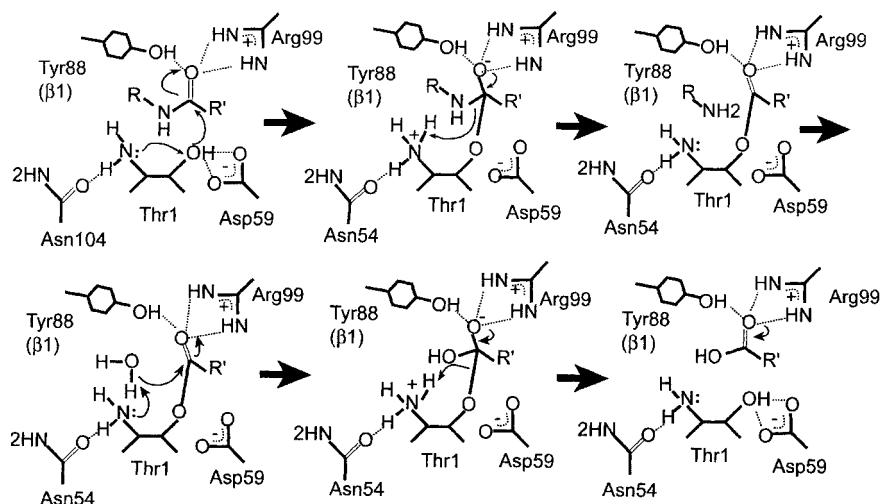


Fig. 3.5 Putative catalytic mechanism of the novel active site in $\beta 7$. The reaction begins when the nucleophilic oxygen of Thr-1 donates its proton to its own α -amino group and attacks the carbonyl carbon of the substrate. The negatively charged tetrahedral intermediate is stabilized by hydrogen bonding. The acylation step is complete when the α -amino group of Thr donates a proton to the nitrogen of the scissile peptide bond. A covalent bond is formed between the substrate

and the enzyme, and part of the substrate is released. The deacylation step begins when the hydroxyl group of water attacks the carbonyl carbon of the acyl-enzyme product and the basic α -amino group of the nucleophile accepts a proton from the water molecule. The negatively charged intermediate is stabilized, as in the acylation step. The reaction is completed when the α -amino group donates a proton to the nucleophile.

and His-114 and Ser-120 of the $\beta 2i$ subunit are located in the S1 pocket of the $\beta 1i$ active site instead of Thr-20, Thr-31, Thr-35 and Arg-45 of the $\beta 1$ subunit, and Tyr-114 and Asp-120 of the $\beta 2$ subunit. The S1 pocket of $\beta 1i$ is less polar than that of $\beta 1$ (Fig. 3.6a, b); thus, it is expected that PGPH activity would be reduced and chymotrypsin-like activity would be enhanced in the IFN γ -induced 20S proteasome.

In the S1 pockets of $\beta 2$, no such large structural changes are noted upon the introduction of IFN γ -induced subunits ($\beta 2i$), as was the case with $\beta 1i$. Ser-32 and Asp-53 of $\beta 2$ are replaced by glutamine in the $\beta 2i$ subunit. It is conceivable that the trypsin-like activity of the $\beta 2i$ subunit is enhanced by the acidic Glu residues, which would stabilize the basic substrate residue in the pocket.

At present there is general agreement that subunit exchanges induced by IFN γ are likely to confer functional alterations in the immunoproteasome; the induced enzyme was shown to have increased trypsin- and chymotrypsin-like activity in studies using fluorogenic substrates, and reduced PGPH activity towards peptide substrates [3, 25, 27]. On the other hand, contradictory reports have also been made [28, 29]. The present structural model of the immunoproteasome appears to fit with the former observations. In fact, such changes in peptidase activities suggest that the immunoproteasome should generate more peptides with hydropho-

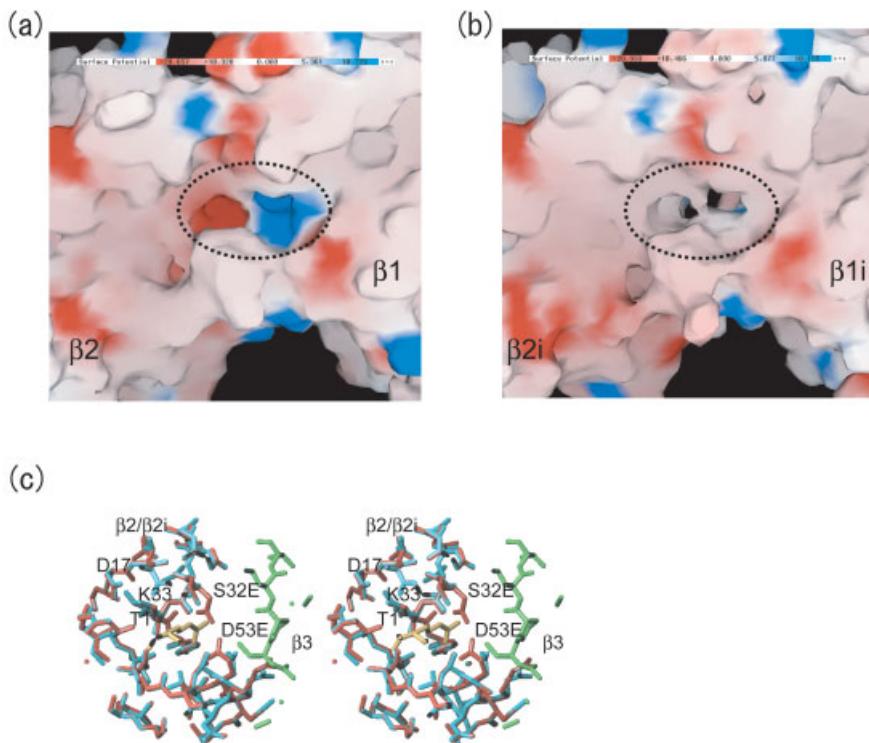


Fig. 3.6 Structure comparison of the active centers of the constitutive proteasome with the active centers of the immunoproteasome. (a), (b) Diagrams of electrostatic surface potential of S1 pockets for the β_1 and β_{1i} active sites. Red and blue indicate negative and positive potentials, respectively. The S1 pocket is surrounded by a dotted circle. The S1 pocket of the β_1 active center is formed by both β_1 and β_2 subunits. (a) The S1 pocket of the β_1 active center is charged positively at the β_2 subunit side and negatively at the β_1 subunit

side. (b) The S1 pocket of the β_{1i} active center of the immunoproteasome is almost completely nonpolar. (c) Structure comparison of the β_2 and β_{2i} active centers in a stereo pair. Amino acid residues are represented by single letter notations with corresponding subunit residue numbers. Both Ser-32 and Asp-53 in the constitutive subunit β_2 (sky blue) are replaced by Glu in inducible subunit β_{2i} (red). The green peptide is part of the β_3 subunit and the yellow one is a model substrate.

bic or basic carboxyl termini and fewer peptides with acidic ones. Thus, formation of the immunoproteasome should facilitate the production of MHC class I-binding peptides, because hydrophobic or basic carboxyl terminal residues normally serve as anchors for binding to MHC class I molecules [3].

3.4**Proteasome Inhibitors**

The proteasome has multiple active sites. Three activities, chymotrypsin-like, trypsin-like and PGPH, are classified by their respective substrate specificities; they all differ in the preferred amino acid in the P1 position adjacent to the scissile amide bond. Proteasome inhibitors can be divided into several groups based on pharmacophores (Tab. 3.1, Fig. 3.7).

Peptide aldehydes were the first proteasome inhibitors to be developed and are still the most widely used inhibitors [41, 42]. Aldehyde inhibitors are slow-binding [42], but they enter cells rapidly and are reversible. These inhibitors are well-known inhibitors of cysteine and serine proteases, and thus can inhibit calpains and cathepsins *in vivo*. Since MG132 and PSI can all inhibit calpains and cathepsins in addition to the proteasome, when using these inhibitors in cell cultures it is important to perform control experiments to confirm that the observed effects are truly due to proteasome inhibition. The dipeptide aldehyde CEP1612 appears at least as good as MG132 in potency and selectivity.

Peptide boronates are much more potent proteasome inhibitors than aldehydes [43]. Boronate-proteasome adducts have much more slower dissociation rates than proteasome–aldehyde adducts, and although boronates are considered reversible inhibitors, the inhibition is practically irreversible over a period of hours. In addition, PS-341 inhibits serine proteases 1000-fold weaker than it does the proteasome [43]. This combination of potency, selectivity and metabolic stability makes the peptide boronates better drug candidates than other classes of proteasome inhibitors. *In vitro* and mouse xenograft studies of PS-341 have shown anti-tumor activity in a variety of tumor types. DFLB and PS-273 are useful fluorescent active site probes; binding of these inhibitors enhances the fluorescence of their environment-sensitive dansyl and naphthyl moieties [44].

Lactacystin is a *Streptomyces* metabolite that selectively modifies the $\beta 5$ subunit of the mammalian proteasome and irreversibly blocks its activity. Other proteolytic sites of the proteasome are also modified and reversibly inhibited, but at much slower rates [45, 46].

Tab. 3.1 Proteasome inhibitors.

| Class of compound | Inhibitor | Other targets |
|-----------------------------|---|-------------------------|
| Peptide aldehydes | MG132, LLnV, PSI, ALLN, ALLnM, CEP1612, Z-LLF | Calpain I, Cathepsin B |
| Peptide boronates | MG262, PS341, PS273 | |
| Lactacystin and derivatives | Lactacystin, β -Lactone | Cathepsin A, TPPII |
| Peptide vinyl sulfones | NLVS, YLVS | Cathepsin S and B |
| Peptide epoxyketones | Dihydroeponemycin Epoxomicin, YU101 | Cathepsin B (very weak) |

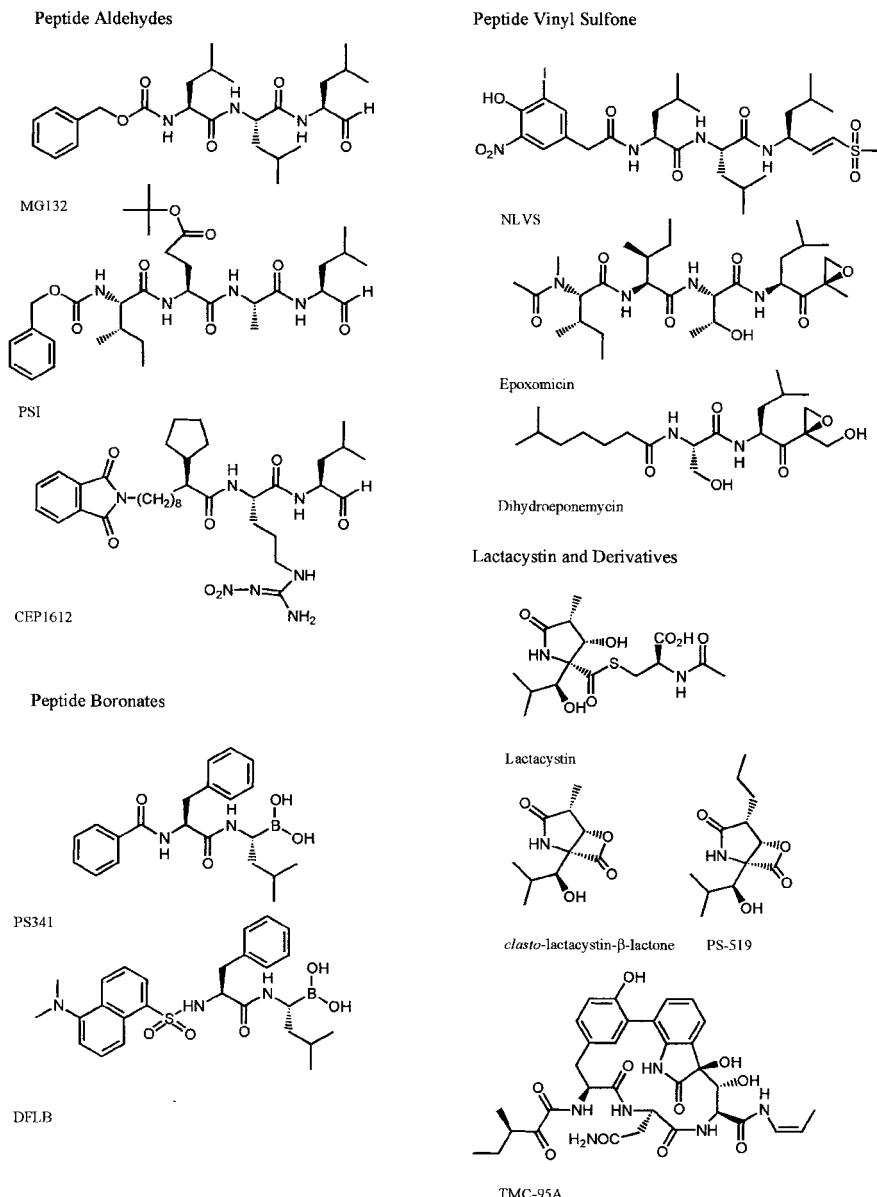


Fig. 3.7 Structures of major proteasome inhibitors.

Peptide vinyl sulfones are synthetic irreversible proteasome inhibitors that covalently modify its catalytic β subunits [47]. They are easier to synthesize than other irreversible proteasome inhibitors, and they do not inhibit serine proteases.

Natural epoxyketones were isolated based on their anti-tumor activity in mice and exert their biological effects by proteasome inhibition [48, 49]. These com-

pounds act by a unique mechanism, in which they react with both the hydroxyl and amino groups of the catalytic threonine of the proteasome.

Proteasome inhibitors have helped to advance our understanding of proteasome biology as well as being effective in treating inflammation and cancer. Moreover, because the proteasome plays a dominant role in the generation of antigenic peptides as ligands for MHC class I molecules, substrates which block this activity are possible candidates for immunosuppressive drugs.

3.4.1

Structure of Proteasome Inhibitor Complexes

Several structures of proteasome inhibitor complexes have been determined by X-ray crystal analysis [22, 50, 51]. Calpain inhibitor I (acetyl-L-L-norleucinal) was covalently bound to Thr-1 of the $\beta 1$, $\beta 2$ and $\beta 5$ subunits, respectively. This peptide aldehyde inhibitor is attacked by the threonyl O γ to form a hemiacetal. The norleucine side chain projects into a pocket with an opening in its side towards a tunnel leading to the particle surface. The leucine side chain at P2 is not in contact with the protein, whilst the leucine side chain at P3 is in contact with the adjacent β -type subunit.

Lactacytin is covalently bound to $\beta 5$. Its dimethyl side chain at C10 projects into S1 like a valine or leucine side chain, but not so deeply as the norleucine side chain of the calpain inhibitor. Lactacytin forms a host of hydrogen bonds with protein main-chain atoms. These hydrogen-bonding interactions could also be present in $\beta 1$ and $\beta 2$.

Structural analysis of the yeast 20S proteasome complexed with epoxomicin showed only adduct formation with the $\beta 5$ subunit. Epoxomicin displays a high degree of selectivity for inhibition of the chymotrypsin-like activity of the proteasome at lower concentrations. Epoxomicin and peptide aldehyde (acetyl-L-L-norleucinal) bind similarly to the catalytic subunits, completing an antiparallel β -sheet. However, epoxomicin complexed with $\beta 5$ forms a unique 6-atom ring. This morpholino derivative results from adduct formation between the α',β' -epoxyketone pharmacophore of epoxomicin and the amino terminal threonyl O γ and N of the $\beta 5$ subunit.

TMC-95A, a cyclic peptide metabolite from *Aplospora montagnei*, is a potent competitive inhibitor of all active sites and forms characteristic hydrogen bonds with the protein backbone. The crystal structure of the yeast 20S proteasome in complex with TMC-95A indicates a non-covalent linkage to the active β -subunits; the N-terminal threonine residues are not modified. The TMC-95A backbone adopts a β -conformation and extends the β -strand S1 by the generation of an antiparallel β -sheet. This structure is similar to that seen with the aldehyde and epoxyketone inhibitors. All interactions of TMC-95A are formed with main-chain atoms and strictly conserved residues of the 20S proteasome.

3.5**Conclusions**

There is no doubt that the Ub-proteasome system plays an indispensable role not only in cell-cycle progression but also in apoptosis. Moreover, this proteolytic pathway also appears to be involved in the immune response. In considering the important contribution of the proteasome in the regulation of these biological processes, it may be extremely useful to use those agents capable of modulating specific functions of ubiquitination and proteasomes as tools in medical interventions, especially in cancer and autoimmune disease. The crystal structure of the 20S proteasome and its inhibitor complexes revealed the intriguing selectivity of proteasome inhibitors. This will form the basis for the development of synthetic selective proteasome inhibitors for use as anti-tumor or anti-inflammatory drugs.

3.6**Acknowledgements**

Supported in part by Grants-in-Aid for Scientific Research (13359003) from the Ministry of Education, Culture, Sports, Science and Technology of Japan (T.T.).

3.7**References**

- 1 HERSHKO, A., CIECHANOVER, A., The ubiquitin system. *Annu. Rev. Biochem.* **1998**, *67*, 425–479.
- 2 PAGANO, M., TAM, S. W., THEODORAS, A. M., BEER-ROMERO, P., DEL SAL, G., CHAU, V., YEW, P. R., DRAETTA, G. F., ROLFE, M., Role of the ubiquitin-proteasome pathway in regulating abundance of the cyclin-dependent kinase inhibitor p27. *Science* **1995**, *269*, 682–685.
- 3 ROCK, K. L., GOLDBERG, A. L., Degradation of cell proteins and the generation of MHC class I-presented peptides. *Annu. Rev. Immunol.* **1999**, *17*, 739–779.
- 4 MITCH, W. E., GOLDBERG, A. L., Mechanisms of muscle wasting. The role of the ubiquitin-proteasome pathway. *N. Engl. J. Med.* **1996**, *335*, 1897–1905.
- 5 HAMPTON, R. Y., GARDNER, R. G., RINE, J., Role of 26S proteasome and HRD genes in the degradation of 3-hydroxy-3-methylglutaryl-CoA reductase, an integral endoplasmic reticulum membrane protein. *Mol. Biol. Cell.* **1996**, *7*, 2029–2044.
- 6 MURAKAMI, Y., MATSU FUJI, S., KAMEJI, T., HAYASHI, S., IGARASHI, K., TAMURA, T., TANAKA, K., ICHIHARA, A., Ornithine decarboxylase is degraded by the 26S proteasome without ubiquitination. *Nature* **1992**, *360*, 597–599.
- 7 TURNER, G. C., DU, F., VARSHAVSKY, A., Peptides accelerate their uptake by activating a ubiquitin-dependent proteolytic pathway. *Nature* **2000**, *405*, 579–583.
- 8 ALMOND, J. B., COHEN, G. M., The proteasome: a novel target for cancer chemotherapy. *Leukemia* **2002**, *16*, 433–443.
- 9 KISSELEV, A. F., GOLDBERG, A. L., Proteasome inhibitors: from research tools to drug candidates. *Chem. Biol.* **2001**, *8*, 739–758.
- 10 HAAS, A. L., SIEPMANN, T. J., Pathways of ubiquitin conjugation. *Faseb. J.* **1997**, *11*, 1257–1268.

- 11** HERSHKO, A., CIECHANOVER, A., The ubiquitin system for protein degradation. *Annu. Rev. Biochem.* **1992**, *61*, 761–807.
- 12** VARSHAVSKY, A., The ubiquitin system. *Trends Biochem. Sci.* **1997**, *22*, 383–387.
- 13** HOCHSTRASSER, M., Ubiquitin-dependent protein degradation. *Annu. Rev. Genet.* **1996**, *30*, 405–439.
- 14** ORLOWSKI, R.Z., The role of the ubiquitin-proteasome pathway in apoptosis. *Cell Death Differ.* **1999**, *6*, 303–313.
- 15** BAUMEISTER, W., WALZ, J., ZUHL, F., SEE-MULLER, E., The proteasome: paradigm of a self-compartmentalizing protease. *Cell* **1998**, *92*, 367–380.
- 16** DEMARTINO, G.N., SLAUGHTER, C.A., The proteasome, a novel protease regulated by multiple mechanisms. *J. Biol. Chem.* **1999**, *274*, 22123–22126.
- 17** BRAUN, B.C., GLICKMAN, M., KRAFT, R., DAHLMANN, B., KLOETZEL, P.M., FINLEY, D., SCHMIDT, M., The base of the proteasome regulatory particle exhibits chaperone-like activity. *Nat. Cell Biol.* **1999**, *1*, 221–226.
- 18** HORWICH, A.L., WEBER-BAN, E.U., FINLEY, D., Chaperone rings in protein folding and degradation. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 11033–11040.
- 19** BRANNIGAN, J.A., DODSON, G., DUGGLEBY, H.J., MOODY, P.C., SMITH, J.L., TOM-CHICK, D.R., MURZIN, A.G., A protein catalytic framework with an N-terminal nucleophile is capable of self-activation. *Nature* **1995**, *378*, 416–419.
- 20** LOWE, J., STOCK, D., JAP, B., ZWICKL, P., BAUMEISTER, W., HUBER, R., Crystal structure of the 20S proteasome from the archaeon *T. acidophilum* at 3.4 Å resolution. *Science* **1995**, *268*, 533–539.
- 21** SEE-MULLER, E., LUPAS, A., BAUMEISTER, W., Autocatalytic processing of the 20S proteasome. *Nature* **1996**, *382*, 468–471.
- 22** GROLL, M., DITZEL, L., LOWE, J., STOCK, D., BOCHTLER, M., BARTUNIK, H.D., HUBER, R., Structure of 20S proteasome from yeast at 2.4 Å resolution. *Nature* **1997**, *386*, 463–471.
- 23** BOCHTLER, M., DITZEL, L., GROLL, M., HARTMANN, C., HUBER, R., The proteasome. *Annu. Rev. Biophys. Biomol. Struct.* **1999**, *28*, 295–317.
- 24** TANAKA, K., Molecular biology of the proteasome. *Biochem. Biophys. Res. Commun.* **1998**, *247*, 537–541.
- 25** FRUH, K., YANG, Y., Antigen presentation by MHC class I and its regulation by interferon gamma. *Curr. Opin. Immunol.* **1999**, *11*, 76–81.
- 26** MONACO, J.J., NANDI, D., The genetics of proteasomes and antigen processing. *Annu. Rev. Genet.* **1995**, *29*, 729–754.
- 27** TANAKA, K., KASAHARA, M., The MHC class I ligand-generating system: roles of immunoproteasomes and the interferon-gamma-inducible proteasome activator PA28. *Immunol. Rev.* **1998**, *163*, 161–176.
- 28** KLOETZEL, P.M., Antigen processing by the proteasome. *Nat. Rev. Mol. Cell Biol.* **2001**, *2*, 179–187.
- 29** NIEMERMAN, G., GEIER, E., LUCCHIARI-HARTZ, M., HITZIGER, N., RAMSPERGER, A., EICHMANN, K., The specificity of proteasomes: impact on MHC class I processing and presentation of antigens. *Immunol. Rev.* **1999**, *172*, 29–48.
- 30** PAMER, E., CRESSWELL, P., Mechanisms of MHC class I-restricted antigen processing. *Annu. Rev. Immunol.* **1998**, *16*, 323–358.
- 31** FLAJNIK, M.F., KASAHARA, M., Comparative genomics of the MHC: glimpses into the evolution of the adaptive immune system. *Immunity* **2001**, *15*, 351–362.
- 32** RECHSTEINER, M., REALINI, C., USTRELL, V., The proteasome activator 11 S REG (PA28) and class I antigen presentation. *Biochem. J.* **2000**, *345 Pt 1*, 1–15.
- 33** GRAY, C.W., SLAUGHTER, C.A., DEMARTINO, G.N., PA28 activator protein forms regulatory caps on proteasome stacked rings. *J. Mol. Biol.* **1994**, *236*, 7–15.
- 34** KOPP, F., DAHLMANN, B., KUEHN, L., Reconstitution of hybrid proteasomes from purified PA700-20 S complexes and PA28alpha/beta activator: ultrastructure and peptidase activities. *J. Mol. Biol.* **2001**, *313*, 465–471.
- 35** TANAHASHI, N., MURAKAMI, Y., MINAMI, Y., SHIMBARA, N., HENDIL, K.B., TANAKA, K., Hybrid proteasomes. Induction by interferon-gamma and contribution to ATP-dependent proteolysis. *J. Biol. Chem.* **2000**, *275*, 14336–14345.

- 36** HENDIL, K.B., KHAN, S., TANAKA, K., Simultaneous Binding of PA28 and PA700 activators to 20S proteasomes. *Biochem. J.* **1998**, *332*, 749–754.
- 37** DICK, T.P., RUPPERT, T., GROETTRUP, M., KLOETZEL, P.M., KUEHN, L., KOSZINOWSKI, U.H., STEVANOVIC, S., SCHILD, H., RAMMENSEE, H.G., Coordinated dual cleavages induced by the proteasome regulator PA28 lead to dominant MHC ligands. *Cell* **1996**, *86*, 253–262.
- 38** UNNO, M., MIZUSHIMA, T., MORIMOTO, Y., TOMISUGI, Y., TANAKA, K., YASUOKA, N., TSUKIHARA, T., The structure of the mammalian 20S proteasome at 2.75 Å. *Resolut. Struct. (Camb.)* **2002**, *10*, 609–618.
- 39** OINONEN, C., ROUVINEN, J., Structural comparison of Ntn-hydrolases. *Protein Sci.* **2000**, *9*, 2329–2337.
- 40** ORLOWSKI, M., CARDOZO, C., MICHAUD, C., Evidence for the presence of five distinct proteolytic components in the pituitary multicatalytic proteinase complex. Properties of two components cleaving bonds on the carboxyl side of branched chain and small neutral amino acids. *Biochemistry* **1993**, *32*, 1563–1572.
- 41** ROCK, K.L., GRAMM, C., ROTHSTEIN, L., CLARK, K., STEIN, R., DICK, L., HWANG, D., GOLDBERG, A.L., Inhibitors of the proteasome block the degradation of most cell proteins and the generation of peptides presented on MHC class I molecules. *Cell* **1994**, *78*, 761–771.
- 42** VINITSKY, A., MICHAUD, C., POWERS, J.C., ORLOWSKI, M., Inhibition of the chymotrypsin-like activity of the pituitary multicatalytic proteinase complex. *Biochemistry* **1992**, *31*, 9421–9428.
- 43** ADAMS, J., BEHNKE, M., CHEN, S., CRUICKSHANK, A.A., DICK, L.R., GRENIER, L., KLUNDER, J.M., MA, Y.T., PLAMONDON, L., STEIN, R.L., Potent and selective inhibitors of the proteasome: dipeptidyl boronic acids. *Bioorg. Med. Chem. Lett.* **1998**, *8*, 333–338.
- 44** MCCORMACK, T., BAUMEISTER, W., GRENIER, L., MOOMAW, C., PLAMONDON, L., PRAMANIK, B., SLAUGHTER, C., SOUCY, F., STEIN, R., ZUHL, F., DICK, L., Active site-directed inhibitors of Rhodococcus 20 S proteasome. Kinetics and mechanism. *J. Biol. Chem.* **1997**, *272*, 26103–26109.
- 45** CRAIU, A., GACZYNKA, M., AKOPIAN, T., GRAMM, C.F., FENTEANY, G., GOLDBERG, A.L., ROCK, K.L., Lactacystin and clasto-lactacystin beta-lactone modify multiple proteasome beta-subunits and inhibit intracellular protein degradation and major histocompatibility complex class I antigen presentation. *J. Biol. Chem.* **1997**, *272*, 13437–13445.
- 46** FENTEANY, G., STANDAERT, R.F., LANE, W.S., CHOI, S., COREY, E.J., SCHREIBER, S.L., Inhibition of proteasome activities and subunit-specific amino-terminal threonine modification by lactacystin. *Science* **1995**, *268*, 726–731.
- 47** BOGYO, M., McMMASTER, J.S., GACZYNKA, M., TORTORELLA, D., GOLDBERG, A.L., PLOEGH, H., Covalent modification of the active site threonine of proteasomal beta subunits and the Escherichia coli homolog HslV by a new class of inhibitors. *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 6629–6634.
- 48** MENG, L., KWOK, B.H., SIN, N., CREWS, C.M., Eponemycin exerts its antitumor effect through the inhibition of proteasome function. *Cancer Res.* **1999**, *59*, 2798–2801.
- 49** MENG, L., MOHAN, R., KWOK, B.H., ELOFSSON, M., SIN, N., CREWS, C.M., Epoxomicin, a potent and selective proteasome inhibitor, exhibits in vivo anti-inflammatory activity. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 10403–10408.
- 50** GROLL, M., BAJOREK, M., KOHLER, A., MORODER, L., RUBIN, D.M., HUBER, R., GLICKMAN, M.H., FINLEY, D., A gated channel into the proteasome core particle. *Nat. Struct. Biol.* **2000**, *7*, 1062–1067.
- 51** GROLL, M., KOGUCHI, Y., HUBER, R., KOHNO, J., Crystal structure of the 20 S proteasome:TMC-95A complex: a non-covalent proteasome inhibitor. *J. Mol. Biol.* **2001**, *311*, 543–548.

4

Antibiotics and the Ribosome

JEFFREY L. HANSEN

4.1

Introduction

Many antibiotics kill bacteria by binding to the ribosome and thereby inhibiting protein synthesis. After decades of antibiotic use, many pathogens have become resistant to antibiotics, and thus new antibiotics are needed to treat bacterial diseases. Rational drug design is expected to play an important role in meeting this need.

Rational efforts to design new antibiotics that target the ribosome became feasible when the crystal structures of both ribosomal subunits were solved at atomic resolution in 2000 [1–3]. Since then, structures of about 20 different antibiotics bound to ribosomes have been published (Tab. 4.1) [4–12]. These structures of antibiotic/ribosomal complexes provide insights into the mechanisms by which antibiotics inhibit protein synthesis and by which mutations confer resistance. Thus, a sound basis now exists for designing new antibiotics that may circumvent resistance.

4.2

The Ribosome

4.2.1

Introduction

The ribosome is the enzyme that catalyzes peptide bond formation. The bacterial ribosome is a large 2500 kDa ribonucleic acid/protein complex comprised of a large subunit (LSU or 50S subunit) and a small subunit (SSU or 30S subunit) (Fig. 4.1). The small ribosomal subunit binds to messenger RNA (mRNA) and reads the genetic code by aligning its base triplet codons with anticodons of transfer RNA molecules (tRNA). The large ribosomal subunit binds to opposite ends of tRNA molecules and catalyzes peptide bond formation.

Tab. 4.1 Antibiotic/ribosome crystal structures.

| Classification of antibiotic | Antibiotic | Ribosome | Reso-lution | FreeR | PDB code | Reference |
|-------------------------------------|-------------------|-----------------|--------------------|--------------|-----------------|------------------|
| 50S | None (native) | <i>Hm</i> | 2.4 | 26.1 | 1JJ2 | 9 |
| 50S | None (native) | <i>Dr</i> | 3.1 | 27.4 | 1LNR | 10 |
| MLS_B | | | | | | |
| Macrolide-16 | Carbomycin | <i>Hm</i> | 3.0 | 26.5 | 1K8A | 8 |
| | Spiramycin | <i>Dr</i> | 3.0 | 26.9 | 1KD1 | 8 |
| | Tylosin | <i>Hm</i> | 3.0 | 26.2 | 1K9M | 8 |
| Macrolide-15 | Azithromycin | <i>Hm</i> | 3.2 | 25.0 | 1M1K | 8 |
| Macrolide-14 | Clarithromycin | <i>Dr</i> | 3.5 | 32.3 | 1J5A | 4 |
| | Erythromycin A | <i>Dr</i> | 3.5 | 30.1 | 1JZY | 4 |
| | Roxithromycin | <i>Dr</i> | 3.8 | 27.4 | 1JZZ | 4 |
| Lincosamide | Clindamycin | <i>Dr</i> | 3.1 | 30.3 | 1JZX | 4 |
| Streptogramin-A | Virginiamycin M | <i>Hm</i> | 3.0 | 24.2 | 1N8R | 7 |
| Streptogramin-B | Not available | | | | | |
| A-site | Anisomycin | <i>Hm</i> | 3.0 | 24.6 | 1K73 | 7 |
| | Chloramphenicol | <i>Dr</i> | 3.5 | 32.1 | 1KO1 | 4 |
| | Chloramphenicol | <i>Hm</i> | 3.0 | 20.9 | 1NJ1 | 7 |
| Nucleotide analog | Puromycin | <i>Hm</i> | | | 1FGO | 11 |
| | Sparsomycin | <i>Hm</i> | 2.8 | 22.2 | 1M190 | 7 |
| P-site | Blasticidin S | <i>Hm</i> | 3.0 | 23.5 | 1KC8 | 7 |
| 30S | None (native) | <i>Tt</i> | 3.0 | 25.2 | 1J5E | 2 |
| | None (native) | <i>Tt</i> | 3.2 | 26.1 | 1HRO | 2 |
| | None (native) | <i>Tt</i> | 3.3 | 30.5 | 1FKA | |
| | Pactamycin | <i>Tt</i> | 3.4 | 28.0 | 1HNX | 5 |
| | Hygromycin B | <i>Tt</i> | 3.3 | 26.1 | 1HNZ | 5 |
| | Paromomycin | <i>Tt</i> | 3.0 | 25.5 | 1FJG | 6 |
| | Paromomycin | <i>Tt</i> | 3.0 | 27.0 | 1N32 | 12 |
| | Paromomycin | <i>Tt</i> | 3.1 | 27.5 | 1IBL | 12 |
| | Paromomycin | <i>Tt</i> | 3.3 | 28.2 | 1IBK | 12 |
| | Paromomycin | <i>Tt</i> | 3.3 | 28.4 | 1N33 | 12 |
| | Spectinomycin | <i>Tt</i> | 3.0 | 25.5 | 1FJG | 6 |
| | Streptomycin | <i>Tt</i> | 3.0 | 25.5 | 1FJG | 6 |
| | Edeine | <i>Tt</i> | 3.2 | 24.5 | 1I94 | 52 |
| | Edeine | <i>Tt</i> | 4.5 | 24.4 | 1I95 | 52 |
| | Tetracycline | <i>Tt</i> | 3.2 | 24.5 | 1I94 | 52 |
| | Tetracycline | <i>Tt</i> | 3.4 | 26.4 | 1HNW | 5 |
| | Tetracycline | <i>Tt</i> | 4.5 | 25.4 | 1I97 | 52 |

4.2.2

Binding of tRNA

Several sites on the ribosome interact with tRNA (Fig. 4.1) and also are targeted by antibiotics. The peptidyl-tRNA binding site (P-site) of the large subunit binds to the 3' end of peptidyl-tRNA. The aminoacyl-tRNA binding site (A-site) of the

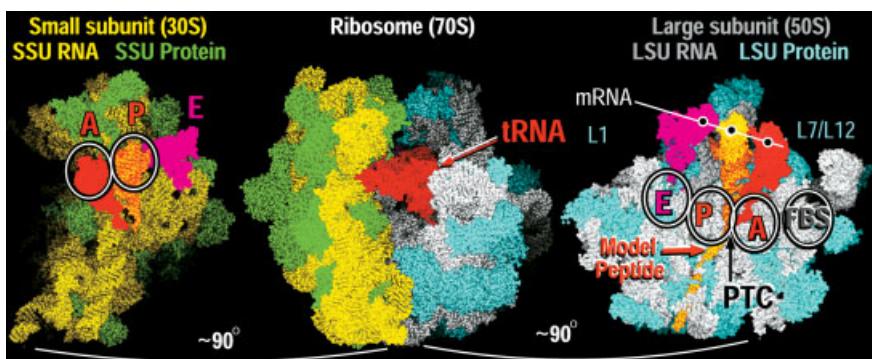


Fig. 4.1 Overview of ribosome. Center: These space filled representations are based on docking of the large ribosomal subunit [1] (grey RNA and blue protein) onto the small subunit [2] (yellow RNA and green protein) based on a protein alpha carbon and rRNA phosphate backbone trace of the 70S ribosome bound with tRNA [53]. Functional sites are circled and labeled according to the corresponding bound tRNA molecules; A-site (red), P-site (orange), E-site (purple) and factor binding site (FBS). Peptide bond formation is catalyzed at the peptidyl transferase center (PTC) that is located on the large subunit between the A-site and P-site. Left: The

small subunit with bound tRNA was rotated to the left to also show its interface surface. Right: The large subunit with bound tRNA was rotated to the right, and its interface surface was clipped, to show a model of a peptide attached to the P-site tRNA (orange) and extending down the exit tunnel. Messenger RNA (mRNA), which binds to the 30S subunit, is schematically represented (straight white line) on the 50S subunit with its codons (black circles) interacting with anticodons of tRNA. With respect to the large subunit (right), tRNA moves from right to left during translation.

large subunit binds to the 3' end of aminoacyl-tRNA. The tRNA exit site (E-site) of the large subunit binds to the 3' end of deacylated tRNA. Meanwhile, corresponding sites on the small subunit interact with the opposite ends of tRNA molecules that contain the anticodon stems.

4.2.3

Peptidyl Transferase Activity

In the first step of the peptidyl transferase reaction, a peptidyl tRNA molecule is bound in the P-site with its nascent peptide extending down the peptide exit tunnel (Fig. 4.1). An elongation factor binds to a factor binding site (FBS) and positions an aminoacyl-tRNA in the A-site. The α amino group of the aminoacyl-tRNA nucleophilically attacks the ester bond which connects the peptide to the tRNA bound in the P-site (Fig. 4.2). The ester bond is broken as an amide bond forms, and the peptide becomes one amino acid longer, and is now attached to the tRNA that in the A-site. Translocation of the products follows peptide bond formation, as the newly formed deacylated- tRNA of the P-site moves into the E-site, and as the newly elongated peptidyl-tRNA moves from the A-site into the P-site.

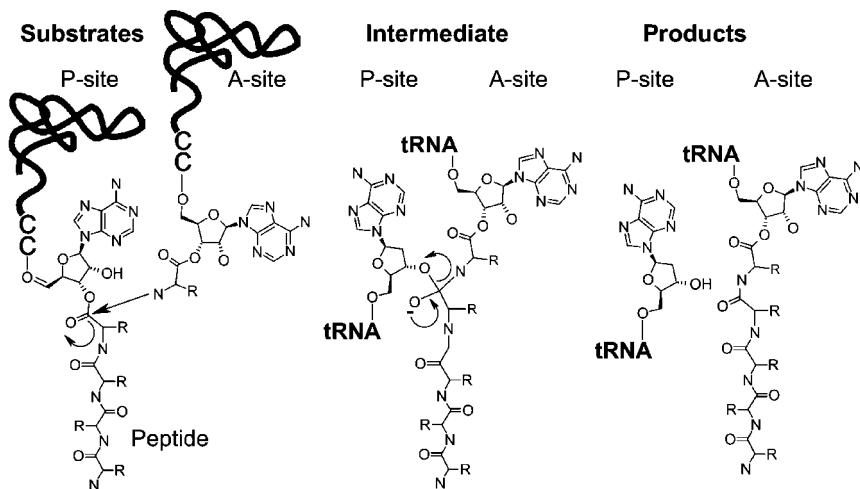


Fig. 4.2 Peptidyl transferase reaction. Left: The α amino group of an A-site substrate, attacks (arrow) the ester bond that links a P-site substrate tRNA to its nascent peptide chain. The first 73 nucleotides of tRNA are represented by ribbons, C74 and C75 are represented by the letter C, and A76 and the peptide are represented by chemical drawings. Center: During the nucleophilic attack,

a tetrahedral carbon intermediate with an oxy-anion is formed. The entire tRNA portion of the substrates except for A76 is represented by (tRNA). Right: The intermediate resolves forming two products; a deacylated tRNA bound to the P-site, and a peptidyl-tRNA that is one amino acid longer and is still bound to the A-site.

4.2.4

Structure of the Ribosome

Prior to the availability of atomic resolution crystal structures of ribosomal subunits, the secondary structure of ribosomal RNA (rRNA) was determined [13, 14] (Fig. 4.3). When the X-ray structures of the 50S [1] and 30S [2, 3] subunits were solved, the predicted secondary structure and many predicted tertiary interactions [15] were found to be quite reliable.

Within the secondary structural diagrams, regions containing highly conserved sequence motifs were identified and were found to be implicated in various functions of the ribosomes (Figs. 4.3 and 4.4). Nucleotides implicated in the peptidyl transferase reaction were located in the central loop of Domain V, which was therefore designated as the peptidyl transferase center (PTC). Other nucleotides that were implicated in factor binding and in an associated GTPase activity were located in a loop of Domain II. Nucleotides in these same regions of secondary structure were also implicated in antibiotic binding because they either were protected by antibiotics from chemical modification or upon mutation-conferred resistance to antibiotics. This chapter will focus on antibiotics that bind to the peptidyl transferase center, because no X-ray structures are available of antibiotics bound to the factor binding site.

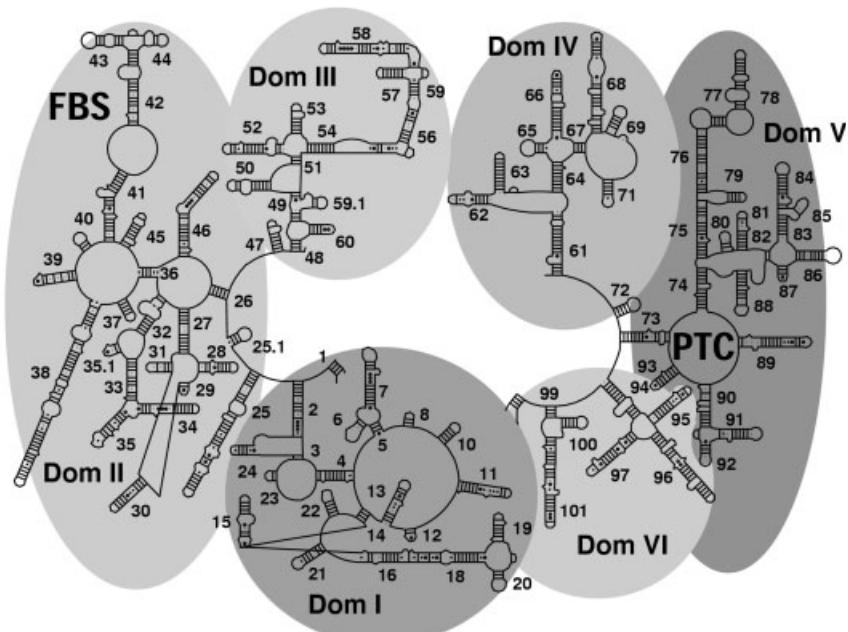


Fig. 4.3 Secondary structure of large ribosomal subunit. The secondary structure of rRNA has been divided into six domains (Dom I to Dom VI) [14]. Various functional sights have

been located within these domains. The peptidyl transferase center (PTC) was located in domain V. The factor binding site was located in domain II. Fig. modified from [1].

4.3 Antibiotics

4.3.1

Introduction

Anti-ribosomal antibiotics are enzyme inhibitors, and the ones that are clinically useful inhibit bacterial ribosomes far more effectively than they inhibit eukaryotic ribosomes. Many enzyme inhibitors exert their effects by binding to the active site of enzymes and thereby prevent the binding of substrates. Others block enzyme function by inhibiting conformational changes essential for enzyme activity. Anti-ribosomal antibiotics are unexceptional in this regard. They bind to the sites on both subunits to which tRNA also binds. Some block the interactions of substrates with the ribosome. Others block the tunnel and thereby prevent the nascent peptide from extending.

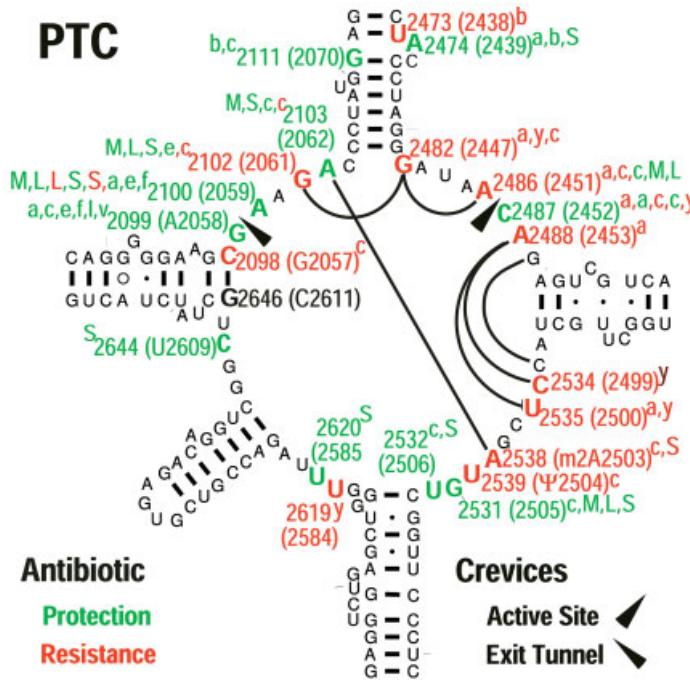


Fig. 4.4 Secondary structure of the PTC and antibiotic information. Nucleotide numbering is from *H. marismortui* (*Hm*) and is followed by the standard corresponding *E. coli* (*Ec*) numbering in parentheses. Nucleotides with superscripts have been implicated in antibiotic interactions by nucleotide protection studies (green) [54–57] or by mutations (orange) that confer resistance to: carbomycin A (M); lincomamides (L); streptogramin A [58] (S); chlor-

amphenicol [59–64] (c); blasticidin S [48] (b); anisomycin [51, 65] (a); or sparsomycin [65, 66] (y). Two hydrophobic crevices (gray triangles) are located at the ends of two RNA helices (center right and center left of figure). Nucleotides that form canonical or non-canonical base pairs or base triples that are involved in the structure of either crevice are connected with gray lines. Fig. modified from [7].

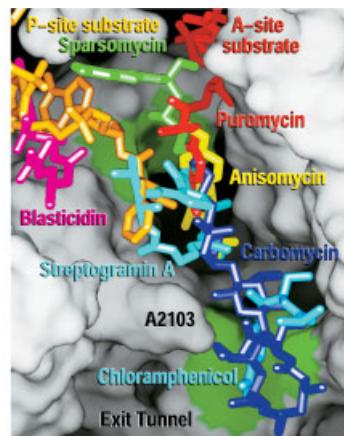
4.3.2

Antibiotics that Bind to the 50S Subunit

Most antibiotics that inhibit the function of the 50S subunit bind near its peptidyl transferase center (Fig. 4.4) and block peptide bond formation. Crystal structures are available for several such antibiotics bound to the ribosome (Fig. 4.5). They appear to inhibit the peptidyl transferase reaction either by competing directly with its substrates for binding, or indirectly by blocking the exit tunnel.

Two hydrophobic crevices near the peptidyl transferase center appear to be particularly important for antibiotic binding (Figs. 4.4 and 4.5). The first hydrophobic crevice, which is at the peptidyl transferase center, normally functions as the binding site for amino acid side chains of A-site substrates [11, 16]. Antibiotics that bind to this site will block the binding with A-site substrates and thereby directly

Fig. 4.5 Overview of antibiotics bound at the peptidyl transferase center. A surface representation of the large subunit of *H. marismortui* includes the P-site, A-site and entrance to the peptide exit tunnel. Most of these antibiotics contact either the active site hydrophobic crevice (green surface, upper center) or the hydrophobic crevice at the entrance to the exit tunnel (green surface, lower right). In addition, many of these antibiotics occupy an elongated pocket (dark surface, center) in the wall of the exit tunnel between these two crevices. The antibiotics shown are all from complexes with *H. marismortui* ribosomes and overlap the binding site of A-site substrates (red sticks) or of a P-site substrates (orange sticks). Fig. modified from [7].



prevent peptide bond formation. The second hydrophobic crevice is located at the entrance to the peptide exit tunnel, and its normal function in translation is not yet known. Nevertheless, binding of antibiotics to the second crevice appears to inhibit peptide bond formation by interfering with the passage of an elongating polypeptide down the peptide exit tunnel. A few of the antibiotics that target the 50S subunit bind near the factor binding site.

4.3.3

MLS_B Antibiotics

The largest class of antibiotics that inhibit the function of the 50S subunit is the MLS_B group (Macrolide, Lincosamide and Streptogramin B) (Fig. 4.6) many of which are used to treat human diseases [17]. Macrolides, Lincosamides and Streptogramin B antibiotics are considered as a group, not based on any chemical similarity, but because their effects on the ribosome are similar. For example, when ribosomes are exposed to any MLS_B antibiotic, a specific group of nucleotides becomes protected from chemical modification (Fig. 4.4). Mutation of many of these same nucleotides confers resistance to antibiotics from all three groups (Fig. 4.4). Prominent in this group of implicated nucleotides are *Ec*¹⁾ A2058 and A2059 which are strongly protected by most MLS_B antibiotics. Finally, mutation or modification of *Ec* A2058 confers strong resistance to MLS_B antibiotics. These observations suggest that *Ec* A2058 and A2059 may be near the center of the region to which all MLS_B antibiotics bind.

X-ray structures of several macrolides [4, 8] and of one lincosamide [4] have been solved in complexes with ribosomes. Although no structure of a streptogra-

1) Throughout the text, nucleotide numbers will correspond to the organism being described and will be preceded by its initials: *Hm*=*Ha-loarcula marismortui*, *Ec*=*Escherichia coli*, *Tt*=

Thermus thermophilus, *Dr*=*Deinococcus radiodurans*. The corresponding standard nucleotide numbering from *E. coli* will normally follow in parentheses.

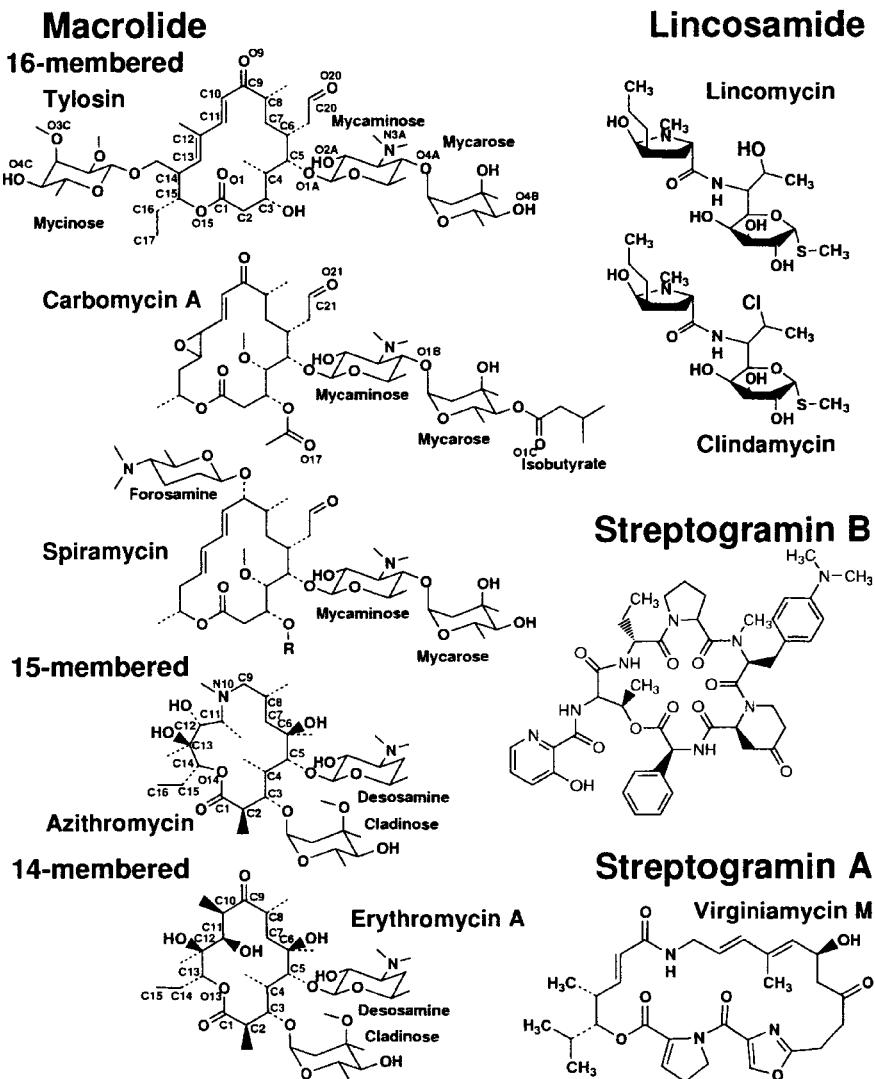


Fig. 4.6 Chemical structures of MLS_B antibiotics. Macrolides are comprised of a central lactone ring of 14, 15 or 16 atoms, from which extend various sugar groups and functional groups. Lincosamides are comprised of two

sugars linked by an amide bond. Like macrolides, streptogramins contain a large central ring. Streptogramin A is included because it binds the ribosome cooperatively with streptogramin B. Fig. combined from [7, 8].

min B in a complex with the ribosome is yet available, there is a structure of a streptogramin A, and it provides considerable insight into the biochemistry of streptogramins. Consistent with the biochemical data, the antibiotics bind near the entrance to the peptide exit tunnel and interact with the hydrophobic crevice formed by *Ec* A2058 and A2059 (Fig. 4.5).

4.3.4

Macrolides

Macrolide antibiotics consist of a central lactone ring from which extend various functional groups and sugar substituents (Fig. 4.6). Macrolides are divided into three subgroups depending on the number of atoms in the lactone ring: 14-membered, 15-membered and 16-membered. Inhibition of protein translation by macrolides has two distinct characteristics. Firstly, macrolides will neither bind to [18] nor inhibit [19] a ribosome that is already translating mRNA, that is, a ribosome that already has a nascent peptide in its exit tunnel. Secondly, macrolides do not directly inhibit formation of a peptide bond. Instead, when macrolides are added to a translational system, di-peptides, tri-peptides and even tetra-peptides accumulate [20]. However, formation of longer peptides is inhibited.

Structures for three 16-membered macrolides and one 15-membered macrolide bound to the archaeal ribosome of *Haloarcula marismortui* (*Hm*) are available [8], as are structures for four 14-membered macrolides bound to the bacterial ribosome of *Deinococcus radiodurans* (*Dr*) [4]. Interestingly, it appears that the 14-membered macrolides do not bind to the bacterial ribosomes in the same way that the 15- or 16-membered macrolides bind to the archaeal ribosomes. Nevertheless, in both systems the macrolides bind to the same site, at the entrance to the peptide exit tunnel, consistent with biochemical inferences about the way that macrolides interfere with elongating peptides. The mode of binding observed for 15- and 16-membered macrolides will be discussed first.

4.3.4.1 Macrolides, 15- and 16-Membered

In *H. marismortui*, the 15-membered and 16-membered macrolides studied so far bind to the ribosome almost identically. When rRNA portions of these structures are superimposed, the lactone rings of the macrolides become superimposed on an almost atom by atom basis (Fig. 4.7). At the center of the macrolide binding site is the hydrophobic crevice at the entrance to the peptide exit tunnel between *Hm* G2099 and A2100 (*Ec* A2058 and A2059) (Fig. 4.8).

The bound macrolides almost completely occlude the peptide exit tunnel, which explains the two distinctive characteristics of macrolide inhibition. Firstly, the reason that macrolides do not inhibit ribosomes that are already actively making protein is that the nascent peptide in the exit tunnel blocks access to the macrolide binding site. Secondly, the reason macrolides do not directly inhibit the peptidyl transferase reaction is that they bind near to, but not directly at, the active site. Only after a few peptide bonds are formed will the peptide contact the bound macrolide and be blocked from further elongation. As a result, short di-, tri- and tetra-peptides will accumulate.

Interestingly, these crystal structures also make it possible to understand more subtle differences among the inhibitory effects of different macrolides. As a peptide elongates, the portion of the macrolide it first encounters will be the sugar branch that extends from C5 of the lactone ring (Fig. 4.6) towards the peptidyl

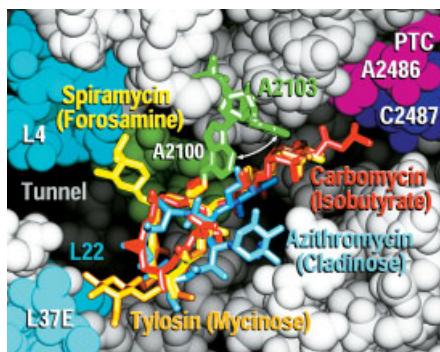


Fig. 4.7 Superposition of macrolides. A cutaway view of a space-filled representation of rRNA (gray) and protein (light blue) show the peptide exit tunnel (left) and the peptidyl transferase center (upper right) of *H. marismortui*. The lactone rings of tylosin (orange sticks), carbomycin A (red sticks), spiramycin (yellow sticks) and azithromycin (light blue sticks) become superimposed when rRNA is superimposed among these structures. The lactone rings bind to the hydrophobic crevice

between *Hm* G2099 and A2100 (green spheres) at the entrance to the peptide exit tunnel. *Hm* A2103 (dark green sticks) changes conformation (white arrow and light green sticks) and forms a covalent bond to 16-membered macrolides. One distinct sugar group (in parentheses) extends from the lactone ring of each antibiotic. The isobutyrate group of carbomycin extends into the active site crevice (purple and blue spheres) between *Hm* A2486 and C2487. Fig. from [8].

transferase active site (Fig. 4.7). Thus, those macrolides that have only a monosaccharide branch extending from C5 allow the synthesis of a longer oligo-peptide than those with longer branches extending from C5. For example, in the presence of erythromycin, which has only a monosaccharide at C5, tetra-peptides accumulate [20]. In the presence of tylosin and spiramycin that have longer disaccharide branches at C5, only di-peptides accumulate. Finally, carbomycin A has an isobutyrate group that extends from its disaccharide branch at C5 all the way into the active site hydrophobic crevice, and strongly inhibits formation of even di-peptides.

4.3.4.2 Binding Interactions Between the Lactone Ring and the Ribosome

Hydrophobic interactions dominate the binding of the 15- and 16-membered lactone rings of macrolide antibiotics to the ribosome. One face of the lactone ring is hydrophobic, because it is comprised entirely of carbon atoms, with the exception of a single hydrogen bond acceptor group, which is the linking oxygen of the ester bond (Figs. 4.6 and 4.8). In contrast, the opposite face contains many hydrophilic groups. Crystal structures of the 16-membered macrolide antibiotics tylosin, carbomycin M and spiramycin, and of the 15-membered macrolide antibiotic azithromycin show that the hydrophobic face of the lactone ring lies almost flat against the wall of the peptide exit tunnel at a site where it is able to form van der Waals contacts with hydrophobic portions of 23S rRNA (Fig. 4.8). For exam-

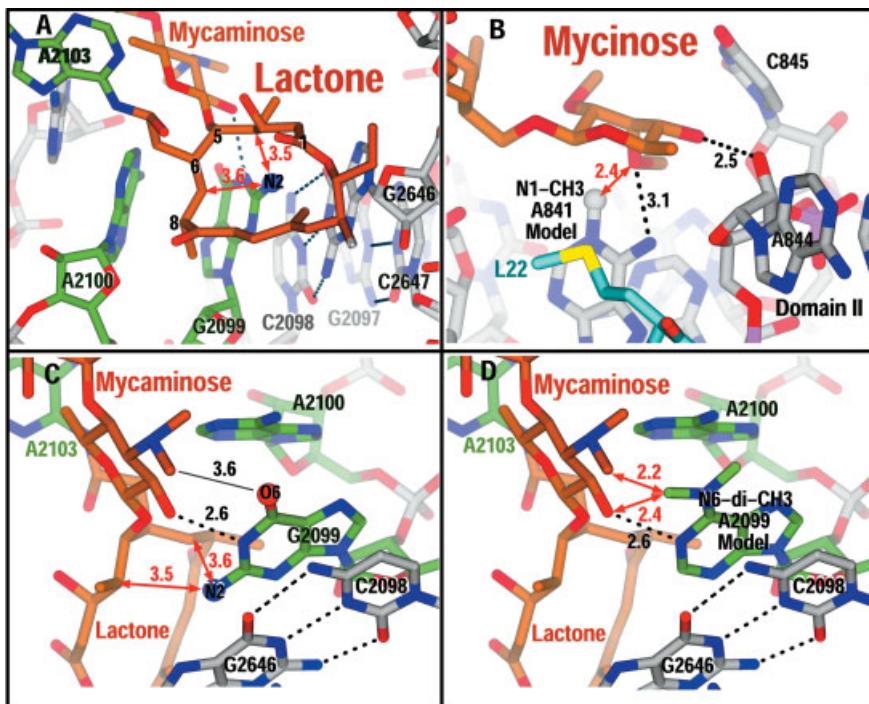


Fig. 4.8 Macrolide binding interactions and antibiotic resistance. The lactone ring of a macrolide (orange), illustrated by tylosin, interacts with the hydrophobic crevice at the entrance to the exit tunnel between *Hm* G2099 and A2100 (*Ec* A2058 and 2059) and with the helix terminating base pair *Hm* C2098-G2646 (*Ec* G2057-C2611). The sugars form hydrogen bonds (dotted lines) to ribosomal RNA. (A) The exocyclic amine (N2) of *Hm* G2099 is positioned at the center of these hydrophobic interactions, and forms unfavorable contacts

(red arrows) with C4 and C7 of the lactone ring. (B) The mycinose sugar of tylosin interacts with domain II (gray sticks) of rRNA and directly contacts large ribosomal subunit protein L22 (light blue). A model of an N1 methyl group (N1-CH3) has been added to the crystal structure. (C) Alternative view of the hydrophobic face of tylosin in binding site. (D) Model building by changing G2099 to dimethyl-adenine (N6-di-CH3), the base common to macrolide producing organisms.

ple, one edge of the lactone ring (atoms C5 to C8) follows a hydrophobic crevice between *Hm* G2099 and A2100 (*Ec* 2058 and 2059), and the other edge of the lactone ring forms van der Waals contacts with the hydrophobic face of *Hm* G2646 (*Ec* C2611) which is exposed because it forms a helix-terminating base-pair with *Hm* C2098 (*Ec* G2057). The hydrophilic groups on the opposite face of the lactone ring are exposed to solvent and do not interact with the ribosome.

The two nucleotides that form the hydrophobic crevice at the entrance to the exit tunnel, *Hm* G2099 and A2100 (*Ec* A2058 and A2059), are also the two nucleotides that are biochemically most implicated in MLS_B binding. Not only are they strongly protected by MLS_B antibiotics from chemical modification, but mutations

of these nucleotides confer MLS_B resistance (Fig. 4.4). Furthermore, bacterial ribosomes that have a highly conserved adenine at *Ec* 2058 are far more sensitive to MLS_B antibiotics than are archaeal ribosomes that in contrast have a highly conserved guanine in that position. Mutation of *Ec* A2058 to guanine in bacteria confers strong resistance to macrolides, especially 14-membered macrolides.

Even using crystal structures of macrolides bound to the less sensitive archaeal ribosomes, the mechanism of the resistance conferred by the *Ec* A2058 to G2058 mutation can be understood. Adenine differs chemically from guanine in only two ways. In guanine, the obligate hydrogen bond donor, N6 of adenine, is replaced by the obligate hydrogen bond acceptor, O6, and adenine lacks the exocyclic amine N2 of guanine. It appears that resistance is caused specifically by the exocyclic N2 of guanine, not by the replacement of N6 by O6. The hydrophilic N2 group of *Hm* G2099 (*Ec* A2058 to G) is placed in the middle of the hydrophobic lactone ring binding site (Fig. 4.8A,B). As a result, the hydrophilic N2 group is unable to form hydrogen bonds with solvent, and the binding site is reduced both in size and in hydrophobicity. The N2 group of *Hm* G2099 contacts both C4 and C7 of the lactone ring. Removal of the hydrophilic N2 of *Hm* G2099 (*Ec* A2058 to G) would enable the lactone ring to fit more tightly against its binding site and to form additional van der Waals contacts. Despite the presence of the N2 group, the 16- and 15-membered macrolides do bind to these less sensitive archaeal ribosomes. This may be due in part to the larger size of their lactone rings which allows more room for the intrusive N2 of *Hm* G2099, but in the case of 16-membered macrolides, other binding features may also come into play (see below).

4.3.4.3 Sugar Interactions with the Ribosome

The sugar branch that extends from C5 of the lactone ring of all macrolides and extends towards the peptidyl transferase center, also contributes to binding affinity (Figs. 4.7 and 4.8). In the case of 16-membered macrolides the first sugar of this branch is mycaminose, and a hydrogen bond forms between its O2A and the N1 of *Hm* G2099 (*Ec* A2058), which also interacts hydrophobically with the lactone ring of macrolides. It is not surprising that G2099 is the nucleotide that is most strongly protected by macrolide binding from chemical modification.

Interestingly, resistance to macrolides can be conferred not only by mutation of *Ec* A2058 to guanine, but also by di-methylation of *Ec* A2058 at N6. This modification is observed in most organisms that produce macrolides. It is easy to examine the effects of this modification by a model building exercise in which *Hm* G2099 is replaced by dimethyl A2099 in the crystal structure (Fig. 4.8C,D). This reveals that one of the added methyl groups will interfere with the atom O2A of the mycaminose sugar (Fig. 4.8C,D), and by pushing the sugar away from its normal position will prevent formation of the hydrogen bond between N1 of *Hm* G2099 and O2A of the macrolide.

When tylosin binds to the ribosome, the mycinose extension from C14 of the lactone ring extends down the exit tunnel and interacts with domain II of rRNA (Fig. 4.8B). Nucleotide modification by N1 methylation of *Ec* G748 (*Hm* A841) in

domain II, which confers tylosin resistance, is at the bottom of the pocket in which the mycinose sugar binds. Model building suggests that the added methyl group will clash with O4C of the tylosin mycinose, and prevent it from forming hydrogen bonds to the ribosome (Fig. 4.8B). Not surprisingly, this modification does not alter the affinity of other macrolides for the ribosome, because they do not have a mycinose at C14. In a rational drug design experiment, removal of both O4C and its methyl group might reduce the level of resistance conferred by the methyl modification of *Ec* G748, by reducing the steric clash, even though one hydrogen bond would still be lost.

4.3.4.4 A Covalent Bond

The crystal structures of the three 16-membered macrolides reveals that each forms a covalent bond with 23S rRNA (Figs. 4.7 and 4.8A). In the absence of substrates and of antibiotics, the base of *Hm* A2103 (*Ec* 2062) lies flat against the wall of the exit tunnel of archaeal ribosomes, and forms a non-canonical base pair with *Hm* A2538 (*Ec* 2503). However, upon binding of tylosin, carbomycin M, or spiramycin the adenine base of *Hm* A2103 (*Ec* 2062) rotates approximately 90°. *Hm* A2103 extends out from the wall of the exit tunnel and forms a carbinolamine bond between its N6 and the aldehyde group extending from C6 of the macrolide (Fig. 4.9).

Formation of this carbinolamine bond is revealed by the continuous electron density that connects the aldehyde group of the antibiotic to *Hm* A2103 in the unbiased electron density maps obtained for each of the 16-membered macrolides studied to date, each of which has an aldehyde group extending from C6. Not surprisingly, that continuous electron density feature is absent in the unbiased electron density map calculated for the structure of the 15-membered macrolide azithromycin which has no aldehyde group at C6. The electron density feature is consistent with formation of a carbinolamine bond and not with a Schiff base, consistent with the expected chemical reactivity between exocyclic amines and aldehydes (Fig. 4.9) [21].

The carbinolamine bond observed in these crystal structures has not been detected biochemically. Nevertheless, its formation may explain the results of a number of previous biochemical experiments. Modification or removal of the aldehyde group from a macrolide results in 100-fold increases in minimum inhibitory concentrations (MIC) [22–28]. Furthermore, mutation of *Hm* A2103 (*Ec* 2062) to guanine (which has an O6 instead of N6) confers resistance specifically to macrolides that have an aldehyde group at C6, but not to others [29]. It seems likely that the carbinolamine bonds observed in these crystal structures also form *in vivo*, and are physiologically relevant to the inhibitory effects of macrolides.

This novel covalent interaction between antibiotics and the ribosome might be exploited in rational drug design experiments. The ribosome is full of exocyclic amines many of which will be found in crystal structures to be close to bound antibiotics. Perhaps an aldehyde group could be added to one of these antibiotics and be designed to form a carbinolamine bond with a nearby exocyclic amine and to thereby increase binding affinity. Of course, the aldehyde group itself might undergo unanticipated reactions.

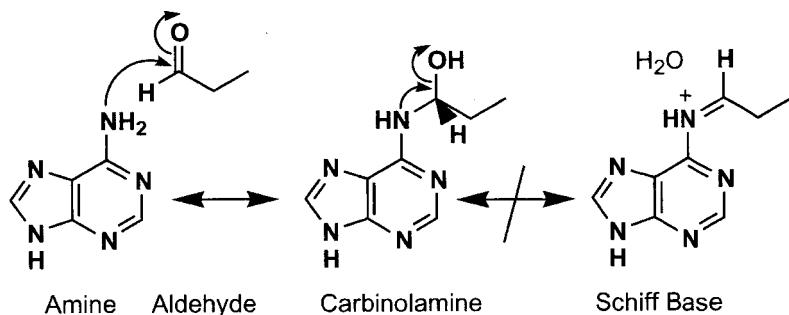


Fig. 4.9 Carbinolamine bond. An amine can react with an aldehyde (left) to form a carbinolamine intermediate (center) that normally dehydrates to form a Schiff base (right). However, the exocyclic amine of a nucleoside base can form only the carbinolamine bond and not a Schiff base.

4.3.4.5 Macrolides, 14-Membered

X-ray crystal structures of the three 14-membered macrolides, erythromycin, clarithromycin and roxithromycin have been solved bound to the *Deinococcus radiodurans* (*Dr*) large subunit [4]. They bind at about the same location as the 15- and 16-membered macrolides, near *Ec* A2058 and A2059 (*Hm* G2099 and 2100) at the entrance to the exit tunnel. The desosamine sugar of the 14-membered macrolides is positioned very much like the corresponding desosamine sugar of azithromycin and the mycaminose sugar of the 16-membered macrolides in complexes with *H. marismortui* ribosomes.

The difference between these two classes of structures is in the way their lactone rings interact with the ribosome. Hydrophobic interactions dominate the binding of 15- or 16-membered lactone rings (Fig. 4.8) which bind with their hydrophobic faces flat against the wall of the exit tunnel. In contrast, the lactone ring of each of the three 14-member macrolides is oriented approximately perpendicular to that of 15- and 16-membered lactone rings, which exposes their hydrophobic faces to solution. Instead, of hydrophobic interactions, it is proposed that hydrophilic interactions mediate the binding of 14-membered macrolides to the ribosome. Only one of the six hydrogen bonds proposed to stabilize binding of 14-membered lactone rings is observed to mediate binding of 15- and 16-membered macrolides to the ribosome. That hydrogen bond, which connects O2A of the mycaminose sugar to N1 of *Hm* G2099, may be important to macrolide binding, because it is observed in all of these structures regardless of the other differences in binding. Finally, the binding of the 14-membered lactone rings differs from that of the 15- and 16-membered rings because the former [4] assume a “folded-in” conformation, and the latter [8] assume a “folded-out” conformation [30].

4.3.4.6 Rational Drug Design of Macrolides

Understanding the mechanism of resistance conferred by a specific mutation is expected to assist in rational drug design experiments. As described above, crystal

structures reveal the mechanisms by which the *Ec* A2058 to G2058 mutation confers resistance to macrolide antibiotics. However, knowledge of that specific resistance mechanism might not be useful in rational drug design experiments. Guanine at *Ec* 2058 is highly conserved in eukaryotes, including humans, and so a macrolide redesigned to overcome that resistance mutation might also bind more tightly to human ribosomes.

The structures of macrolide antibiotics bound to the ribosomes seem to provide numerous possibilities for rational drug design experiments. However, a few of these experiments may require an understanding of what causes the 14-membered macrolides to bind to the ribosome differently than do the 15- and 16-membered macrolides. With respect to the ribosome, some atoms of the 14-membered macrolides are located over 6 Å away from the corresponding atoms of the larger macrolides. Knowledge of which mode of binding a given antibiotic assumes would be crucial to a rational drug design experiment that involves those atoms.

One possible cause of the different macrolide binding modes is that the smaller size of the 14-membered lactone ring might cause them to bind differently than do the larger ones. If so, rational drug design experiments on 14-membered macrolides may need to preserve and supplement the relevant hydrophilic interactions, such as the six hydrogen bonds that mediate their binding. In contrast, rational drug design experiments on 15- or 16-membered macrolides may need to focus instead on the hydrophobic interactions that mediate the binding of their larger lactone rings.

A second possibility is that the two distinct modes of macrolide binding are caused by the conserved differences between archaeal and bacterial ribosomes. A good candidate for such a conserved difference is the nucleotide *Hm* G2099 (*Ec* A2058). *Hm* G2099 is located at the center of the macrolide binding site and is highly conserved as guanine in archaea and as adenine in bacterial ribosomes. Furthermore, mutation of *Ec* 2058 to G2058 affects macrolide binding and confers resistance to macrolides. If the G2099 to A change does cause the observed differences in binding between these two systems, then the structure of a macrolide bound to the bacterial ribosome might be the most relevant for rational drug design purposes. An archaeal pathogen has not yet been described.

A third possibility is that the differences in macrolide binding between these sets of experiments result from inaccuracies in the models. Relevant to this possibility, the 14-membered macrolide structures are based on lower resolution data and have higher free *R* values than do the other macrolide structures.

These alternate modes of macrolide binding not only pose a challenge to rational drug design but potentially provide new opportunities in this area. Perhaps features from both modes of binding could be exploited by a single antibiotic. For example, the binding affinity of a 15- or 16-membered macrolide antibiotic might be increased by attaching a hydrogen bonding group that is designed to extend from its hydrophilic face towards one of the hydrogen bonding partners that was observed to mediate binding of a 14-membered macrolide.

Many rational drug design experiments may be possible even prior to understanding the cause of these differences in binding. Since the desosamine sugar of

the 14-membered macrolides binds similarly to the corresponding mycaminose sugars of the larger macrolides, rational drug design experiments involving that sugar might be independent of the mode of binding. Accordingly, the differences in binding may be irrelevant to experiments that involve any of the atoms of the C5 sugar branch.

4.3.5

Lincosamides

Currently, one structure of a lincosamide antibiotic bound to the ribosome is available for analysis [4]. Like the macrolide antibiotics, clindamycin binds near the hydrophobic crevice at the entrance to the peptide exit tunnel. As with the macrolide carbomycin A, clindamycin interacts not only with the hydrophobic crevice at the entrance to the peptide exit tunnel, but also with the active site hydrophobic crevice. The nucleotides that surround the clindamycin binding site were previously implicated in binding of lincosamides based on nucleotide protection studies and on the analysis of mutations conferred by resistance (Fig. 4.4).

Since lincosamide and macrolide antibiotics share binding sites that only partially overlap, it might be possible to design a combination antibiotic. As this experiment would involve a macrolide, uncertainty over which of the two binding modes should be used for the macrolide might complicate the experiment. Nevertheless, it may be possible to apply this approach of combining the binding features of two antibiotics into one among the various antibiotics that share overlapped binding sites near the peptidyl transferase center (Fig. 4.5).

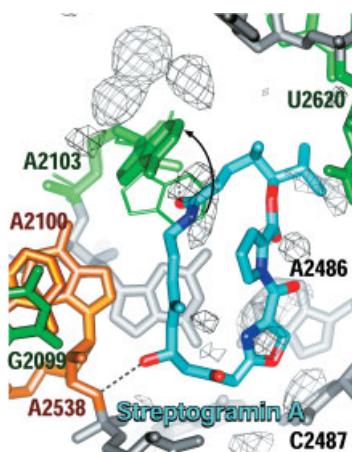
4.3.6

Streptogramins

The chemical structures of streptogramin antibiotics are unlike those of either macrolides or lincosamides (Fig. 4.6) and are unique in that organisms that synthesize them invariably make two different compounds, one being a streptogramin A and the other a streptogramin B. Streptogramin A and streptogramin B antibiotics are independently effective as antibiotics, but are much more effective when used in combination, because they bind cooperatively to the ribosome [18, 31]. Apparently, streptogramin A causes a conformational change of the ribosome near the peptidyl transferase center that increases the affinity of the ribosome for streptogramin B [32].

Although no structure of a streptogramin B bound to the ribosome is available, there is a structure for streptogramin A bound to the ribosome [7]. Streptogramin A is a large antibiotic that binds in a location that overlaps both the A-site and the P-site (Fig. 4.5), consistent with reports that streptogramin A inhibits binding of substrates to both sites [32, 33]. When streptogramin A binds to the ribosome, the base of *Hm* A2103 (*Ec* A2062) rotates about 90° away from the wall of the exit tunnel (Fig. 4.10) much as it does when 16-membered macrolides bind. This conformational change allows the base of *Hm* A2103 (*Ec* 2062) to interact hydropho-

Fig. 4.10 Streptogramin A. Streptogramin A (light blue) binds to rRNA (gray) among nucleotides that are implicated in its binding by chemical modification (green) or resistance mutation (orange) studies. In response to streptogramin A binding, the *Hm* A2103 (*Ec* A2062) changes conformation (black arrow) from its native position against the wall of the exit tunnel (thin green sticks) and a position that extends into the exit tunnel (thick green sticks). In this extended conformation, a face of the base of *Hm* A2103 stacks against the conjugated amide group of streptogramin A, and the ribose sugar forms a hydrogen bond (dotted line) with the same amide group.



bically with the flat side of the conjugated amide bond of streptogramin A, and the sugar of *Hm* A2103 to hydrogen bond to the oxygen of the same amide bond. In addition, the nucleotides of the active site crevice undergo smaller conformational changes [7]. It appears that the far side of the base of *Hm* A2103 forms part of the binding site for streptogramin B.

4.3.7

Chloramphenicol

Chloramphenicol is widely used in research and in the clinic, although it is slightly toxic to human mitochondria [34]. Its chemical structure (Fig. 4.11) has been compared to nucleoside analogues such as puromycin, but biochemically the inhibition of protein translation by chloramphenicol is more like that of MLS_B antibiotics, a group with which it shares no chemical similarity. For example, many of the nucleotides that are protected from chemical modification by MLS_B antibiotics (e.g., *Ec* A2058 and A2059) are also protected by chloramphenicol, and the same mutations confer resistance to both types of antibiotics (Fig. 4.4). Furthermore, oligo-peptides are reported to accumulate when protein translation is inhibited by chloramphenicol or by macrolides [35]. However, chloramphenicol binding also protects *Ec* A2451 and C2452 at the active site hydrophobic crevice, and the only macrolide that protects *Ec* A2451 is Carbomycin A, evidently because of its isobutyrate extension. Furthermore there are nucleotides protected by MLS_B antibiotics that are not protected by chloramphenicol.

The interactions of chloramphenicol with the ribosome have been studied for decades, and the some of the results are confusing. For example, the binding of chloramphenicol to the ribosome has been reported variously as competitive [36–40], independent [41], and cooperative [40] with respect to macrolide binding. These inconsistencies may reflect the binding of chloramphenicol to two sites on the large ribosomal subunit that differ in affinity [42].

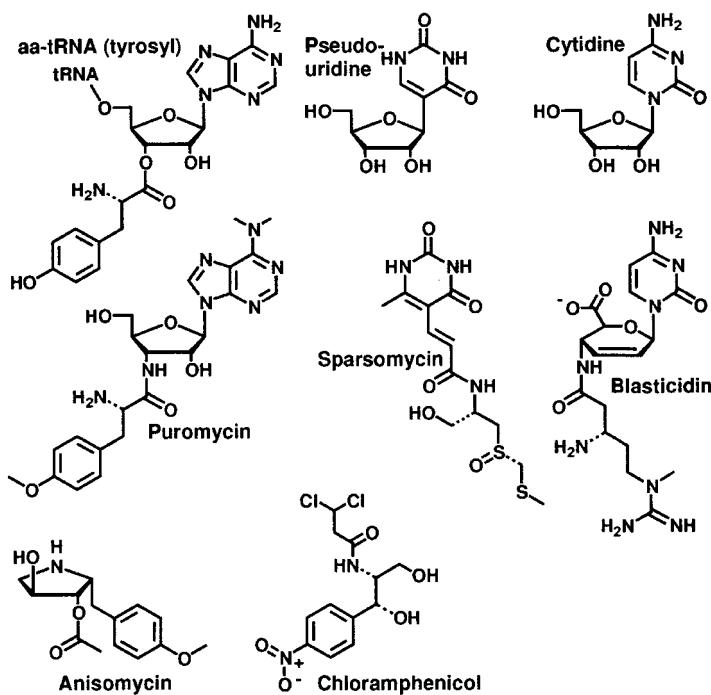
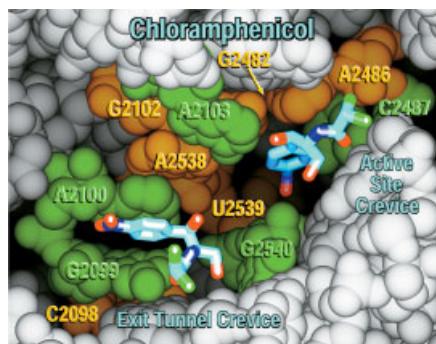


Fig. 4.11 Chemical structures of antibiotics. Chemical structures of nucleosides are provided in the top row for comparison with the antibiotics shown in the second row. The third row shows chemical structures of antibiotics that only weakly resemble the nucleosides to which they have been compared.
Fig. modified from [7].

Two crystal structures of chloramphenicol bound to the ribosome are available. In one structure, chloramphenicol is observed to bind only at the active site hydrophobic crevice of the bacterial (*D. radiodurans*) ribosome [4]. In the other structure chloramphenicol binds only at the hydrophobic crevice at the entrance to the exit tunnel of an archaeal (*H. marismortui*) ribosome [7]. Both of these sites are surrounded by nucleotides implicated in chloramphenicol binding either by nucleotide protection studies or by mutational studies (Fig. 4.12). They probably correspond to the two sites inferred from biochemical experiments.

Currently, it seems likely that the high affinity binding site of chloramphenicol in bacterial ribosomes is the active site hydrophobic crevice and that the low affinity site corresponds to the crevice at the entrance to the peptide exit tunnel. In addition, it seems reasonable that binding of chloramphenicol at the active site crevice might have a greater inhibitory effect on protein translation than binding at the more distant site.

Fig. 4.12 Chloramphenicol. A cutaway view of a space filled representation of the *H. marismortui* ribosome chloramphenicol bound at two sites, the active site hydrophobic crevice [4] and at the hydrophobic crevice at the entrance to the exit tunnel [7]. Both of these binding sites are surrounded by nucleotides that upon mutation confer resistance to chloramphenicol (orange spheres) or that are protected by chloramphenicol from chemical modification (green spheres). Fig. from [7].



4.3.8

Nucleoside Analogue Antibiotics

Many antibiotics have been described as nucleoside analogues, because they contain chemical groups that resemble portions of an aminoacylated nucleoside (Fig. 4.11). However, few of those antibiotics actually bind to the ribosome in a manner analogous with nucleosides. For example, chloramphenicol has been compared with the aminoacyl-nucleoside substrate puromycin. Its aromatic group has been compared both with the base of a nucleoside [43] and with an aromatic amino acid side chain [44]. Although, the aromatic group of chloramphenicol approaches the active site crevice from the same direction as the tyrosyl side chain of puromycin (Fig. 4.12), the binding of chloramphenicol to the ribosome is neither like that of an amino acid side chain nor a nucleoside base (Fig. 4.12). Nevertheless, a few antibiotics contain chemical groups that are similar to portions of an aminoacylated nucleoside in chemical structure and in their binding interactions with the ribosome.

4.3.8.1 Puromycin

The prototypical aminoacylated nucleoside analogue antibiotic is puromycin which inhibits the protein translation in all three domains of life. The chemical structure of puromycin is the same as that of tyrosylated adenosine, except for the presence of three added methyl groups and the replacement of an ester bond with an amide bond (Fig. 4.11). Puromycin mimics tyrosyl-tRNA so well that it binds to the A-site and gets incorporated into an elongating peptide. This leads to termination of translation because puromycin terminated peptides fall off the ribosome. Puromycin derivatives have been used crystallographically as peptidyl transferase substrates and have contributed to our understanding of the structure of the peptidyl transferase site (Fig. 4.5) [11, 16, 45].

4.3.8.2 Aminoacyl-4-aminohexosyl-cytosine Antibiotics

The aminoacyl-4-aminohexosyl-cytosine group of antibiotics is classified based on chemical structures (Fig. 4.11) each of which contains a cytosine base. Included in this group are blasticidin S, gougerotin, amicetin and bamicetin [46]. These antibiotics are not used therapeutically. However blasticidin S is an important antifungal used on rice crops. Only the crystal structure of blasticidin S bound to the ribosome is available. The chemical structure of blasticidin S resembles that of peptidyl-cytidine, although not as closely as puromycin resembles a tyrosylated adenine (Fig. 4.11).

In the crystal structure, blasticidin S binds to two sites on the ribosome (Fig. 4.13) and forms base-pairs with *Hm* G2284 and G2285 (*Ec* 2251 and 2252). These two guanines are almost totally conserved residues of the P-loop, and are known to base pair with the CCA end of tRNA in order to position the P-site substrate in the peptidyl transferase center [47].

The blasticidin S molecule that base pairs with *Hm* G2284 is the better ordered of the two, and its structure can be fit completely into unbiased electron density maps. The other blasticidin S molecule is represented by less complete electron density, and only its base and sugar groups can be modeled unambiguously. When the phosphates of rRNA are superimposed between one structure containing blasticidin S and another structure containing a P-site substrate bound to the ribosome, both the base and sugar groups of the well ordered blasticidin S molecule superimpose on C75 of a P-site substrate bound to the ribosome (Fig. 4.13). For the less ordered blasticidin S molecule that base pairs with *Hm* G2285, the superimposition with C74 of the P-site substrate is limited to only the nucleoside base. Therefore, blasticidin S, like puromycin, acts in the same manner as a nucleoside analogue in its interaction with the ribosome.

The guanidinium tail of the well ordered blasticidin S binds hydrophobically to *Hm* A2474, which is protected by blasticidin S from chemical modification (Fig. 4.13). In addition, the nitrogens of its guanidinium tail hydrogen bond with the phosphate that connects *Hm* A2474 to U2473, a nucleotide that upon mutation confers resistance to blasticidin S [48]. These biochemical observations sug-

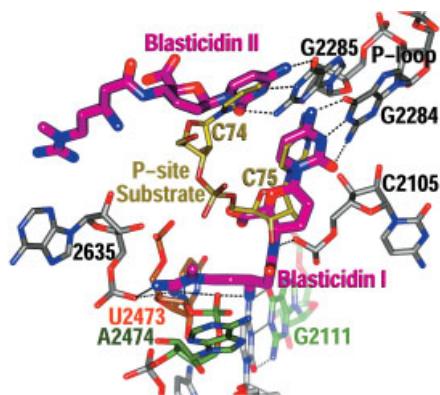


Fig. 4.13 Blasticidin S. Blasticidin S (purple) base pairs to the P-loop of rRNA (grey). Its guanidinium tail forms hydrogen bonds (dotted lines) to the phosphate and stacks hydrophobically onto the base of *Hm* A2474, a nucleotide that is protected from chemical modification (green) by blasticidin S. Superimposition of rRNA between crystal structures causes blasticidin S to become superimposed on C74 and C75 (yellow) of the analogue of a CCA end of tRNA. Fig. from [7].

gest that the blasticidin S molecule that is best ordered in the structure corresponds to the blasticidin S molecule that is also the most active physiologically. Whether or not binding of blasticidin S at the second site is also physiological relevant warrants further investigation.

4.3.9

Other Antibiotics that Bind to the 50S Subunit

4.3.9.1 Sparsomycin

Sparsomycin inhibits protein synthesis in all organisms and consequently is not used as an antibiotic. However, it has been tested extensively for its antitumor activity [49]. Sparsomycin contains a pseudouracil base with a conjugated link to a sulfur containing tail (Fig. 4.11). This chemical structure has been compared with that of puromycin, and it has been suggested that sparsomycin binds to the A-site much like puromycin [50]. Sparsomycin does not bind to the ribosome unless a P-site substrate is also present.

The structure of sparsomycin and a P-site substrate analogue simultaneously bound to the 50S ribosomal subunit is available. The binding of sparsomycin to the ribosome bears no similarity to the binding of puromycin. The pseudouracil group of sparsomycin binds on top of a P-site substrate analogue and is sandwiched between it and the mobile nucleotide *Hm* A2637 (Fig. 4.14). Sparsomycin's sulfur containing tail extends from above the P-site into the A-site hydrophobic crevice, where it should compete for binding with an A-site substrate. These binding characteristics account for the failure of sparsomycin to bind the ribosome in the absence of a P-site substrate, its cooperative binding with a P-site substrate and for its inhibition of peptide bond formation. Each of the functional groups on the pseudouracil group forms favorable interactions with the P-site substrate or with the ribosome.

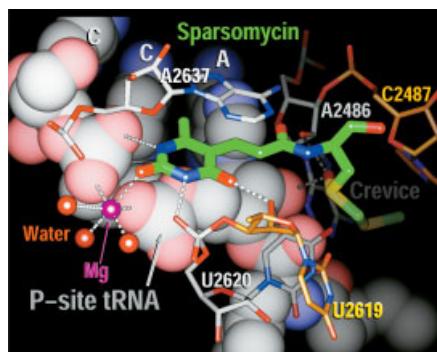


Fig. 4.14 Sparsomycin. Sparsomycin (green sticks) is sandwiched between a tRNA analogue (gray spheres) and *Hm* A2637 (white sticks) and forms numerous hydrogen or ionic bonds (dotted lines) with tRNA or rRNA. Fig. from [7].

4.3.9.2 Anisomycin

Like sparsomycin, anisomycin inhibits eukaryotic ribosomes and has been tested as an antitumor drug. Also like sparsomycin, the chemical structure of anisomycin has been compared with that of puromycin [44]. The crystal structure of anisomycin bound to the ribosome shows that anisomycin binds to the active site hydrophobic crevice (Fig. 4.15). The methoxyphenyl group of anisomycin, like that of puromycin, inserts into the active site hydrophobic crevice. However, the binding of anisomycin is unlike that of puromycin, because it approaches the active hydrophobic crevice from the opposite direction (Fig. 4.5). As the aromatic group of anisomycin stacks on *Hm* C2487 (*Ec* 2452), its sugar group occupies the oblong pocket between the two hydrophobic crevices (Figs. 4.6 and 4.15). The nitrogen of the sugar group forms a hydrogen bond with N3 of *Hm* C2487. When *Hm* C2487 is mutated to uridine, halophiles become resistant to anisomycin [51]. The simplest explanation is that this mutation eliminates the hydrogen bond by replacing the N4 hydrogen bond acceptor of cytidine with an N4 hydrogen bond donor of uridine. This hypothesis could be tested to provide insight into rational drug design principles by replacing N3 of anisomycin with an oxygen atom. This alternate form of anisomycin should preferentially bind the mutant ribosome and, at the same time, should be ineffective against the wild type ribosome.

4.4

Prospects for Rational Drug Design of Antibiotics that Bind to the Ribosome

The time for rational design of antibiotics that bind the ribosome has arrived. Currently, structures of 20 antibiotic/ribosome complexes are readily available in the Protein Data Bank (<http://www.rcsb.org/pdb/>). This chapter has discussed only half of them, the ones that bind to the 50S subunit.

It is anticipated that rational drug design experiments will succeed in the near future, assuming the necessary precautions are taken. First, neither any X-ray nor any NMR structure should be accepted without question. The reliability of any given structure must be verified prior to being used as the basis for a rational drug design experiment. Analysis of X-ray crystallographic statistics will reveal the extent of precision that can be expected of a structure. Evaluating the stereochem-

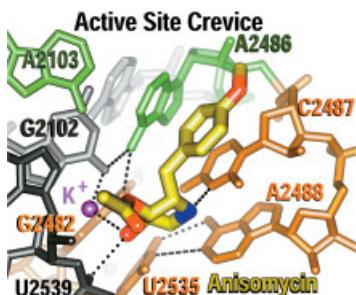


Fig. 4.15 Anisomycin. Anisomycin (yellow) binds to the active site hydrophobic crevice forming numerous hydrogen or ionic bonds (dotted lines) with rRNA. Fig. from 7].

istry of a model by checking bond angles, bond distances, dihedral angles and van der Waals distances will reveal what level of confidence can be placed in a given structure. Ideally, such analysis will also include a complete set of X-ray diffraction amplitudes, although such data are not always released.

The possibility of errors in deposited structures must be considered. Errors in chirality may be prevalent in high resolution small molecule structures. For example, the small molecule structure of streptogramin A is represented by its enantiomer in the Cambridge Structural Data Base. The effort to solve the structure of streptogramin A bound to the ribosome was initially hindered by relying on that incorrect small molecule structure. Likewise, the structure of anisomycin is incorrectly diagrammed as its enantiomer in most of the ribosomal literature. Fortunately, these chirality errors were identified when solving structures of these antibiotics bound to macromolecules. In fact, the identification of these errors may increase confidence in the reliability of these structures of complexes between ribosomes and antibiotics.

Knowledge of the precision of a particular X-ray structure may also be important in deciding whether or not a particular rational drug design experiment is feasible. For example, at a resolution of 3.0 Å, the rotamers for the links between the 16-membered macrolides and sugars are reliable, with the possible exception of the forosamine sugar of spiramycin. However, the rotamers of smaller groups, such as the dimethyl amine of mycaminose (Fig. 4.9), are ambiguous. For most rational drug design experiments, this ambiguity will probably be irrelevant. However, for a rational drug design experiment involving that dimethylamino group, awareness that its conformation is ambiguous could be crucial.

Another factor that may require analysis is the observation that some antibiotics bind to multiple sites. Tetracycline provides the most extreme example and binds to six separate sites on the small ribosomal subunit [52]. The existence of multiple binding sites raises the same challenges and opportunities that are raised by antibiotics that have multiple modes of binding to a single site, such as is reported for macrolides. Knowledge of which of these sites or which of these modes of binding is most physiologically relevant to the inhibitory effects of an antibiotic would be essential for certain rational drug design experiments. At the same time, identifying the less relevant sites to which an antibiotic binds at lower affinity may provide for an alternative strategy. A site which is physiologically irrelevant because its affinity for an antibiotic is low, may become physiologically relevant if an antibiotic can be redesigned to bind with higher affinity.

Many structures of antibiotics bound to the ribosome are available and are of sufficient quality (Tab. 4.1) to warrant rational drug design experiments. If the proper precautions are taken, it is expected that some of these experiments will succeed in the near future.

4.5**Acknowledgements**

I thank Thomas A. Steitz and Peter B. Moore for directing my research on antibiotics bound to the ribosome and for reading this manuscript, Nenad Ban and Poul Nissen for earlier work on the structure of the 50S subunit, Betty Freeborn for technical help and useful scientific discussions, Jimin Wang for help with crystallographic software, Joe Ippolito and Martin Schmeing for help with some antibiotic experiments, Nukri Sanishvili and Andres Joachimiak for help with data collection at ID19 at APS. Much of the research that this chapter is based on was supported by National Institutes of Health Grant (GM22778) to Thomas A. Steitz, and by a grant from Agouron Institute to T.A.S. and Peter B. Moore.

4.6**References**

- 1** BAN, N., NISSEN, P., HANSEN, J., MOORE, P., STEITZ, T., The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science* **2000**, *289*, 905–920.
- 2** WIMBERLY, B.T., BRODERSEN, D.E., CLEMONS, W.M., MORGAN-WARREN, R.J., CARTER, A.P., VONRHEIN, C., HARTSCH, T., RAMAKRISHNAN, V., Structure of the 30S ribosomal subunit. *Nature* **2000**, *407*, 327–339.
- 3** SCHLUENZEN, F., TOCILJ, A., ZARIVACH, R., HARMS, J., GLUEHMANN, M., JANELL, D., BASHAN, A., BARTELS, H., AGMON, I., FRANCESCHI, F., YONATH, A., Structure of functionally activated small ribosomal subunit at 3.3 angstroms resolution. *Cell* **2000**, *102*, 615–623.
- 4** SCHLUENZEN, F., ZARIVACH, R., HARMS, J., BASHAN, A., TOCILJ, A., ALBRECHT, R., YONATH, A., FRANCESCHI, F., Structural basis for the interaction of antibiotics with the peptidyl transferase centre in eubacteria. *Nature* **2001**, *413*, 814–821.
- 5** BRODERSEN, D.E., CLEMONS, W.M., CARTER, A.P., MORGAN-WARREN, R.J., WIMBERLY, B.T., RAMAKRISHNAN, V., The structural basis for the action of the antibiotics tetracycline, pactamycin, and hygromycin B on the 30S ribosomal subunit. *Cell* **2000**, *103*, 1143–1154.
- 6** CARTER, A.P., CLEMONS, W. M. J., BRODERSEN, D. E., MORGAN-WARREN, R. J., WIMBERLY, B. T., RAMAKRISHNAN, V., Functional insights from the structure of the 30S ribosomal subunit and its interactions with antibiotics. *Nature* **2000**, *407*, 306–307.
- 7** HANSEN, J.L., MOORE, P., STEITZ, T., Co-crystal structures of five antibiotics bound at the peptidyl transferase center of the 50S ribosomal subunit. *J. Mol. Biol.* **2003**, *330*, 1061–1075.
- 8** HANSEN, J., IPPOLITO, J., BAN, N., NISSEN, P., MOORE, P., STEITZ, T., The structures of four macrolide antibiotics bound to the large ribosomal subunit. *Mol. Cell* **2002**, *10*, 1–20.
- 9** KLEIN, D.J., SCHMEING, T.M., MOORE, P.B., STEITZ, T.A., The kink-turn; a new RNA motif. *EMBO J.* **2001**, *20*, 4214–4221.
- 10** HARMS, J., SCHLUENZEN, F., ZARIVACH, R., BASHAN, A., GAT, S., AGMON, I., BARTELS, H., FRANCESCHI, F., YONATH, A., High resolution structure of the large ribosomal subunit from a mesophilic eubacterium. *Cell* **2001**, *107*, 679–688.
- 11** NISSEN, P., HANSEN, J., BAN, N., MOORE, P., STEITZ, T., The structural basis of ribosome activity in peptide bond synthesis. *Science* **2000**, *289*, 920–930.
- 12** OGLE, J.M., BRODERSEN, D.E., CLEMONS, W.M., JR., TARRY, M.J., CARTER, A.P., RAMAKRISHNAN, V., Recognition of cognate

- transfer RNA by the 30S ribosomal sub-unit. *Science* **2001**, *292*, 897–902.
- 13** NOLLER, H. F., WOESE, C. R., Secondary structure of 16S ribosomal RNA. *Science* **1981**, *212*, 403–411.
- 14** NOLLER, H. F., KOP, J., WHEATON, V., BROSIUS, J., GUTELL, R. R., KOPYLOV, A. M., DOHME, F., HERR, W., STAHL, D. A., GUPTA, R., WAESER, C. R., Secondary structure model for 23S ribosomal RNA. *Nucleic Acids Res.* **1981**, *9*, 6167–6189.
- 15** CANNONE, J. J., SUBRAMANIAN, S., SCHNARE, M. N., COLLETT, J. R., D'SOUZA, L. M., DU, Y., FENG, B., LIN, N., MADABASI, L. V., MULLER, K. M., PANDE, N., SHANG, Z., YU, N., GUTELL, R. R., The comparative RNA web (CRW) site: an online database of comparative sequence and structure information for ribosomal intron, and other RNA's. *BMC Bioinformatics* **2002**, *3*, 2–31.
- 16** SCHMEING, T. M., SEILA, A. C., HANSEN, J. L., FREEBORN, B., SOUKUP, J. K., SCRANGE, S. A., STROBEL, S. A., MOORE, P. B., STEITZ, T. A., A pre-translocation intermediate in protein synthesis observed in crystals of enzymatically active 50S subunits. *Nat. Struct. Biol.* **2002**, *9*, 225–230.
- 17** SCHONFELD, W., KIRST, H. A., (eds.) *Macrolide Antibiotics. Milestones in Drug Therapy*. **2002**, Birkhauser, Basel.
- 18** CONTRERAS, A., VAZQUEZ, D., Cooperative and antagonistic interactions of peptidyl-tRNA and antibiotics with bacterial ribosomes. *Eur. J. Biochem.* **1977**, *74*, 539–547.
- 19** PESTKA, S., Studies on transfer ribonucleic acid-ribosome complexes. Effect of antibiotics on peptidyl puromycin synthesis on polyribosomes from *Escherichia coli*. *J. Biol. Chem.* **1972**, *247*, 4669–4678.
- 20** MAO, J. C. H., ROBISHAW, E. E., Effects of macrolides on peptide-bond formation and translocation. *Biochemistry* **1971**, *10*, 2054–2061.
- 21** MCGHEE, J., von HIPPEL, P., Formaldehyde as a probe of DNA structure. I. Reaction with exocyclic amino groups of DNA bases. *Biochemistry* **1977**, *14*, 1281–1303.
- 22** KIRST, H., TOOTH, J. E., DEBONO, M., WILLARD, K. E., TRUEDELL, B. A., OTT, J. L., COUNTER, F. T., FELTY-DUCKWORTH, A. M., PEKAREK, R. S., Synthesis and evaluation of tylosin-related macrolides modified at the aldehyde function: a new series of orally effective antibiotics. *J. Med. Chem.* **1988**, *31*, 1631–1641.
- 23** KIRST, H. A., WILLARD, K. E., DEBONO, M., TOOTH, J. E., TRUEDELL, B. A., LEEDS, J. P., OTT, J. L., FELTY-DUCKWORTH, A. M., COUNTER, F. T., Structure-activity studies of 20-deoxo-20-amino derivatives of tylosin-related macrolides. *J. Antibiot.* **1989**, *42*, 1673–1683.
- 24** NARANDJA, A., SUSKOVIC, B., KELNERIC, Z., DJOKIC, S., Structure-activity relationship among polyhydro derivatives of tylosin. *J. Antibiot.* **1993**, *47*, 581–587.
- 25** OMURA, S., KATAGIRI, M., UMEZA, I., KOMIYAMA, K., MAEKAWA, T., SEKIKAWA, K., MATSUMAE, A., HATA, T., Structure-biological activities relationships among leucomycins and their derivatives. *J. Antibiot.* **1968**, *21*, 532–538.
- 26** OMURA, S., MIYANO, K., MATSUBARA, H., NAKAGAWA, A., Novel dimeric derivatives of leucomycins and tylosin, sixteen-membered macrolides. *J. Med. Chem.* **1982**, *25*, 271–275.
- 27** OMURA, S., TISHLER, M., NAKAGAWA, Y., HIRONAKA, Y., HATA, T., Relationship of structures and microbiological activities of the 16-membered macrolides. *J. Med. Chem.* **1972**, *15*, 1011–1015.
- 28** WHALEY, H. A., PATTERSON, E. L., DORNBUSH, A. C., BACKUS, E. J., BOHONOS, N., Isolation and characterization of relomycin a new antibiotic. *Antimicrob. Agents Chemother.* **1964**, *1963*, 45.
- 29** DEPARDIEU, F., COURVALIN, P., Mutation in rRNA responsible for resistance to 16-membered macrolides and streptogrammins in *Streptococcus pneumoniae*. *Antimicrob. Agents Chemother.* **2001**, *45*, 319–323.
- 30** BERTHO, G., LADAM, P., GHARBI-BENAROUS, J., DELAFORGE, M., GIRAUT, J., Solution conformation of methylated macrolide antibiotics roxithromycin and erythromycin using NMR and molecular modelling. Ribosome-bound conformation determined by TRNOE and formation of cytochrome P450-metabolite com-

- plex. *Int. J. Biolog. Macromol.* **1998**, *22*, 103–127.
- 31** DI GIAMBATTISTA, M., HUMMEL, H., BOCK, A., COCITO, C., Action of synergimycins and macrolides on in vivo and in vitro protein synthesis in archaebacteria. *Mol. General Genetics* **1985**, *199*, 323–329.
- 32** CHINALI, G., MOUREAU, P., COCITO, C.G., The action of virginiamycin M on the acceptor, donor, and catalytic sites of peptidyltransferase. *J. Biol. Chem.* **1984**, *259*, 9563–9568.
- 33** PESTKA, S., Studies on the formation of transfer ribonucleic acid-ribosome complexes. XI. Antibiotic effects on phenylalanyl-oligonucleotide binding to ribosomes. *Proc. Natl. Acad. Sci. USA* **1969**, *64*, 709–714.
- 34** LAM, R. F., LAI, J. S., NG, J. S., RAO, S. K., LAW, R. W., LAM, D. S., Topical chloramphenicol for eye infections. *Hong Kong Med. J.* **2002**, *8*, 44–47.
- 35** RHEINBERGER, H. J., NIERHAUS, K. H., Partial release of AcPhe-Phe-tRNA from ribosomes during poly(U)-dependent poly(Phe) synthesis and the effects of chloramphenicol. *Eur. J. Biochem.* **1990**, *193*, 643–650.
- 36** WOLFE, A. D., HAHN, F. E., Mode of action of chloramphenicol: IX. Effects of chloramphenicol upon a ribosomal amino acid polymerization system and its binding to bacterial ribosome. *Biochim. Biophys. Acta* **1965**, *95*, 146–155.
- 37** GRIVELL, L. A., NETTER, P., BORST, P., SLONIMSKI, P., Mitochondrial antibiotic resistance in yeast: ribosomal mutants resistant to chloramphenicol, erythromycin and spiramycin. *Biochim. Biophys. Acta* **1973**, *312*, 358–367.
- 38** VAZQUES, D., Antibiotics which affect protein synthesis: the uptake of ¹⁴C-chloramphenicol by bacteria. *Biochem. Biophys. Res. Commun.* **1963**, *12*, 409–413.
- 39** VAZQUES, D., Antibiotics affecting chloramphenicol uptake by bacteria, their effect on amino acid incorporation in a cell-free system. *Biochim. Biophys. Acta* **1966**, *114*, 289–295.
- 40** LANGLOIS, R., CANTOR, C. R., VINCE, R., PESTKA, S., Interaction between erythromycin and chloramphenicol binding sites on the *Escherichia coli* ribosome. *Biochemistry* **1977**, *16*, 2349–2355.
- 41** OLEINICK, N. L., WILHELM, J. M., CORCORAN, J. W., Nonidentity of the site of action of erythromycin A and chloramphenicol on *Bacillus subtilis* ribosomes. *Biochim. Biophys. Acta* **1967**, *155*, 290–292.
- 42** YUKIOKA, M., HATAYAMA, T., MORISAWA, S., Affinity labeling of the ribonucleic acid component adjacent to the peptidyl recognition center of peptidyl transferase in *Escherichia coli* ribosomes. *Biochim. Biophys. Acta* **1975**, *390*, 192–208.
- 43** COUTSOGEORGOPoulos, C., Amino acyl-aminonucleoside inhibitors of protein synthesis. The effects of amino acyl ribonucleic acid on the inhibition. *Biochemistry* **1967**, *6*, 1704–1711.
- 44** KIRILLOV, S., PORSE, B. T., VESTER, B., WOOLLEY, P., GARRETT, R. A., Movement of the 3'-end of tRNA through the peptidyl transferase centre and its inhibition by antibiotics. *FEBS Lett.* **1997**, *406*, 223–233.
- 45** HANSEN, J. L., SCHMEING, T. M., MOORE, P. B., STEITZ, T. A., Structural insights into peptide bond formation. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 11670–11765.
- 46** LICHTENTHALER, F. W., TRUMMLITZ, G., Structural basis for inhibition of protein synthesis by the aminoacyl-aminohexosyl-cytosine group of antibiotics. *FEBS Lett.* **1974**, *38*, 237–242.
- 47** SAMAHA, R. R., GREEN, R., NOLLER, H. F., A base pair between tRNA and 23S rRNA in the peptidyl transferase centre of the ribosome. *Nature* **1995**, *377*, 309–314.
- 48** PORSE, B., RODRIGUEZ-FONSECA, C., LEVIEV, I., GARRETT, R. A., Antibiotic inhibition of the movement of tRNA substrates through a peptidyl transferase cavity. *Biochem. Cell Biol.* **1995**, *73*, 877–885.
- 49** HOFS, H. P., WAGENER, D. J., DE VALK-BAKKER, V., VAN RENNES, H., DE VOS, D., DOESBURG, W. H., OTTENHEIJM, H. C., DE GRIP, W. J., Schedule-dependent enhancement of antitumor activity of ethyldeshydroxy-sparsomycin in combination with classical antineoplastic agents. *Anticancer Drugs* **1995**, *6*, 277–284.

- 50 FLYNN, G. A., ASH, R. J., Necessity of the sulfoxide moiety for the biochemical and biological properties of an analog of sparsomycin. *Biochem. Biophys. Res. Commun.* **1983**, *114*, 1–7.
- 51 HUMMEL, H., BOCK, A., 23S ribosomal RNA mutations in halobacteria conferring resistance to the anti-80S ribosome targeted antibiotic anisomycin. *Nucleic Acids Res.* **1987**, *15*, 2431–2443.
- 52 PIOLETTI, M., SCHLUENZEN, F., HARMS, J., ZARIVACH, R., GLUHMANN, M., AVILA, H., BASHAN, A., BARTELS, H., AUERBACK, T., JACOBI, C., HARTSCH, T., YONATH, A., Crystal structures of complexes of the small ribosomal subunit with tetracycline, edeine and IF3. *EMBO J.* **2001**, *20*, 1829–1839.
- 53 YUSUPOV, M. M., YUSUPOVA, G. Z., BAUCOM, A., LIEBERMAN, K., EARNEST, T., CATE, J., NOLLER, H. F., Crystal structure of the ribosome at 5.5 Å resolution. *Science* **2001**, *292*, 883–896.
- 54 DOUTHWAITE, S., Interaction of the antibiotics clindamycin and lincomycin with *Escherichia coli* 23S ribosomal RNA. *Nucleic Acids Res.* **1992**, *20*, 4717–4720.
- 55 MOAZED, D., NOLLER, H. F., Chloramphenicol, erythromycin, carbomycin, and vernamycin B protect overlapping sites in the peptidyl transferase region of 23S ribosomal RNA. *Biochimie* **1987**, *69*, 879–884.
- 56 RODRIGUEZ-FONSECA, C., AMILS, R., GARRETT, R. A., Fine structure of the peptidyl transferase centre on 23S-like rRNAs deduced from chemical probing of antibiotic-ribosome complexes. *J. Mol. Biol.* **1995**, *247*, 224–235.
- 57 MOAZED, D., NOLLER, H. F., Sites of interaction of the -CCA end of peptidyl-tRNA with 23S rRNA. *Proc. Natl. Acad. Sci. USA* **1991**, *88*, 3725–3728.
- 58 PORSE, B. T., GARRETT, R. A., Sites of interaction of streptogramin A and B antibiotics in the peptidyl transferase loop of 23S rRNA and the synergism of their inhibitory mechanisms. *J. Mol. Biol.* **1998**, *286*, 375–387.
- 59 KEARSEY, S. E., CRAIG, I. W., Altered ribosomal RNA genes in mitochondria from mammalian cells with chloramphenicol resistance. *Nature* **1981**, *290*, 607–608.
- 60 BLANC, H., WRIGHT, C. T., BIBB, M. J., WALLACE, D. C., CLAYTON, D. A., Mitochondrial DNA of chloramphenicol-resistant mouse cells contains a single nucleotide change in the region encoding the 3' end of the large ribosomal RNA. *Proc. Natl. Acad. Sci. USA* **1981**, *78*, 3789–3793.
- 61 ETAYEBI, M. S., PRASAD, S. M., MORGAN, E. A., Chloramphenicol-erythromycin resistance mutations in a 23S rRNA gene of *Escherichia coli*. *J. Bacteriol.* **1985**, *162*, 551–557.
- 62 VESTER, B., GARRETT, R. A., The importance of highly conserved nucleotides in the binding region of chloramphenicol at the peptidyl transferase centre of *Escherichia coli* 23S ribosomal RNA. *EMBO J.* **1988**, *7*, 3577–3587.
- 63 DUJON, B., Sequence of the intron and flanking exons of the mitochondrial 21S rRNA gene of yeast strains having different alleles at the *rib-1* loci. *Cell* **1980**, *20*, 185–197.
- 64 BLANC, H., ADAMS, C. W., WALLACE, D. C., Different nucleotide changes in the large rRNA gene of the mitochondrial DNA confer chloramphenicol resistance on two human cell lines. *Nucleic Acids Res.* **1981**, *9*, 5785–5795.
- 65 TAN, G. T., DEBLASIO, A., MANKIN, A. S., Mutations in the peptidyl transferase center of 23S rRNA reveal the site of action of sparsomycin, a universal inhibitor of translation. *J. Mol. Biol.* **1996**, *261*, 222–230.
- 66 LAZARO, E., RODRIGUEZ-FONSECA, C., PORSE, B., URENA, D., GARRETT, R. A., BALLESTA, J. P. G., A sparsomycin-resistant mutant of *Halobacterium salinarium* lacks a modification at nucleotide U2603 in the peptidyl transferase centre of 23S rRNA. *J. Mol. Biol.* **1996**, *261*, 231–238.

5

Structure-Based Design of Cathepsin K Inhibitors

DANIEL F. VEBER and MAXWELL D. CUMMINGS

5.1

Introduction

The use of structural information derived from protein X-ray crystallography for the design of therapeutically useful molecules has long been viewed as an ideal approach. We have had the opportunity to participate in a structure-based medicinal chemistry project directed at the development of inhibitors of the bone-specific cysteine protease cathepsin K, as a treatment for bone loss diseases such as osteoporosis [1]. The origins of this project were in a high-profile effort at SmithKline Beecham Pharmaceutical Laboratories (now GlaxoSmithKline) built on attempts to utilize newly developed gene sequence data to define therapeutic targets [2]. This gave the program an exceptionally high level of resource allocation including medicinal chemistry, molecular modeling and X-ray crystallography even in advance of the full purification and large-scale isolation of the enzyme in the quantities needed for crystallography studies. The high level of support was maintained over the extended period of time commonly required to address the pharmacokinetic issues that so often plague protease inhibitor projects. This project therefore offers a unique perspective on the application of structure-based design throughout the course of an extended drug discovery effort.

Many factors can limit the effectiveness of structure-based design in a drug development project. For example, a delay in the supply of protein for crystallography can hamper the full application of structure-based design toward a medicinal chemistry effort. Available structures of homologues are undoubtedly useful, but small differences in protein structures may be significant in analogue design. When lead discovery and development outpaces structural characterization, the opportunity for structure-based design to impact the early stages of lead optimization can be reduced. This proved not to be the case for the study of cathepsin K inhibitors in our laboratories; although protein supplies were initially an issue, it quickly became possible to use structure-based design throughout much of the project. Additionally, in some cases X-ray crystallography may not be able to unequivocally distinguish the presence or absence of a covalent bond between protein and bound ligand. Within the cysteine protease inhibitor project reviewed here, we see that even at 2.3 Å resolution the presence or absence of covalent in-

hibitor binding is ambiguous, but nonetheless this is clearly a critical design issue. The more recent ability of crystallographers to define structures at 2.0 Å resolution and better on a consistent basis has done much to address this issue.

In the absence of high-resolution structural data detailing the analogue series and target protein of interest, related structural information can have a significant positive impact on a structure-based design effort. Prior to 1995 there were ten papain crystal structures and two cathepsin B crystal structures on deposit at the Protein Data Bank [3] (PDB) comprising complexes with a variety of active site inhibitors. A current search of the PDB (www.rcsb.org) reveals a wealth of structural information related to cathepsin K and close homologues. There are 43 deposited structures involving cathepsins B/K/L/S or papain, and many of these have bound active site inhibitors. Fourteen of these structures were deposited by our colleagues at SmithKline Beecham over the course of the cathepsin K discovery project (PDB ID codes 1AU0, 1AU2, 1AU3, 1AU4, 1BG0, 1AYU, 1AYV, 1AYW, 1BP4, 1BP8, 1BQI, 1NLJ, 1NL6, 1ATK); additional cathepsin K-inhibitor structures have been described by us [4]. The structures from this project range in resolution from 1.9 Å to 2.6 Å, with most of the structures between 2.2 Å and 2.5 Å resolution. Over the course of the cathepsin K drug discovery and development project both proprietary and public three-dimensional structural information has played a key role in the medicinal chemistry effort.

The structural information developed in this project and molecular modeling studies based thereon proved to be very valuable in guiding the medicinal chemistry effort directed toward the development of therapeutically useful cathepsin K inhibitors, as will be discussed here. Through this work it has also been possible to highlight some of the present limitations on the use of structural information, the recognition of which can serve to guide the advances necessary to make it even more productive in the future.

5.2

Background and Issues to be Addressed Using Protein Structure

Structural information on protease-bound inhibitors can offer design insights in several ways. It can directly prove or indirectly suggest the likely mechanism by which the inhibitor is acting. It can be used to suggest sites on the protein for improved interactions with inhibitors and suggest design approaches for potential enhancement of potency; similarly, this type of analysis can suggest avenues for improvement of selectivity relative to enzyme homologues, the structure of which may also be known by crystallography or inferred by homology modeling. Identification of inhibitor sites not critical to binding may be exploited in the alteration of various physicochemical properties of pharmaceutical relevance. Finally, it can be used to suggest ways to rigidify inhibitor conformation, a property now recognized as important for the development of oral bioavailability in new drug candidates [5].

Cysteine proteases hydrolyze protein amide bonds through formation of a thioester bond with an active site cysteine thiol. Cathepsin K is a member of the pa-

pain super-family of cysteine proteases. Other members of this family include cathepsins L, S and B, all of which are structurally homologous and utilize the same mechanistic pathway for amide bond hydrolysis. It is therefore not surprising that many of the inhibitors of members of this enzyme class contain an electrophilic group that reacts with the nucleophilic cysteine thiol. Indeed, inhibitors of cysteine proteases have been recognized as being separable into at least five distinct classes based on the mechanism of binding to the protease. In undertaking systematic structure-activity studies, it is important to check continuously that there has not been a shift in mechanism of action as this can confuse the conclusions being drawn.

The categorization of inhibitors of the papain super-family into five distinct groups has been reviewed in some detail and examples of each type have been cited [6]. These groups range in mechanism of action from those interacting in a purely non-covalent fashion to those that rely primarily on a group that is irreversibly reactive toward the active site thiol. A brief summary of these categories is presented here, with some illustrative examples; the reader is directed to reference 6 for more detail.

1. Category I inhibitors act by purely non-covalent forces. The binding can involve interactions that relate to substrate recognition sites, but this is not a strict requirement as they may also involve ancillary, non-active site elements of the protease.
2. Category II inhibitors act by formation of a reversible covalent bond between some group on the enzyme and a functional group on the inhibitor. In the case of cysteine proteases, the enzyme functional group is generally the active site cysteine thiol. This category is most specifically characterized by inhibitors having an electrophile that is not reactive with thiol in any context other than the enzyme-active site. The ketones discussed extensively here are a specific example of this category. They show uniquely low activation energy for formation of the covalent linkage, with the enzyme on-rate being effectively diffusion controlled. The reverse process, being microscopically reversible, will also have low activation energy, making category II inhibitors functionally equivalent to non-covalent inhibitors.
3. Category III inhibitors are an extension of category II inhibitors, but rely more heavily on the chemical reactivity of the inhibitor electrophile for formation of the reversible covalent bond between the enzyme and inhibitor. Owing to this higher intrinsic reactivity, members of this class are more likely to react non-specifically with nucleophiles unassociated with the target enzyme, including water. The greater potential for non-specific reaction with other enzymes or receptors makes this group less desirable than categories I and II for drug development. Important members of this category include the peptide aldehydes which were a well-known inhibitor type long before the initiation of the cathepsin K project at SmithKline Beecham. It was the availability of structural data on the enzyme-bound form of this inhibitor class that first allowed the design of category II inhibitors of the ketone type (see below).

4. Category IV inhibitors utilize the mechanistic machinery of the target protease to form a covalent bond with the protease that is either irreversible or requires a further chemical reaction such as hydrolysis to regenerate the active enzyme. The enzyme-regenerating reaction may be either a catalyzed or spontaneous chemical process. So-called suicide substrates, quiescent inhibitors and slow turnover substrates are all representative of this category. The very slow enzyme off-rates characteristic of this inhibitor class mean that any short term selectivity for the desired target is likely to disappear in the long term exposure required for *in vivo* drug action. The example of diacyl-carbohydrazides discussed below belongs in this category. These compounds have proven to be especially problematic for mechanistic characterization using protein X-ray crystallography.
5. Category V inhibitors also rely on the formation of an irreversible covalent bond with the enzyme, but the formation of this bond relies more on the intrinsic chemical reactivity of the inhibitor electrophile than on the reactivity of the enzyme-active site. On-rates therefore tend to be faster than for category IV inhibitors. Again, short term selectivity can appear quite good because of large kinetic differences in the rates of inhibition of related enzymes even though they are all fast (seconds to minutes) on the time scale of *in vivo* drug action (hours to days). However, the lack of an off-rate for every inhibited enzyme means that all inhibited enzymes become functionally inactivated. Even for relatively slowly inhibited enzymes, selectivity may not translate to the relevant *in vivo* situation.

It is noteworthy that the off-rate of category IV and V inhibitors can be slow relative to the time involved in crystallization and X-ray data collection. As a result of this, the nature of the interactions of the inhibitor with the enzyme seen in the X-ray structure may not bear any relationship to the interactions that occur in driving the rate of inhibition. Thus, the structure-based optimization of inhibitors may be flawed when using structures of complexes involving these types of inhibitors.

Inhibitors of categories III, IV and V are, in general, less desirable as drug candidates, especially where long term drug administration is planned. Nonetheless, category III and IV inhibitors can be desirable in short term therapy and have been features of antimicrobials and anticancer therapeutics. However, the irreversible derivatization of proteins and their degradation fragments can produce an immune response to the drug and to the proteins that have been derivatized. Either outcome is undesirable. It is critical to structure-activity studies to know that the mechanism of action of a drug in categories I and II has not suddenly evolved to a category III, IV or V type, over the progression of structural modification. The recognition of such a change or lack of change can be an important contribution from structural studies of enzyme-inhibitor complexes.

5.3

Cysteine Protease Inhibitors: Historical Perspective

Papain and related cysteine proteinases are among the best studied enzymes from both structural and mechanistic viewpoints. Papain, a protease from papaya, has historical status as the “parent” of this cysteine protease super-family. The term “papain” was first introduced in the 19th century [7] and the purified enzyme of this name was first crystallized in 1937 [8]. Enzymes of this class have since been recognized as being widely distributed in microbes and animals. The unifying features are a high degree of sequence homology and an active site cysteine that forms a transient thioester with the cleavage site carboxyl (Fig. 5.1). The active site cysteine thiol is thought to be activated by a proximate histidine imidazole group for nucleophilic attack on the substrate’s scissile amide bond. The unique character of this thiol as a nucleophile has been the basis for the design of many types of inhibitors containing an electrophilic “warhead” group intended to act by reversible or irreversible covalent bond formation. Detailed aspects of both the mechanism of proteolysis and the design of inhibitors have been widely reviewed [9–11]; further discussion of these aspects is not warranted here. An additional common feature of members of this enzyme super-family is their synthesis as inactive proenzymes. Specific proteolytic cleavage of this proenzyme results in release of the inhibitory pro-peptide segment and thereby yields the mature active protease [9].

X-ray crystallographic studies on proenzymes, enzymes and inhibited forms of papain super-family members have done much to establish a structural rationale for both the mechanistic aspects of hydrolysis and the nature of inhibition. Berger and Schechter [10] insightfully introduced the nomenclature defining the “unprimed” (S) side, N-terminal to the cleavage site, and the “prime” (S') side, C-terminal to the cleavage site. Early crystallographic studies with papain-chloromethylketone [11] and papain-E64 [15] complexes were consistent with this proposal, serving to validate this envisioned mode of active site substrate alignment. However, the direction of binding of the pro-segment (Fig. 2a) of the proenzyme

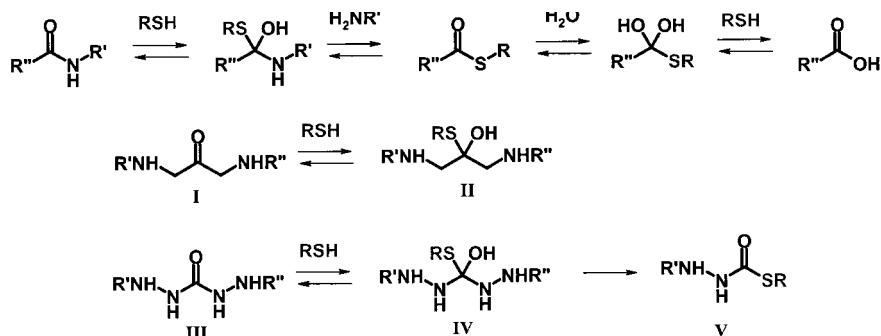


Fig. 5.1 Equilibria for papain family, cysteine protease (RSH): (top) catalyzed hydrolysis of an amide bond; (middle) inhibition by a

ketone or aldehyde; (bottom) inhibition by a diacyl-carbohydrazide.

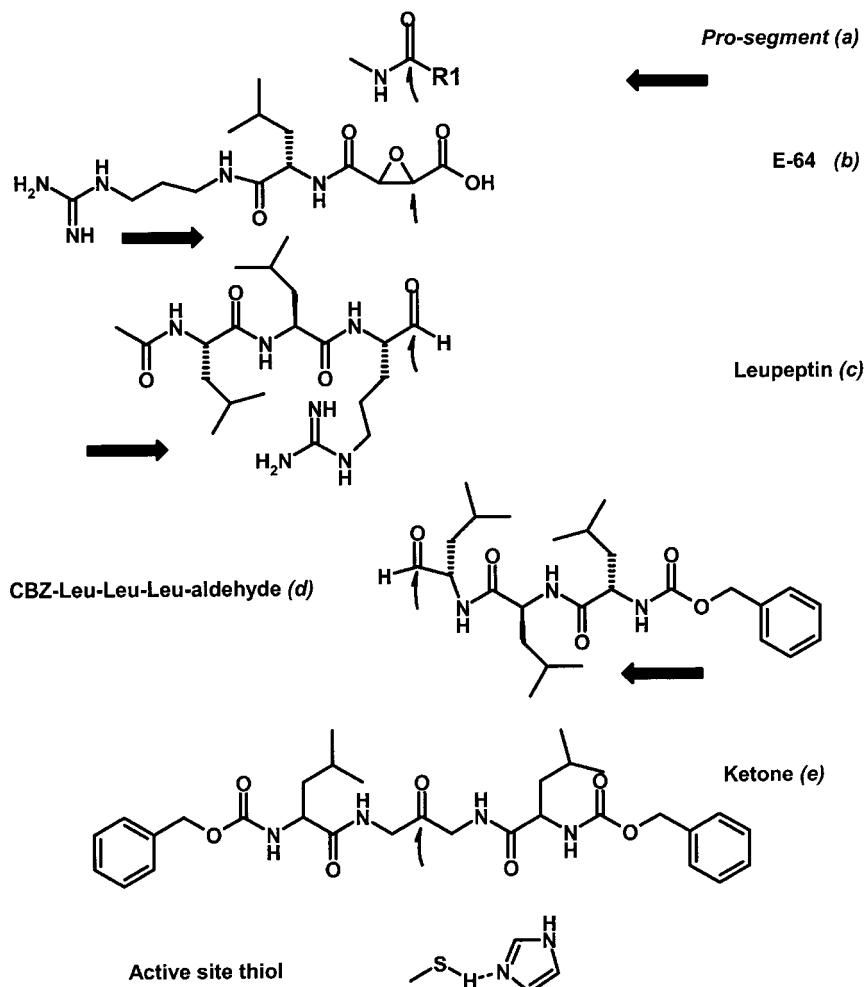


Fig. 5.2 Inhibitors of papain family cysteine proteases. The heavy arrows indicate the direction of binding of the peptidal inhibitor relative to the active site cysteine thiol. The

curved arrows indicate the site of thiol attack if it is thought to occur. R1 of pro-segment (a) is the specific sequence for any given papain family pro-enzyme.

is opposite to that observed for these two inhibitors (for cathepsin K see [12]). It has been assumed that this reversed binding direction is responsible for the inhibitory properties of the pro-segment, and that peptide cleavage cannot occur when a peptide binds in this direction. This assumption is difficult to either substantiate or disprove and therefore remains a matter of conjecture. Clearly, however, high affinity can be realized when a peptide binds in either direction in the active site.

Efforts to initiate structure-based design of cathepsin K inhibitors at SmithKline Beecham during the early 1990s were hampered by a very limited supply of the active form of the enzyme. We learned quite early in the project that the classic types of inhibitors for papain such as leupeptin, E-64 and Cbz-Leu-Leu-Leu-aldehyde were also very effective inhibitors of cathepsin K [13], but timely structural studies were not possible because of the lack of enzyme. The crystallography group therefore decided to build an experience base with a number of inhibited forms of papain [14] (46% sequence identity with human cathepsin K). E-64- and leupeptin-inhibited papain gave structures consistent with those that had been reported previously; in both of these cases the peptide bound on the unprime side of the active site, with modification of the active site cysteine thiol as a thiohemiacetal (leupeptin; see Fig. 5.1 structure II and Fig. 5.2b) or an alkylated thiol (E-64; see Fig. 5.2c). The structure of papain-bound Cbz-Leu-Leu-Leu-aldehyde revealed an unexpected binding mode. The Cbz group occupies the prime side of the active site, with the peptide extending toward the active site cysteine in the same manner as the pro-segment of the proenzyme and forming a thiohemiacetal with the active site thiol in a substrate-like manner [14]. From the structures of leupeptin and Cbz-Leu-Leu-Leu-aldehyde bound to papain it was recognized that favorable binding interactions for similar inhibitors could be achieved on both sides of the active site cleft, despite the overall differences between the prime and unprime sides. The affinity for similar recognition elements on both sides of the active site and concomitant ease with which binding direction can be reversed is a critical and recurring issue for the evaluation of structure-activity studies of all of the inhibitor types that we have studied in our laboratories. For inhibitors of this enzyme class, the potentially ambiguous nature of the definitions of “prime” and “unprime” becomes clear since peptide chains are able to track almost equally well in either direction through the binding cleft. Either direction of binding can accommodate the positioning of a carbonyl mimic of the substrate amide bond in a position to react with the active site cysteine thiol as exemplified by the two aldehydes (Fig. 5.2c,d). This ambiguity of definition may suggest that Nature has used both directions in evolving the substrate selectivity profiles of the various members of this protease family. This flexibility of binding mode can complicate the development of a predictive SAR for inhibitors and it is therefore crucial to the track binding mode with structural data whenever possible, and to supplement this with modeling studies when structural data are unavailable.

5.4

Diaminoketone-Based Inhibitors

The observation of two distinct active site binding directions had critical impact on the design of novel class II type inhibitors in the course of the SmithKline Beecham medicinal chemistry effort. Our long term strategy for developing improved cathepsin K inhibitors had been to increase inhibitor selectivity by using electrophilic groups of lower intrinsic chemical reactivity than the aldehydes. It

also included a desire to enhance selectivity by extension of binding elements into the prime and unprime sites simultaneously. Both of these goals were achieved through observations arising from a superposition of the bound forms of the two aldehyde-inhibited papain structures. This suggested the design and synthesis of ketone-based inhibitors (e.g., Fig. 5.2e). The first of these to be synthesized proved to be a 22 nM inhibitor of cathepsin K [15].

Crystallographic analysis of the cathepsin K-bound form of the symmetric 1,3-bis(acyl)diaminoketone (Fig. 5.2e), confirmed the predicted involvement of both prime and unprime side binding (exemplified by Fig. 5.3 showing compound (a) depicted in Fig. 5.4) coupled with the close proximity of the ketone carbonyl to the active site cysteine thiol. Whether the close proximity of the carbonyl carbon and thiol sulfur is fully a carbon-sulfur bond (1.8 Å) must remain ambiguous even with our highest resolution structures (1.9 Å).

Notably, the symmetric ketone (Fig. 5.2e) is a relatively poor inhibitor of papain, the enzyme structure from which it was designed. This has subsequently been recognized as a consequence of a steric interaction in the S3 binding region that was not a concern in the design studies [16]. It remains clear, however, that sound structure-based ideas should be probed by the synthesis of many related analogues in order to avoid missing an unforeseen and potentially advantageous path forward.

This designed ketone was encouraging in that an important goal of potent *in vitro* inhibition of cathepsin K had been achieved. It served as the beginning of a long term search for a suitable therapeutic candidate. Ultimately, improved *in vitro* potency, improved cell-based activity and improved pharmacokinetics were attained. The first steps toward these goals involved attempts to remove peptide-like elements from the inhibitors. This was achieved through a combination of classical SAR and structure-based design.

A key step in this direction involved the synthesis of analogues having an arylsulfonamide as replacement for one of the Cbz-L-leucine groups as seen in ke-

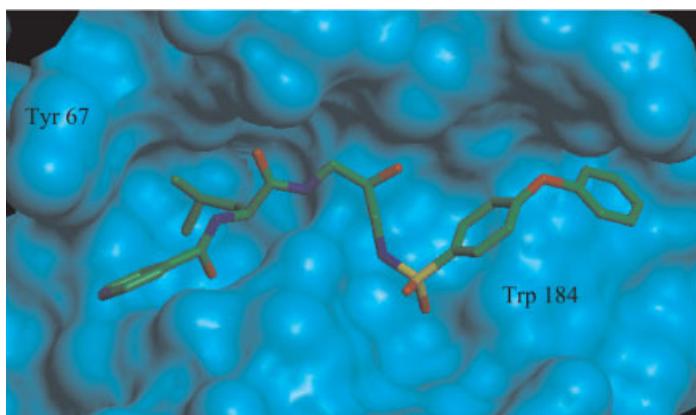


Fig. 5.3 Crystal structure of acyclic diaminoketone (Fig. 5.4a) bound in the active site of human cathepsin K (PDB ID code 1AU2) [15].

tone (a) of Fig. 5.4. In this case, the 4-phenoxy-phenyl sulfonamide replaces one entire Cbz-L-leucine and is associated with a 10-fold increase in potency as well as the removal of one chiral center. X-ray crystallography established that the aryl sulfonamide occupies the prime side substrate binding site (Fig. 5.3) [15]. Key binding interactions for this inhibitor include aromatic–aromatic interactions with Tyr-67 (unprime) and Trp-184 (prime), positioning of an *iso*-butyl side chain in the hy-

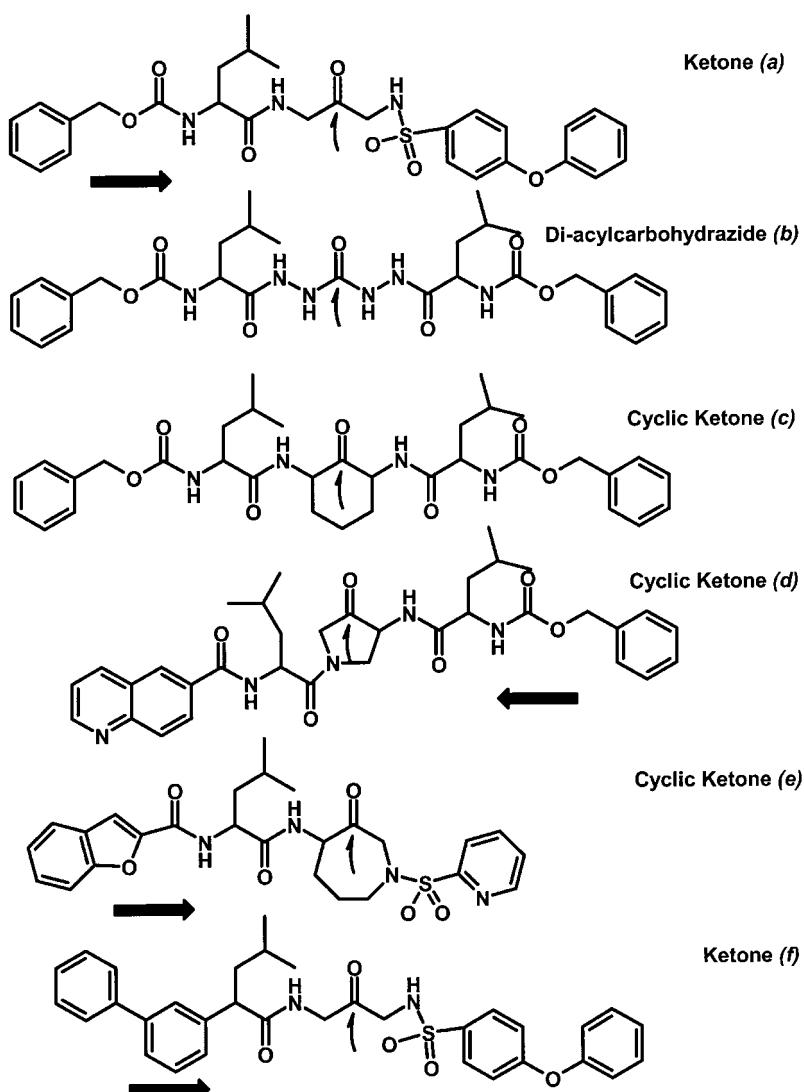


Fig. 5.4 Inhibitors of papain family cysteine proteases. The heavy arrows indicate the direction of binding of the peptidal inhibitor relative to the active site cysteine thiol. The curved arrows indicate the site of thiol attack if it is thought to occur.

dropobic S2 pocket, and β -sheet-like hydrogen bonding involving Gly-66 and Asn-161. The structure is consistent with Cys-25 thiohemiketal formation. To date an aryl sulfonamide moiety at this position relative to the active ketone has not been observed bound to the unprime side of the active site.

A modeling study led to a designed non-peptide replacement for the remaining Cbz-L-leucine group of compound Fig. 5.4a (see Fig. 5.3) by 2-*meta*-biphenyl-(R)-2-(4-methylvaleryl)- to give ketone Fig. 5.4f, which retained good inhibitory potency and provided markedly improved selectivity relative to close enzyme homologues [17]. X-ray crystallography of the cathepsin K-bound inhibitor established two key points. Firstly, as in the previous compound (Fig. 5.4a), the shorter terminal aryl sulfonamide group binds on the prime side of the active site. Secondly, the *meta*-biphenyl group is of the R configuration. The binding mode of this configuration positions the iso-butyl side chain in the same binding pocket as the L-leucine side chain of the parent inhibitor and the *meta*-biphenyl isostERICALLY replaces the Cbz-group of the parent inhibitor. This latter point was a critical outcome of the modeling study, as the initial *para*-biphenyl isomer was dramatically less active as a cathepsin K inhibitor despite representing a reasonable Cbz-Leu mimic. More detailed modeling analysis indicated that the *meta*-biphenyl would better complement the binding site, and this prediction was confirmed by subsequent testing and crystallographic analysis [17].

5.5

Cyclic Diaminoketone Inhibitors

Analysis of crystal structures of cathepsin K complexes with acyclic inhibitors bound to cathepsin K and molecular modeling studies led to the postulation of a preferred mode of cyclization around the active ketone (i.e., Fig. 5.4c). Introduction of conformational constraint in this manner was thought to offer additional opportunity to further reduce the peptidic nature of the compounds, and also might lead to increased potency by “locking in” a biologically relevant conformation. Subsequent systematic synthesis and testing of 5- and 6-membered ring analogues established a strong and unexpected preference for the N-to-C cyclization as exemplified by cyclic ketones (d) and (e) of Fig. 5.4 [18]. Retrospective analysis of relevant crystal structures shows a reasonable overlap between analogous acyclic and cyclic analogues in the region of the added constraint (Fig. 5.5); it was postulated that the dramatically lower affinity for the C-to-C cyclized compounds might be due to increased steric hindrance around the active carbonyl [18]. In addition to highlighting a reasonable correspondence between the atoms of the acyclic and cyclic molecules in the region of the added constraint, Fig. 5.5 shows that while multiple distinct binding modes are accommodated that allow for aromatic-aromatic interactions with Tyr-67, the positioning of the iso-butyl Leu mimic in S2 is strictly conserved. Similarly, multiple binding modes that terminate with aromatic-aromatic interactions involving Trp-184 are possible on the prime side of the active site.

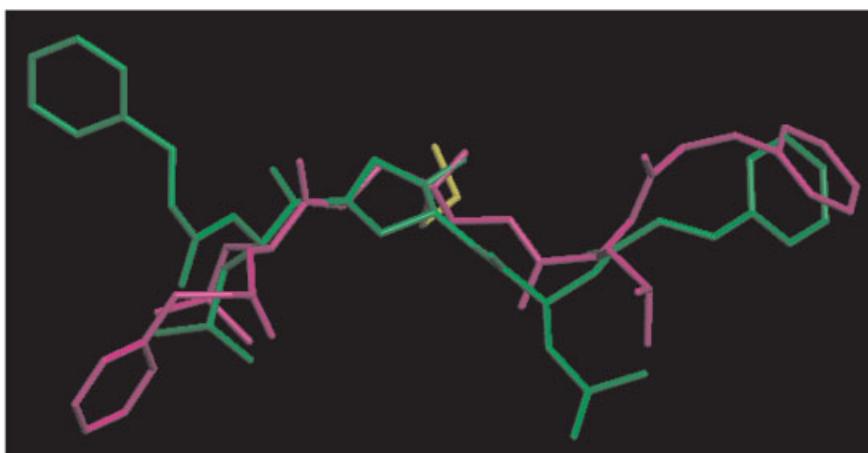
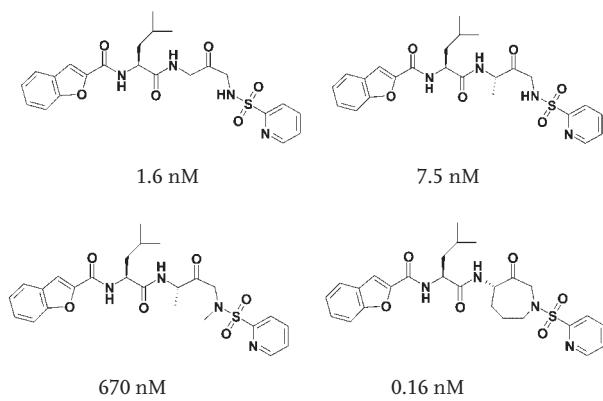


Fig. 5.5 Superimposed structures of acyclic and cyclic diaminoketones bound in the active site of human cathepsin K. Acyclic molecule (purple): Fig. 5.2e, $n=0$ in Tab. 5.2, PDB ID code 1AU0 [15]; cyclic molecule (green):

$n=1$ in Tab. 5.2, PDB ID code 1AU3 [18]. Protein backbone atoms were used for superposition; all protein atoms except 1AU0:Cys-25:Ca-S γ omitted from view.

The pyrrolidinone and piperidinones (Tab. 5.2, $n=1$ and $n=2$, respectively) had similarly impressive potencies, establishing that such cyclization led to increased potency, and also that inhibition was not extremely sensitive to ring size in this series. As described above, the binding modes observed for the pyrrolidinones (Tab. 5.2, $n=1$) in the cathepsin K complex crystal structures seem fairly consistent with previous structures observed with acyclic compounds bound (Fig. 5.5). Unfortunately the 5- and 6-membered cyclic diaminoketones shared the undesir-

Tab. 5.1 $K_{i,\text{app}}$ values versus human cathepsin K for acyclic and constrained analogues [19].



Tab. 5.2 $K_{i,app}$ values versus human cathepsin K for acyclic and cyclic analogues [15, 18, 19].

| Ring | $K_{i,app}$ (nM) |
|----------------------------------|------------------|
| $n=0$ (acyclic) | 22 |
| $n=1$ (mixture of diastereomers) | 2.3 |
| $n=2$ (mixture of diastereomers) | 2.6 |
| $n=3$ diastereomer A | 2.0 |
| $n=3$ diastereomer B | 15 |

able property of instability of configuration at the aminoketone chiral center, making them unsuitable for further development.

A path forward was ultimately realized with 7-membered ring azepanones [19] (Fig. 5.4e; Tab. 5.2, $n=3$; Fig. 5.6). Studies had established that epimerization was negligible in the acyclic 1,3-diaminoketones, and that the rate of ketone enolization slowed as the ring of the cyclic ketones was expanded from 5- to 6-membered [19]. More detailed studies of epimerization in the azepanone series indicated that these compounds were configurationally stable over a pharmaceutically relevant time-scale [19]. Structure-activity relationships developed in previous series were applied successfully to the azepanone scaffold, yielding extremely potent enzyme inhibitors that exhibited good diastereomeric selectivity as well as reasonable selectivity versus cathepsin K homologues.

As part of an effort to explore substitution of the azepanone scaffold more fully, a molecular modeling-based design effort was undertaken. This computational study proved more challenging than it might otherwise have been, due to an initial difficulty encountered in obtaining crystals of cathepsin K with an azepanone bound in the active site, and the fact that we did not know the diastereomeric preference at the outset of these studies. Development of a binding model for the azepanone-Leu scaffold involved detailed inspection of previous crystal structures, conformational analysis of a relevant azepanone model system and manual model building [19]. Ultimately we established a binding model comprising the *S* diastereomer in the less favorable axial conformation, with the pyridyl sulfonamide positioned on the prime side of the active site (see also previous section). Conformational analysis had predicted a preference for the equatorial conformer, but it had proved impossible to develop a plausible binding model based on this conformer for either diastereomer. For example, positioning of the azepanone and prime side pyridyl sulfonamide appropriately does not allow for S2 occupancy by

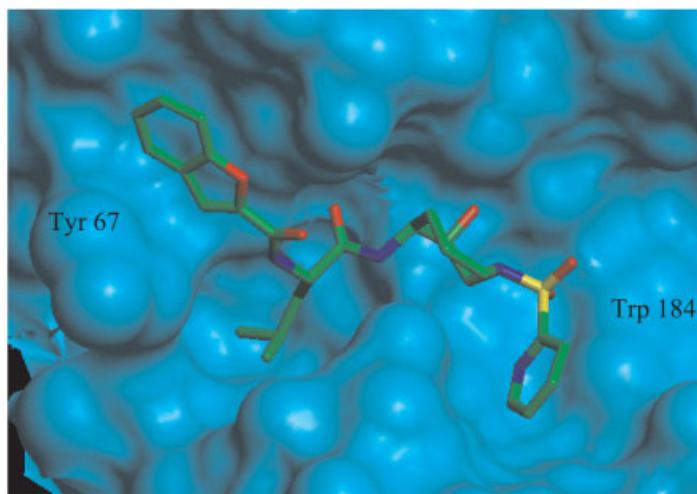


Fig. 5.6 Crystal structure of azepanone (Fig. 5.4e) bound in the active site of human cathepsin K (PDB ID code 1NLJ) [19].

the *iso*-butyl group, and leads to significant clashes with the enzyme (Fig. 5.7). During the course of the modeling study we obtained small molecule crystal structures of both diastereomers, establishing that the active diastereomer was of the *S* configuration (in agreement with our prediction), and that both diastereomers adopted an equatorial conformation (in disagreement with the scaffold used in our modeling study). At the end of a lengthy effort to obtain enzyme-inhibitor co-crystals, we determined the crystal structure of a cathepsin K-azepanone complex, which served, amongst other purposes, to validate our modeling prediction [19]. In this case, despite a wealth of apparently closely related structural precedents, development of a satisfactory binding model remained challenging.

Crystallographic analysis of the complex of the azepanone (Fig. 5.4e) bound to cathepsin K established that, as predicted, the benzofuran maintained the interaction with Tyr-67 and β -sheet-like hydrogen bonds with Gly-66 and Asn-161, the *iso*-butyl side chain occupied the S2 pocket and the pyridyl sulfonamide was positioned in S2' and formed a hydrogen bond with the indole NH of Trp-184 (Fig. 5.6). As in previous structures the proximity of Cys-25:S γ is consistent with a mechanism of action involving thiohemiketal formation [19].

Our studies with cyclic and acyclic diaminoketones provide useful examples of the impact of conformational constraint on inhibitory potency. Tabs. 5.1 and 5.2 show two analogue series that highlight the sometimes dramatic effects that cyclization can have. This modification can have equally dramatic effects on pharmacokinetic properties of the molecules as well. Conformationally restricted analogues have long been a goal of medicinal chemistry efforts such as the one reviewed here. Only recently has sufficient pharmacokinetic data become available to reveal the critical importance of constraint, in the form of reduced rotatable bond count, for improving oral bioavailability [5] as has been seen with compound (Fig. 5.4e) [19].

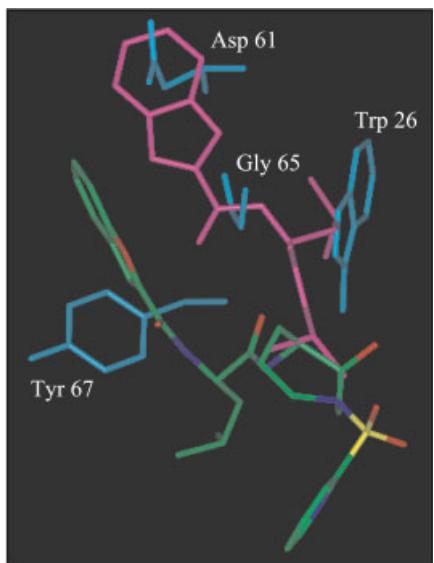


Fig. 5.7 Crystal structure of azepanone (Fig. 5.4e) bound in the active site of human cathepsin K (PDB ID code 1NLJ) [19]. The two aromatic ring systems of the bound azepanone have been altered slightly from the crystal structure, to make these groups completely planar. One possible equatorial conformer of the *S* diastereomer was generated, and the azepanone of this model (purple) was superimposed onto that of the bound inhibitor (color-by-atom). The azepanone and the pyridyl sulfonamide of the model (not shown for clarity) overlay reasonably well with the bound inhibitor, but the unprime side does not. Selected atoms of cathepsin K residues that clash with the model are shown in cyan.

5.6

Alkoxymethyl and Thiomethyl Dipeptidyl Ketone-Based Inhibitors

Alkoxymethyl and thiomethyl ketone derivatives are compound classes that were very early subjects of study in our laboratory as inhibitors of cathepsin K [20]. They had a proven record as inhibitors of other cysteine and serine proteases [21]. They also fit our strategic objectives relating to a desire for reduced intrinsic chemical reactivity compared with the aldehydes, as well as providing the potential for including binding elements capable of reaching sites on both the prime and unprime sides of the active site. Contrary to our straightforward mechanistic expectations of reversible ketone addition to the enzyme, mechanisms of inhibition that fit the definitions of inhibitor classes I, II and IV were determined based on data from X-ray crystallography and mass spectrometry. Unpredicted modes of binding were also revealed in the numerous crystallography studies that followed.

The simplest member of this family of inhibitors, the methoxy-methyl ketone (Fig. 5.8a), has been studied as a complex with papain [14]. This enzyme-inhibitor complex shows binding of the peptide portion of the inhibitor on the prime side of the active site in a manner similar to that seen for Cbz-Leu-Leu-Leu-aldehyde (Fig. 5.2d) bound to papain. In contrast to the thiohemiacetal seen with the aldehyde, the carbonyl of the methoxy-methyl ketone (Fig. 5.8a) is quite distant from the active site thiol, with no possibility of covalent interaction. Therefore, this inhibitor meets the criteria for class I inhibition. The *n*-propyloxy ketone (Fig. 5.8b), which also binds on the prime side of the active site, has the ketone in close proximity to the active site thiol of cathepsin K, as seen in the structure of the inhibitor-enzyme complex [20]. Covalent attachment appears to have followed from

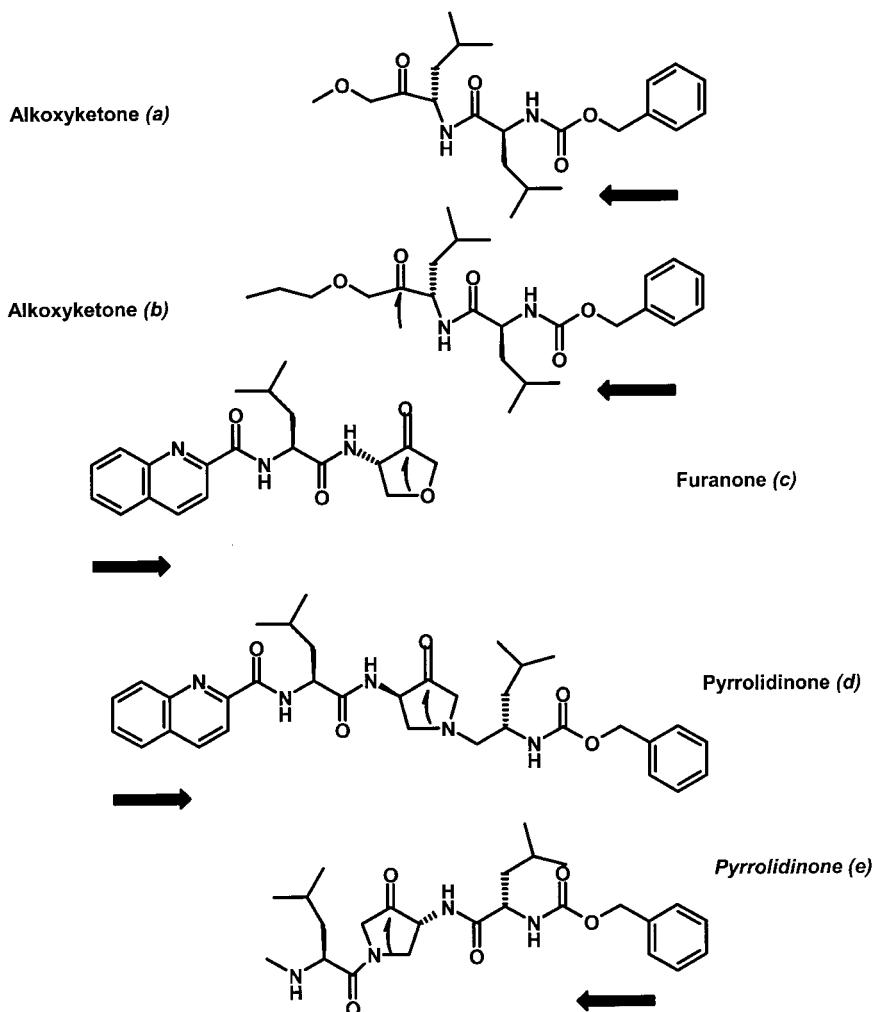


Fig. 5.8 Inhibitors of papain family cysteine proteases. The heavy arrows indicate the direction of binding of the peptidal inhibitor relative to the active site cysteine thiol. The curved arrows indicate the site of thiol attack if it is thought to occur.

the simple addition of two methylene groups, allowing the alkyl chain to extend somewhat into the binding site for the side chain of leucine that is normally present on the unprimed side of the substrates. The consequence is a change from class I to class II inhibition of cysteine proteases upon extension of a methyl to propyl ether.

Further studies in this compound series revealed that replacement of the ether oxygen with sulfur consistently gave inhibitors that behaved as class IV inhibitors. In each case covalent attachment to the enzyme thiol was detected by mass spec-

trometry and kinetic studies indicated that inhibition was irreversible [20]. The nature of the reaction has not been confirmed by X-ray crystallographic analysis for any of these thiomethyl ketones.

What is seen for this series of inhibitors is that small structural modifications of the inhibitor can completely change the mechanism of inhibition. Clearly, structure-activity conclusions are very difficult to draw correctly in the absence of structural data in such a case. In general, the need to assess this type of mechanistic promiscuity by crystallography and/or mass spectrometry can become an excessively demanding resource drain in a medicinal chemistry project.

The synthesis of a series of furanones (Fig. 5.8c) and pyrrolidinones (Tab. 5.2, $n=1$) has revealed equally complicated binding interactions that have been recognized through X-ray crystallography of enzyme-inhibitor complexes [4]. For these compounds the class II mechanism of inhibition remains constant. However, in addition to the observation of unpredictable changes in binding direction within the active site cleft, a new issue arose relating to diastereomer preference for the chiral center in the furanone or pyrrolidinone ring. This issue carries some importance because of the relative ease of epimerization at this chiral center. These compounds are difficult to isolate and maintain in a epimerically pure form. Thus, co-crystallization of inhibitor with the enzyme is the best way to assure that the most potent isomer is identified and its absolute configuration is defined. The pyrrolidinone (Fig. 5.8d) has the expected *R* configuration when bound to cathepsin K. This preference is most common for inhibitors that make significant binding interactions on the unprime side. With the structurally similar furanone (Fig. 5.8c) the *S* configuration in the 5-membered ring is observed, with the portion that is identical to that of the pyrrolidinone (Fig. 5.8d), similarly binding on the unprime side. This contrasts with the smaller pyrrolidinone (Fig. 5.8e) which binds in the opposite direction and for which the *R* configuration in the pyrrolidinone ring is observed. This binding direction would appear to result from the potential for interaction of the added Cbz group with Trp-184 of cathepsin K.

5.7

Diacylcarbohydrazides

The discovery of potent diaminoketone-based inhibitors of cathepsin K (Fig. 5.2e) was rapidly followed up by analogue synthesis aimed at altering electronic, steric and binding features. A very radical modification that simultaneously addressed nearly all of these properties was a bis-aza replacement of the two methylene groups adjacent to the ketone to give the diacyl-carbohydrazide of the type shown in Fig. 5.4b. This two atom modification of ketone (Fig. 5.2e) had a profound impact on the mechanism of inhibition and presented unique and unexpected challenges in the structural studies [22]. The structural analysis of an enzyme-inhibitor complex for the inhibitor (Fig. 5.4b), with cathepsin K showed the inhibitor to be bound in a manner quite similar to that seen for the corresponding ketone-based inhibitor (Fig. 5.2e), with the proximity of the active site thiol to the central

carbonyl leading to the assumption of covalent attachment. In this light, the resulting tetrahedral adduct would normally be viewed as an intermediate in a hydrolytic pathway. The initial assumption was that a molecule frozen at this point might represent an ideal transition state or intermediate analogue, consistent with the very high inhibition potency observed for this compound class. However, mechanistic studies in solution involving detailed kinetic analysis of the inhibition reaction coupled with mass spectral analysis of the inhibited enzyme revealed quite a different picture [23]. The urea of the diacyl-carbohydrazide was seen to undergo rapid partial hydrolysis with formation of a thiocarbamate linkage to the active site cysteine thiol (Fig. 5.1, structure V). Deacylation of this inhibited form of the enzyme was quite slow, and occurred at the same rate for all of the thiol proteases studied in the papain family [23]. This type of inhibitor is, in effect, behaving as a class IV type inhibitor by forming a thiocarbamate (Fig. 5.1, structure V), in contrast to the class II type (Fig. 5.1, structure IV) inferred from the structural analysis. The difference between these two classes is important since selectivity seen for type IV inhibitors in *in vitro* assays may disappear in an *in vivo* setting while that seen *in vitro* for a type II inhibitor should be expected to relate directly to the *in vivo* situation. Alternate explanations for the observation of intact inhibitor in the X-ray analysis are: 1.) The potential for equal population or rapidly equilibrating populations of unprime and prime side binding after reaction and cleavage of the inhibitor, which could give a false appearance of intact inhibitor. 2.) In the solid state it is possible that the cleavage product (an acyl-hydrazide) remains bound to the acylated enzyme and in close proximity to the thio-ester site, thereby allowing for rapid reformation of the diacylcarbohydrazide. Therefore, the local high concentration of the two cleavage components in the crystal structure studies compared to those in solution studies may be the reason for the differences in conclusions drawn in the two types of studies. Explanation 2 is most consistent with the data presented in reference 27.

A class of thiazole hydrazide inhibitors of cathepsin K has also been evaluated in references 26 and 27. Similar problems to those described in the preceding paragraph related to enzyme acylation and slow turnover of inhibitor in solution were observed in trying to determine the mechanism of protease inhibition based solely on the X-ray structure analysis.

5.8 Conclusions

The integrated effort of crystallography, modeling and medicinal chemistry as exemplified in this chapter, reveals advanced approaches whereby the progression from drug discovery leads to compounds with therapeutic potential have been significantly accelerated. The early use of structure data from a readily available enzyme such as papain which is homologous with a novel target (cathepsin K in this instance) can produce useful data during the initial discovery phases when the novel target enzyme has not yet been produced in the quantities needed for

crystallization studies. The importance of this point in establishing an early collaboration of crystallography with the discovery team cannot be over-emphasized. The ability to generate structural data on less than perfect chemical lead classes such as the aldehydes described here can result in unexpected observations that can lead to creative advances in the design of improved molecules. The increased availability of rare proteins such as cathepsin K through the techniques of molecular biology facilitates the development of an extensive database of complex structures with increasingly improved inhibitors over the course of a medicinal chemistry project. Certainly structure-based studies would have been impossible had osteoclasts been the only source of cathepsin K for crystallography. Indeed we are not aware of any reports of the isolation of purified cathepsin K for more than mass spectral and immunologic identification. Improved resolution of X-ray structures (1.9–2.5 Å resolution) have added accuracy to design compared with past experiences at lower resolution (2.5–3.0 Å) where a chiral center preference was not predictable in modeling studies [24]. Faced with the reality that protein crystallography gives less than atomic resolution, advanced modeling techniques can be used to search for optimal structural solutions more systematically, and to help resolve conformational and configurational questions. Conclusions that can be readily drawn about inhibitors based on structural data available today include mechanistic inferences and determination of inhibitor-binding modes. These are critically important issues for valid predictions based on structure activity conclusions.

In spite of real improvements in the availability of structure-based data, the present limitations as seen in this chapter indicate areas for further increases in both the accuracy and the speed of molecular design. Distinct improvements could arise from further gains in the resolution of protein structures. Resolution that could unequivocally define bond distances would add to the ability to elucidate mechanistic detail. A second major area for improvement is in the prediction of how protein structure can adapt to accommodate modified ligands. Advances in the prediction and evaluation of both structural changes and altered intermolecular interactions during docking calculations will enhance our ability to design and prioritize new molecules. This remains an active area of investigation in many research groups. It is also important to further facilitate the generation of crystals of protein-bound ligands and the subsequent generation of structure data. We see that detailed analysis and design for improved selectivity requires not only the knowledge of the target-bound ligand but also for ligands bound to non-target proteins. How can we hope to rationally improve selectivity of, for example, a cathepsin K inhibitor in the absence of data for binding of that same inhibitor to cathepsins L, S and B or any other protein that could be the source of drug side effects? Progress beyond our present state could come from more rapid and facile isolation and crystallization of these enzymes. It might also arise from improvements in predictive methods for protein structure and for binding interactions. Our present abilities have resulted from stretching the limits of currently available methodologies and this paradigm is a likely model for future advances.

5.9

References

- 1 D. F. VEBER, F. H. DRAKE, M. GOWEN, *Curr. Opin. Chem. Biol.* **1997**, *1*, 151–156.
- 2 F. H. DRAKE, R. A. DODDS, I. E. JAMES, J. R. CONNOR, C. DEBOUCK, S. RICHARDSON, E. LEE-RYKACZEWSKI, L. COLEMAN, D. RIEMAN, R. BARTHLOW, G. HASTINGS, M. GOWEN, *J. Biol. Chem.* **1996**, *271*, 12511–12516.
- 3 H. M. BERMAN, J. WESTBROOK, Z. FENG, G. GILLILAND, T. N. BHAT, H. WEISSIG, I. N. SHINDYALOV, P. E. BOURNE, *Nucl. Acids Res.* **2000**, *28*, 235–242.
- 4 R. W. MARQUIS, Y. RU, J. ZENG, R. E. LEE TROUT, S. M. LoCASTRO, A. D. GRIBBLE, J. WITHERINGTON, A. E. FENWICK, B. GARNIER, T. TOMASZEK, D. TEW, M. E. HEMLING, C. J. QUINN, W. W. SMITH, B. ZHAO, M. S. MCQUENEY, C. A. JANSON, K. D'ALESSIO, D. F. VEBER, *J. Med. Chem.* **2001**, *44*, 725–736.
- 5 D. F. VEBER, S. R. JOHNSON, H.-Y. CHENG, B. R. SMITH, K. W. WARD, K. D. KOPPLE, *J. Med. Chem.* **2002**, *45*, 2615–2623.
- 6 D. F. VEBER, S. K. THOMPSON, *Curr. Opin. Drug Discovery* **2000**, *3*, 362–369.
- 7 A. WURTZ, E. BOUCHUT, *Compt. Rend. Acad. Sci.* **1879**, *89*, 425.
- 8 A. K. BALLS, H. LINEWEAVER, R. R. THOMPSON, *Science* **1937**, *86*, 379.
- 9 B. TURK, D. TURK, V. TURK, *Biochim. Biophys. Acta* **2000**, *1477*, 98–111.
- 10 D. LEUNG, G. ABBENANTE, D. P. FAIRLIE, *J. Med. Chem.* **2000**, *43*, 305–341.
- 11 R. W. MARQUIS, *Ann. Rep. Med. Chem.* **2000**, *35*, 309–320.
- 12 See J. LALONDE, B. ZHAO, C. A. JANSON, K. J. D'ALESSIO, M. S. MCQUENEY, M. J. ORSINI, C. M. DEBOUCK, W. W. SMITH, *Biochemistry* **1999**, *38*, 862–869 and references cited therein.
- 13 A. BERGER, I. SCHECTER, *Philos. Trans. R. Soc. London* **1970**, *B257*, 249–264.
- 14 J. DRENTH, K. H. KALK, H. M. SWEN, *Biochemistry* **1976**, *15*, 3731–3738.
- 15 K. I. VARUGHESI, F. R. AHMED, P. R. CAREY, S. HASNAIN, P. HUBER, A. C. STORER, *Biochemistry* **1989**, *28*, 1220–1232; and D. YAMAMOTO, K. MATSUMOTO, H. ONISHI, T. ISHIDA, M. INOUYE, K. KITAMURA, H. MIZUNO, *J. Biol. Chem.* **1991**, *266*, 14771–14777.
- 16 J. LALONDE, B. ZHAO, C. A. JANSON, K. J. D'ALESSIO, M. S. MCQUENEY, M. J. ORSINI, C. M. DEBOUCK, W. W. SMITH, *Biochemistry* **1999**, *38*, 862–869.
- 17 B. VOTTA, M. A. LEVY, A. BADGER, J. BRADBEER, R. A. DODDS, I. E. JAMES, S. THOMPSON, M. J. BOSSARD, T. CARR, J. R. CONNOR, T. A. TOMASZEK, L. SZEWCUK, F. H. DRAKE, D. F. VEBER, M. GOWEN, *J. Bone Mineral Res.* **1997**, *12*, 1396–1406.
- 18 J. M. LALONDE, B. ZHAO, W. W. SMITH, C. A. JANSON, R. L. DESJARLAIS, T. A. TOMASZEK, T. J. CARR, S. K. THOMPSON, H.-J. OH, D. S. YAMASHITA, D. F. VEBER, S. S. ABDEL-MEGUID, *J. Med. Chem.* **1998**, *41*, 4567–4576.
- 19 D. S. YAMASHITA, W. W. SMITH, B. ZHAO, C. A. JANSON, T. A. TOMASZEK, M. J. BOSSARD, M. A. LEVY, H.-J. OH, T. J. CARR, S. K. THOMPSON, C. F. IJAMES, S. A. CARR, M. MCQUENEY, K. J. D'ALESSIO, B. Y. AMEGADZIE, C. R. HANNING, S. ABDEL-MEGUID, R. L. DESJARLAIS, J. G. GLEASON, D. F. VEBER, *J. Am. Chem. Soc.* **1997**, *119*, 11351–11352.
- 20 D. S. YAMASHITA, R. A. DODDS, *Curr. Pharm. Design* **2000**, *6*, 1–24.
- 21 R. L. DESJARLAIS, D. S. YAMASHITA, H.-J. OH, I. N. UZINSKAS, K. F. ERHARD, A. C. ALLEN, R. C. HALTIWANGER, B. ZHAO, W. W. SMITH, S. ABDEL-MEGUID, K. J. D'ALESSIO, C. A. JANSON, M. S. MCQUENEY, T. A. TOMASZEK, M. A. LEVY, D. F. VEBER, *J. Am. Chem. Soc.* **1998**, *120*, 9114–9115.
- 22 R. W. MARQUIS, D. S. YAMASHITA, YU RU, S. LoCASTRO, H.-J. OH, K. F. ERHARD, R. L. DESJARLAIS, M. S. HEAD, W. W. SMITH, B. ZHAO, C. A. JANSON, S. S. ABDEL-MEGUID, T. A. TOMASZEK, M. A. LEVY, D. F. VEBER, *J. Med. Chem.* **1998**, *41*, 3563–3567.
- 23 R. W. MARQUIS, Y. RU, S. M. LoCASTRO, J. ZENG, D. S. YAMASHITA, H.-J. OH, K. F. ERHARD, L. D. DAVIS, T. A. TOMASZEK, D. TEW, K. SAYERS, J. PROKSCH, K. WARD, B. SMITH, M. LEVY, M. D. CUMMINGS, R. C. HALTIWANGER, G. TRESCHER, B.

- WANG, M. E. HEMLING, C. J. QUINN, H.-Y. CHENG, F. LIN, W. W. SMITH, C. A. JANSON, B. ZHAO, M. S. MCQUENEY, K. D'ALESSIO, C.-P. LEE, A. MARZULLI, R. A. DODDS, S. BLAKE, S.-M. HWANG, I. E. JAMES, C. J. GRESS, B. R. BRADLEY, M. W. LARK, M. GOWEN, D. F. VEBER, *J. Med. Chem.* 2001, 44, 1380–1395.
- 24 R. W. MARQUIS, Y. RU, D. S. YAMASHITA, H.-J. OH, T. J. CARR, M. A. LEVY, T. A. TOMASZEK, C. F. IJAMES, W. W. SMITH, B. ZHAO, C. A. JANSON, S. S. ABDEL-MEGUID, K. J. D'ALESSIO, M. S. MCQUENEY, D. F. VEBER, *Bioorg. Med. Chem.* 1999, 7, 581–588.
- 25 D. M. JONES, B. ATRASH, H. RYDER, A. C. TEGER-NILSSON, E. GYZANDER, M. SZELKE, *J. Enzym. Inhib.* 1995, 9, 43–60; A. M. MAJALLI, K. T. CHAPMAN, M. MACCOSS, N. A. THORNBERRY, E. P. PETERSON, *Bioorg. Med. Chem. Lett.* 1994, 4, 1965–1968.
- 26 S. K. THOMPSON, S. M. HALBERT, M. J. BOSSARD, T. A. TOMASZEK, M. A. LEVY, B. ZHAO, W. W. SMITH, S. S. ABDEL-MEGUID, C. A. JANSON, K. J. D'ALESSIO, M. S. MCQUENEY, B. Y. AMEGADZIE, C. R. HANNING, R. L. DESJARLAIS, J. BRIAND, S. K. SARKAR, M. J. HUDDLESTON, C. F. IJAMES, S. A. CARR, K. T. GARNES, A. SHU, J. R. HEYS, J. BRADBEER, D. ZEMBRYKI, L. LEE-RYKACZEWSKI, I. E. JAMES, M. W. LARK, F. H. DRAKE, M. GOWEN, J. G. GLEASON, D. F. VEBER, *Proc. Natl. Acad. Sci.* 1997, 94, 14249–14254.
- 27 M. J. BOSSARD, T. A. TOMASZEK, M. A. LEVY, C. F. IJAMES, M. J. HUDDLESTON, J. BRIAND, S. THOMPSON, S. M. HALBERT, D. F. VEBER, S. A. CARR, T. D. MEEK, D. G. TEW, *Biochemistry* 1999, 38 (48), 15893–15902.
- 28 D. F. VEBER, M. G. BOCK, S. F. BRADY, E. H. ULM, D. W. COCHRAN, G. M. SMITH, B. I. LAMONT, R. M. DiPARDO, M. POE, R. M. FREIDINGER, B. E. EVANS, J. BOGER, *Biochem. Soc. Trans.* 1984, 12, 956–959.

6**Structure-Based Design of Potent and Selective Cdk4 Inhibitors**

TERUKI HONMA

6.1**Introduction¹⁾**

Cyclins and cyclin-dependent kinases (Cdks) play important roles in the regulation of the cell cycle [1, 2]. Each Cdk associates with a corresponding cyclin regulatory subunit to activate its phosphorylation activity. Different cyclin–Cdk pairs are active during each phase of the cell cycle (Fig. 6.1)[3]. In particular, the D type cyclins associate with Cdk4 or Cdk6 and are thought to regulate the restriction point in the early G₁ phase [4–7]. The restriction point represents a key stage in the life cycle of a cell [8]. Once cells have traversed this point, they proceed to a cell cycle entry without further input of growth factors for cell cycle progression.

The loss of function or deletion of p16 (natural specific inhibitor of cyclin D–Cdk4/6) or pRb (substrate of cyclin D–Cdk4/6) frequently occurs in clinical cancer cells [9]. In addition, the overexpression of Cdk4 or cyclin D is also observed in clinical cancer cells [10]. These facts suggest that alterations of the Cdk4-pRb pathway would be a major factor involved in carcinogenesis. Therefore, Cdk4/6-specific inhibitors are expected to be useful as a new class of “non-cytotoxic” anticancer agents [11].

Several Cdk inhibitors such as UCN-01 [12], flavopiridol [13, 14], purvalanol B [15], pyrido[2,3-d]pyrimidin-7-one [16] and indenopyrazole [17] have already been developed. Among these, the non-specific Cdk inhibitors UCN-01 and flavopiridol have been assessed in clinical trials. The next generation of Cdk inhibitors, selective inhibitors of only one target Cdk, are expected specifically to cause cell cycle arrest [11]. Suppression of tumor growth by G₁ arrest is thought to reduce the stress for normal cells to a greater extent than that in other phases because normal cells are usually resting in the G₀–G₁ phase. Thus, we tried to design Cdk4-selective inhibitors that cause specific cell cycle arrest in the G₁ phase.

In order to identify highly selective and potent Cdk4 inhibitors, we have applied a structure-based design approach consisting of the following two steps: (1) lead generation of a new class of Cdk4 inhibitors using a new de novo design strategy

¹⁾ Abbreviations: Cdk, cyclin-dependent kinase; pRb, retinoblastoma protein; PDB, Protein Data Bank; MW, molecular weight; SAR, structure-activity relationship; ACD, Available Chemicals Directory

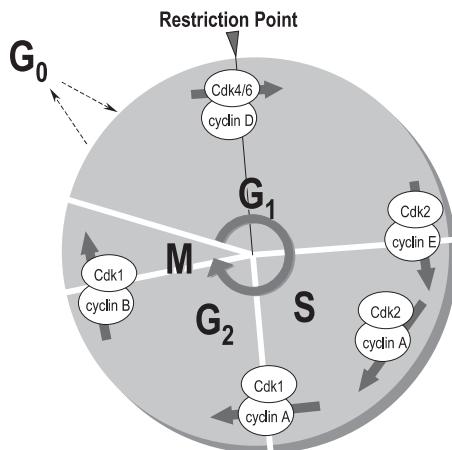


Fig. 6.1 Cell cycle and regulation of cell cycle progression by cyclin-dependent kinases [3]. G₁, Gap1 phase between M and S phases; S, DNA synthesis phase; G₂: gap 2 phase between S and M phases; and M, mitosis phase.

[18, 19] and (2) enhancement of Cdk4 selectivity of lead compounds over Cdk1/Cdk2 and the other kinases on the basis of the binding mode and structural differences between Cdk4 and other kinases [20].

This chapter describes the course of the research and discusses the strategy for the structure-based design in each step.

6.2

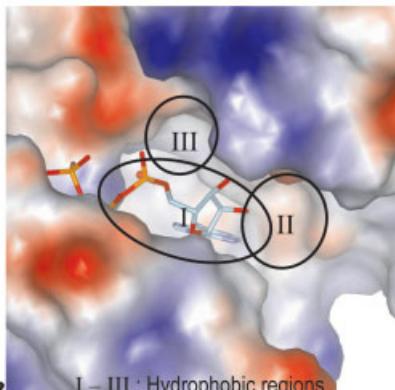
Homology Modeling of Cdk4

When we began designing a novel class of Cdk4 inhibitors in the late 1990s, an X-ray structure of the target protein, Cdk4, had not yet been solved (the X-ray structure of Cdk4 has not yet been reported to PDB as of December 2002). Therefore, homology modeling of the target protein based on X-ray structures of homologous proteins must be performed. One of the most important factors to construct homology models for structure-based design is the selection of a template X-ray structure. When selecting the template X-ray structure, both sequence identity with the target protein and whether or not the X-ray structure is an activated form should be taken into account. In many proteins, X-ray structures of an activated form are different from those of an inactivated form, and the alterations extend to putative drug binding sites. It is necessary to understand the activation mechanism of family proteins and to examine the literature of the X-ray structures of template candidates. With respect to the homology modeling of Cdk4, there are many template candidates that have different sequence identities and activation/inactivation states. We conclusively constructed a Cdk4 homology model using an X-ray structure of an activated form of Cdk2 as a template, which is described below.

The X-ray analyses of the Cdk family are summarized in Fig. 6.2A. In 1993, the first X-ray analysis of monomeric Cdk2 (PDB ID: 1HCK) was reported by De Bondt et al. [21]. In 1996, the structure of the Thr160-phosphorylated Cdk2-cyclin

| PDB ID | Molecule | Hydrogen-bonding sites used by small ligands | | | | | | | | | | | |
|--------|---|--|-----------|-----------|--------------------|-------------------|-----------|----------------------|--------------------|--------------------|------------|--------------------|--------------------|
| | | E81 CO | L83 NH | L83 CO | K33 N ζ H | D86 O δ | D86 NH | Q131 O ϵ | N132 O δ | D145 O δ | D145 NH | K89 N ζ H | Wat 74 α |
| 1B39 | Cdk2-ATP | NH | N | | PO | | | | OH | | | | |
| 1HCK | Cdk2-ATP | NH | N | | PO | OH | | | OH | | | | |
| 1FIN | Cdk2-ATP-cyclinA $_2$ | NH | N | | | | | | | | | | |
| 1JST | Cdk2-ATP-cyclinA $_2$ | NH | N | | PO | | | | | | | | PO |
| 1QMZ | Cdk2-ATP-cyclinA $_2$ -substrate $^{\beta}$ | NH | N | | PO | OH | | | OH | | | | |
| 1DM2 | Cdk2-hymenialdisine | NH | CO | NH | | | | | | NH | | | CO |
| 1CKP | Cdk2-Purvalanol B | | | N NH | | | | | | | | | CO- |
| | Cdk2-L828276 $^{\alpha}$ | OH | CO | | OH | | | | | NH | | | OH |
| | Cdk2-isopentenyladenyne | NH | N | | | N | | | | | | | |
| | Cdk2-olomoucine | | N NH | | | | | | OH | | | | |
| | Cdk2-roscovitine | | N NH | | | | | | | | | | |
| 1D18 | Cdk2-dimethoxyquinazoline | | N | | OH | | | | | OH | | | |
| 1E1V | Cdk2-v2058 | NH | N NH | | | | | | | | | | |
| 1E1X | Cdk2-v6027 | NH | N NH | | OH | | | | | | | | OH |
| 1E9H | Cdk2-indirubin-cyclinA $_2$ | NH | CO | NH | SO | | | | | | | | SO |
| 1FVT | Cdk2-oxyindole(BPDB) | NH | CO | NH | | NH | | | | | | | |
| 1FVV | Cdk2-oxyindole(EPDB)-cyclinA $_2$ | NH | CO | NH | N | | | | | | | | SO |
| 1G5S | Cdk2-H717 | | | N NH | | | | | | NH | NH | | |
| 1JSV | Cdk2-aminopyrimidine | | N NH | N NH | | SO | | | | | | | SO |
| 1AQ1 | Cdk2-staurosporin | NH | CO | | | NH | | | | | | | |
| 1GIH | Cdk2-diarylurea | | CO | NH | CO | | | | | | | | |
| 1GJL | Cdk2-diarylurea | | CO | NH | CO | NH | | | | | | | |
| 1BUH | Cdk2-Ckshs | | | | | | | | | | | | |
| 1JSU | Cdk2-cyclinA $_2$ -p27 | | | | | | | | | | | | |
| 1BI7 | Cdk6-p16 | | | | | | | | | | | | |
| 1B18 | Cdk6-p19 | | | | | | | | | | | | |
| 1BLX | Cdk6-p19 | | | | | | | | | | | | |

A



B I – III : Hydrophobic regions

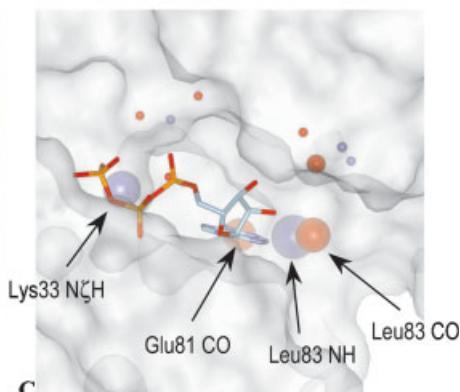
C
Iated Cdk2; (d) substrate peptide (HHASPRK); (e) des-chloroflavopiridol. (B)
The ATP binding pocket of Cdk2 (1JST). (C)
Locations of hydrogen-bonding sites used by
small ligands.

Fig. 6.2 X-ray structures of the Cdk family. (A) Reported X-ray structures of the Cdk family and hydrogen-bonding sites used by small ligands. (a) Wat74 is a conserved water molecule in 1JST; (b)ATP γ S; (c) Thr160-phosphory-

Iated Cdk2; (d) substrate peptide (HHASPRK); (e) des-chloroflavopiridol. (B) The ATP binding pocket of Cdk2 (1JST). (C) Locations of hydrogen-bonding sites used by small ligands.

A-ATP γ S complex (1JST), which is the final activated form of Cdk2, was clarified by Russo and co-workers [22–24]. They discovered the activation mechanism of Cdk2 by analyzing conformational changes in combination with cyclin A as well as the phosphorylation of Thr160 as shown in Fig. 6.3. In the X-ray structure of

the monomeric Cdk2-ATP complex (1HCK, Fig. 6.3 A), the T-loop inhibits pRb from approaching the pRb-binding region in Cdk2. This T-loop is sequentially moved out of the pRb binding region by binding with cyclin A (1FIN, Fig. 6.3 B) and by the phosphorylation of Thr160 (1JST, Fig. 6.3 C).

Although the X-ray structures of the Cdk6-p16 complex (1BI7) and Cdk6-p19 complex (1BI8 and 1BLX) are already known, their cyclin binding regions and the ATP-binding pockets in Cdk6 are both significantly altered by the endogenous inhibitors, p16 or p19 [25, 26]. In particular, alterations of locations of the T-loop and Gly-rich loop have been observed, and these alterations inhibit binding with ATP. In fact, the binding affinity of ATP to the Cdk6-p16 complex is dramatically decreased [25]. For this reason, in spite of a very high sequence identity between Cdk4 and Cdk6 (70%), we were not able to use these structural data when designing new Cdk4 inhibitors. However, the X-ray structure of the Cdk6-p16 complex revealed the location of the p16-binding region. p16 is a Cdk4/6 specific inhibitor

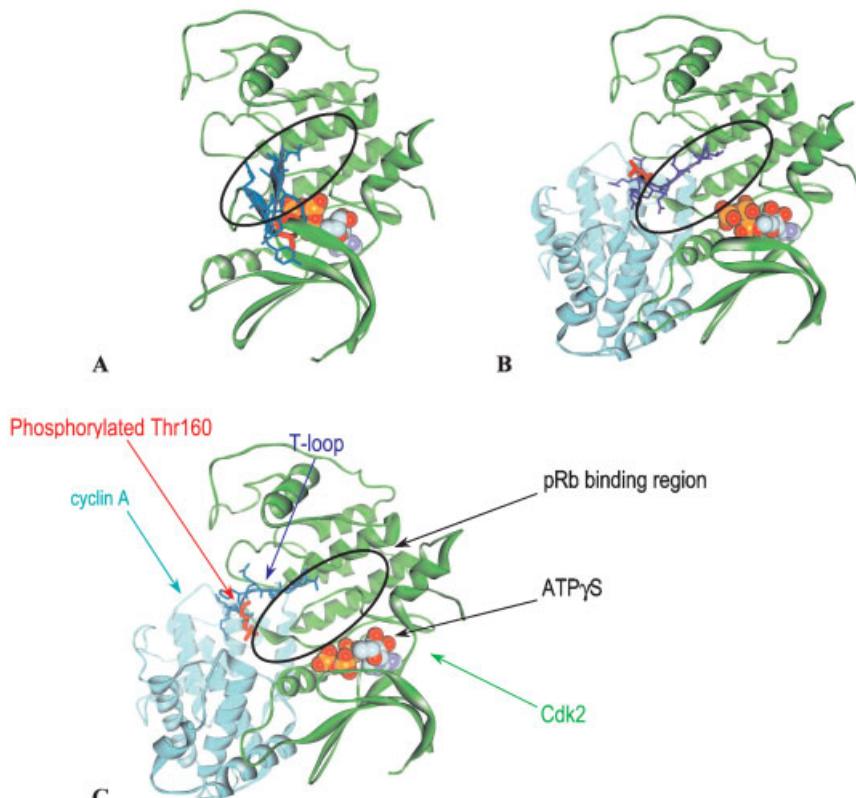


Fig. 6.3 Activation mechanism of Cdk2 by binding with cyclin A and by phosphorylation of Thr160. (A) X-ray structure of Cdk2-ATP

(1HCK). (B) X-ray structure of Cdk2-cyclin A-ATP (1FIN). (C) X-ray structure of Thr160 phosphorylated Cdk2-cyclin A-ATP γ S (1JST).

protein and does not inhibit Cdk1 and Cdk2 [28]. This information contributed to enhancing Cdk4 selectivity over Cdk1/2 as described in Section 6.6.

On the other hand, it is known that the level of sequence identity between Cdk4 and Cdk2 is quite high (45%) [29–32]. Therefore, we thought that it seemed appropriate to construct a Cdk4 model based on the final activated form of Cdk2 (1JST). According to the sequence alignments by Hanks et al. [29] we constructed an initial Cdk4 model using the modeling software, BIOCES[E] [33]. The model was minimized using the CHARMM (Chemistry at Harvard Molecular mechanics) force field [34] with the exception of the conserved region in Cdk4 and Cdk2. This minimized structure was used for the structure-based design of Cdk4 inhibitors.

6.3

Analysis of the ATP Binding Pocket

Before designing new inhibitors using the constructed homology model of the target protein, it is necessary to identify location(s) of drug-binding site(s) and to analyze their properties.

Many methods for drug-binding site detection have already been reported [35–38]. Most of these methods identify reasonably sized concave and hydrophobic features on the protein surface as putative drug-binding sites. Recently, some programs such as SiteID [39] and Alpha Site Finder [40] have become commercially available for this purpose. In the case of Cdk4, we were able to use the ATP binding pocket as one of the drug-binding sites because X-ray structures of the Cdk2-ATP complex and Cdk2-ATP-competitive inhibitor complexes have already been solved. With respect to ATP competitive inhibitors, a problem of target kinase selectivity arises because sequence identities between ATP binding pockets of most Ser/Thr/Tyr kinases are significantly higher than those of other regions [20, 29]. However, the development of cyclin- and pRb-competitive inhibitors also have the following problems. The cyclin binding region and the pRb binding region are so flat that the discovery of small, potent organic molecules that inhibit strong protein–protein binding is predicted to be difficult. Because sequence identities between Cdk4 and Cdk1/2 around these regions are not particularly low compared with those around the ATP binding pockets, a problem regarding selectivity of Cdk4 over Cdk1/2 remains an unresolved problem. For these reasons, we decided to use the ATP binding pocket as a drug binding site and solved the selectivity problem by introducing substituents toward the neighboring Cdk4-specific region (p16 binding region) [20].

In order to establish the properties of the ATP binding pocket, such as shape of the cavity, hydrogen-bonding sites and hydrophobic regions, we examined X-ray structures of Cdk2-small ligand complexes. Fig. 6.2B shows the solvent-accessible surface of the ATP binding pocket of Cdk2 (1JST) colored by partial charges. According to Fig. 6.2B, we can recognize the shape of the cavity, locations of hydrogen-bonding sites (colored sites) and hydrophobic binding regions (white regions). Furthermore, the importance of these hydrogen-bonding sites and hydrophobic re-

gions can be predicted by the comparative analysis of X-ray structures of the Cdk2-ligand complexes.

Hydrogen-bonding sites used by small ligands in the X-ray structures of Cdk2-ligand complexes are shown in Fig. 6.2A. Although there are more than 20 hydrogen-bonding sites in the ATP binding pocket that can be used by small ligands, only ten hydrogen-bonding sites were actually used and are listed in Fig. 6.2A. Among them, the main chain CO of Glu81, the main chain NH and CO of Leu83, and the side chain N ζ H are frequently used. In particular, the main chain NH of Leu83 is the most important because it serves as a hydrogen-bonding donor in every structure reported thus far. Locations of used hydrogen-bonding sites appear in Fig. 6.2C as colored spheres. Radii of the spheres depend on the frequency of use. The three most frequently used hydrogen-bonding sites (Glu81 NH, Leu83 NH and CO) are located in the deep area of the ATP binding pocket. The less frequently used hydrogen-bonding sites are located closer to the outside of the cavity.

In terms of hydrophobicity, some deep areas (hydrophobic regions I and II in Fig. 6.2A) in the ATP binding pocket around Phe80, Glu81 and Leu83 are occupied by the flat aromatic rings of most Cdk2 inhibitors such as L-828276 [41], staurosporin [42] and roscovitine [43]. Indeed, these areas are surrounded by the hydrophobic amino acid residues Ile10, Val18, Ala31, Val64, Phe80, Phe82, Leu134 and Ala144, and form a narrow cavity.

In Cdk4, most of the residues that are important for hydrogen bonds or hydrophobic interactions are conserved according to Hanks' multiple alignment [29]. Among the altered amino acid residues, those between Leu83 in Cdk2 and Val96 in Cdk4 are not critical because only the main chain is used for hydrogen bonds. Therefore, we determined that the information regarding these structural requirements would also be useful for discovering a new class of Cdk4 inhibitors [18].

6.4

Strategies for Structure-Based Lead Identification: Virtual Screening of Known Compounds and de Novo Design

Lead identification based on structural information can be divided into two strategies. The most popular strategy is virtual screening (also called *in silico* screening or 3D database search) of known compounds by docking into the target protein and scoring of the docking models [44, 45]. Another strategy is the design of novel structures fitted to the drug-binding site (*de novo* design) [46]. The features of these two strategies are briefly summarized below.

1. Virtual screening (*in silico* screening, 3D database search) of known compounds
 - *Representative programs:* DOCK [47, 48], ADAM & EVE [49], Autodock [50], PRO-SELECT [51], FlexX [52] and LigandFit [53].
 - *Advantage:* Selected compounds are available from in-house chemical collections, commercial reagents or known synthetic methods.

- *Disadvantage:* Novel structures that do not exist in known compound databases cannot be generated.
- 2. de Novo design
 - *Representative programs:* LEGEND [54, 55], LUDI [56], SPROUT [57], HOOK [58], GrowMol [59], PRO-LIGAND [60], CONCERTS [61] and LeapFrog [62].
 - *Advantage:* Novel structures that do not exist in known compound databases can be generated.
 - *Disadvantage:* Most output structures are not available from commercial reagents or known synthetic methods.

Apart from the disadvantages listed above, the following two problems commonly arise: (1) protein flexibility is not taken into sufficient consideration, and (2) the prediction of binding affinities between the proteins and ligands is not particularly accurate. Since both methods have advantages and disadvantages, the selection of the strategy depends on objectives of the lead identification and the situation. To discover lead compounds efficiently from in-house chemical collections or commercial reagents, virtual screening of known compounds is suitable and successful examples have been shown in many recent reports [44, 45]. On the other hand, to identify a novel class of lead compounds, we would select the de novo design strategy. In particular, situations where only known compounds have been used and attractive lead compounds have not been obtained, we have to try the de novo design.

6.4.1

New de Novo Design Strategies

As already mentioned, the de novo design strategy has the following critical disadvantages: (1) de novo design programs do not consider synthetic feasibility when constructing structures and (2) the prediction of binding affinities for the designed structures is not particularly accurate. For these reasons, in syntheses, structures designed by the de novo design strategy have been problematic with insufficient or non-existent potency. Therefore, in practice, it has been difficult to obtain novel lead compounds that are appropriate for medicinal chemists in actual drug discovery projects.

To overcome these obstacles, since the late 1990s, validation methods for small fragments that could be substructures of the whole structure have been developed. These “fragment-based” methods use biophysical experimental techniques such as NMR spectroscopy [63], X-ray analysis [64] and MS spectroscopy [65] to validate whether or not each small fragment can bind to the active site of the target protein. Validated small fragments (called small binders) have been developed into full-sized novel lead compounds by connecting them or introducing other suitable substituents. Because these methods can be used to construct whole structures using small binders for which binding affinities and binding modes are confirmed, the success rate for the design of potent and novel lead compounds will be greatly improved.

A different approach to improving the efficiency of de novo design is prioritization by chemical availability (commercial availability and synthetic feasibility) [66, 67]. From a chemist's point-of-view, the synthetic feasibility of various types of structures designed by the de novo design strategy would be broadly distributed. For example, there are commercially available structures that do not require any synthetic effort, and there are synthetically feasible structures that can be synthesized within one week. On the other hand, there are synthetically problematic structures that take over a year to synthesize, even by post-doctoral synthetic chemists. In addition, there are also structures that cannot be obtained by currently available synthetic reactions and structures that are not stable under physiological conditions. For these reasons, when chemists select synthetic targets from all of the structures designed by the de novo design strategy, information concerning synthetic accessibility is critical. The synthetic accessibility of thousands of output structures should be provided automatically by a computer-aided system because manual research as to the synthetic feasibility of a large number of structures by chemists is impossible.

When performing de novo design, the key to success is to reduce the risk and cost of each step as much as possible using one or both of these methods. Fig. 6.4 summarizes a new strategy of de novo lead generation combined with fragment-based methods and with the evaluation of chemical availability. Firstly, small structures that are predicted to interact with the most important hydrogen-bonding sites and hydrophobic regions are designed using de novo design programs. Considering the chemical availability of each structure, candidate structures are selected and obtained through purchase or synthesis. Novel scaffolds of lead com-

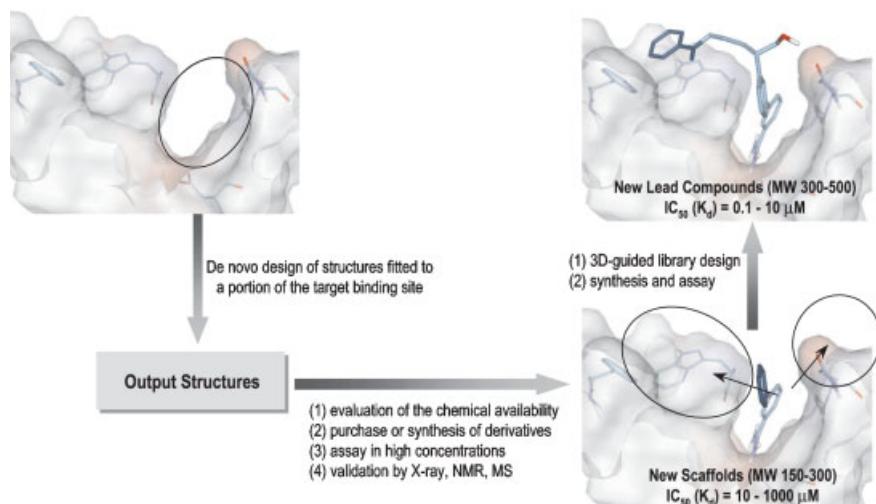


Fig. 6.4 Schematic figure of new de novo design strategy.

pounds are identified by assay in high concentrations and are validated using X-ray, NMR or MS. Ideally, scaffolds should have a MW ranging from 150 to 300 and an IC₅₀ (K_d) ranging from 10 to 1000 μM . We then design full-sized compounds (MW ranging from 300 to 500) by the optimization of the scaffold moiety and/or the introduction of substituents fitted to neighboring binding sites.

6.4.2

Evaluation of Chemical Availability by SEEDS

To evaluate chemical availability of output structures obtained through de novo design programs, we constructed a supporting system for chemical structure selection called SEEDS (System for Evaluation of availability of Essential structures generated by De novo design programs) [67]. This system was constructed by taking advantage of database searches. The SEEDS system plays three different important roles and are as follows: (1) SEEDS automatically picks out an essential core substructure of each output obtained from de novo design programs, (2) based on each essential part, several queries for searching compound databases and reaction databases are made automatically, (3) using these queries, SEEDS searches for commercially available derivatives and/or synthetically feasible derivatives of the essential part, and makes a list of all commercially available derivatives as well as lists of the chemical stability and synthetic accessibility of all outputs (Fig. 6.5).

The chemical stability and synthetic accessibility are evaluated using the results of a database search with various types of queries. The rules for extracting essential parts and making queries for searching compound databases are shown in an example (Fig. 6.6). Structure 1 is one of the output structures generated by LEGEND. If we searched compound databases using structure 1 itself as a query, no commercially available derivatives would be obtained because of its relatively simple alkyl chains, which will not form hydrogen bonds. Therefore, SEEDS extracts only the essential part (structure 2) of each output. Essential parts include rings and atoms that form hydrogen bonds. Atoms that form hydrogen bonds are more important than other parts because hydrogen bonds are one of the major determinants of binding affinity between the protein and the small ligand. Rings are more important than other parts because they constrain the atoms significantly, which form hydrogen bonds, in appropriate directions. Rings are also important for hydrophobic interactions with the hydrophobic surface of the protein. Subsequently, SEEDS assigns an appropriate atom type (i.e., atom list, ring bond count and substitution count) to each atom, and also assigns a suitable bond type (i.e., bond list and reaction center) to each bond [68]. Query 3 is one of the examples used for searching compound databases. In fact, SEEDS automatically makes 3 queries when searching compound databases and 15 queries when searching reaction databases. Chemical stability and synthetic accessibility are evaluated using the results of a database search with all 18 queries.

By using SEEDS for output structures generated through de novo design programs, we can obtain the derivatives with minimal chemical efforts and can eluci-

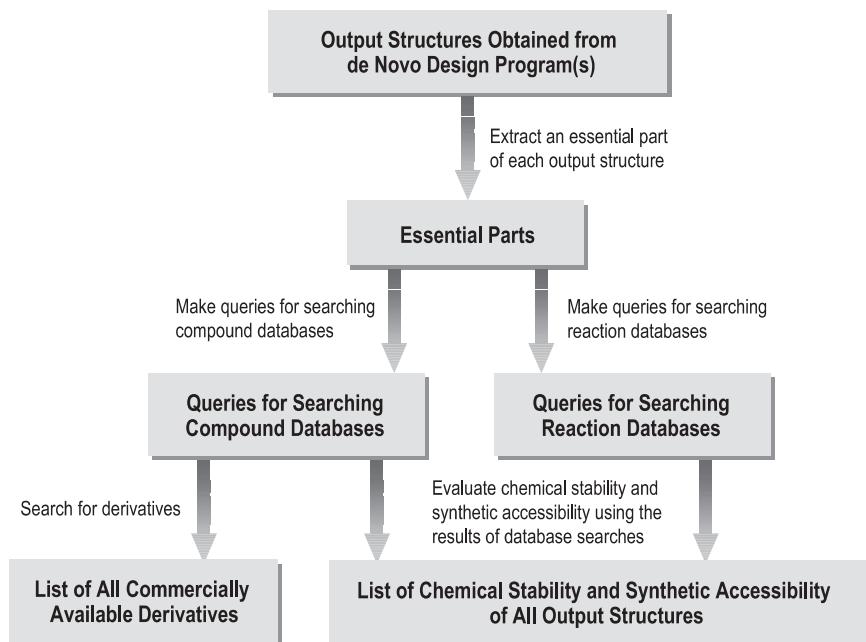


Fig. 6.5 Schematic figure of SEEDS.

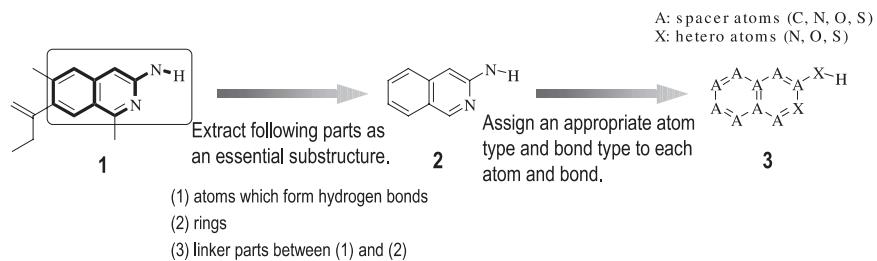


Fig. 6.6 Rules for extracting essential parts and making queries for searching compound databases. **1** An example of output structures

generated by LEGEND. **2** An essential part of structure **1**. **3** One of the queries for searching compound databases made by SEEDS.

date the potentialities with a biological assay in high concentrations. As previously described, SEEDS provides both estimates of synthetic accessibility for each output and commercially available derivatives of the essential part of each output. Taking into consideration the ability of binding affinity predictions for ligands designed by the de novo design strategy, commercially available derivatives should be purchased and assayed because syntheses of novel structures are costly alternatives even if the structures are synthetically feasible. When attractive structures are not identified from commercially available derivatives, the next synthetically accessible derivatives should be considered.

6.5

Structure-Based Generation of a New Class of Potent Cdk4 Inhibitors

We used the de novo design program LEGEND with SEEDS to generate novel scaffold candidates of Cdk4 inhibitors with MW of less than 350. After the discovery of some promising scaffold candidates, it is important to validate whether the scaffold candidates bind to Cdk4 in the desired ways. In order to validate the scaffold candidates, it is critical to use experimental techniques such as X-ray analysis and NMR spectroscopy. However, both X-ray analysis and NMR measurement of Cdk4 have proven to be difficult because a crystal of Cdk4 is unstable and the MW of Cdk4 (44 kDa) is too large to be measured by NMR. Therefore, we synthesized several informer libraries of the scaffold candidates. Using the preliminary SAR, we validated the binding mode by molecular modeling methods. Subsequently, modifications based on the predicted binding mode were implemented in order to develop potent lead compounds from the identified scaffold.

6.5.1

Identification of New Scaffold Candidates Using LEGEND and SEEDS

To identify new scaffolds that satisfy structural requirements (Fig. 6.2), we applied the de novo design program, LEGEND. This program is based on the atom-by-atom approach [46] and is suitable for generation of drug molecules in a deep, narrow cavity just like the ATP binding pocket. With this program, we can specify the direction of growth, the size of the molecules that will be generated and the hydrogen bonds that will be formed during calculations. The specification of calculations performed by LEGEND and the process of scaffold generation are shown in Fig. 6.7.

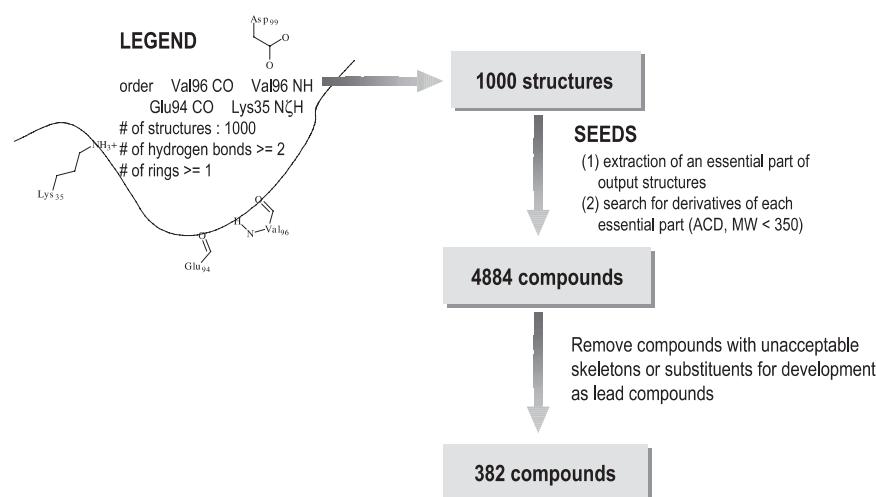


Fig. 6.7 Process of scaffold identification using LEGEND and SEEDS [18].

First, we obtained 1000 output structures, however, most of these structures were not commercially available or synthetically feasible. As previously mentioned, this problem appears to apply to most de novo design programs that build structures sequentially.

Using SEEDS, we initially searched for commercially available derivatives. On searching the ACD [69], 4884 compounds with MW of less than 350 were selected. Compounds with skeletons or functional groups that were unacceptable for the development of lead compounds were omitted, 382 commercially available compounds remained. The compounds were purchased for screening in cyclin D-Cdk4 assays at concentrations of up to 1 mM. We obtained 18 compounds with IC₅₀ values of under 500 μM as hits. Most of the hits had large, flat aromatic rings and neighboring hydrogen-bonding donors and acceptors. Twelve of the 18 hits were clustered into one of the following four classes according to their structural profile: diarylurea, cyclic urea and thiourea, 4-aryl-2-aminopyridine and triazine, and pteridine (Fig. 6.8).

In these four classes, our interest was directed at the diarylurea scaffold because there were five potent compounds with a diarylurea structure, which is appropriate for the rapid construction of structurally diverse libraries. Therefore, we constructed diarylurea informer libraries by parallel synthesis to validate the potential of the scaffold and to obtain preliminary SAR.

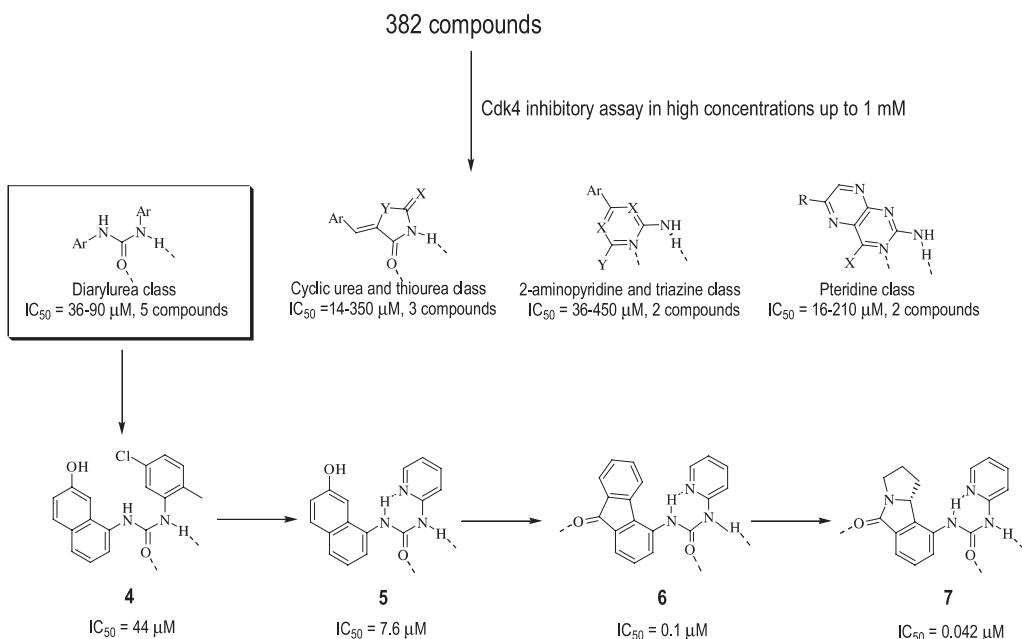


Fig. 6.8 From scaffold candidates to lead compounds.

6.5.2

From the Identified Scaffold to Lead Compounds

The data obtained through LEGEND suggested that neighboring NH and CO in the diarylurea scaffolds form hydrogen bonds with the main chains of Glu94 and/or Val96. Aromatic rings are supposed to be located in the hydrophobic regions as shown in Fig. 6.2B. Based on these insights, we designed informer libraries as described below (Fig. 6.8).

Firstly we searched for alternative substituents in place of the 5-chloro-2-methylphenyl group in diarylurea **4**, leaving the 7-hydroxynaphthyl group intact. Considering the deep, narrow shape of the ATP binding pocket, we mainly selected flat aromatic amines as building blocks to synthesize diarylurea derivatives. Fifty-five urea compounds were synthesized in moderate yields (30–70%) from 8-isocyanato-2-naphthol and alkylamines. Substitutions with 2-pyridinyl (**5**) gave IC₅₀ values of 7.6 μM, while substitutions with 3-pyridinyl and 4-pyridinyl groups showed poor inhibitory activities [18]. It seemed that the relative position of the nitrogen atom in the aromatic ring might be important for the binding of the inhibitor. None of the compounds substituted with aliphatic amines showed enhanced potency [18].

Next, we kept the 2-pyridinyl group intact and modified the 7-hydroxynaphthyl group in compound **5**. The 7-hydroxynaphthyl group was predicted to be located in the hydrophobic, narrow and deep area of the ATP binding pocket. Therefore, as in the case of the first informer library, we mainly selected aromatic amines as building blocks. Using these building blocks, 410 urea compounds were synthesized and tested in Cdk4 inhibitory assays. This area required an aromatic ring, because none of the aliphatic substituted compounds showed inhibitory activity [18]. The existence of a hydrogen-bonding acceptor in the left part of compound **5** led to a significant improvement in potency. This was also the case with compound **6** (IC₅₀=0.1 μM) [18]. Compound **6** was selected as the lead compound for further modifications.

6.5.3

Confirmation of the Binding Mode

In order to investigate the binding mode of **6** more comprehensively, we performed a docking study between the Cdk4 model and **6**. As a result, four patterns of binding modes were suggested [18]. Considering the structural requirements for Cdk2/4 inhibitors (Fig. 6.2) and the SAR of the informer libraries, we selected the binding mode candidate shown in Fig. 6.9B. In this binding mode, there is some degree of steric repulsion between the pyridine ring and the terminal benzene ring in the fluorenone. On the other hand, the terminal benzene ring in the fluorenone is supposed to be important for binding affinity since the terminal benzene ring is appropriately located on the hydrophobic surface of Cdk4.

Based on these assumptions, we attempted to replace the terminal benzene ring with various types of 5-membered rings in order to eliminate unnecessary intramolecular steric repulsion but while maintaining favorable hydrophobic interac-

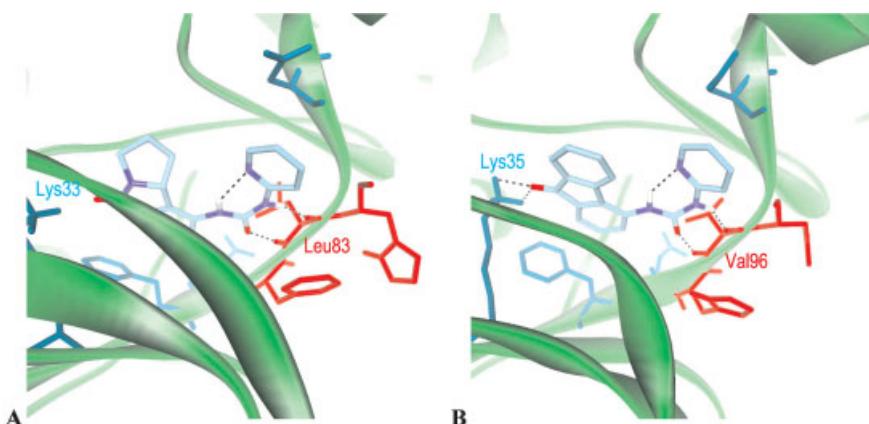


Fig. 6.9 The X-ray structure of Cdk2-7 complex (A) and modeling prediction of Cdk4-6 complex (B) [18, 19].

tions. These modifications provided more potent compounds [18]. Representative compound 7 [IC_{50} (Cdk4)=0.042 μM] is shown in Fig. 6.8.

For further validation of the binding mode, we tried X-ray analysis of the Cdk2-6 complex using the soaking method. However, compound 6 was so insoluble (solubility at pH 7.4 in Tris-Cl buffer, <0.01 $\mu\text{g mL}^{-1}$) that soaking was not appropriate. During modifications of 6, the relatively soluble compound 7 (solubility at pH 7.4 in Tris-Cl buffer, 7.18 $\mu\text{g mL}^{-1}$) with a potent binding affinity was obtained. Using compound 7, we succeeded in X-ray analysis of the Cdk2-7 complex (PDB ID: 1GIH) [19]. A comparison between the modeling prediction and the X-ray structure is shown in Fig. 6.9. The X-ray analysis showed that 7 formed hydrogen bonds with the main chain NH and CO groups of Leu83 (corresponding to Val96 in Cdk4) and adopted a U-shaped conformation in the ATP binding pocket of Cdk2. The binding mode of 7 in Cdk2 (Fig. 6.9 A) was found to be consistent with the predicted binding mode of 6 in Cdk4 (Fig. 6.9 B).

Based on the results of the docking study with the Cdk4 model, X-ray analysis of the Cdk2 complex, and the SAR obtained by the Cdk4 inhibitory assays, we concluded that mode A was the most preferable binding mode of the diarylurea series of lead compounds in the activated form of the cyclin D-Cdk4 complex.

6.6

Structure-Based Design of Cdk4 Selective Inhibitors

Cdk4 inhibitor 7, which was obtained from the new de novo design strategy, is moderately selective for Cdk4 over 14 representative kinases except for the Cdk family; however, compound 7 inhibits Cdk1 and Cdk2 in addition to Cdk4 with almost same potency [20].

The enhancement of the selectivity for the target protein over its highly homologous counterparts is not feasible because there are few differences in the protein surfaces of the active sites. Recently, successful examples of the development of such selective compounds have been described [70–73], these processes were developed by applying structural information concerning critical differences in amino acid residues and/or conformational changes between the target proteins (MMP-12, PTP-1B and COX-2) and their counterparts. However, in the case of protein Ser/Thr/Tyr kinases, it is much more difficult to improve the selectivity for a target protein because there are hundreds of homologous kinases in the superfamily [29–32]. In order to overcome this problem, we initially identified Cdk4-specific amino acid residues around the ATP binding pocket by comparing the amino acid sequences of 390 representative kinases. Subsequently, a chemical library was designed using this information regarding the locations of these amino acid residues and the binding mode of compound 7 to achieve specific interactions with Cdk4 and to discriminate against interactions with other kinases.

6.6.1

Identification of Cdk4-Specific Amino Acid Residues

First, we performed a 3D superposition of the structural model of Cdk4 bound with compound 7 and the X-ray structure of Thr160-phosphorylated Cdk2-cyclin A-ATP γ S (PDB ID: 1JST). We then chose 39 amino acid residues in the Cdk4 model that are located within 7 Å of the ATP γ S molecule in the Cdk2.

The multiple alignment of 390 representative kinase sequences by Hanks and Quinn has been published on the internet [29]. Using multiple alignments, we calculated the following: (1) the conserved ratios for each of the 39 residues of Cdk4 in the 390 kinases, and (2) the amino acid identities in the region confined within the 39 residues around the ATP binding pocket. The multiple alignments among Cdk4, Cdk6, Cdk1 and Cdk2 and the results of the calculations are summarized in Fig. 6.10A. From these data, we were able to identify frequently altered amino acid residues in the 390 kinases and altered amino acid residues in Cdk1/2.

Among the frequently altered residues, those with side chains not facing inhibitors were not useful for designing new inhibitors. We had to use the differences in chemical properties and/or shape of the side chains of the residues that faced compound 7 to enhance Cdk4 selectivity. For this purpose, 26 residues that were directly in contact with or accessible to ATP γ S or compound 7 were selected by considering the 3D structures of the complexes. These residues are shown in the boxes in Fig. 6.10A. Thirteen residues (Ile12, Val14, Ala16, Tyr17, Phe93, His95, Val96, Asp97, Gln98, Asp99, Thr102, Glu144 and Ala57) that were <40% conserved in 390 kinases and in contact with or accessible to ATP γ S or compound 7 are indicated by yellow circles in Fig. 6.10A. The side chains of Phe93, His95, Val96 and Ala157 were already predicted to have contact with compound 7 by modeling and the X-ray structure of the Cdk2-7 complex. The moderately higher selectivity of compound 7 for Cdk4 over other kinases is likely to be due to specific interactions with these residues.

| no. in Cdk4 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 32 | 33 | 34 | 35 | 56 | 72 | 73 | 92 | 93 | 94 |
|---------------------------|----|----|-----|----|-----|----|----|-----|----|-----|----|----|----|----|-----|-----|----|----|----|----|----|
| Cdk4 | E | I | G | V | G | A | Y | G | T | V | Y | V | A | L | K | E | V | R | V | F | E |
| Cdk6 | E | I | G | E | G | A | Y | G | K | V | F | V | A | L | K | E | V | R | V | F | E |
| Cdk2 | K | I | G | E | G | T | Y | G | V | V | Y | V | A | L | K | E | V | K | V | F | E |
| Cdk1 | K | I | G | E | G | T | Y | G | V | V | Y | V | A | M | K | E | V | S | I | F | E |
| AGC group (71 kinases) | 0 | 18 | 99 | 4 | 100 | 6 | 8 | 94 | 6 | 99 | 18 | 1 | 99 | 10 | 100 | 100 | 58 | 11 | 70 | 3 | 68 |
| CaMK group (42 kinases) | 21 | 21 | 100 | 7 | 100 | 24 | 12 | 55 | 7 | 81 | 12 | 33 | 93 | 7 | 100 | 100 | 40 | 36 | 69 | 26 | 71 |
| CMGC group (81 kinases) | 5 | 51 | 98 | 1 | 100 | 30 | 70 | 77 | 14 | 100 | 33 | 79 | 86 | 32 | 100 | 99 | 47 | 16 | 67 | 58 | 86 |
| PTK group (104 kinases) | 16 | 25 | 99 | 2 | 98 | 24 | 5 | 100 | 7 | 97 | 29 | 91 | 96 | 1 | 100 | 98 | 58 | 31 | 42 | 9 | 82 |
| OPK group (92 kinases) | 11 | 48 | 88 | 0 | 97 | 13 | 13 | 78 | 15 | 84 | 47 | 55 | 91 | 8 | 99 | 93 | 51 | 16 | 43 | 0 | 64 |
| all (390 kinases) | 10 | 34 | 96 | 2 | 99 | 19 | 22 | 84 | 10 | 93 | 30 | 58 | 93 | 11 | 100 | 98 | 52 | 21 | 56 | 18 | 75 |

| no. in Cdk4 | 95 | 96 | 97 | 98 | 99 | 102 | 140 | 142 | 143 | 144 | 145 | 146 | 147 | 157 | 158 | 159 | 160 | 161 | %ATP + %full + | |
|---------------------------|----|----|----|----|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|----------------|-----|
| Cdk4 | H | V | D | Q | D | T | D | K | P | E | N | I | L | A | D | F | G | L | 100 | 100 |
| Cdk6 | H | V | D | Q | D | T | D | K | P | Q | N | I | L | A | D | F | G | L | 90 | 70 |
| Cdk2 | F | L | H | Q | D | K | D | K | P | Q | N | L | L | A | D | F | G | L | 72 | 45 |
| Cdk1 | F | L | S | M | D | K | D | K | P | Q | N | L | L | A | D | F | G | L | 64 | 43 |
| AGC group (71 kinases) | 0 | 41 | 0 | 0 | 59 | 24 | 100 | 100 | 69 | 65 | 100 | 46 | 85 | 24 | 100 | 82 | 90 | 49 | | |
| CaMK group (42 kinases) | 0 | 38 | 10 | 0 | 7 | 0 | 100 | 98 | 79 | 93 | 100 | 38 | 86 | 48 | 98 | 100 | 100 | 62 | | |
| CMGC group (81 kinases) | 9 | 17 | 41 | 14 | 74 | 6 | 100 | 99 | 94 | 15 | 100 | 22 | 83 | 35 | 98 | 85 | 99 | 79 | | |
| PTK group (104 kinases) | 2 | 10 | 2 | 3 | 32 | 5 | 99 | 0 | 0 | 0 | 100 | 37 | 88 | 40 | 98 | 95 | 97 | 82 | | |
| OPK group (92 kinases) | 3 | 7 | 13 | 1 | 14 | 2 | 99 | 95 | 45 | 10 | 100 | 63 | 72 | 22 | 9 | 89 | 99 | 59 | | |
| all (390 kinases) | 3 | 19 | 13 | 4 | 39 | 7 | 99 | 72 | 51 | 27 | 100 | 42 | 82 | 33 | 99 | 90 | 97 | 68 | | |

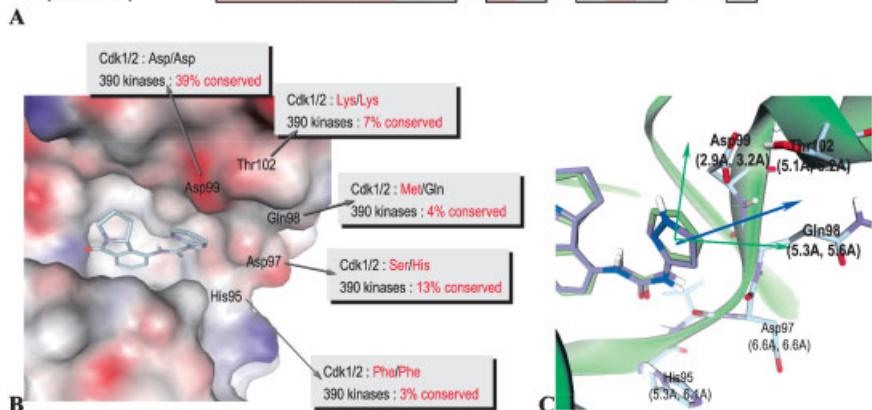


Fig. 6.10 Identification of Cdk4-specific amino acid residues [20]. (A) Amino acid residues around the ATP binding pocket.
 (a) Sequence identity confined to around the ATP binding pocket of Cdk4; (b) sequence

identity based on full sequences. (B) Binding mode of compound 23 in the Cdk4 homology model and altered/frequently altered residues in the p16-binding region. (C) Suitable positions to introduce substituents.

Subsequently, we tried to identify altered amino acid residues in Cdk4 and Cdk1/2. In the region of the ATP binding pocket of Cdk4, the residue identities were found to be higher with respect to Cdk1/2 (64%/72%), while the identities throughout the full sequence were not as high (43%/45%). Among the residues around the ATP binding pocket, we identified eight residues (Val4, Ala16, His95, Val96, Asp97, Gln98, Thr102 and Glu144) that were altered in Cdk1/2 and have contact or are accessible to ATPγS or compound 7 (these residues are indicated by red circles and a red open circle in Fig. 6.10A). Except for Gln98, all of these resi-

dues were altered in both Cdk1 and Cdk2. Interestingly, Gln98 was conserved in Cdk2 as shown in Fig. 6.10A and B. It is worth noting that five of the eight residues were distributed on one side of the ATP binding pocket. According to X-ray analysis of Cdk6-p16/p19 complexes, this side is the p16-binding region [25–27]. p16 is a Cdk4/6-specific, endogenous inhibitor protein that is thought to suppress carcinogenesis [9, 74]. Therefore, the amino acid residues in the p16 binding region of Cdk4 are likely to interact with p16 in a specific way.

By analyzing the binding mode of compound 7, we found that the pyridine ring of compound 7 is directed toward the amino acid residues in the p16 binding region (Fig. 6.10B). The terminal of the side chain of Val96 cannot come into contact with the inhibitor because it is hidden by the side chain of Leu147. Ile12 is replaced by Leu in most of the kinases, this causes only small changes in the protein surface. For these reasons, we removed Val96 and Ile12 from the target residues. On the other hand, Asp97 is replaced by His (Cdk2) or Ser (Cdk1), and Thr102 is replaced by Lys (Cdk1/2) in most kinases. These alterations cause dramatic changes in the chemical properties and shape of the protein surface. Amino acid residues such as Asp99, Gln98 and His95 are also useful. Asp99 is replaced by Ala, Ser, Asn, Pro or Cys in most protein tyrosine kinases. Gln98 is replaced by Met in Cdk1 and by various other residues in other kinases. His95 is replaced by Phe in Cdk1/2 and by Phe or Tyr in most kinases. Therefore, in order to enhance Cdk4 selectivity as compared with Cdk1/2 and other kinases, the introduction of substituents that face the side chains of these residues was expected to be more effective.

6.6.2

Library Design Based on the Locations of Cdk4-Specific Amino Acid Residues

To select the first target residues to interact with substituents on our lead compounds, we calculated the shortest distances between the substitutable positions of the inhibitors and the side chain of each residue in the p16 binding region (Fig. 6.10C). We found that compound 8 [blue sticks in Fig. 6.10C, IC₅₀ (Cdk4)=0.021 μM], which has a 5-membered pyrazole ring, is useful as a lead in addition to compound 7 and makes it possible to approach these residues from various directions. Asp99, Thr102 and Gln98 were more accessible than the other residues. Based on the binding modes, the 5-position of the pyrazole ring in compound 8 (blue bold arrow in Fig. 6.10C) was predicted to be the most appropriate position and direction to make contact with the side chains of the three residues. Therefore, we constructed a 5-substituted pyrazole-3-yl urea library to enhance Cdk4 selectivity.

In order to design suitable substituents, we employed the de novo design programs, LUDI and LeapFrog. Although the output structures obtained from the de novo design programs are often not synthetically feasible, these structures have suitable properties such as a hydrogen-bonding donor, acceptor and hydrophobicity and have an appropriate length of substituents. The alkylaminomethyl groups and their cyclized groups were provided as typical structures by both LUDI and

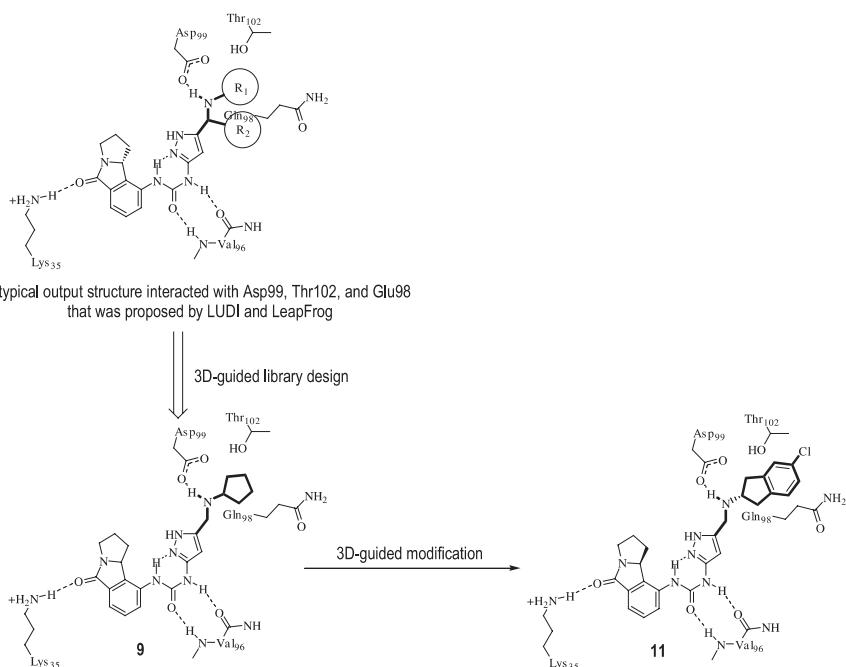


Fig. 6.11 Library design based on locations of Cdk4-specific amino acid residues using de novo design programs.

LeapFrog (Fig. 6.11). The NHs of these amino groups were predicted to interact with the carboxyl group of Asp99 and/or the side chain OH of Thr102. The OH of Thr usually functions as both a hydrogen-bonding donor and acceptor. However, in this case, the OH of Thr102 was already predicted to make a hydrogen bond with the carboxyl group of Asp99, and the side chain of Thr102 was expected to work as a hydrogen-bonding acceptor and to be a small hydrophobic portion for our lead compounds. With respect to alkyl groups via the amino-methyl group, hydrophobic substituents were thought to be appropriate considering the hydrophobic environment surrounded by the side chain methyl group of Thr102 and the side chain methylene group of Gln98 (Fig. 6.10B). These alkyl-amino groups can easily be introduced by reductive amination with the corresponding aldehyde of compound 8. Thus, we constructed a chemical library of 8 featuring a pyrazole ring with alkylaminomethyl substituents at the 5-position.

Compounds in the first alkylamino library (64 compounds) were tested by both cyclin D-Cdk4 and cyclin A-Cdk2 inhibitory assays. Interestingly, the introduction of the amino groups caused a 100- to 200-fold decrease in the Cdk2-inhibitory activity compared with that of compound 7 [IC_{50} (Cdk2)=0.078 μM] [20]. Although the addition of these amino groups also caused a minor reduction in Cdk4-inhibitory activity in general, the compounds substituted with cyclopentyl (9, Fig. 6.11) had almost comparable inhibition potencies [IC_{50} (Cdk4)=0.065 μM] with that of

compound 7. Consequently, compound 9 showed much better selectivity for Cdk4 over Cdk2 (150-fold). The SAR and structural information around the p16 binding region suggested the following interactions: (1) the NH_2^+ moieties on the alkylamino groups formed a hydrogen bond with the carboxyl group of Asp99 (Asp86 in Cdk2), (2) in Cdk2, the alkylamino groups constrained by the hydrogen bond with Asp86 caused electrostatic and/or steric repulsion with the side chain of Lys89, and (3) in Cdk4, the hydrophobic and bulky substituents of 9 fit into the hydrophobic region surrounded by the side chains of Thr102 and Gln98.

From these facts, we designed 10 (Fig. 6.12 A) with a pyrrolidine ring to restrict the substituent, and 11 (Fig. 6.11) with a 5-chloroindan-2-ylaminomethyl group to enhance the hydrophobic interaction with Cdk4 and selectivity over Cdk2. Compound 10 showed comparable selectivity over Cdk2 (120-fold) with that of 9 [20]. In order to investigate the structural basis of selective Cdk4 inhibition over Cdk2, we carried out X-ray analysis of these Cdk4-selective inhibitors complexed with Cdk4. However, we have not yet succeeded because of the low stability of Cdk4. Accordingly, we made mutant Cdk2, which has F82H, L83V and K89T and mimics the p16-binding region in Cdk4. In addition, we have successfully completed the X-ray analysis of the mutant Cdk2-10 complex (PDB ID: 1GIJ) using the soaking method (Fig. 6.12 A) [19]. In this X-ray structure, the NH_2^+ group of the pyrrolidine in 10 formed a hydrogen bond with the carboxyl group of Asp86. This proved our hypothesis.

On the other hand, 11 showed a 30- to 40-fold enhancement in Cdk4 inhibitory activity ($\text{IC}_{50}=0.0023\ \mu\text{M}$) compared with that of 9. Compound 11 also showed good selectivity over Cdk1/2 (Fig. 6.13 B). The docking studies between Cdk4/Cdk2 and 11 predicted steric repulsion between the large chloroindanyl group constrained by the hydrogen bond with Asp86 and the side chain of Lys89 in Cdk2 (Fig. 6.12 B). In addition, it was predicted that the chloroindanyl group in 11 would make good van der Waals contact with the side chains of Thr102 and

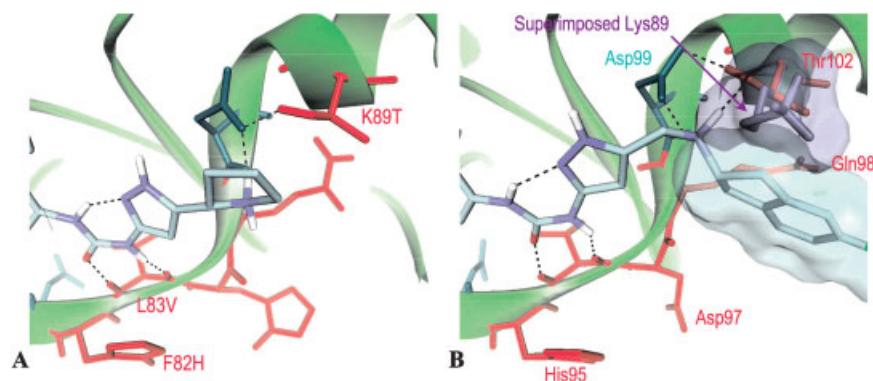


Fig. 6.12 Structural basis of selective Cdk4 inhibition by compound 11 [20]. (A) Binding mode of 10 in Cdk4 mimic Cdk2 (X-ray,

PDB ID: 1GIJ). (B) Binding mode of 11 in Cdk4 (model).

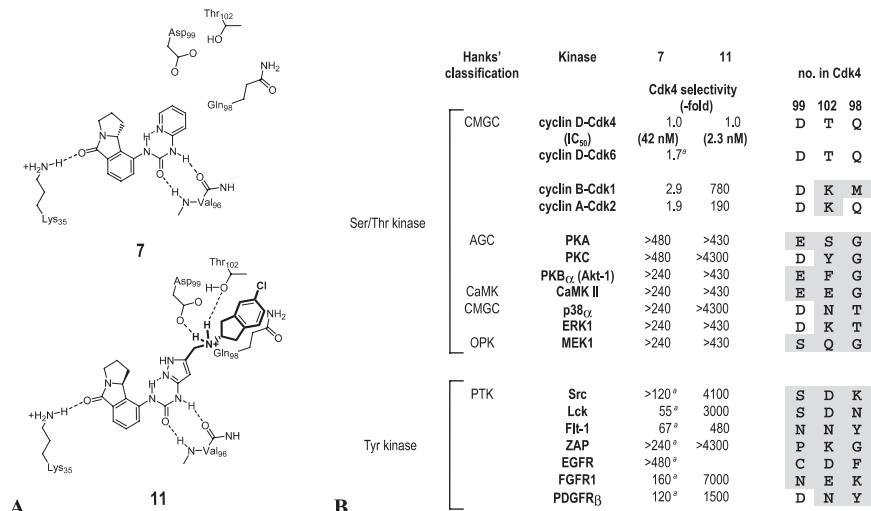


Fig. 6.13 Cdk4 selectivity of compounds **7** and **11** [20]. (A) Schematic figures of binding modes of compounds **7** and **11**. (B) Cdk4 se-

lectivity of compounds **7** and **11** and sequence alignment (Asp99, Thr102 and Gln98).
(a) These data are based on racemic **7**.

Gln98 in the Cdk4 model. This contact between hydrophobic surfaces possibly contributes to the enhanced binding affinity of **11** to Cdk4. Interestingly, **11** showed higher selectivity over Cdk1 (780-fold) than over Cdk2 (190-fold), while **10** showed almost equal selectivity over both Cdk1 (130-fold) and Cdk2 (120-fold) [20]. As already shown in Fig. 6.10A and B, Gln98 in Cdk4 is replaced by Met in Cdk1 but is conserved in Cdk2. This difference between Cdk1 and Cdk2 may cause differences in the inhibitory activities of **11** between Cdk1 (0.44 μM) and Cdk2 (1.8 μM).

Finally, we investigated the Cdk4 selectivity of **11** over other representative kinases. Schematic figures of the binding modes of the key compounds **7** and **11**, their Cdk4 selectivity over the representative kinases and the sequence alignment (Asp99, Thr102 and Gln98) are summarized in Fig. 6.13A and B. The selectivity of **11** over Cdk1/2 was remarkably higher than that of compound **7**, while the selectivity of **11** over other Ser/Thr kinases and Tyr kinases remained at the same level or increased. The NH₂⁺ of **11**, which is assumed to interact with Asp99, would be critical for the higher Cdk4 selectivity not only over Cdk1/2 but also over other kinases. As shown in Fig. 6.13B, Asp99 is replaced by non-acidic residues in most tyrosine kinases. Thr102 and Gln98 are also frequently replaced by other residues in both Ser/Thr kinases and tyrosine kinases. The interactions between these three residues and **11** were assumed to contribute to the improvement in the selectivity for Cdk4 over Lck (3000-fold), Flt-1 (480-fold), FGFR1 (7000-fold) and PDGFR_β (1500-fold) compared with those of compound **7**.

6.7 Conclusion

In the generation of lead compounds, we employed a structure-based design consisting of the following steps: (1) homology modeling of the target protein, (2) scaffold generation with the de novo design program, LEGEND and SEEDS, (3) construction of informer libraries and (4) structural modification based on the predicted binding mode. From the results of these steps, we developed a novel diarylurea class of potent inhibitor, **7** ($IC_{50}=0.042\text{ }\mu\text{M}$). The combination of LEGEND and SEEDS was so effective that we were able to overcome the difficulty of utilizing output structures from the de novo design program and could identify some promising scaffolds including diarylurea. The construction of informer libraries, the prediction of the binding mode by molecular modeling, and modification based on the predicted binding mode efficiently led to potent lead compounds (in the order of $0.01\text{ }\mu\text{nM}$) from initial compounds (in the order of $10\text{--}100\text{ }\mu\text{M}$).

The identification of Cdk4-specific amino acid residues allowed us to focus on target residues that interact with substituents on lead compounds. Library design based on the locations of these residues and the binding modes of lead compounds enabled us to develop potent and selective compounds efficiently. Among them, compound **11** showed excellent selectivity not only over Cdk1/2 (780-fold/190-fold) but also over many other kinases ($>430\text{-fold}$) investigated thus far. Further studies have revealed that **11** causes specific G_1 arrest in the Rb(+) cancer cell line (MOLT-4) at concentrations of between 0.1 and $0.5\text{ }\mu\text{M}$.

The kinase superfamily consists of numerous kinases that have a common folding pattern. Therefore, it is generally difficult to improve the target kinase selectivity with respect to all the other kinases at the same time. However, protein kinases play various important roles in the human body, and they are also very attractive targets for drug discovery. We believe that this approach should be useful for the more systematic and efficient development of highly specific kinase inhibitors.

6.8 Acknowledgements

I would like to express my sincere thanks to Dr. Yoshikazu Iwasawa, Dr. Takashi Hayama, Dr. Kyoko Hayashi, Mr. Tetsuya Aoyama, Dr. Noriaki Hashimoto, Mr. Takanori Yoshizumi, Dr. Ikuko Suzuki-Takahashi, and other colleagues associated with the Cdk4 project. I also wish to thank Dr. Hiroshi Funabashi, Mr. Toshiharu Iwama, Ms. Chinatsu Ikeura, Dr. Tetsuya Takimura (of the modeling group) and Mari Ikuta (of the X-ray group) for their technical assistance and exciting discussions. I am grateful to Ms. Jocelyn Winward for her critical reading of this manuscript. I am indebted to Dr. Hajime Morishima, Dr. Susumu Nishimura, and Dr. Akio Murai for their useful suggestions and heartfelt encouragement.

6.9

References

- 1 PINES, J., *Trends Biochem. Sci.* **1993**, *18*, 195–197.
- 2 PINES, J., *Adv. Cancer Res.* **1995**, *55*, 181–212.
- 3 HUNTER, T., PINES, J., *Cell* **1994**, *79*, 573–583.
- 4 TAYA, Y., *Molecular Cells* **1995**, *5*, 191–195.
- 5 WEINBERG, R.A., *Cell* **1995**, *81*, 323–330.
- 6 KATO, J., MATSUSHIME, H., HIEBERT, S.W., EWEN, M.E., SHERR, C.J., *Genes Dev.* **1993**, *7*, 331–342.
- 7 BALDIN, V., LUKAS, J., MARCOTE, M.J., PAGANO, M., DRAETTA, G., *Genes Dev.* **1993**, *7*, 812–821.
- 8 PLANAS-SILVA, M.D., WEINBERG, R.A., *Curr. Opin. Cell Biol.* **1997**, *9*, 768–772.
- 9 KAMB, A., *Trends Genet.* **1995**, *11*, 136–140.
- 10 SHERR, C.J., *Trends Biochem. Sci.* **1995**, *20*, 187–190.
- 11 FRY, D.W., GARRET, M.D., *Curr. Opin. Oncol. Endocrine Metab. Invest. Drugs* **2000**, *2*, 40–59.
- 12 AKINAGA, S., SUGIYAMA, K., AKIYAMA, T., *Anti-Cancer Drug Des.* **2000**, *15*, 43–52.
- 13 SEDLACEK, H.H., CZECH, J., NAIK, R., KAUR, G., WORLAND, P., LOSIEWICZ, M., PARKER, B., CARLSON, B., SMITH, A., SENDEROWICZ, A., SAUSVILLE, E., *Int. J. Oncol.* **1996**, *9*, 1143–1168.
- 14 SENDEROWICZ, A.M., HEADLEE, D., STINSON, S.F., LUSH, R.M., KALIL, N., VILLALBA, L., HILL, K., STEINBERG, S.M., FIGG, W.D., TOMPKINS, A., ARBUCK, S.G., SAUSVILLE, E.A., *J. Clin. Oncol.* **1998**, *16*, 2986–2999.
- 15 GLAY, N.S., WODICKA, L., THUNNISSEN, A.-M., NORMAN, T.C., KWON, S., ESPINOZA, F.H., MORGAN, D.O., BARNES, G., LECLERC, S., MEIJER, L., KIM, S.-H., LOCKHART, D.J., SCHULTZ, P.G., *Science* **1998**, *281*, 533–538.
- 16 BARVIAN, M., BOSCHELI, D., COSSROW, J., DOBRUSIN, E., FATTAEY, A., FRITSCH, A., FRY, D., HARVEY, P., KELLER, P., GARRETT, M., LA, F., LEOPOLD, W., McNAMARA, D., QUIN, M., TRUMPP-KALLMEYER, S., TOOGOOD, P., WU, Z., ZHANG, E., *J. Med. Chem.* **2000**, *43*, 4606–4616.
- 17 NUGIEL, D.A., ETZKORN, A.-M., VIDWANS, A., BENFIELD, P.A., BOISCLAIR, M., BURTON, C.R., COX, S., CZERNIAK, P.M., DOLENIAK, D., SEITZ, S.P., *J. Med. Chem.* **2001**, *44*, 1334–1336.
- 18 HONMA, T., HAYASHI, K., AOYAMA, T., HASHIMOTO, N., MACHIDA, T., FUKASAWA, K., IWAMA, T., IKEURA, C., IKUTA, M., SUZUKI-TAKAHASHI, I., IWASAWA, Y., HAYAMA, T., NISHIMURA, S., MORISHIMA, H., *J. Med. Chem.* **2001**, *44*, 4615–4627.
- 19 IKUTA, M., KAMATA, K., FUKASAWA, K., HONMA, T., MACHIDA, T., HIRAI, H., SUZUKI-TAKAHASHI, I., HAYAMA, T., NISHIMURA, J., *Biol. Chem.* **2001**, *276*, 27548–27554.
- 20 HONMA, T., YOSHIZUMI, T., HASHIMOTO, N., HAYASHI, K., KAWANISHI, N., FUKASAWA, K., TAKAKI, T., IKEURA, C., IKUTA, M., SUZUKI-TAKAHASHI, I., HAYAMA, T., NISHIMURA, S., MORISHIMA, H., *J. Med. Chem.* **2001**, *44*, 4628–4640.
- 21 DE BONDT, H.L., ROSENBLATT, J., JANCA-RIK, J., JONES, H.D., MORGAN, D.O., KIM, S.-H., *Nature* **1993**, *363*, 595–602.
- 22 JEFFREY, P.D., RUSSO, A.A., POLYAK, K., GIBBS, E., HURWITZ, J., MASSAGUE, J., PAVLETICH, N.P., *Nature* **1995**, *376*, 313–320.
- 23 RUSSO, A.A., JEFFREY, P.D., PAVLETICH, N.P., *Nat. Struct. Biol.* **1996**, *3*, 696–700.
- 24 PAVLETICH, N.P., *J. Mol. Biol.* **1999**, *287*, 821–828.
- 25 RUSSO, A.A., TONG, L., LEE, J.-O., JEFFREY, P.D., PAVLETICH, N.P., *Nature* **1998**, *395*, 237–243.
- 26 JEFFREY, P.D., TONG, L., PAVLETICH, N.P., *Genes Dev.* **2000**, *14*, 3115–3125.
- 27 BROTHERTON, D.H., DHANARAJ, V., WICK, S., LEONARDO, B., DOMAILLE, P.J., VOLYANIK, E., XU, X., PARISINI, E., SMITH, B.O., ARCHHER, S.J., SERRANO, M., BRENNER, S.L., BLUNDELL, T.L., LAUE, E.D., *Nature* **1998**, *395*, 244–250.
- 28 SERRANO, M., HANNON, G.J., BEACH, D.A., *Nature* **1993**, *366*, 704–707.
- 29 http://www.nih.go.jp/mirror/Kinases/pkr/pk_catalytic/pk_hanks_seq_align_long.html

- 30** HANKS, S.K., QUINN, A.M., HUNTER, T., *Science* **1988**, *241*, 42–52.
- 31** HANKS, S.K., QUINN, A.M., *Methods Enzymol.* **1991**, *200*, 38–62.
- 32** HANKS, S.K., HUNTER, T., *FASEB J.* **1995**, *9*, 576–596.
- 33** BIOCES[E], developed by NEC Co., Tokyo, Japan.
- 34** BROOKS, B.R., BRUCCOLERI, R.E., OLAFSON, B.D., STATES, D.J., SWAMINATHAN, S., KARPLUS, M., *J. Comput. Chem.* **1983**, *4*, 187–217.
- 35** GOODFORD, P.J., *J. Med. Chem.* **1985**, *28*, 849–856.
- 36** MIRANKER, A., KARPLUS, M., *Proteins: Struct. Funct. Genet.* **1991**, *11*, 29–34.
- 37** DEL CARPIO, C.A., TAKAHASHI, Y., SASAKI, S., *J. Mol. Graph.* **1993**, *11*, 23–42.
- 38** HENDRICH, M., RIPPMA, F., BARNICKEL, G., *J. Mol. Graph.* **1997**, *15*, 359–363.
- 39** SiteID is distributed by TRIPPOS Associates, St. Louis, MO, USA.
- 40** Alpha Site Finder is distributed by Chemical Computing Group Inc., Montreal, Quebec, Canada.
- 41** DE AZEVEDO, W.F., JR., MUELLER-DIECKMANN, H.-J., SCHULZE-GAHMEN, U., WORLAND, P.J., SAUSVILLE, E., KIM, S.-H., *Proc. Natl. Acad. Sci.* **1996**, *93*, 2735–2740.
- 42** LAWRIE, A.M., NOBLE, M.E.M., TUNNAH, P., BROWN, N.R., JOHNSON, L.N., ENDICOTT, J.A., *Nat. Struct. Biol.* **1997**, *4*, 796–801.
- 43** DE AZEVEDO, W.F. Jr., LECLERC, S., MEIJER, L., HAVLICEK, L., STRNAD, M., KIM, S.-H., *Eur. J. Biochem.* **1997**, *243*, 518–526.
- 44** KLEBE, G., MARTIN, Y.C., eds., *Perspective in Drug Discovery and Design*, Vol. 20, KLUWER/ESCOM, **2000**.
- 45** BOEHM, H.-J., SCHNEIDER, G. (eds.), *Methods and Principles in Medicinal Chemistry*, Vol. 10, Wiley-VCH, Weinheim, **2000**.
- 46** MURCKO, M.A., An Introduction to de Novo Ligand Design in *Practical Application of Computer-Aided Drug Design* CHARIFSON, P.S. (ed.), Marcel Dekker Inc., New York, **1997**.
- 47** KUNTZ, I.D., BLANEY, J.M., OATLEY, S.J., LANGRIDGE, R., FERRIN, T., *J. Mol. Biol.* **1982**, *161*, 269–288.
- 48** EWING, T.J.A., MAKINO, S., SKILLMAN, A.G., KUNTZ, I.D., *J. Compt.-Aided. Mol. Design* **2001**, *15*, 411–428.
- 49** YAMADA-MIZUTANI, M., TOMIOKA, N., ITAI, A., *J. Mol. Biol.* **1994**, *243*, 310–326.
- 50** MORRIS, G.M., GOODSELL, D.S., HALLIDAY, R.S., HUEY, R., HART, W.E., BELEW, R.K., *J. Comput. Chem.* **1998**, *19*, 1639–1662.
- 51** MURRAY, C.W., CLARK, D.E., AUTON, T.R., FIRTH, M.A., LI, J., SYKES, R.A., WASKOWYCZ, B., WESTHEAD, D.R., YOUNG, S., C., *J. Compt.-Aided Mol. Design* **1996**, *11*, 193–207.
- 52** RAREY, M., KRAMER, B., LENGAUER, T., KLEBE, G., *J. Mol. Biol.* **1996**, *261*, 470–489.
- 53** LigandFit is distributed by Accelrys Inc., San Diego, CA, USA.
- 54** NISHIBATA, Y., ITAI, A., *Tetrahedron* **1991**, *47*, 8985–8990.
- 55** NISHIBATA, Y., ITAI, A., *J. Med. Chem.* **1993**, *36*, 2921–2928.
- 56** BOEHM, H.J., *J. Comput.-Aided Mol. Design* **1992**, *6*, 61–78.
- 57** BOEHM, H.J., *J. Comput.-Aided Mol. Design* **1992**, *6*, 593–606.
- 58** GILLET, V.J., JOHNSON, A.P., MATA, P., SIKE, S., WILLIAMS, P., *J. Comput.-Aided Mol. Design* **1993**, *7*, 127–153.
- 59** EISEN, M.B., WILEY, D.C., KARPLUS, M., HUBBARD, R.E., *Proteins Struct. Funct. Genet.* **1994**, *19*, 199–221.
- 60** CLARK, D.E., FRENKEL, D., LEVY, S.A., LI, J., MURRAY, C.W., ROBSON, B., WASZKOWYCZ, B., WESTHEAD, D.R., *J. Comput.-Aided Mol. Design* **1995**, *9*, 13–32.
- 61** PEARLMAN, D.A., MURCKO, M.A., *J. Med. Chem.* **1996**, *39*, 1651–1663.
- 62** LeapFrog is distributed by TRIPPOS Associates, St. Louis, MO, USA.
- 63** WYSS, D.F., MCCOY, M.A., SENIOR, M.M., *Curr. Opin. Drug Discov. Develop.* **2002**, *5*, 630–647.
- 64** CARR, R., JHOTI, H., *Drug Discov. Today* **2002**, *7*, 522–527.
- 65** SWAYZE, E.E., JEFFERSON, E.A., SANNELOWERY, K.A., BLYN, L.B., RISEN, L.M., ARAKAWA, S., OSGOOD, S.A., HOFSTADLER, S.A., GRIFFEY, R.H., *J. Med. Chem.* **2002**, *45*, 3816–3819.

- 66** GILLET, V.J., MYATT, G., ZSOLDOS, Z., JOHNSON, A.P., *Perspect. Drug Discov. Design* **1995**, *3*, 34–50.
- 67** IWAMA, T., HONMA, T., IKEURA, C., FUNABASHI, H., MORISHIMA H., unpublished results.
- 68** Database searches were carried out using ISIS/base. In ISIS/base, the author can assign the atom list, ring bond count and substitution count to each atom, and also assign the bond list to each bond when making queries for searching compound databases. In addition, the reaction center can be assigned to each bond when making queries for searching reaction databases. The atom list and bond list indicate atomic species and bond orders that are allowed. The ring bond count shows numbers of additional rings that are allowed. The substitution count indicates numbers of additional substituents that are allowed. The reaction center is the changing pattern of the bond between the reactant and the product on searching reaction databases.
- 69** The available chemicals directory is distributed by MDL Information Systems, San Leandro CA, USA.
- 70** CHEN, J.M., NELSON, F.C., LEVIN, J.I., MOBILIO, D., MOY, F.J., NILAKANTAN, R., ZASK, A., POWERS, R., *J. Am. Chem. Soc.* **2000**, *122*, 9648–9654.
- 71** IVERSEN, L.F., ANDERSEN, H.S., BRANNER, S., MORTENSEN, S.B., PETERS, G.H., NORRIS, K., OLSEN, O.H., JEPPESEN, C.B., LUNDT, B.F., RIPKA, W., MOLLER, K.B., MOLLER, N.P.H., *J. Biol. Chem.* **2000**, *275*, 10300–10307.
- 72** BAYLY, C.I., BLACK, W.C., LEGER, S., OUMET, N., OUELLET, M., PERCIVAL, M.D., *Bioorg. Med. Chem. Lett.* **1999**, *9*, 307–312.
- 73** KURUMBAIL, R.G., STEVENS, A.M., GIERSE, J.K., McDONALD, J.J., STEGEMAN, R.A., PAK, J.Y., GILDEHAUS, D., MIYASHIRO, J.M., PENNING, T.D., SEIBERT, K., ISAKSON, P.C., STALLINGS, W.C., *Nature* **1996**, *384*, 644–648.
- 74** MEDEMA, R.H., HERRERA, R.E., LAM, F., WEINBERG, R.A., *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 6289–6293.

7**Crystallization and Analysis of Serine Proteases with Ecotin**

SANDRA M. WAUGH and ROBERT J. FLETTERICK

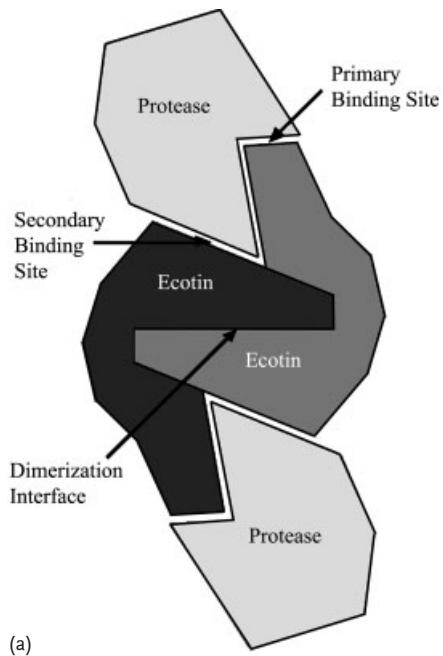
Abstract

Ecotin is a macromolecular inhibitor of serine proteases. It is remarkable in binding and blocking activity of virtually all serine proteases with the canonical Asp-His-Ser catalytic triad, regardless of the amino acid sequence or substrate specificity. It forms a heterotetramer that inhibits the protease by presenting a substrate-like, 11 amino acid loop to the active site. Since ecotin mimics a substrate, it defines the amino acids of the protease that are important for substrate binding and potentially for drug design. The ecotin-protease complex, itself, is well suited for crystal growth and three-dimensional structure determination. The 13 currently available three-dimensional structures of ecotin molecules are reviewed here, along with the protocols for expression, purification and formation of an inhibited tetramer. An analysis of ecotin and protease contacts suggests how the ecotin-defined active site cleft can be used for drug design.

7.1**What is Ecotin?**

Ecotin (eco) is a potent inhibitor of serine proteases that is derived from *Escherichia coli*. It was originally named for its ability to inhibit trypsin (*E. coli* trypsin inhibitor), but it is known to interact with and inhibit virtually all characterized trypsin-fold serine proteases. It is insensitive to the active site P1 preference of the protease (the amino acid N-terminal to the cleaved or scissile bond¹⁾) and inhibits proteases with specificity towards basic, large hydrophobic, small aliphatic and acidic amino acids [2]. This remarkable breadth of inhibition classifies eco as a fold-specific inhibitor. It forms a unique tetrameric complex consisting of two protease molecules and two inhibitor molecules (the E₂P₂ complex), binding in a bidentate manner with two surface loop regions known as the primary and secondary sites (3) (Fig. 7.1). Eco itself is a 142 amino acid protein that forms a stable

¹⁾ Serine protease substrate recognition sites are labeled according to the method of Schechter and Berger [1].



(a)

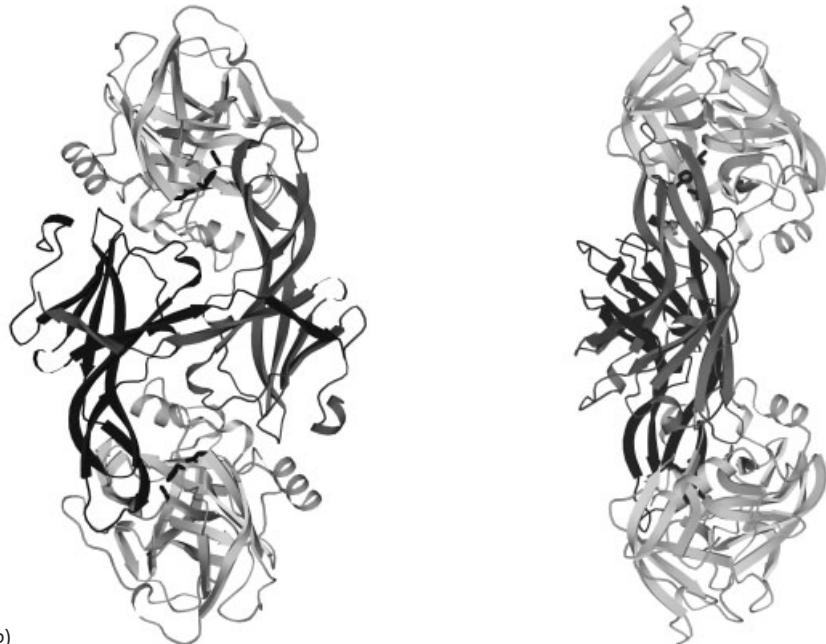


Fig. 7.1 The tetramer of eco bound to a serine protease. Visualized as a cartoon of the canonical protease and eco interaction (a), and (b), as two views of the three dimensional solution of D102N trypsin in complex with eco [3]. Each eco molecule has three protein–protein interaction surfaces. The C-terminus forms an anti-parallel β ribbon to complete the ecotin dimer interface. The 80's and 50's loops form the primary binding site by interacting with the protease at the active site cleft in a substrate-like β -sheet conformation. The 60's and 100's loops of eco form the secondary binding site by interacting with the C-terminal α -helix of the protease. Note that each eco molecule contacts both of the protease molecules. Two eco molecules (black and medium grey) form a pair of interactions each with two protease molecules (light grey). The catalytic triad residues Ser-195, Asp-102 and His-57 are in black ball and stick representation. This figure was made with Molscript [37] and Raster 3D [38].

17 kDa dimer held by an arm-in-arm anti-parallel β ribbon interaction between the two C-termini. The extensive dimer interface characterizes eco as a domain swapped dimer [4]. The three-dimensional structure of eco is composed entirely of β strands and has been described as a modified jellyroll structure [3]. Separate segments of the eco chain form two spatially independent binding sites to bind to two protease molecules [3, 5].

The primary site binds in a manner that mimics a protease substrate, forming between three and eight main chain hydrogen bonds in a β -sheet conformation along the active site cleft. This substrate-like binding defines the substrate binding cleft in seven serine proteases to date. The 80's loop of eco makes an extended interaction with the protease while the 50's loop of eco has two functions. It stabilizes the 80's loop through the formation of hydrogen bonds and a disulfide bridge between Cys-51 and Cys-86, and interacts directly with the protease. Eco is a potent inhibitor due to the primary site loops, but the method of the inhibition is still undefined. It is proposed that the protease is inhibited by excluding water from the active site and preventing deacylation of the acyl-enzyme intermediate, the second step in hydrolysis [6]. The eco 80's loop C-terminal to the cleaved bond may also stay strongly associated with the active site and increase the reverse, ligating reaction [7].

The secondary site of eco binds to the protease over 20 Å away from the active site and forms up to 30 van der Waals interactions and up to five additional hydrogen bonds. An additional important source of binding energy and association, the secondary site is composed of the 60's and 100's loops of eco and a hydrophobic patch near the protease residues 91 to 94 and the C-terminal α helix amino acids 236 to 242. This patch and helix separated from the 80's loop accounts for the fold specificity of eco. Each inhibitor molecule forms an interaction with both proteases of the tetramer in a clamp configuration that can be adapted to fit most serine proteases.

Because of the large number of proteases it inhibits, WT (wild type) eco has become useful as a biochemical tool. As a selective inhibitor of serine proteases, eco has been used in cell culture assays to probe enzymatic activity as well as to titrate the percentage of catalytically active molecules of a previously uncharacterized serine protease [8]. Immobilized eco has been used for the direct purification of trypsinogen [9]. Finally, eco selectively identified a novel protease implicated in prostate cancer, membrane type serine protease I (MT-SP1) [10].

7.2 Methods

7.2.1 Expression of Wild Type Ecotin

The following protocol for the expression and purification of WT eco and many of its variants is a compilation of current strategies from the Craik, Fletterick and La-

zarus groups [4, 11, 12]. Eco is expressed in the bacterial strains JM101 or BL21, or an *E. coli* ecotin knockout strain, IMΔecoJ, generated in the Craik Lab (Ian Murray, unpublished results). Eco variants are generated in the pBS vector using the Kunkle mutagenesis protocol [13], cut with *Bam*HI and *Hind*III (New England BioLabs), and ligated into the *Bam*HI/*Hind*III pTacTacEcotin plasmid. IMΔecoJ cells are transformed with the pTacTacEcotin plasmid and selected for growth on LB/Ampicillin plates. A single colony is grown overnight in LB buffer plus 60 µg mL⁻¹ ampicillin, and then diluted into 1 L of the same buffer. The culture is grown with shaking at 37°C for 1 h when eco expression is induced by the addition of IPTG to 200 mM. Protein expression continues overnight and up to 12 h.

7.2.2

Purification of Wild Type Ecotin

The cells are harvested by centrifugation at 6000 g for 30 min, and the supernatant discarded. Osmotic shock is used to selectively lyse the cells and free the periplasm, by resuspending the cell pellet in 10 mM Tris pH 8.0, 25% sucrose plus 15 mg mL⁻¹ lysozyme (Sigma) and incubating with shaking for 1 h. The supernatant is retained for the next step where, due to the incredible stability of eco, it can be separated from many of the remaining proteins by acidification. The solution is dropped to pH 2.8 by the gradual addition of 1 M HCl, incubated for 30 min and centrifuged. The acid-soluble supernatant is returned to neutral pH by the addition of 1 M Tris pH 8.0 and followed by the addition of NaCl to 0.3 M. Neutralization is followed by immersion in boiling water for 10 min. After the appearance of visible white protein precipitate, the supernatant is isolated by centrifugation and dialyzed into distilled water. The resulting eco solution is concentrated and purified to homogeneity by reversed phase HPLC chromatography using a C4 peptide and protein column (Vydac, 5 µm, 4.6×250 mm) with a 20–60% gradient of 0.1% trifluoroacetic acid and 0.08% trifluoroacetic acid, 95% acetonitrile. Eco generally elutes at a solution content of 35% acetonitrile. The now pure eco is lyophilized and resuspended in distilled water, or a storage buffer composed of 50 mM Tris pH 8.0, 10 mM NaCl. The eco concentration is verified by absorbance at 280 nm and SDS-PAGE. Eco has a calculated extinction coefficient at 280 nm of 21860 M⁻¹ cm⁻¹, and runs as a monomer of 16 kDa on SDS-PAGE (Fig. 7.2). Typical yields for a preparation of wild type eco range from 80 to 150 mg L⁻¹. Measurement of eco by MALDI-MS confirms that the molecular mass is 16090 kDa.

Several modifications of the above protocol have been described for eco purification. These are very useful for the purification of eco variants as they consist of a gentler series of purification steps. Following the acidification step, the protein is dialyzed into distilled water and subjected to a 65% (w/v) ammonium sulfate precipitation. The insoluble fraction is resuspended in distilled water, dialyzed into 10 mM Tris pH 8.0, and concentrated. The resulting protein is purified to homogeneity by FPLC chromatography on a Mono-Q anion exchange column (Pharmacia) with a slow gradient of 0–15% NaCl.

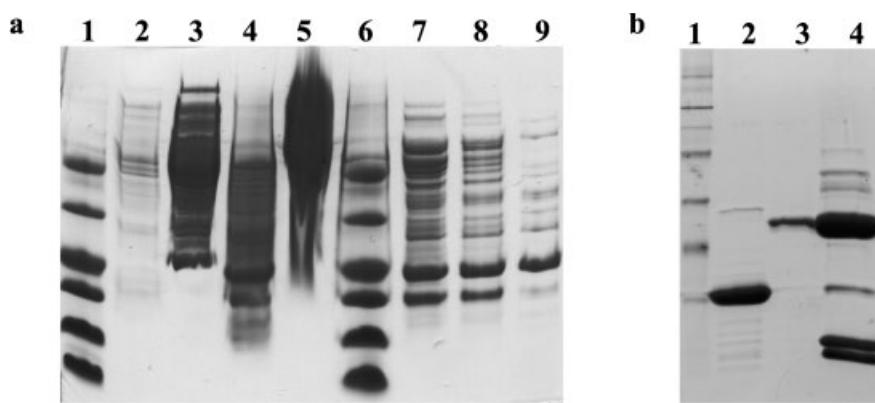


Fig. 7.2 Wild type eco purification followed by SDS-PAGE. (a) Consecutive steps of eco purification. Column 1, molecular weight marker from the bottom: 3, 5.6, 14.8, 20.1, 29.3, 44.7 kDa; column 2, pre-induction; column 3, periplasmic fraction; column 4, supernatant following acidification; column 5, insoluble pellet following acidification step; column 6, molecular weight marker; column 7, neutralized supernatant; column 8, soluble fraction after boiling; column 9, purified eco before the column chromatography step. (b) Complex formation with purified eco. Column 1, molecular weight marker from the bottom: 6.5, 16, 25, 32.5, 47.5, 62, 83, and 175 kDa; column 2, purified eco [81–84 IEPD]; column 3, purified GrB [N66Q]; column 4, purified complex of inhibitor and protease.

7.2.3

Crystallization of Ecotin and Protease Complexes

Eco forms trimers of two inhibitors to one protease molecule when excess eco is present [4]. To form the tetramer preferentially, mix the ecotin and protease in a 0.75:1 ratio. An additional purification step should be performed to remove any excess protease. The complex can be purified to homogeneity by a repetition of the final protease column purification or by FPLC chromatography on a Superdex 75 gel filtration column [4] (Pharmacia).

7.3

Representative Examples of Ecotin and Protease Structures

Eco is a powerful tool for defining the active sites of serine protease due to the extended substrate-like interaction that it makes with the protease. The three-dimensional structure of a complex with eco has many advantages that a structure of a protease alone or bound to a small molecule inhibitor does not have. Eco can be used to take a molecular impression of the serine protease active site and reveal features that determine substrate preference. These features are used to design specific inhibitors with therapeutic prospects. Often, a small molecule inhibitor is used to define a protease active site cleft, but the resulting structures have particular drawbacks. Typically, a small molecule inhibitor lacks the prime side interac-

tions within the substrate-binding cleft and does not extend to greater than the S4 pocket on the non-prime side. In the following paragraphs, we will briefly review three structures to highlight how eco defines the extended substrate binding of a serine protease. The fiddler crab collagenase structure defines the limits of the extended substrate binding [14], the rat granzyme B (grB) structure highlights an unusually accurate substrate-like interaction [15], and the factor Xa structure defines potential inhibitor pockets and visualizes a previously unstructured domain [15 a].

7.3.1

Ecotin Defines the S7 Through S4' Subsites of Collagenase

Fiddler crab collagenase is a unique collagenase of the chymotrypsin fold. While it is not a potential therapeutic target, the structure with eco defines the largest interaction surface between eco and a protease to date, and facilitates modeling of a collagen substrate into the active site. Collagenase cleaves collagen into the signature one-fourth and three-fourths-length fragments [16], hydrolyzes tetrapeptide substrates containing a variety of P1 amino acids including Tyr, His, Leu, Arg and Lys [17, 18], and is inhibited by WT eco with an inhibition constant of 510 ± 60 pM. It was hypothesized that the specificity determinants are very broad but selective for extended amino acids without a Pro or hydroxy-Pro. This selectivity would isolate portions of the collagen that have lower melting energies making them more suitable for proteolysis. The three-dimensional structure of collagenase and WT eco defines the largest primary site interaction to date between a serine protease and an eco molecule by filling the S7 to S4' subsites of the enzyme and burying 1950 \AA^2 of surface area [14]. Eco forms three additional main chain hydrogen bonds with collagenase at the P6 and P7 amino acids that are not found in the trypsin complex. At the secondary site, eco buries 1060 \AA^2 and makes two additional hydrogen bonds between the Glu-92 of the protease and the 60's loop of eco.

The additional interaction at the collagenase extended sites is offset by the loss of three hydrogen bonds in the core region of ecotin. These probably contribute to the flexibility needed to fit the 80's loop so deeply into the protease binding cleft. In this complex structure, the eco molecule is rotated several degrees compared with the trypsin complex. Eco monomers generally rotate and translate up to 5 \AA relative to each other between solved complexes [19]. This is attributed to a hinge motion around the dimer interface of eco, as if eco were a clamp that adjusts to each of the protease targets presented to it. The secondary site C-terminal helix of collagenase is almost 1.5 \AA from the comparable trypsin helix. Adjustments in the eco molecule to accommodate this movement are the reason for the rotation and translation of the dimer and the loss of three stabilizing hydrogen bonds in the eco protein core.

The structure of collagenase and eco highlights an important aspect of the eco-protease interaction—flexibility. The eco dimer is extremely accommodating of the differences between serine protease molecules because it binds at two relatively invariant regions, the substrate binding cleft and the C-terminal α helix–90's loop.

These interactions are found throughout the family of serine proteases and represent side chain-insensitive substrate binding surfaces.

7.3.2

Ecotin as a Tight Binding Substrate

The structure of the complex of rat grB [N66Q] and eco [81–84 IEPD] [15] shows the mechanism of aspartic acid recognition, a unique specificity in serine proteases. GrB is a serine protease with an important initiating role in apoptosis. It is packaged in the granules of cytotoxic lymphocytes with a family of serine protease named granzymes, the putative pore forming protein perforin and proteoglycans [20]. In these secretory granules, the proteases are activated by the protease dipeptidyl protease II [21, 22]. Upon detection of a target cell, the granules are released directionally and enter the target cell with the aid of perforin. In the target cell, grB is responsible for the activating cleavage of caspases, and the direct proteolysis of substrates important for the apoptotic cascade [20]. GrB has also been implicated in the generation of autoimmunogenic peptides [23] and is found active and at high levels in the synovial fluid of patients with rheumatoid arthritis [24, 25] making it a potential therapeutic target. Despite a strict requirement for acidic residues at P1, grB is still inhibited by eco with an inhibition constant of 2 ± 5 nM [8].

To generate an inhibitor mirroring the substrate specificity of grB for crystallization, a series of eco variants were engineered by sequentially mutating the 81 through 86 loop positions. Mutations at a single site, eco [M84D], encompassing the P4–P1 sites, eco [81–84 IEPD], and the P4–P2' sites, eco [81–86 IEPDWG], were constructed. Inhibition constants were calculated according to tight binding inhibitor theory [26], and are all in the 1 to 10 nM range (Fig. 7.2). When visualized by SDS-PAGE, two distinct degradation bands appear at approximately 8 and 9 kDa for the variant eco complexes. These fragments are the expected size if cleavage by grB occurs at the active site 80's loop (Fig. 7.3). Although the inhibitor is cleaved, proteolytic activity measurements confirm the inhibition. All of the eco variants in complex with grB-formed crystals, and eco [81–84 IEPD] produced crystals of appropriate resolution for structure determination. Eco [81–84 IEPD] was constructed to reflect the specificity at the P1 to P4 sites of grB. The wild type eco P1'-Met and P2'-Ala already fit the grB substrate profile. Previous alterations to eco that altered the primary binding site loop to a substrate for the protease have resulted in an inhibition incompetent eco. It is possible that eco forms a transient covalent complex and that the 50's loop interaction prevents the attack of water that relieves the complex. The secondary site 60's and 100's loops may also contribute energy to the stabilization of the tetramer.

The structure of grB and eco illustrates the amino acids in the active site of grB that make contact with an ideal substrate, and provide a template for use in computational drug design efforts. A set of ten amino acids found along the binding site cleft of grB includes sites identified through mutational studies on other proteases, such as structural positions 192, 226, 189, 99 and 215, and sites not previously considered a part of the extended binding site region including 171, 174

| Ecotin | Inhibition, K_i (nM) |
|---------------|--|
| WT | 1±3 |
| M84D | 1±5 |
| 81–84 IEPD | 5±2 |
| 81–86 IEPDWG | 11±4 |

(a)

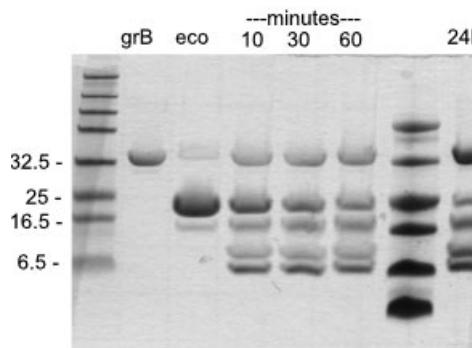


Fig. 7.3 SDS-PAGE analysis of the substrate-like features of eco. (a) Inhibition constants for a panel of ecotin mutants where the ecotin 80's loop is gradually altered to match the ideal grB substrate specificity, and (b) SDS-PAGE analysis of the [81–86 IEPDWG] eco when reacted with WT grB. The eco band is cut into fragments of 8 and 9 kDa corresponding to cleavage at the D/W bond of eco. The E₂P₂ complex is further purified before crystallization.

and 218. These structural determinants are now under investigation to understand their role in the unique extended specificity of grB (S. M. Waugh, R. J. Fletterick, and C. S. Craik submitted).

7.3.3

Ecotin Defines Regions Distal to the Factor Xa Protease Domain

Eco is a potent inhibitor of many members of the blood coagulation cascade including plasma kallikrein, factor XIIa [27], factor IXa and factor Xa [11]. Factor Xa acts as the adaptor of clotting signals from the extrinsic and intrinsic pathways, amplifying the signal by binding to factor Va and proteolytically activating thrombin. Factor Xa is composed of a serine protease heavy chain and a disulfide-linked light chain composed of a pair of epidermal growth factor (EGF) repeat regions and a γ-carboxy-glutamic acid-rich domain (Gla domain). The protease domain is activated by the factor IXa–factor VIII complex or the tissue factor–factor VIIa complex. Factor Xa has been the focus of many drug design programs for this important role in the generation and maintenance of blood clots. The first crystal structure of factor Xa was solved without a peptide-derived small molecule inhibitor bound in the active site [28]. The uninhibited protease is a good starting point for drug design, but missing possible substrate interactions. WT eco is the most potent macromolecular inhibitor of factor Xa to date, and the M84R mutation lowers the inhibition constant from 54 pM to 11 pM. Recently, the X-ray crystallographic structure of factor Xa bound to M84R eco has been solved. Not only does the ecotin primary site interaction define the specificity determinants of the pro-

tease-binding site from P6 to P2', but more of the light chain is ordered than in any other solved structure [15a].

At the primary interaction site, M84Reco defines the amino acids that contact the substrate-like 80's loop. A hydrophobic P4 interaction occurs between Trp-215, Tyr-99 and Phe-174 of the protease with Val-84 of eco. Gln-192 primarily mediates the P3 interaction. Previous structures of factor Xa bound to small molecule inhibitors also implicated these amino acids [29, 30]. The unusual preference at the P2 substrate position for Gly and Tyr-Trp-Phe is clarified in this structure. Tyr-99 rotates over 120° from the previous position, opening up the P2 pocket to the eco residue Thr-83, and closing down the S4 pocket. This conformational change at Tyr-99 explains the factor Xa preference for large hydrophobic P2 amino acids that was so puzzling in the initial structures.

The structure with eco also visualizes the Gla domain in a previously unseen conformation. One EGF domain is not visible in the structure, so the Gla domain seems to float in space distant from the protease (Fig. 7.4). Mg²⁺ ions present in the crystallization buffer bind to the Gla domain in an intermediate conformation, filling the three high affinity and low specificity sites. This creates a domain that is primed for formation of the omega membrane-binding loop with the addition of Ca²⁺ to the remaining low affinity sites.

In this structure, eco defines an active site binding surface of over 1900 Å², and explains the unusual kinetic results at the P2 substrate position. It also orders a

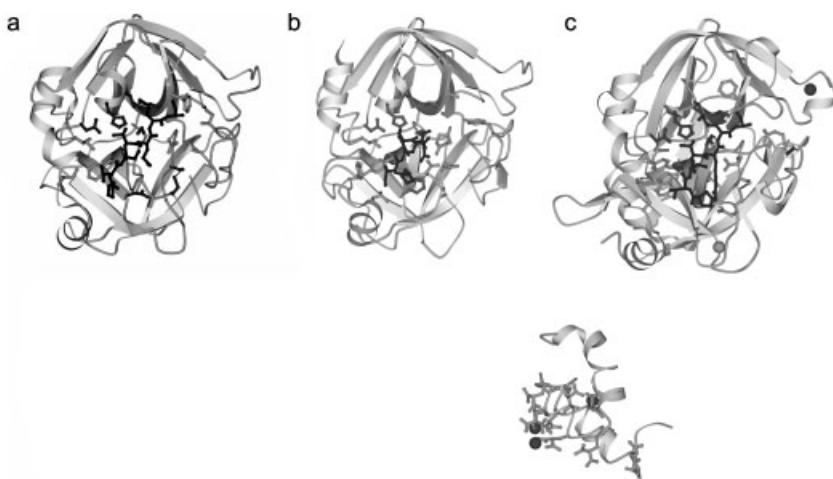


Fig. 7.4 The three-dimensional fold of three serine proteases. (a) Fiddler crab collagenase defined in complex with WT eco, (b) rat granzyme B in complex with [81–84 IEPD] eco, and (c) human factor Xa in complex with M84R eco. Each protease is shown in grey. The catalytic triad and amino acids contacting the ecotin molecule in the active site are shown in grey ball and stick representation. The ecotin mol-

ecule representing the substrate positions P4 through P2' is shown in black ball and stick representation. The EGF and Gla domain of factor Xa are shown in medium grey. Note that the EGF domain is hidden behind the protease domain. The magnesium ions are colored dark grey and the calcium ion is colored light grey. This figure was made with Molscript [37] and Raster 3D [38].

portion of the protease not previously visible to crystallography, the Gla domain, and visualizes it in a membrane-bound conformation. The new, eco bound structure reveals important portions of the molecule that were poorly understood, despite the presence of multiple factor Xa structures in the literature.

7.4

Crystallization and Structure Determination of E₂P₂ Complexes

Eco is a remarkable inhibitor. It binds to and inhibits proteases with a broad range of pI values from 3.92 (collagenase) to 9.5 (grB) forming an E₂P₂ complex that typically buries between 1900 and 2500 Å². This large and strongly associated complex is able to form crystals at a range of pH values from pH 5.0 to pH 8.8 (Tab. 7.1). To bind serine proteases efficiently with many different substrate prefer-

Tab. 7.1a Structural characteristics of solved eco structures. (a) Eco structures with no protease bound, or in complex with rat trypsin

| Complex Eco | Protease | Crystallization conditions | Asymmetric unit contents | Space group |
|--------------------------------|----------------------------------|---|-------------------------------|--|
| WT 1ecy.pdb | None | 50 mM Tris pH 8.5 20–36% PEG 1500 0.1% (w/v) β-octyl glucoside | E ₂ | P 2 ₁ 2 ₁ 2 ₁ |
| WT 1ecz.pdb | None | 50 mM Tris pH 8.5 20–36% PEG 1500 0.1% (w/v) trehalose | E | P 4 ₃ 2 ₁ 2 |
| W130-[ADG] 1ifg.pdb | None | 100 mM Tris pH 8.0 200 mM acetate 28–30% PEG 4000 15% glycerol | E | P 4 ₃ 2 ₁ 2 |
| WT [3] | Rat trypsin [D102N] | 150 mM Na cacodylate pH 6.5 300 mM Na acetate 16.5% PEG 4000 | EP | P 2 ₁ |
| A86H 1slu.pdb ^{a)} | Rat trypsin [N143H, E151H] | 100 mM Tris pH 8.0 300 mM Na acetate 18–22% PEG 4000 | EP | C1 2 1 |
| Y69F, D70P 1ezu.pdb | Rat trypsin [D102N] | 150 mM Na cacodylate pH 6.2 300 mM Na acetate 14% PEG 4000 | E ₂ P ₂ | P 2 ₁ |
| M84R, 60's loop 1ezs.pdb | Rat trypsin | 150 mM Na cacodylate pH 6.2 300 mM Na acetate 14% PEG 4000 | E ₂ P ₂ | P 1 |

^{a)} Additional structural solutions with bound metal ions: Cu, 1slv.pdb, Ni, 1slw.pdb, and Zn, 1slx.pdb.

Tab. 7.1b Eco structures in complex with a serine protease. The following abbreviations are used for the complexes: E, an ecotin monomer; E_2 , an ecotin dimer; EP, an ecotin and protease dimer; E_2P_2 , a tetramer of two eco and two proteases

| Complex Eco | Protease | Crystallization conditions | Asymmetric unit contents | Space group |
|-------------------------|-----------------------------|---|--------------------------------|-----------------------------------|
| WT 1azz.pdb | Fiddler crab collagenase | 100 mM tricine pH 8.2 to 8.8 31–38% PEG 3350 | E_2P_2 | P 3 ₂ 2 1 |
| 81–84 IEPD 1 fi8.pdb | Rat granzyme B [N66Q] | 100 mM Na acetate pH 5.8 100 mM ammonium sulfate 25% PEGmme2000 | E_2P_2 | P 2 ₁ |
| M84R 1id5.pdb | Thrombin | 100 mM citric acid pH 5.0 10 mM sperminetetrahydrochloride 15% PEG 6000 | EP | C 2 2 2 ₁ |
| WT 1n80.pdb | Bovine chymo- trypsin | 100 mM Na acetate pH 4.0 1.3 M ammonium sulfate 1% PEG 200 | E_2P_2 | P 2 ₁ |
| WT ^{b)} | Factor Xa | 200 mM sodium tartrate, pH 7.1 20% PEG 3350 2% glycerol | EP | I 2 2 2 |
| WT ^{c)} C | Rat granzyme C | 100 mM Tris pH 8.0 125 mM Na acetate 26% PEG 4000 | EP | P 2 ₁ 2 ₁ 2 |

a) see Ref. [15 a]

c) unpublished data.

ences, eco must present a generic, yet acceptable substrate-like binding loop to all of the protease-active sites. Protease substrates and the eco 80's loop bind in an anti-parallel β -sheet conformation and form up to five main chain to main chain intermolecular hydrogen bonds [31, 32], but additional side chain interactions are necessary. Val, Thr and Met prime the eco 80's loop for β -sheet interactions with the protease because they have a high likelihood of occurring in β -sheet conformations [34], and are present at the protease positions P4, P2, P1 and P1'. At P4, a broad preference for hydrophobic amino acids occurs in positional scanning synthetic combinatorial substrate library results for thrombin, trypsin and chymotrypsin [18]. This suggests that ecoVal-81 is an important substrate recognition element, helping eco to fit many active sites. The two Met residues at the substrate P1 and P1' positions also contribute to a generic fit. They form van der Waals contacts within the S1 and S1' pockets that anchor eco without forming electrostatic interactions repulsive to another protease binding pocket.

WT eco can be engineered for greater selectivity and binding affinity. A single mutation, M84R, at the protease P1 position, increases the affinity of eco towards

Tab. 7.2 Inhibition constants for WT, M84R and M84R/M85R eco variants against a panel of serine proteases. Ecotin can become a highly selective and potent inhibitor of serine proteases through engineering with phage display [34] and molecular design [36].

| Protease | Ecotin | Inhibition, K_i (nM) | Ecotin | Inhibition, K_i (nM) |
|------------------|--------|---------------------------|-----------|---------------------------|
| Rat trypsin (34) | WT | 0.31 ± 0.06 | M84R | 0.38 ± 0.10 |
| Collagenase (14) | WT | 0.51 ± 0.06 | M84R | Not determined |
| uPA (34) | WT | 2800 ± 160 | M84R | 3.6 ± 0.6 |
| Thrombin (11) | WT | 3000 | M84R | 1.5 |
| Factor Xa (11) | WT | 0.054 ± 0.013 | M84R | 11 |
| MT-SP1 (10) | WT | 0.782 ± 0.092 | M84R/M85R | 0.0098 ± 0.0015 |

most trypsin-like proteases (Tab. 7.2). The blood coagulation protease thrombin has the greatest increase in affinity to the M84R eco mutation. WT eco is unable to inhibit the protease but M84R eco has an inhibition constant of 1.5 nM an increase of 2×10^4 times [12]. Using the method of phage display, eco variants can be generated that are extremely potent against such proteases as trypsin [34], urokinase type plasminogen activator (uPA) [35], MT-SP1 ([10], Factor XIIa and plasma kallikrein (Allart Stoop, personal communication). The mutations indicate that targeting both the active site substrate-binding regions and distant surface regions like the eco secondary binding site can discover selectivity [36].

There are ten structures of an eco molecule bound to a protease molecule available (Tab. 7.1). Five of the complexes contain the E₂P₂ tetramer in the unit cell and five occur as an EP dimer related across a crystallographic axis to form the tetramer. Of the nine complexes solved, seven crystal lattices are represented. Each complex forms a unique set of crystal contacts upon formation of a stable, diffraction quality crystal. Inspection of the protein structures indicates that less than 30% of the crystal contacts are between two eco molecules. Instead, contacts between the molecules in the unit cell are between loops of the protease, and they are distributed across the entire protease surface. No region of every protease always forms a crystal contact, but the 60's loop, 30's loop and Pro-222 occur in many of the complexes (Fig. 7.5). Crystal contacts range in size from a few hundred Å², to over 1000 Å² indicating that the E₂P₂ tetramer forms diffraction quality crystals in many crystal lattice conformations. The E₂P₂ associations and configurations change depending on the protease, but the solved structures with ecotin provide a range of models for molecular replacement methods. Most structures have been solved with an experimentally determined EP complex or an EP model made from the wild type protease and an ecotin molecule.

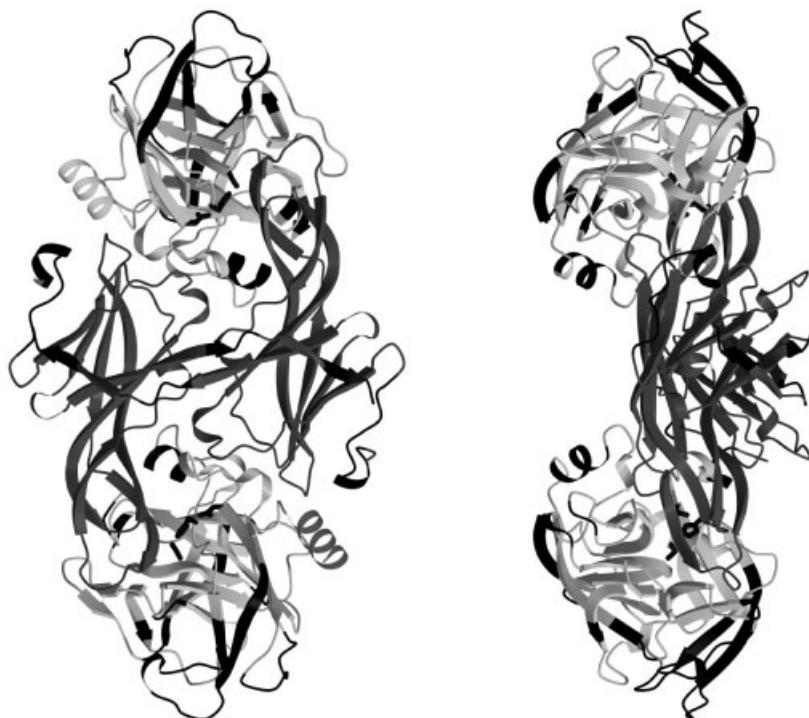


Fig. 7.5 The distribution of all E₂P₂ complex crystal contacts visualized on the E₂P₂ tetramer. The ecoin dimer is colored medium grey, the serine proteases are colored light grey and the crystal contact loops are colored black. The catalytic triad of the protease is co-

lored black in ball and stick representation. Note that the contacts occur all over the available surface of the protease and less so on the eco molecule. This figure was made with Molscript [37] and Raster 3D [38].

7.5 Conclusion

For two reasons, eco is useful as a tool for drug design. Firstly, eco forms a primary site interaction that defines the amino acids in contact with the P4 to P3' amino acids of a substrate. An active site defined in this manner reflects a physiological form of the binding cleft that is much larger than one defined with a small molecule inhibitor and extends to the prime side determinants. Secondly, eco forms a stable complex with trypsin-fold serine proteases that can form diffraction quality crystals. The variety and distribution of crystal contacts over the surface of the E₂P₂ complex suggests it is stable in a larger range of pH, salt and precipitating agents than the protease alone. Therefore, eco facilitates crystal growth through the formation of a complex that samples a large number of crystallization conditions and increases the likelihood of productive crystal associations over the protease alone. In the future, eco will continue to be important in

generating X-ray crystallographic solutions of trypsin-fold serine proteases. It may also have an important role in defining protease–protein interactions that occur distal to the active site of the protein. Eco, with its unique ability to form and crystallize stable complexes in a manner that maps the determinants of enzyme specificity, should be considered for the structure determination of any serine protease.

7.6

References

- 1 I. SCHECHTER, A. BERGER, On the active site of proteases. 3. Mapping the active site of papain; specific peptide inhibitors of papain. *Biochem. Biophys. Res. Commun.* **1968**, *32*, 898–902.
- 2 C. H. CHUNG, H. E. IVES, S. ALMEDA, A. L. GOLDBERG, Purification from *Escherichia coli* of a periplasmic protein that is a potent inhibitor of pancreatic proteases. *J. Biol. Chem.* **1983**, *258*, 11032–11038.
- 3 M. E. MCGRATH, T. ERPEL, C. BYSTROFF, R. J. FLETTERICK, Macromolecular chelation as an improved mechanism of protease inhibition: structure of the ecotin-trypsin complex. *Embo J.* **1994**, *13*, 1502–1507.
- 4 C. T. EGGERS, S. X. WANG, R. J. FLETTERICK, C. S. CRAIK, The role of ecotin dimerization in protease inhibition. *J. Mol. Biol.* **2001**, *308*, 975–991.
- 5 D. H. SHIN, H. K. SONG, I. S. SEONG, C. S. LEE, C. H. CHUNG, S. W. SUH, Crystal structure analyses of uncomplexed ecotin in two crystal forms: implications for its function and stability. *Protein Sci.* **1996**, *5*, 2236–2247.
- 6 I. S. SEONG, H. R. LEE, J. H. SEOL, S. K. PARK, C. S. LEE, S. W. SUH, Y. M. HONG, M. S. KANG, D. B. HA, C. H. CHUNG, The P-1 Reactive Site Methionine Residue of Ecotin Is Not Crucial for Its Specificity on Target Proteases – a Potent Inhibitor of Pancreatic Serine Proteases from *Escherichia Coli*. *J. Biol. Chem.* **1994**, *269*, 21915–21918.
- 7 E. S. RADISKY, D. E. KOSHLAND, Jr., A clogged gutter mechanism for protease inhibitors. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 10316–10321.
- 8 J. L. HARRIS, E. P. PETERSON, D. HUDIG, N. A. THORNBERRY, C. S. CRAIK, Definition and redesign of the extended substrate specificity of granzyme B. *J. Biol. Chem.* **1998**, *273*, 27364–27373.
- 9 Z. LENGYEL, G. PAL, M. SAHINTOTH, Affinity purification of recombinant trypsinogen using immobilized ecotin. *Protein Express. Purif.* **1998**, *12*, 291–294.
- 10 T. TAKEUCHI, M. A. SHUMAN, C. S. CRAIK, Reverse biochemistry: use of macromolecular protease inhibitors to dissect complex biological processes and identify a membrane-type serine protease in epithelial cancer and normal tissue. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 11054–11061.
- 11 J. L. SEYMOUR, R. N. LINDQUIST, M. S. DENNIS, B. MOFFAT, D. YANSURA, D. REILLY, M. E. WESSINGER, R. A. LAZARUS, Ecotin is a potent anticoagulant and reversible tight-binding inhibitor of factor Xa. *Biochemistry* **1994**, *33*, 3949–3958.
- 12 S. Q. YANG, C. I. WANG, S. A. GILLMOR, R. J. FLETTERICK, C. S. CRAIK, Ecotin: a serine protease inhibitor with two distinct and interacting binding sites. *J. Mol. Biol.* **1998**, *279*, 945–957.
- 13 T. A. KUNKEL, Rapid and efficient site-specific mutagenesis without phenotypic selection. *Proc. Natl. Acad. Sci. USA*, **1985**, *82*, 488–492.
- 14 J. J. PERONA, C. A. TSU, C. S. CRAIK, R. J. FLETTERICK, Crystal structure of an ecotin-collagenase complex suggests a model for recognition and cleavage of the collagen triple helix. *Biochemistry* **1997**, *36*, 5381–5392.
- 15 S. M. WAUGH, J. L. HARRIS, R. FLETTERICK, C. S. CRAIK, The structure of the pro-apoptotic protease granzyme B reveals the molecular determinants of its

- specificity. *Nature Struct. Biol.* **2000**, *7*, 762–765.
- 15a** S.X. WANG, E. HUR, C.A. SOUSA, L. BRINEN, E.J. SLIVKA, R.J. FLETTERICK, The extended interactions and Gla domain of blood coagulation factor Xa. *Biochemistry* **2003**, *42*, 7959–7966.
- 16** C.A. TSU, J.J. PERONA, V. SCHELLENBERGER, C.W. TURCK, C.S. CRAIK, The substrate specificity of Uca pugnator collagenolytic serine protease 1 correlates with the bovine type I collagen cleavage sites. *J. Biol. Chem.* **1994**, *269*, 19565–19572.
- 17** C.A. TSU, C.S. CRAIK, Substrate recognition by recombinant serine collagenase 1 from Uca pugnator. *J. Biol. Chem.* **1996**, *271*, 11563–11570.
- 18** J.L. HARRIS, B.J. BACKES, F. LEONETTI, S. MAHRUS, J.A. ELLMAN, C.S. CRAIK, Rapid and general profiling of protease specificity by using combinatorial fluorogenic substrate libraries. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 7754–7759.
- 19** S.A. GILLMOR, T. TAKEUCHI, S.Q. YANG, C.S. CRAIK, R.J. FLETTERICK, Compromise and accommodation in ecotin, a dimeric macromolecular inhibitor of serine proteases. *J. Mol. Biol.* **2000**, *299*, 993–1003.
- 20** J.A. TRAPANI, Granzymes: a family of lymphocyte granule serine proteases. *Genome Biol.* **2001**, *2*.
- 21** M.J. SMYTH, M.J. MCGUIRE, K.Y. THIA, Expression of recombinant human granzyme B. A processing and activation role for dipeptidyl peptidase I. *J. Immunol.* **1995**, *154*, 6299–6305.
- 22** C.T. PHAM, T.J. LEY, Dipeptidyl peptidase I is required for the processing and activation of granzymes A and B in vivo. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 8627–8632.
- 23** L. CASCIOIA-ROSEN, F. ANDRADE, D. UIALNET, W.B. WONG, A. ROSEN, Cleavage by granzyme B is strongly predictive of autoantigen status: implications for initiation of autoimmunity. *J. Exp. Med.* **1999**, *190*, 815–826.
- 24** L.E. SOWER, G.R. KLIMPEL, W. HANNA, C.J. FROELICH, Extracellular activities of human granzymes. I. Granzyme A induces IL6 and IL8 production in fibroblast and epithelial cell lines. *Cell Immunol.* **1996**, *171*, 159–163.
- 25** P.P. TAK, L. SPAENY-DEKKING, M.C. KRAAN, F.C. BREEDVELD, C.J., FROELICH, C.E. HACK, The levels of soluble granzyme A and B are elevated in plasma and synovial fluid of patients with rheumatoid arthritis (RA). *Clin. Exp. Immunol.* **1999**, *116*, 366–370.
- 26** J.W. WILLIAMS, J.F. MORRISON, The kinetics of reversible tight-binding inhibition. *Methods Enzymol.* **1979**, *63*, 437–467.
- 27** J.S. ULMER, R.N. LINDQUIST, M.S. DENNIS, R.A. LAZARUS, Ecotin is a potent inhibitor of the contact system proteases factor XIIa and plasma kallikrein. *Febs. Lett.* **1995**, *365*, 159–163.
- 28** K. PADMANABHAN, K.P. PADMANABHAN, A. TULINSKY, C.H. PARK, W. BODE, R. HUBER, D.T. BLANKENSHIP, A.D. CARDIN, W. KISIEL, Structure of human des(1–45) factor-Xa at 2.2-angstrom resolution. *J. Mol. Biol.* **1993**, *232*, 947–966.
- 29** H. BRANDSTETTER, A. KUHNE, W. BODE, R. HUBER, W. VONDERSAAL, K. WIRTHENSOHN, R.A. ENGH, X-ray structure of active site-inhibited clotting factor Xa – implications for drug design and substrate recognition. *J. Biol. Chem.* **1996**, *271*, 29988–29992.
- 30** K. KAMATA, H. KAWAMOTO, T. HONMA, T. IWAMA, S.H. KIM, Structural basis for chemical inhibition of human blood coagulation factor Xa. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 6630–6635.
- 31** L. HEDSTROM, Trypsin: a case study in the structural determinants of enzyme specificity. *Biol. Chem.* **1996**, *377*, 465–470.
- 32** J.J. PERONA, C.S. CRAIK, Evolutionary divergence of substrate specificity within the chymotrypsin-like serine protease fold. *J. Biol. Chem.* **1997**, *272*, 29987–29990.
- 33** R.W. WILLIAMS, A. CHANG, D. JURETIC, S. LOUGHREN, Secondary structure predictions and medium range interactions, *Biochim. Biophys. Acta* **1987**, *916*, 200–204.
- 34** S.Q. YANG, C.S. CRAIK, Engineering bidentate macromolecular inhibitors for trypsin and urokinase-type plasminogen

- activator. *J. Mol. Biol.* **1998**, *279*, 1001–1011.
- 35** C. I. WANG, Q. YANG, C. S. CRAIK, Isolation of a high affinity inhibitor of urokinase-type plasminogen activator by phage display of ecotin. *J. Biol. Chem.* **1998**, *270*, 12250–12256.
- 36** M. C. A. LABOISIERE, M. M. YOUNG, R. G. PINHO, S. TODD, R. J. FLETTERICK, I. KUNTZ, C. S. CRAIK, Computer-assisted mutagenesis of ecotin to engineer its secondary binding site for urokinase inhibition. *J. Biol. Chem.* **2002**, *277*, 26623–26631.
- 37** P. J. KRAULIS, Molscript: A program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **1991**, *24*, 946–950.
- 38** E. A. MERRITT, M. E. P. MURPHY, Raster 3D *Acta Crystallogr., Sect. D, Biol. Crystallogr.* **1994**, *50*, 869–973.

8

X-ray Crystallography in the Development of Orthogonal Ligand–Receptor Pairs

DONALD F. DOYLE and LAUREN J. SCHWIMMER

8.1 Introduction

Interactions between small molecules and proteins are ubiquitous in nature. They are involved in almost every cellular process, from phosphates regulating signal transduction, to enzymes catalyzing reactions, to hormones regulating gene transcription. Manipulation of naturally occurring ligand–protein interactions can lead to the creation of an orthogonal ligand–receptor pair (OLRP). An orthogonal ligand–receptor pair is one in which the ligand is not bound by the endogenous, wild type (wt) receptor and the orthogonal receptor is mutated so that it will not bind the endogenous ligand. In other words, a protein is mutated so that it no longer binds its natural ligand, but now binds a new, synthetic ligand (Fig. 8.1). The creation of orthogonal ligand–receptor pairs can be useful in deconvoluting some cellular processes and also provide useful tools in the area of chemical genetics.

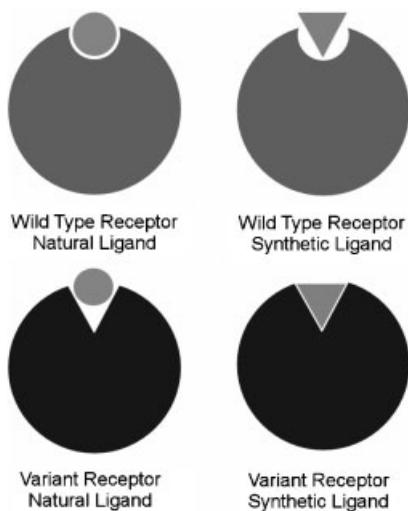


Fig. 8.1 Schematic diagram showing how a naturally occurring small molecule ligand–protein interaction is changed to create an orthogonal ligand–receptor pair.

8.2

Applications of OLRPs

Orthogonal ligand–receptor pairs have multiple applications for basic cellular and biochemical research. They also have practical applications that could soon be in use in the pharmaceutical and biotechnology industries.

8.2.1

Applications in Basic Research

Orthogonal ligand–receptor pairs offer the ability to control transcription with small, drug-like molecules. The ability to control transcription with exogenous small molecules would provide powerful tools for research. Genome projects have revealed tens of thousands of genes with unknown functions. Orthogonal ligand–receptor pairs controlling transcription could be used to study the function of these genes by redirecting the receptor to bind near the promoter of the gene, delivering the gene for the orthogonal receptor to the cell or animal, and activating transcription of the target gene at the desired time by adding the orthogonal ligand. If transcription of a set of genes in a cell line or animal were under the control of multiple ligand–receptor pairs, simply applying the appropriate cocktail of ligands could control multiple genes. Indeed, the levels of expression of each gene could be titrated, not only turned just on or off.

A general method of activating transcription with small molecules would provide useful research tools for creating conditional knockout cell lines and animals. Genetic knockouts can result in embryonic lethality, which does not allow the gene function in a mature organism to be studied. Spatio-temporal control of gene expression through orthogonal ligand–receptor pairs would allow researchers to stop or start the expression of a gene at any point in development. This conditional gene expression would give a better picture of the gene's function and be useful in elucidating signal transduction pathways.

Another method of achieving spatio-temporal control of protein function is to inhibit the protein once it is expressed, creating conditional alleles. Genetic knockouts can have effects on organism that do not allow for direct study of the gene that is removed. Sometimes the cell or organism can compensate for the loss of the gene through redundant pathways. The genes for orthogonal proteins can be incorporated, or substituted for the wild type protein, in the genome of an organism. An orthogonal ligand for this protein can then be used to inhibit the protein and the effect of this inhibition can be observed. This type of orthogonal ligand–enzyme pair can also be used to elucidate direct substrates of the enzyme. If part of the ligand (e.g., a phosphate group) is transferred to another substrate, this part can be radioactively labeled, resulting in a labeled product that can be detected.

Signal transduction pathways can also be elucidated by incorporating variant proteins, which require chemical inducers of dimerization to function. These small molecules can cause two proteins to come together to initiate the signal.

They can also be used to block a signaling pathway by capturing a protein and not allowing it to interact with its natural substrates.

8.2.2

Applications in Applied Research

For practical applications, orthogonal ligand–receptor pairs controlling transcription could potentially have a profound impact for gene therapy. The orthogonal receptor could be directed to activate a gene *at its natural locus in the genome*, which would address the cause of the disease or induce apoptosis to kill the diseased cell. Another possibility for gene therapy is that the orthogonal ligand–receptor pair can be used to control the expression of another *delivered* therapeutic gene (e.g., CFTR for cystic fibrosis). When the therapeutic gene is delivered, the gene for the orthogonal receptor would be delivered with it. The receptor would be expressed under a cell-type specific promoter, bind to the synthetic promoter for the therapeutic gene and express the therapeutic gene to the desired level by the appropriate dose of the orthogonal ligand. This could potentially circumvent a current problem with gene therapy – controlling the amount of expression of the therapeutic protein. Under the described control, if a low efficiency of gene delivery is achieved, then the appropriate amount of protein can be achieved by increasing the dose of the orthogonal ligand (within limits). Alternatively, administering a low ligand concentration would compensate for efficient gene delivery.

Nuclear receptors function in plants. Agricultural applications of orthogonal ligand–nuclear receptor pairs include conditional expression of any RNA or protein, including short interfering RNAs to block gene expression, human vaccine proteins, insecticidal proteins and disease resistance genes. Expressing insecticidal proteins and disease resistance genes only at specific times (rather than constitutively) would likely decrease the occurrence of resistance developing in the environment.

In signal transduction, after the orthogonal ligand–receptor pairs are used to elucidate the pathways, they can be used to manipulate the pathways or to build new ones. Application of chemical inducers of dimerization to this system would permit activation of signal transduction only in cells expressing the orthogonal receptor. This system, however, could not be directed to specific target genes and would further be limited to signal transduction pathways that already exist within the cell. If this system were linked to a conditional allele system, new signal transduction pathways within the cell could be engineered. A chemical inducer of dimerization could be used to activate a signal transduction pathway consisting of orthogonal ligand–receptor pairs.

8.3

Early Work

Hwang and Miller are pioneers in this area. They changed the specificity of *E. coli* elongation factor Tu from guanosine 5'-diphosphate (GDP) to xanthosine 5'-di-

phosphate (XDP) with a single amino acid change [1]. The mutation of aspartate, which normally forms a salt bridge with the amine at the 2-position of GDP, to asparagine allows hydrogen bonding with the 2-position carbonyl of XDP. Because the salt bridge with GDP can no longer occur, orthogonality is introduced.

The term “orthogonal ligand–receptor pair” was first used by Schreiber and co-workers [2]. They utilized the crystal structure of cyclophilin A (CyP) and its ligand, cyclosporin A (CsA), to guide the mutagenesis of CyP and synthesis of CsA derivatives. Space created by a phenylalanine to alanine and serine to threonine double mutation in CyP complemented the addition of a methyl group to CsA. The methylated CsA does not bind the wild type CyP, but has a low nanomolar binding affinity with the double variant. These first two examples show the viability of creating orthogonal ligand–receptor pairs and that structure-guided mutagenesis is useful in creating orthogonal-ligand–receptor pairs.

An orthogonal ligand–receptor pair using the progesterone receptor (PR) was created in a different way. O’Malley and co-workers [3, 4] found that a PR truncated by 42 residues from the C-terminus (PR-LBDA lost affinity for progesterone and other endogenous steroids. However, PR-LBDA retained affinity for the anti-progesterone and antiglucocorticoid small molecule mifepristone (RU486). In fact, when PR-LBDA was fused with the Gal-4 DNA-binding domain (GAL4-DBD) and a transcriptional activation domain, this receptor was activated at concentrations as low as 0.01 nM. This concentration is well below (>1000-fold) the concentration required for mifepristone to show progesterone antagonist activity. The ability of this mutated receptor to bind a synthetic ligand at concentrations below those which cause biological effects and not bind the natural ligands makes this a functionally orthogonal ligand–receptor pair. This system has been shown to be an effective ligand-inducible transcriptional regulator both *in vitro* and in a mouse model.

8.4

Structure-Guided Mutagenesis and OLRLPs

With the advent of the genomic and proteomic eras, the availability of a wide variety of crystal structures has made structure-guided mutagenesis a viable method for the creation of orthogonal ligand–receptor pairs. By examining crystal structures of proteins with and without their natural ligand(s), researchers can identify residues in contact with the ligand. There are two categories of mutations that can be made: binding pocket mutations and distant mutations. Binding pocket mutations are those mutations made with residues that directly contact the bound ligand or define the binding pocket. Distant mutations are mutations that occur further away from the binding pocket; they do not directly contact the ligand and do not define the binding pocket.

8.4.1

Binding Pocket Mutations

Binding pocket mutations have proven to be an effective location for structure-guided mutagenesis. The interactions between the ligand and the protein can be visualized. The residues that line the binding pocket not only define the size and shape of the binding pocket, but also contribute to the electrostatics and hydrophobicity of this space. Mutations to these residues can, somewhat predictably, alter the shape, size or environment of the binding pocket and, therefore, change the binding specificity of the protein. Large amino acids could be changed to small amino acids to open holes in the binding pocket to allow larger ligands to bind or *vice versa*. Also, polar amino acids could be changed to non-polar amino acids to make the binding pocket more hydrophobic and so forth. These types of changes have recently been reviewed [5].

There are many examples in the literature of using binding pocket mutations to make orthogonal ligand–receptor pairs. For the purpose of brevity, the emphasis of this chapter will be on nuclear receptor orthogonal ligand–receptor pairs.

8.4.1.1 Nuclear Receptors

Nuclear receptors are a superfamily of ligand-activated transcription factors that control a broad range of physiological processes [6–8]. They are modular proteins consisting of two functional domains, a ligand binding domain (LBD) and a DNA-binding domain (DBD), which are connected by a hinge region. The DNA binding domain consists of two zinc finger modules that bind six to eight base pairs of DNA in a response element (enhancer) regulating the target gene. The ligand-binding domain binds a small molecule ligand. Upon ligand binding, the nuclear receptor undergoes a conformational change that recruits the transcription machinery (Fig. 8.2). Nuclear receptors function as either homodimers or heterodimers. Orthogonal ligand–nuclear receptor pairs would allow control of transcription through synthetic, drug-like molecules. This control would be useful in gene therapy, agriculture and metabolic engineering.

Nuclear receptors are good candidates for creating orthogonal ligand–receptor pairs for a variety of reasons. Firstly, many crystal structures are available with a variety of natural and synthetic ligands as well as those that have no ligand present. Therefore, these structures can be examined and used to guide mutagenesis. Also, the binding pockets are predominantly defined by side chain atoms. Side chains are more easily substituted than backbone atoms, which require changing the protein fold.

Natural variations in nuclear receptors reinforce the idea that nuclear receptors are good candidates for the creation of orthogonal ligand–receptor pairs. For example, the human retinoic acid receptor (RAR) has three subtypes: RAR α , RAR β and RAR γ . These three subtypes differ at 87 amino acids but have only three divergent residues in the binding pocket [11] (Tab. 8.1). However, these three divergent residues are responsible for different binding profiles of synthetic retinoids

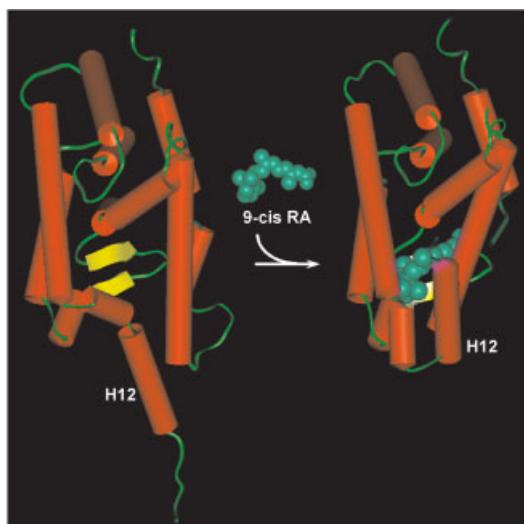


Fig. 8.2 The ligand-binding domain of the nuclear receptor RXR binds the small molecule 9-cis retinoic acid (9-cis RA). In the apo-structure on the left, helix 12 (H12) extends out into the solution. Upon binding 9-cis RA, H12 folds back on top of the ligand and LBD creating a binding site for the transcription machinery. The transcription machinery is thereby recruited to the transcription start site, resulting in activation of transcription. Nuclear receptors generally follow this mechanism of ligand-activated transcription. Coordinates from 1LBD [9] and 1FBY [10].

Tab. 8.1 Divergent RAR-binding pocket residues.

| <i>Receptor</i> | | | |
|-----------------|------|------|------|
| RAR- α | S232 | I270 | V395 |
| RAR- β | A225 | I263 | V388 |
| RAR- γ | A234 | M272 | A397 |

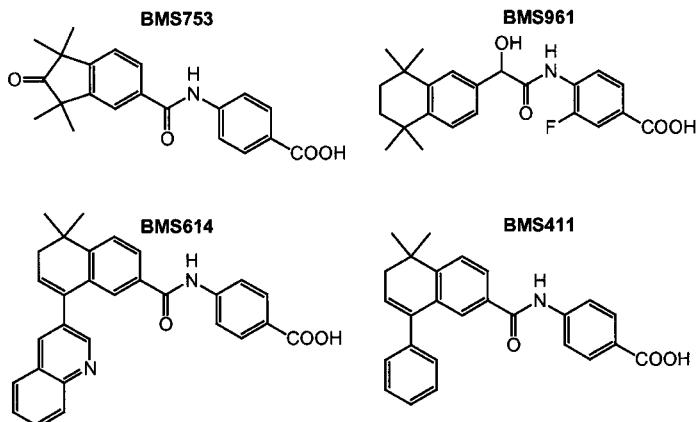


Fig. 8.3 Synthetic RAR ligands that show subtype selectivity.

BMS753, BMS961, BMS614 and BMS411 (Fig. 8.3). Gronemeyer and co-workers showed that mutating the residues of the binding pocket in one subtype to the residues in another subtype (i.e., RAR γ A234S; M272I; A397V to make RAR $\gamma \rightarrow \alpha$) changed the binding specificity of the protein. They observed that the binding profile of the variant matched the binding profile of the subtype naturally containing the binding pocket mutations and not of the parent [12]. This switch in specificity demonstrates that simple binding pocket mutations have an effect on the specificity of the receptor and that specificity changes can be predicted.

8.4.1.2 Estrogen Receptor OLRPs

The creation of orthogonal ligand–receptor pairs using the human estrogen receptor (ER) has been the target of two groups: Katzenellenbogen at the University of Illinois and Koh at the University of Delaware. Both groups focused on the interaction of glutamate-353 with the hydroxyl group on the A-ring of estradiol, ER's endogenous ligand (Fig. 8.4).

Katzenellenbogen and co-workers [15] proposed three structure-based exchanges to create an ER α orthogonal ligand–receptor pair: (1) reversing the roles of the hydrogen bond donor/acceptor, (2) shifting the position of the interaction by one carbon and (3) replacing a polar interaction with a non-polar interaction (Fig. 8.4). They made three variants (E353S, E353D and E353A) and tested them against a variety of ligands with corresponding changes. The third exchange gave the best results. Ligand 1 (Fig. 8.5) has 35-fold greater potency with the E353A variant than with the wild type ER α . Also, estradiol is 400-fold less potent with this variant than with wt-ER α and ligand 1 is 2000-fold less potent with wt-ER α than with this variant. However, there is still significant activation of the E353A at endogenous concentrations of estradiol making this ligand–receptor pair non-orthogonal. This work is an important step toward understanding the ligand–receptor interactions that are necessary for creating a pair that is orthogonal.

Concurrently, Shi and Koh [16] developed a similar orthogonal ligand–receptor pair with ER α G400V (a variant that has less background activity in cell culture). They chose to replace the intra-molecular salt bridge between E353 and R394 with an inter-molecular salt bridge between the ligand and R394 (Fig. 8.4). They also chose to make an E353A mutation to allow additional space at this end of the binding pocket in order to move the E353 half of the salt bridge to the ligand. Using molecular modeling techniques, three carboxylic acid derivatives of estradiol were chosen to be synthesized as potential ligands. They found that the modified estradiol with a vinyl linker, 2 (Fig. 8.5), activated ER α G400V;E353A at a concentration that did not activate ER α G400V. Also, estradiol does not activate the E353A variant at endogenous concentrations. These two conditions make this pair functionally orthogonal.

To further refine their system, Shi and Koh [17] examined this orthogonal ligand–receptor pair in wild type ER α (wtER α) and wild type ER β (wtER β) as well as in the corresponding variants (ER α E353A and ER β E305A). They discovered that ER α E353A is significantly activated by nanomolar concentrations of estradiol;

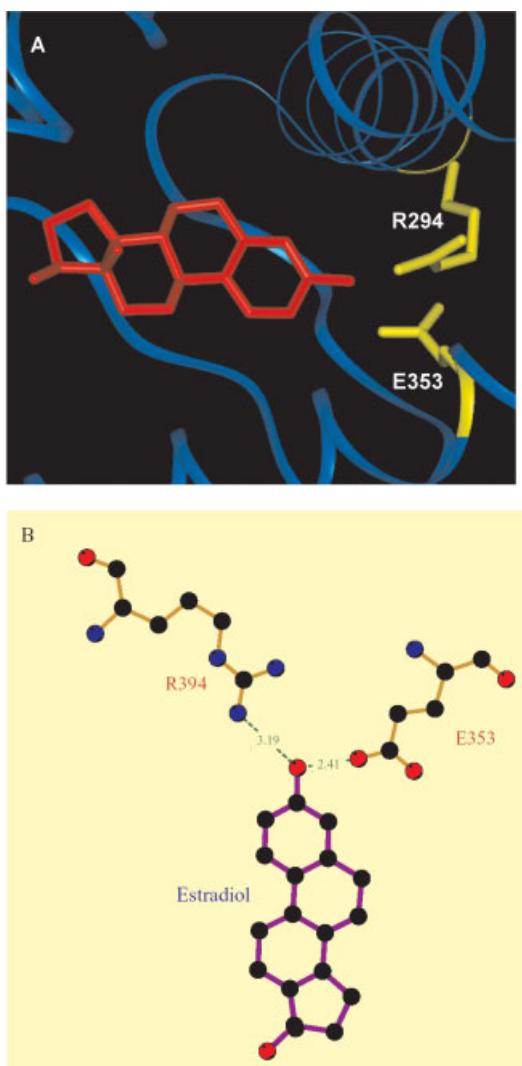


Fig. 8.4 (A) X-ray crystal structure of estradiol bound to human estrogen receptor. Residues glutamate-353 and arginine-394 were modified in attempts to create orthogonal ligand–receptor pairs. (B) Ligplot representation of estradiol interactions with E353 and R394 [13]. Coordinates from 1A52 [14].

therefore the ligand 2-ER α E353A ligand–receptor pair is not orthogonal. However, 2 shows a 400-fold preference for the variant ER β E305A over the wild type ER β and a 158-fold preference for ER β E305A over wtER α . Also, the variant ER β E353A is not significantly activated by endogenous concentrations of estradiol. This 2-ER β E305A ligand–receptor pair, created using structure-guided mutagenesis, is orthogonal to both naturally occurring ER subtypes. It also supports using polar-group exchange as a viable method to create orthogonal ligand–receptor pairs.

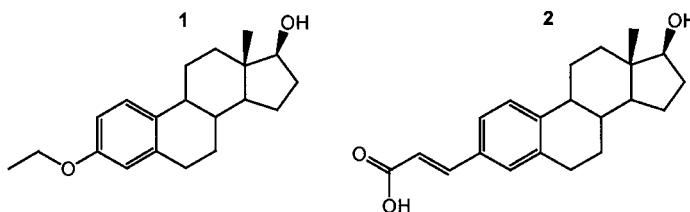


Fig. 8.5 Synthetic ER ligands.

8.4.1.3 Retinoic Acid Receptor OLRPs

Koh and co-workers [18] used a charge reversal approach in an attempt to create an orthogonal ligand–receptor pair with the human retinoic acid receptor- γ (RAR- γ) and a synthetic, positively charged retinoid. RAR- γ is activated by both *all-trans*-retinoic acid (at-RA) and 9-*cis*-retinoic acid, both of which have a carboxylate group. The crystal structures of RAR- γ with these ligands suggest that this negative charge is stabilized by positively charged amino acids at one end of the binding pocket [19, 20] (Fig. 8.6). The authors hypothesized that two variants, S289D and S289G;R278E (both residues are in the positively charged end of the binding pocket), would bind one of five at-RA derivatives and activate the variant RAR- γ , but not activate the wild type receptor. They found that compound 3, which is neutral, and 4, which is positively charged (Fig. 8.7), were able to induce transcription in the S289G;R278E or S289D, respectively, over the wild type receptor. However, there is still significant wild type activity with both of these compounds, thus neither of them can be considered orthogonal. This attempt still represents an important step in creating charge-reversed orthogonal ligand–receptor pairs and increases the understanding of electrostatic interactions in ligand binding. The difficulties of charge reversal were pointed out early on by Hwang and Warshel [21].

8.4.1.4 Retinoid X Receptor OLRPs

Corey and co-workers, in a series of two papers [22, 23], demonstrated that binding pocket mutations have the ability to change the specificity of ligand binding in the human retinoid X receptor- α (RXR- α). In the first paper, the authors selected two sites at which to make a variety of non-polar mutations, F313 and L436 (Fig. 8.8) [22]. They then tested the transcriptional activation by 9-*cis* retinoic acid (9-cRA) and three synthetic RXR ligands. Their results indicated that some variants showed decreased activation by 9-cRA and increased activation by synthetic ligands, while others retained their 9-cRA activation, but were no longer activated by the synthetic ligands. This indicated that these residues were important for ligand specificity and that RXR- α and this class of synthetic ligands showed promise for obtaining an orthogonal ligand–receptor pair.

In the second paper, Corey and co-workers took a different approach to selecting which mutations to make [23]. After comparing the sequences of RXR- α and RAR- γ they chose to mutate ten RXR residues to the corresponding RAR residues

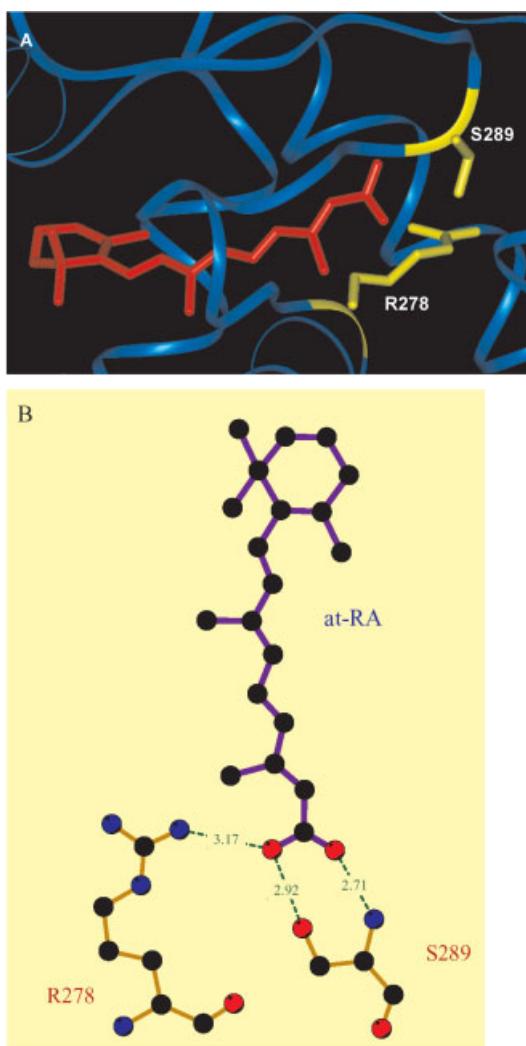
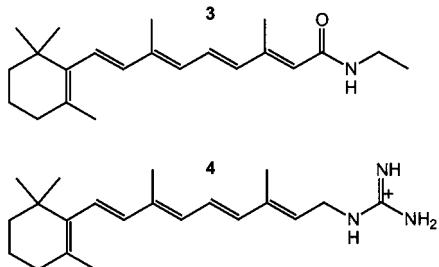


Fig. 8.6 (A) X-ray crystal structure of all-trans retinoic acid (atRA) bound to human RAR- γ . Residues arginine-278 and serine-289 were modified in attempts to make orthogonal ligand–receptor pairs. (B) Ligplot representation of atRA interactions with S289 and R278 [13]. Coordinates from 2LBD [19].

in single, double and triple variant combinations. These were tested for transcriptional activation by 9-cRA, *all-trans* retinoic acid and five synthetic ligands. One combination, Q275C;I310M;F313I, showed low activation by 9-cRA, but high activation by LG335 (Fig. 8.9). This is an orthogonal ligand–receptor pair, discovered through structure-guided mutagenesis. This paper also introduces the concept of “near drugs.” These are compounds synthesized in the drug discovery process that show no activity with the endogenous receptor. This pool of compounds has been shown to be an excellent source of orthogonal ligands for nuclear receptors and could potentially apply to creating orthogonal ligand–receptor pairs for any other drug target and its “near drugs.”

Fig. 8.7 Neutral and positively charged synthetic analogues of atRA.



8.4.2

Distant Mutations

Distant mutations are useful in making subtle changes in the overall structure of the protein. They may cause a chain reaction of slight differences in side chain orientation that propagates to the binding pocket. Also, several mutations, which have no individual effect, may have a synergistic effect to cause a favorable or unfavorable change in the protein's ligand binding ability. The problem is that the effect of these mutations are difficult to predict using current methods.

Using directed evolution, rather than a structure-based approach, Kagamiyama and co-workers [24] found that distant mutations were able to alter substrate specificity of aspartate 2-oxoglutarate aminotransferase (AspAT). They found that the combination of six mutations, N34D, I37M, S139G, N142T, N297S and V387L (Fig. 8.10), decreased k_{cat}/K_m values for AspAT for acidic 2-oxo acids, the natural ligands, by 20-fold and increased the k_{cat}/K_m for β -branched amino acids, which are not normally substrates, by up to 10^5 -fold. Examination of the crystal structure shows that these residues are between 3.5 and 16.1 Å (Tab. 8.2) from the binding pocket. The authors suggested that these mutations combine to reorient a tryptophan, an arginine and a water molecule and remove steric hindrance for the β -branch. However, the N34D mutation is more than 9 Å from the binding pocket and the authors could not speculate on the value of this mutation, even though it proved to be essential.

The six mutations in this example were discovered using directed evolution and not through structure-guided mutagenesis. The difficulty of trying to predict which distant mutations should be combined to have the desired result is immense. Examination of the AspAT crystal structure may have suggested some of the mutations, which are fairly close to the binding pocket, but, using current methods, it would have been impossible to predict the N34D mutation through structure-guided mutagenesis or that the combination of these mutations would lead to such a large change in specificity. This unpredictability is a major disadvantage and currently excludes these types of mutations from being used for initial structure-guided mutagenesis to create orthogonal ligand–receptor pairs. Identifying networks of interacting amino acids in proteins may rectify this problem. A recent leap forward in this area has been made by Ranganathan and co-workers

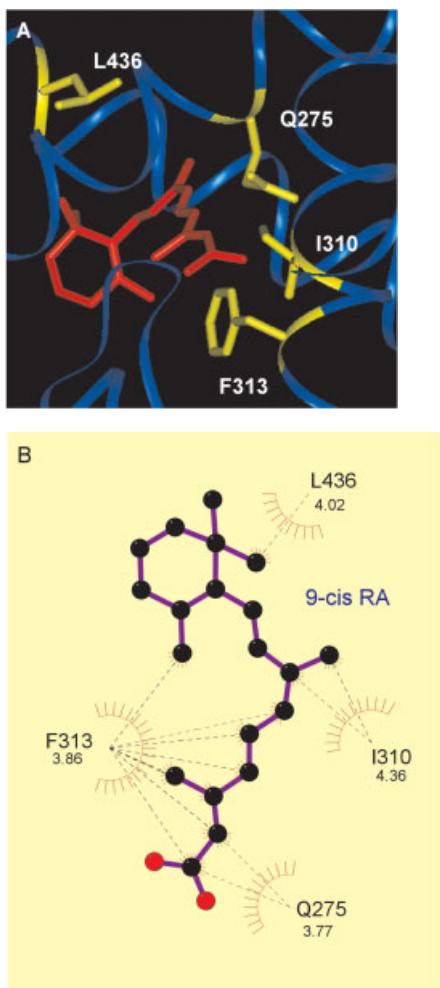


Fig. 8.8 (A) X-ray crystal structure of 9-cis retinoic acid (9-cis RA) bound to human RXR- α . Residues glutamate-275, isoleucine-310, phenylalanine-313 and leucine-436 were modified in attempts to make orthogonal ligand–receptor pairs. Indicated distances are shortest between residue and 9-cis RA. (B) Ligplot representation of 9-cis RA interactions with Q275, I310, F313 and L436 [13]. Coordinates from 1FBY [10].

[26]. They used a statistical thermodynamic method to analyze multiple sequence alignments of protein families. They found that co-evolving amino acids form structurally (and presumably functionally) interacting networks of amino acids. These networks form the basis of allosteric communication within a protein. The identification of interacting amino acid networks suggests distant mutations that could be made in PDZ domains, G-proteins, hemoglobin and serine proteases. The ability to identify allosteric amino acid networks from a single crystal structure rather than a family of proteins could make structure-guided mutagenesis of distant mutations more widely feasible.

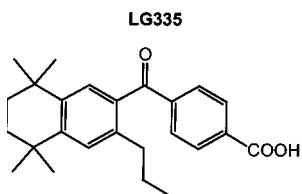


Fig. 8.9 The synthetic “near drug” LG335 is the orthogonal ligand in an orthogonal ligand–receptor pair created from the nuclear receptor RXR- α .

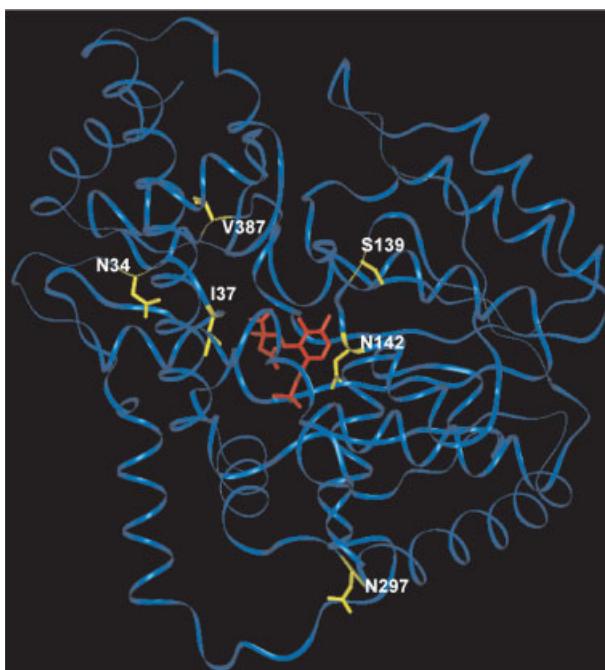


Fig. 8.10 X-ray crystal structure of an aspartate aminotransferase (AspAT) bound to its cofactor pyridoxal 5'-phosphate and aspartate. Directed evolution techniques produced changes in ligand specificity due to substitution of the disparate positions indicated. Coordinates from 1ART [25].

Tab. 8.2 AspAT residues shortest distance to binding pocket.

| <i>Residue</i> | <i>Distance (Å)</i> |
|----------------|---------------------|
| N34 | 9.14 |
| I37 | 3.55 |
| S139 | 7.29 |
| N142 | 7.31 |
| N297 | 16.08 |
| V387 | 9.66 |

8.5

Other Examples of OLRLPs

8.5.1

Chemical Inducers of Dimerization to Control Transcription

Chemical inducers of dimerization (CID) are bivalent small molecules that bind two proteins simultaneously. The purpose of these molecules is to bring the proteins together to induce signal transduction [27]. For brevity, we will limit our discussion to CIDs that directly control transcription. The basic architecture of these systems consists of two chimeric proteins. The first contains a DNA-binding domain (DBD) fused to a ligand-binding domain (LBD) and the second chimera contains an LBD and an activation domain (AD). The small molecule that binds both of these proteins simultaneously induces proximity of the two proteins, resulting in transcription activation (Fig. 8.11).

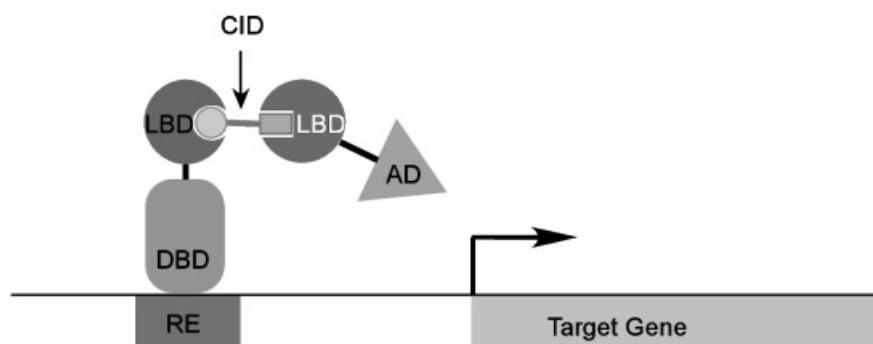


Fig. 8.11 Chemical inducers of dimerization (CID) to control transcription. This system requires two chimeric proteins, one comprising a DNA-binding domain (DBD) fused to a ligand-binding domain (LBD), and the other comprising an LBD fused to a transcriptional

activation domain (AD). The CID is a bivalent small molecule that binds both LBDs, brings the activation domain near the transcription start site, resulting in activation of transcription.

Pollock and co-workers and Cornish and co-workers have both done work in this area. Pollock's system [28] uses a zinc finger domain (ZFD) fused to FK506-binding protein (FKBP) as the DBD/LBD chimera and FKBP rapamycin-associated protein (FRAP) fused with the activation domain S3H as the LBD/AD chimera. Rapamycin binds both FKBP and FRAP simultaneously [29], bringing the AD into the vicinity of the DBD, which activates transcription. Rather than rapamycin, which is an immunosuppressive drug, the CID used in this study is AP21967 (Fig. 8.12), an evidently non-immunosuppressive rapamycin analogue [30]. They also use an FRAP with a T2098L mutation, which enables binding of rapamycin and AP21967. Although FRAP:T2098L binds rapamycin, a molecule that does have biological effects, this molecule is produced by *Streptomyces hygroscopicus*, thus is not endogenous to mammalian cells. The authors reported dose-dependent mRNA and protein production in response to AP21967. Therefore, AP21967 and FRAP:T2098L qualify as an orthogonal ligand–receptor pair to control transcription.

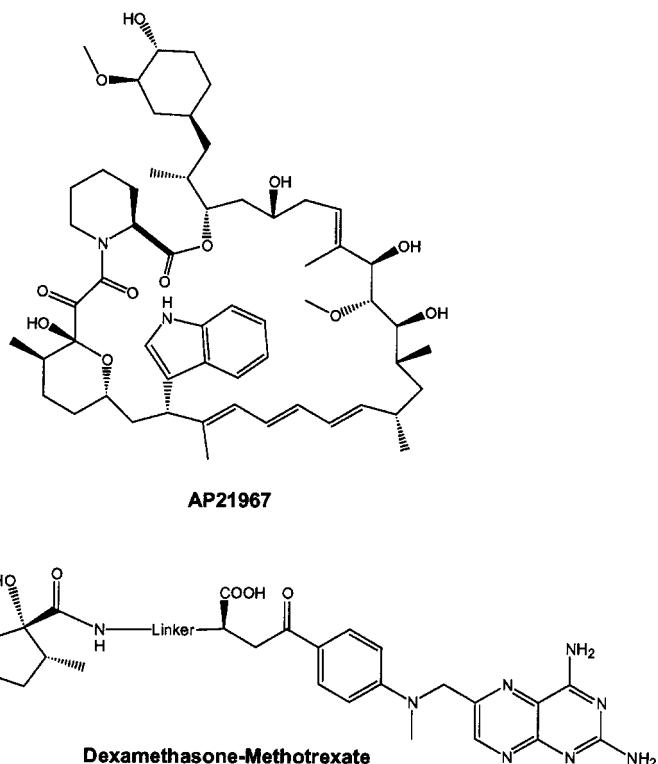


Fig. 8.12 Orthogonal chemical inducers of dimerization. AP21967 is a rapamycin analogue that binds both FKBP and the T2098L variant

of FRAP. Dexamethasone-methotrexate binds both the glucocorticoid ligand-binding domain and dihydrofolate reductase.

Cornish and co-workers [31, 32] developed a system that uses LexA fused with dihydrofolate reductase (DHFR) as the DBD/LBD chimera and the glucocorticoid receptor (GR) fused with B24 as the LBD/AD chimera. The CID consists of methotrexate (Mtx), the DHFR ligand and dexamethasone (Dex), the GR ligand, connected with a linker (Fig. 8.12). Using a *lacZ* reporter, which encodes the enzyme β -galactosidase, the authors were able to detect dose-dependent hydrolysis of the β -galactosidase substrates *o*-nitrophenyl β -D-galactopyranoside (ONPG) in liquid media and X-gal on solid media. Therefore, transcription is occurring in response to this CID.

This system was actually developed to engineer enzymes to catalyze reactions in the linker region of the CID. For this system to function it must be orthogonal to the cell type in which it functions. This system is orthogonal in yeast, the cell type in which it was developed. It could be extended to other cell types, but would not be orthogonal in mammalian cells because the LBDs are endogenous in many systems and in whole animals. The ligands that are used also bind to the endogenous proteins, in fact, methotrexate is used as an anti-cancer drug. This CID system could be developed for transcription-controlling applications in mammalian cells if it contained two orthogonal ligand–receptor pairs. Both of the LBDs used have published crystal structures [33, 34]. Examination of these structures may suggest mutations that could be made in the binding pocket and modifications that could be made in the ligands to impart full orthogonality on the system in mammalian cells.

8.5.2

OLRPs and ATP Analogues

Shokat and co-workers have a different vision for the utility of orthogonal ligand–receptor pairs. They use orthogonal ligand–receptor pairs of protein kinases and adenosine 5'-triphosphate (ATP) analogues to examine kinase signaling pathways. Amino acids in the ATP binding pockets are mutated to accept ATP derivatives, which do not bind the wild type kinases. These ligand–receptor pairs allow individual enzyme function in this complicated signaling pathway to be studied in two different ways.

Firstly, direct kinase protein substrates can be identified by creating synthetic nucleotides that are substrates for the variant kinases. These nucleotides can be $\gamma^{32}\text{P}$ labeled. This labeled phosphate is transferred to the protein substrate of the kinase, which can be identified due to its radioactivity. This application was demonstrated by Ulrich et al. [35] using p38 mitogen-activated protein kinase (p38). The crystal structure of c-Src complexed with adenosine 5'-diphosphate (ADP) was used to identify threonine 255, which is analogous to T106 in p38, as a possible site for mutation [36]. A T106G mutation in p38 would create a pocket into which a benzyl group could fit (Fig. 8.13). Labeled N⁴-(benzyl) AICAR triphosphate (5-amino-imidazole-4-carboxamide ribotide) (Fig. 8.14) was effectively used by p38 T106G to phosphorylate activating transcription factor-2 (ATF-2) and did not bind to the wild type p38. However, p38 T106G still has the ability to use ATP as a sub-

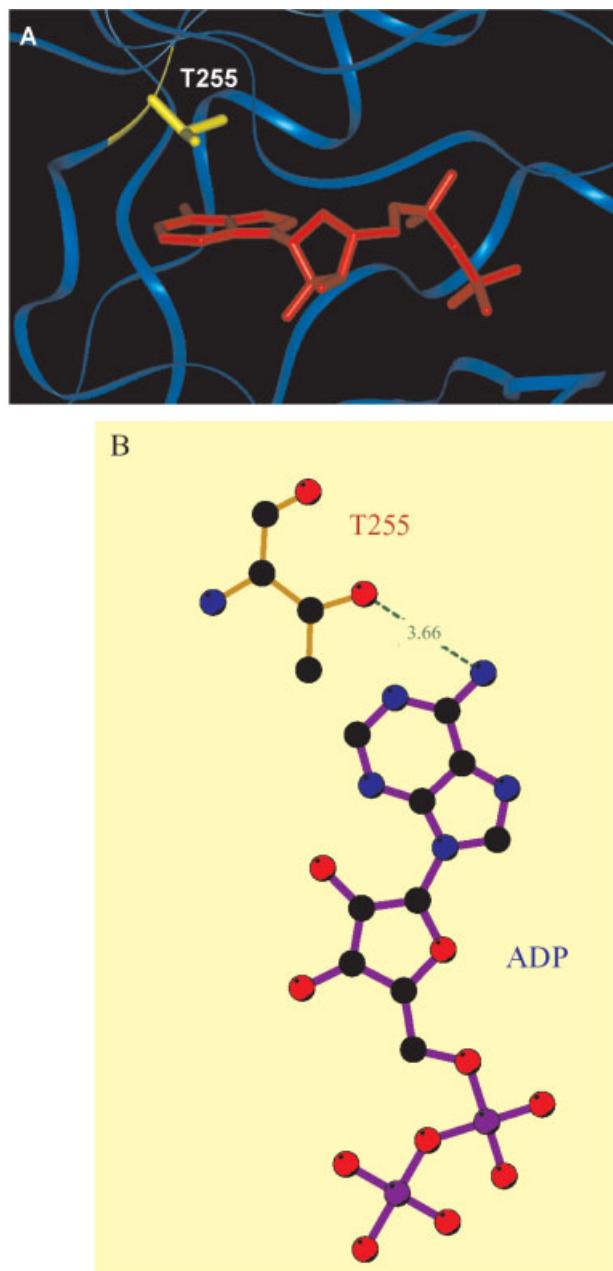


Fig. 8.13 (A) X-ray crystal structure of ADP bound to c-SRC. Threonine-255 is analogous to threonine-106 in mitogen-activated protein kinase p38. T106 was changed to a glycine in

an attempt to create an orthogonal ligand–receptor pair. (B) Ligplot representation of ADP interaction with T255 [13]. Coordinates from 2SRC [36].

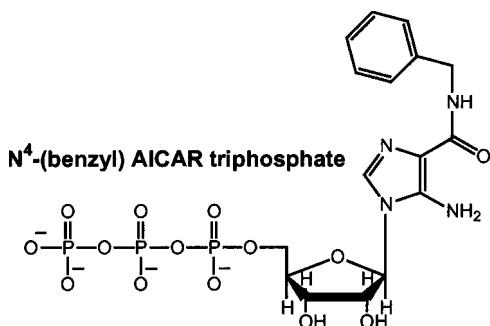


Fig. 8.14 The synthetic orthogonal ligand N⁴-benzyl AICAR triphosphate does not bind the wild type p38, but can be used as a substrate by the T106G variant.

strate. This is the first step in creating an orthogonal ligand–receptor pair for labeling direct substrates of protein kinases. This change in specificity is sufficient to identify protein substrates with radiolabeled orthogonal ligands.

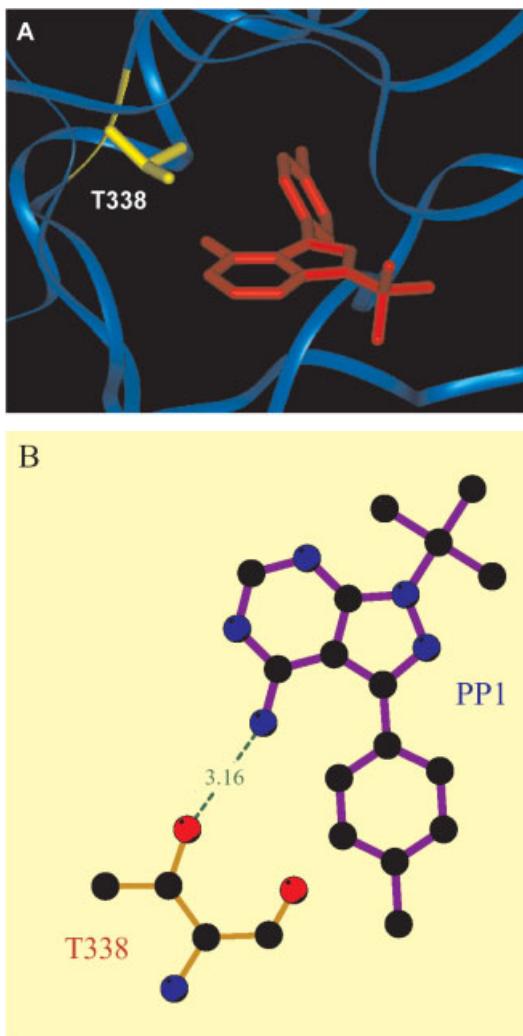
A second application involves creating conditional alleles. This would allow closer study of the function of proteins in the kinase pathway. Variant, but functional, kinases are incorporated into the genome. These variant kinases can then be inhibited by ATP analogues specific for that variant. Therefore, a specific protein can be shutdown at a specific time and the effect can be observed immediately. The kinase pathway is so complicated and redundant that simple genetic knockouts often have no phenotype because the cell compensates for the loss of a single gene. With this chemical system, the cell does not have time to compensate. For example, Bishop et al. [37] used the crystal structure of Hck with tyrosine kinase inhibitor PP1 to identify T338, which is analogous to F88 in cyclin-dependent kinase Cdc28, as a candidate for mutagenesis [38]. The authors hypothesized that an F88G mutation would allow Cdc28 to be inhibited by PP1 analogues (Fig. 8.15). They found that analogue 5 (Fig. 8.16) selectively inhibited the variant Cdc28, but did not inhibit the wild type protein. This orthogonal ligand–receptor pair allowed the authors to determine that Cdc28 caused a pre-mitotic cell-cycle arrest in *Saccharomyces cerevisiae* different from that seen in previous studies [39, 40].

Both of these strategies have broader applications than the kinase signaling pathways. Orthogonal ligand–receptor pairs could, theoretically, be made for any ligand–receptor pair. These pairs could then be used to study metabolic pathways and protein function, to identify small-molecule ligands and other applications.

8.6 Summary

Orthogonal ligand–receptor pairs have only just begun to make an impact on basic research. The future holds much broader impacts in both basic research and practical applications in the pharmaceutical industry and biotechnology. X-ray crystallographic structures have formed the basis for many structure-based

Fig. 8.15 (A) X-ray crystal structure of PP1 bound to Hck. Threonine-338 is analogous to phenylalanine-88 in cyclin-dependent kinase Cdc28. T88 was changed into a glycine in an attempt to create an orthogonal ligand–receptor pair. (B) Ligplot representation of PP1 interaction with T338 [13]. Coordinates from 1QCF [38].



approaches to creating orthogonal ligand–receptor pairs. Currently, binding pocket substitutions have seen the most success, but progress is being made on the ability to use X-ray crystallographic structures to guide more distant substitutions.

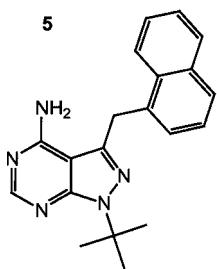


Fig. 8.16 The PP1 analogue that is an orthogonal ligand inhibitor specific for the F88G variant of the cyclin-dependent kinase Cdc28.

8.7

References

- 1 HWANG, Y. W., MILLER, D. L., *J. Biol. Chem.* **1987**, *262*, 13081–13085.
- 2 BELSHAW, P. J., SCHOEPFER, K. L., MORRISON, K. L., SCHREIBER, S. L., *Angew. Chem. Int. Ed. Engl.* **1995**, *34*, 2129–2132.
- 3 VEGETO, E., ALLAN, G. F., SCHRADER, W. T., TSAI, M.-J., McDONNELL, D. P., O'MALLEY, B. W., *Cell* **1992**, *69*, 703–713.
- 4 WANG, Y., O'MALLEY, B. W. Jr., TSAI, S. Y., O'MALLEY, B. W., *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 8180–8184.
- 5 KOH, J. T., *Chem. Biol.* **2002**, *9*, 17–23.
- 6 MANGELSDORF, D. J., THUMMEL, C., BEATO, M., HERRLICH, P., SCHUTZ, G., UMEMONO, K., BLUMBERG, B., KASTNER, P., MARK, M., CHAMBON, P., EVANS, R. M., *Cell* **1995**, *83*, 835–839.
- 7 KATZENELLENBOGEN, J. A., KATZENELLENBOGEN, B. S., *Chem. Biol.* **1996**, *3*, 529–536.
- 8 RENAUD, J. P., MORAS, D., *Cell. Mol. Life Sci.* **2002**, *57*, 1748–1769.
- 9 BOURGUET, W., RUFF, M., CHAMBON, P., GRONEMEYER, H., MORAS, D., *Nature (London)* **1995**, *375*, 377–382.
- 10 EGEA, P. F., MITSCHLER, A., ROCHEL, N., RUFF, M., CHAMBON, P., MORAS, D., *Embo J.* **2000**, *19*, 2592–2601.
- 11 RENAUD, J. P., ROCHEL, N., RUFF, M., VIVAT, V., CHAMBON, P., GRONEMEYER, H., MORAS, D., *Nature (London)* **1995**, *378*, 681–689.
- 12 GÉHIN, M., VIVAT, V., WURTZ, J., LOSSON, R., CHAMBON, P., MORAS, D., GRONEMEYER, H., *Chem. Biol.* **1999**, *6*, 519–529.
- 13 WALLACE, A. C., LASKOWSKI, R. A., THORNTON, J. M., *Prot. Eng.* **1995**, *8*, 127–134.
- 14 TANNENBAUM, D. M., WANG, Y., WILLIAMS, S. P., SIGLER, P. B., *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 5998–6003.
- 15 TEDESCO, R., THOMAS, J. A., KATZENELLENBOGEN, B. S., KATZENELLENBOGEN, J. A., *Chem. Biol.* **2001**, *8*, 277–287.
- 16 SHI, Y., KOH, J. T., *Chem. Biol.* **2001**, *8*, 501–510.
- 17 SHI, Y., KOH, J. T., *J. Am. Chem. Soc.* **2002**, *124*, 6921–6928.
- 18 KOH, J. T., PUTNAM, M., TOMIC-CANIC, M., McDANIEL, C. M., *J. Am. Chem. Soc.* **1999**, *121*, 1984–1985.
- 19 RENAUD, J., ROCHEL, N., RUFF, M., VIVAT, V., CHEMBON, P., GRONEMEYER, H., MORAS, D., *Nature (London)* **1995**, *378*, 681–689.
- 20 KLAHOLZ, B. P., RENAUD, J. P., MITSCHLER, A., ZUSI, C., CHAMBON, P., GRONEMEYER, H., MORAS, D., *Nat. Struct. Biol.* **1998**, *5*, 199–202.
- 21 HWANG, J., WARSHEL, A., *Nature (London)* **1988**, *334*, 270–272.
- 22 PEET, D. J., DOYLE, D. J., COREY, D. R., MANGELSDORF, D. J., *Chem. Biol.* **1998**, *5*, 13–21.
- 23 DOYLE, D. F., BRAASCH, D. A., JACKSON, L. K., WEISS, H. E., BOEHM, M. F., MANGELSDORF, D. J., COREY, D. R., *J. Am. Chem. Soc.* **2001**, *123*, 11367–11371.
- 24 YANO, T., OUE, S., KAGAMIYAMA, H., *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 5511–5515.
- 25 OKAMOTO, A., HIGUCHI, T., HIROTSU, K., KURAMITSU, S., KAGAMIYAMA, H., *J. Biochem.* **1994**, *116*, 95–107.

- 26 SÜEL, G. M., LOCKLESS, S. W., WALL, M. A., RANGANATHAN, R., *Nat. Struct. Biol.* 2003, 10, 59–69.
- 27 SPENCER, D. M., WANDLESS, T. J., SCHREIBER, D. L., CRABTREE, G. R., *Science* 1993, 262, 1019–1024.
- 28 POLLOCK, R., GIEL, M., LINHER, K., CLACKSON, T., *Nature Biotech.* 2002, 20, 729–733.
- 29 CHOI, J., CHEN, J., SCHREIBER, S. L., CLARDY, J., *Science* 1996, 273, 239–242.
- 30 POLLOCK, R., ISSNER, R., ZOLLER, K., NATESAN, S., RIVERA, V. M., CLACKSON, T., *Proc. Natl. Acad. Sci. USA* 2000, 97, 13221–13226.
- 31 ABIDA, W. M., CARTER, B. T., ALTHOFF, E. A., LIN, H., CORNISH, V. W., *Chembiochem* 2002, 3, 887–895.
- 32 BAKER, K., BLECZINSKI, C., LIN, H., SALAZAR-JIMENEZ, G., SENGUPTA, D., KRANE, S., CORNISH, V. W., *Proc. Natl. Acad. Sci. USA* 2002, 99, 16537–16542.
- 33 SAWAYA, M. R., KRAUT, J., *Biochem.* 1997, 36, 586–603.
- 34 BLEDSOE, R. K., MONTANA, V. G., STANLEY, T. B., DELVES, C. J., APOLITO, C. J., McKEE, D. D., CONSLER, T. G., PARKES, D. J., STEWART, E. L., WILSON, T. M., LAMBERT, M. H., MOORE, J. T., PEARCE, K. H., XU, H. E., *Cell* 2002, 110, 93–105.
- 35 ULRICH, S. M., SALLEE, N. A., SHOKAT, K. M., *Bioorg. Med. Chem. Lett.* 2002, 12, 3223–3227.
- 36 XU, W., DOSHI, A., LEI, M., ECK, M. J., HARRISON, S. C., *Mol. Cell* 1999, 3, 629–638.
- 37 BISHOP, A. C., UBERSAX, J. A., PETSCH, D. T., MATHEOS, D. P., GREP, N. S., BLETHROW, J., SHIMIZU, E., TSLEN, J. Z., SCHULTZ, P. S., ROSE, M. D., WOOD, J. L., MORGAN, D. O., SHOKAT, K. M., *Nature (London)* 2000, 407, 395–401.
- 38 SCHINDLER, T., SICHERI, F., PICO, A., GAZIT, A., LEVITZKI, A., KURIYAN, J., *Mol. Cell* 1999, 3, 639–648.
- 39 REED, S. I., *Genetics* 1980, 95, 561–577.
- 40 MENDENHALL, M. D., HODGE, A. E., *Microbiol. Mol. Biol. Rev.* 1998, 62, 1191–1243.

9

Engineering Proteins to Promote Crystallization

LEI JIN and ROBERT E. BABINE

9.1

Introduction

Structure-based drug design has proven to be an important tool for lead identification, optimization and determination of structure-activity relationships in the drug discovery process [1–3]. Detailed three-dimensional structures of protein targets with ligands provide information on exactly how small molecules bind, and enable a more rational approach for medicinal chemists to make new compounds. The binding of a ligand to its target often results in large conformational changes in the protein, which cannot be predicted by current computational tools. Thus, there is a growing need for large numbers of ligand-protein crystal structures to guide the iterative cycles of structure-based drug design.

With recent advances in data collection (in house and at synchrotron sources) and automation software for structure determination, crystal structures can be solved relatively more easily and faster than before. Although crystallization robots and nanodrop technology allow more crystallization conditions to be screened in a timely fashion with small amounts of protein [4], obtaining quality crystals for the target protein and ligand-protein complexes remains a major challenge.

The general approach to crystallizing a protein is to first use commercially available screens containing a broad spectrum of conditions, such as those provided from Hampton Research [5, 6] (<http://www.hamptonresearch.com>) and Emerald BioStructures (<http://www.decode.com/emeraldbiostructures>), and then fine-tune the conditions by systematically adjusting the concentrations of protein and precipitants, varying pH values, additives, temperature and other conditions. However, one of the parameters that is often ignored is the protein sequence itself.

In the past, a common approach was to crystallize homologous proteins from different species if the desired protein from one source failed to crystallize. Kendrew pioneered this approach and selected sperm whale myoglobin for structural studies [7]. Campbell et al. crystallized almost every enzyme in the glycolytic pathway from different organisms [8]. In the classic Agouron paper on iterative protein crystallographic analysis [3], *E. coli* thymidylate synthase was used as a surrogate for the human enzyme due to difficulties in obtaining high quality crystals of the human protein. For proteins from agents that infect human cells, such as

viruses, there are often several serotypes of the same protein. This situation allows the crystallographer to screen several natural proteins for their ability to crystallize. For rhinovirus 3C protease, although the apo-structure was obtained from the protein from serotype 14, subsequent structural studies of the inhibitor complex were carried out mainly with the serotype 2 enzyme due to the relative ease in obtaining co-crystals and better diffraction properties of the serotype 2 crystals [9].

With advances in recombinant DNA technology, another approach gaining popularity is to use protein engineering to promote crystallization and for solving protein structures (reviewed in [10]). This chapter will discuss how protein engineering has become a powerful tool for improving the quality of protein crystals.

9.2

Removing Protein Heterogeneity by Truncation

Multi-domain proteins are often hard to crystallize due to the flexibility of the linkers that connect the distinct domains. A common approach is to express and crystallize each domain independently. In distinct-domain or single-domain proteins, flexible N- and C- termini and loops can also prevent the formation of crystals. Identification of the boundaries of domains and flexible loops is often crucial. This information can be determined from limited protease digestion experiments using proteases of different specificity and identification of proteolytic fragments by N-terminal amino acid sequencing and MALDI mass spectrometry. Secondary structure analysis and sequence alignment of proteins from different species can also help delineate boundary limits. If available, analysis of a low resolution crystal structure can determine, quite accurately, the boundary limits of flexible linkers and loops. DNA engineering techniques are ideal for removing these problem regions from the protein target to promote crystallization.

The ligand-binding domain of the GluR2 receptor produced crystals that diffracted to 1.5 Å resolution only after the domain boundaries were accurately identified by proteolysis and deletion mutagenesis [11]. In the case of the 24 kDa fragment of the DNA gyrase B subunit from *Staphylococcus aureus*, the wild type protein crystallized only in the presence of a ligand and diffracted to 3 Å resolution. Analysis of the structure revealed that there was no electron density for the N-terminal 23 amino acids and a loop comprising residues 109–127. Systematic removal of the residues in the loop was carried out by deletion mutagenesis. The 105–127 deletion mutant was crystallized and diffracted to 2 Å resolution [12]. The ligand-binding domain of human vitamin D receptor (VDR) has an insertion domain connecting helices H1 and H3. The length varies between 72 and 81 residues in the VDR family compared with 15–25 residues for the other nuclear hormone receptors. The sequence homology of this loop region was very low. This loop is accessible to proteases and is not likely to be well ordered based on secondary structure prediction. Removal of the flexible insertion domain led to a more soluble protein and yielded crystals that diffracted to 1.8 Å resolution [13].

Another example, which is directly involved in the process of structure-based drug design, is the re-engineering of human urokinase, a cancer target [14]. Initial crystals of human low molecular weight urokinase in complex with a peptidic inhibitor diffracted to 2.5 Å resolution. The active site is in close contact with another molecule related by a noncrystallographic 2-fold axis. The noncrystallographic and crystallographic packing effectively blocks the active site making it difficult to diffuse small molecules into the active site by soaking methods. Urokinase contains two chains, A and B, connected by a disulfide bond. The B chain is the serine protease domain. In the search for a new crystal form of human urokinase, the A-chain and the C-terminal region of the B-chain that were flexible in the initial structure, indicated by high B-factors, were deleted. This shortened urokinase with two point mutations (which will be discussed later) produced crystals in a different space group with a monomer in the asymmetric unit, an accessible active site for soaking experiments and diffraction to 1.5 Å resolution. This new crystal form was used to determine the co-crystal structures of urokinase with small molecule inhibitors.

9.3

Removing Protein Heterogeneity by Point Mutation

It is well known that post-translational modifications are another source of heterogeneity for proteins. There are at least 150 different kinds of covalent modifications of proteins, such as glycosylation, phosphorylation, S-thiolation, etc. [15]. Although the modification itself may be very small, like the addition of a single phosphate group, it may serve a physiological role and cause substantial conformational changes in the protein. Post-translational modification of a protein usually changes its molecular weight and can also change its charge. Different forms of the same protein can be equivalent to entirely foreign proteins where the crystallization process is concerned. The goal is to achieve the greatest degree of chemical and structural homogeneity of protein. Various enzymes can be used to remove an attached chemical group. For example, alkaline phosphatase can dephosphorylate and glycosidases can remove carbohydrate, but additional purification steps are required. A better way to ensure the homogeneity of the protein is to mutate residues that are subject to post-translational modification. In the urokinase example we mentioned previously [14], two point mutations were constructed. The N302Q mutation removed a glycosylation site, and C279A mutation prevented modification of the free sulfhydryl left when the disulfide bond between the A- and B-chains was broken by truncation of the A-chain. This truncated urokinase with two point mutations produced crystals diffracting to high resolution and suitable for ligand-soaking experiments for structure-based drug design.

The struggle to obtain the crystal structure of angiotensin-converting enzyme (ACE) is another example. ACE converts inactive angiotensin I to active angiotensin II which plays a critical role in the renin-angiotensin system. The many ACE inhibitors in clinical use were developed without knowledge of the ACE structure.

They are a first line of therapy for hypertension, heart failure, myocardial infarction and diabetic nephropathy. ACE is a highly glycosylated protein and has proved hard to crystallize for more than two decades. Unglycosylated ACE was shown to be inactive. There are two ACE isoforms, one in somatic tissues and one in sperm cells. The testis ACE (tACE) is identical to the portion of somatic ACE that is sufficient for its cardiovascular function, except for the N-terminal 36 residues. A 2.0 Å resolution structure of a truncated and chemically modified form of tACE recently became available [16, 17]. The mutant tACE (tACE Δ 36NJ) lacks the N-terminal 36 residues and the C-terminal hydrophobic transmembrane domain, was expressed in the presence of the α -glucosidase I inhibitor N-butyl-deoxynojirimycin and was treated with endoglycosidase H to remove all but the terminal N-acetylglucosamine residues. It was homogeneous on SDS-PAGE and retained full enzymatic activity, but the expression level was very low. In order to identify the minimum glycosylation requirements for the expression of enzymatically active tACE, the glycosylation sites (Asn-X-Ser/Thr) on tACE Δ 36 were systematically mutated by substituting Asn with Gln [18]. Two of the underglycosylated tACE Δ 36 mutants were expressed at significantly high levels and produced crystals that were isomorphous with the tACE Δ 36NJ crystals, although the resolutions were not as good (2.8 and 3.0 Å).

9.4

Improving Crystal Packing by Point Mutation

It is well known that homologous proteins from different species do not behave in the same way with regard to crystallization. Often the differences are a few surface residues. One of the first examples of using point mutation to promote crystallization was the crystallization of human H-chain ferritin [19]. Human ferritin produced only poor quality crystals under conditions in which rat and horse ferritins yielded good crystals. Introduction of a K86Q mutation at the surface of the human ferritin allowed the formation of metal bridges in the crystalline lattice that had been observed in the related rat and horse L-chain ferritin crystal structures and yielded high quality crystals that diffracted to 1.9 Å resolution. A second example is the elegant study from the Villafranca group demonstrating that a single surface amino acid change had a dramatic effect on the crystallization of human thymidylate synthase (TS) [20]. The structure of the apo-enzyme of human TS had a disordered active site and was unusable for drug design. Eleven amino acids were chosen for point mutation based on the structures of the *E. coli* and human TS, as well as sequence alignment over eight species of TS. All the amino acids were on the surface of human TS and were not conserved. The mutations changed either the charge or polarity of the wild type amino acid. Some of the mutants crystallized under different conditions with different space groups to those of the wild type TS and provided ligand-binding information. This was one of the early examples of using site-directed mutagenesis to change the solubility of a particular protein and introduce more favorable crystal contacts.

Carugo and Argos compared the protein-protein contacts in monomeric protein crystal structures with the physiological protein-protein contacts in oligomers [21]. They found that the area of the protein surface involved in packing contacts was generally smaller and its amino acid composition indistinguishable from that of the protein surface accessible to the solvent. Unlike physiological protein-protein recognition processes, protein crystallization depends on seemingly “random” protein-protein interactions. Crystallization trials on bovine pancreatic ribonuclease confirmed that nearly the entire protein surface could be involved in crystal packing contacts [22]. These make the “design” of good crystal packing very difficult, if not impossible. There have been several examples of mutations introduced by error in the PCR process that improved the quality of protein crystals and increased the resolution significantly [23, 24].

Sequence alignment of related proteins from different species can provide valuable information for determining possible residues for mutation. Hydrophobic residues that are not conserved and are probably surface-exposed are good candidates. These residues are often the cause of protein aggregation. In the case of the integrase of human HIV-1, in order to overcome the aggregation problem of the native protein, the hydrophobic residues were systematically replaced with either alanine or lysine [25]. A single Phe-to-Lys mutant was highly soluble and crystallized. It provided a crystal structure to 2.5 Å resolution.

Wild type human leptin aggregates extensively. A single mutation of Trp-to-Glu at residue 100 on the surface of the molecule dramatically improved the solubility, allowed crystallization and provided structural information to 2.4 Å resolution [26]. The tryptophan residue was not conserved in the sequence alignment of leptin from ten different species. Mutagenesis studies on the 24 kDa fragment of the DNA gyrase B subunit from *E. coli* further confirmed that single amino acid changes on the surface can dramatically change crystallization properties [27]. Three out of nine mutants produced crystals with different morphology to that of the wild type protein. The F104Y mutant diffracted to 2.44 Å resolution, an improvement over the 2.9 Å resolution obtained from the crystals of the wild type protein.

A more general approach of using site-directed mutagenesis to improve the crystallizability of a protein has been applied to Rho-specific guanine nucleotide dissociation inhibitor (RhoGDI) [28, 29]. The hypothesis was that surface residues with high conformational entropy, such as lysines and glutamates, are likely to impair the formation of protein-protein contacts in crystal lattices. Replacing lysines and glutamates with alanine-created epitopes that could be incorporated into crystal contacts with lower entropic penalty, as well as removing charges to alter the protein’s isoelectric point and solubility properties. There were no crystals obtained with the wild type protein of RhoGDIΔ66. With single and multiple mutations of RhoGDIΔ66, several novel crystal forms were generated. One lysine mutant diffracted to 2.0 Å resolution and a glutamate mutant diffracted to 1.2 Å resolution. Interestingly, the new crystal contacts often incorporated the epitopes that carried the mutated sites. Lysines and glutamates can be targeted even without knowledge of the tertiary structure of a protein since both amino acids are almost

exclusively located on the protein surface. Clusters of lysine and/or glutamate residues in close proximity in the sequence are good targets for mutagenesis. The successful crystallization of the RGS-like domain from PDZ-RhoGEF is a good example of such an approach. The wild type protein did not crystallize. One triple mutant (K463A, E465A and E466A) out of five cluster mutations that were made was readily crystallized and provided structural information [30, 31].

All of the above examples demonstrate that protein engineering is a powerful tool for improving protein crystallizability. With recent advances in high throughput methods for cloning, expression, purification and crystallization of proteins, multiple truncation constructs with or without random or rationally selected point mutations are easier to obtain. Engineering the target protein should be considered one of the variables for obtaining target structures in addition to more classical screening approaches that vary the solution conditions in the crystallization experiment.

9.5

Acknowledgements

We would like to thank Drs. Joan Gorga and Michael Rynkiewicz for critical reading of the manuscript.

9.6

References

- 1 BABINE, R. E., BENDER, S. L., Molecular recognition of protein-ligand complexes: applications to drug design. *Chem. Rev.* **1997**, *97*, 1359–1472.
- 2 GREER, J., ERICKSON, J. W., BALDWIN, J. J., VARNEY, M. D., Application of the three-dimensional structures of protein target molecules in structure-based drug design. *J. Med. Chem.* **1994**, *37*, 1035–1054.
- 3 APPELT, K., BACQUET, R. J., BARTLETT, C. A., BOOTH, C. L., FREER, S. T., FUHRY, M. A. M., GEHRING, M. R., HERRMANN, S. M., HOWLAND, E. F., JANSON, C. A., JONES, T. R., KAN, C.-C., KATHARDEKAR, V., LEWIS, K. K., MARZONI, G. P., MATTHEWS, D. A., MOHR, C., MOOMAW, E. W., MORSE, C. A., OATLEY, S. J., OGDEN, R. C., REDDY, M. R., REICH, S. H., SCHOETTLIN, W. S., SMITH, W. W., VARNEY, M. D., VILAFRANCA, J. E., WARD, R. W., WEBBER, S., WEBBER, S. E., WELSH, K. M., WHITE, J., Design of enzyme inhibitors using iterative protein crystallographic analysis. *J. Med. Chem.* **1991**, *34*, 1925–1934.
- 4 LESLEY, S. A., KUHN, P., GODZIK, A., DEACON, A. M., MATHEWS, I., KREUSCH, A., SPRAGGON, G., KLOCK, H. E., McMULLAN, D., SHIN, T., VINCENT, J., ROBB, A., BRINNEN, L. S., MILLER, M. D., MCPHILLIPS, T. M., MILLER, M. A., SCHEIBE, D., CANAVES, J. M., GUDA, C., JAROSZEWSKI, L., SELBY, T. L., EISLIGER, M. A., WOOLEY, J., TAYLOR, S. S., HODGSON, K. O., WILSON, I. A., SCHULTZ, P. G., STEVENS, R. C., Structural genomics of the *Thermotoga maritima* proteome implemented in a high-throughput structure determination pipeline. *Proc. Natl. Acad. Sci USA* **2002**, *99*, 11664–11669.
- 5 JANCARIK, J., KIM, S.-H.. Sparse matrix sampling: a screening method for crystallization of proteins. *J. Appl. Cryst.* **1991**, *24*, 409–411.

- 6** CUDNEY, R., PATEL, S., WEISGRABER, K., NEWHOUSE, Y., MCPHERSON, A. Screening and optimization strategies for macromolecular crystal growth. *Acta Cryst.* **1994**, *D50*, 414–423.
- 7** KENDREW, J.C., PARRISH, R.G., MARRACK, J.R., ORLANDS, E.S. *Nature* **1954**, *174*, 946–949.
- 8** CAMPBELL, J.W., DUEE, E., HODGSON, G., MERCER, W.D., STAMMERS, D.K., WENDELL, P.L., MUIRHEAD, H., WATSON, H.C., X-ray diffraction studies on enzymes in the glycolytic pathway. *Cold Spring Harbor Symp. Quant. Biol.* **1972**, *36*, 165–170.
- 9** MATTHEWS, D.A., DRAGOVICH, P.S., WEBBER, S.E., FUHRMAN, S.A., PATICK, A.K., ZALMAN, L.S., HENDRICKSON, T.F., LOVE, R.A., PRINS, T.J., MARAKOVITS, J.T., ZHOU, R., TIKHE, J., FORD, C.E., MEADOR, J.W., FERRE, R.A., BROWN, E.L., BINFORD, S.L., BROTHERS, M.A., DELISLE, D.M., WORLAND, S.T. Structure-assisted design of mechanism-based irreversible inhibitors of human rhinovirus 3C protease with potent antiviral activity against multiple rhinovirus serotypes. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 11000–11007.
- 10** PRICE, S.R., NAGAI, K., Protein engineering as a tool for crystallography. *Curr. Opin. Biotechnol.* **1995**, *6*, 425–430.
- 11** CHEN, G.Q., SUN, Y., JIN, R., GOUAUX, E., Probing the ligand binding domain of the GluR2 receptor by proteolysis and deletion mutagenesis defines domain boundaries and yields a crystallizable construct. *Protein Sci.* **1998**, *7*, 2623–2630.
- 12** DALE, G.E., KOSTREWA, D., GSELL, B., STIEGER, M., D'ARCY, A., Crystal engineering: deletion mutagenesis of the 24 kDa fragment of the DNA gyrase B subunit from *Staphylococcus aureus*. *Acta Cryst.* **1999**, *D55*, 1626–1629.
- 13** ROCHEL, N., WURTZ, J.M., MITSCHLER, A., KLAHOLZ, B., MORAS, D., The crystal structure of the nuclear receptor for vitamin D bound to its natural ligand. *Mol. Cell* **2000**, *5*, 173–179.
- 14** NIENABER, V., WANG, J., DAVIDSON, D., HENKIN, J., Re-engineering of human urokinase provides a system for structure-based drug design at high resolution and reveals a novel structural subsite. *J. Biol. Chem.* **2000**, *275*, 7239–7248.
- 15** HAN, K.K., MARTINAGE, A., Post-translational chemical modification(s) of proteins. *Int. J. Biochem.* **1992**, *24*, 19–28.
- 16** YU, X.C., STURROCK, E.D., WU, Z., BIEGMANN, K., EHLLERS, M.R., RIORDAN, J.F. Identification of N-linked glycosylation sites in human testis angiotensin-converting enzyme and expression of an active deglycosylated form. *J. Biol. Chem.* **1997**, *272*, 3511–3519.
- 17** NATESH, R., SCHWAGER, S.L., STURROCK, E.D., ACHARYA, K.R., Crystal structure of the human angiotensin-converting enzyme-lisinopril complex. *Nature* **2003**, *421*, 551–554.
- 18** GORDON, K., REDELINGHUYS, P., SCHWAGER, S.L., PAPAGEORGIO, A.C., NATESH, R., ACHARYA, K.R., STURROCK, E.D., Deglycosylation, processing, and crystallization of human testis angiotensin-converting enzyme. *Biochem. J.* **2003**, *371*, 437–442.
- 19** LAWSON, D.M., ARTYMIUK, P.J., YEWDALL, S.J., SMITH, J.M., LIVINGSTONE, J.C., TREFFRY, A., LUZZAGO, A., LEVI, S., AROSIO, P., CESARENI, G., THOMAS, C.D., SHAW, W.V., HARRISON, P.M., Solving the structure of human H ferritin by genetically engineering intermolecular crystal contacts. *Nature* **1991**, *349*, 541–544.
- 20** McELROY, H.E., SISSON, G.W., SCHOETTLIN, W.E., AUST, R.M., VILLAFRANCA, J.E. Studies on engineering crystallizability by mutation of surface residues of human thymidylate synthase. *J. Cryst. Growth* **1992**, *122*, 265–272.
- 21** CARUGO, O., ARGOS, P., Protein-protein crystal-packing contacts. *Protein Sci.* **1997**, *6*, 2261–2263.
- 22** CROSIO, M.P., JANIN, J., JULLIEN, M., Crystal packing in six crystal forms of pancreatic ribonuclease. *J. Mol. Biol.* **1992**, *228*, 243–251.
- 23** BRAIG, K., OTWINOWSKI, Z., HEGDE, R., BOISVERT, D.C., JOACHIMIAK, A., HORWICH, A.L., SIGLER, P.B., The crystal structure of the bacterial chaperonin GroEL at 2.8 Å. *Nature* **1994**, *371*, 578–586.

- 24** OUBRIDGE, C., ITO, N., TEO, C. H., FEARNSLEY, I., NAGAI, K., Crystallisation of RNA-protein complexes. II. The application of protein engineering for crystallisation of the U1A protein-RNA complex. *J. Mol. Biol.* **1995**, *249*, 409–423.
- 25** JENKINS, T. M., HICKMAN, A. B., DYDA, F., GHIRLANDO, R., DAVIES, D. R., CRAIGIE, R., Catalytic domain of human immunodeficiency virus type 1 integrase: identification of a soluble mutant by systematic replacement of hydrophobic residues. *Proc. Natl Acad. Sci. USA* **1995**, *92*, 6057–6061.
- 26** ZHANG, F., BASINSKI, M. B., BEALS, J. M., BRIGGS, S. L., CHURGAY, L. M., CLAWSON, D. K., DiMARCHI, R. D., FURMAN, T. C., HALE, J. E., HSIUNG, H. M., SCHONER, B. E., SMITH, D. P., ZHANG, X. Y., WERY, J., SCHEVITZ, R. W., CRYSTAL STRUCTURE OF THE OBESITY PROTEIN LEPTIN-E100. *Nature* **1997**, *387*, 206–209.
- 27** D'ARCY, A., STIHLE, M., KOSTREWA, D., DALE, G. Crystal engineering: a case study using the 24 kDa fragment of the DNA gyrase B subunit from Escherichia coli. *Acta Cryst.* **1999**, *D55*, 1623–1625.
- 28** LONGENECKER, K. L., GARRARD, S. M., SHEFFIELD, P. J., DEREWENDA, Z. S., Protein crystallization by rational mutagenesis of surface residues: Lys to Ala mutations promote crystallization of RhoGDI. *Acta Cryst.* **2001**, *D57*, 679–688.
- 29** MATEJA, A., DEVEDJIEV, Y., KROWARSCH, D., LONGENECKER, K., DAUTER, Z., OTLEWSKI, J., DEREWENDA, Z. S., The impact of Glu → Ala and Glu → Asp mutations on the crystallization properties of RhoGDI: the structure of RhoGDI at 1.3 Å resolution. *Acta Cryst.* **2002**, *D58*, 1983–1991.
- 30** LONGENECKER, K. L., LEWIS, M. E., CHIKUMI, H., GUTKIND, J. S., DEREWENDA, Z. S., Structure of the RGS-like domain from PDZ-RhoGEF: linking heterotrimeric γ protein-coupled signaling to Rho GTPases. *Structure* **2001**, *9*, 559–569.
- 31** GARRARD, S. M., LONGENECKER, K. L., LEWIS, M. E., SHEFFIELD, P. J., DEREWENDA, Z. S., Expression, purification, and crystallization of the RGS-like domain from the Rho nucleotide exchange factor, PDZ-RhoGEF, using the surface entropy reduction approach. *Protein Expr. Purif.* **2001**, *21*, 412–416.

10

High-throughput Crystallography

HARREN JHOTI

10.1

Introduction

In the late 1990s the concept of high-throughput crystallography was first proposed. It was inspired by the significant achievements in the genomics arena in which the DNA of several organisms, including homo sapiens, were being sequenced using high-throughput methodologies [1]. In fact, earlier in the decade the broader application of “high-throughput” methodologies became commonplace within the pharmaceutical industry with the birth of high-throughput chemistry and screening [2]. During this period X-ray crystallography was unable to keep pace and the other drug discovery technologies being performed in a high-throughput mode became the focus of many pharmaceutical companies. More recently, however, there has been a resurgence in interest for using structure-based approaches, driven largely by major technological developments in crystallography, which has resulted in many new crystal structures of therapeutic targets. Furthermore, the ability to rapidly obtain crystal structures of a target protein in complex with small molecules is driving a new wave of structure-based drug design.

The need for “high-throughput crystallography” was originally driven not by the drug discovery sector but by the “structural genomics” initiatives that were set-up with the aim of solving crystal structures of representatives from protein families for which little or no structural information was known [3]. In short, the original goal of the structural genomics projects was to obtain crystal structures for all known protein families. Most of these initiatives focused on different bacterial genomes but they all shared a common need to develop new methods and technologies ranging from molecular biology to protein production and structure determination. In this chapter some of these technology developments will be described briefly and how they have enabled high-throughput X-ray crystallography to be applied within the drug discovery process discussed.

10.2

Technological Advances

There are many areas in which new technologies and methods are being developed to enable high-throughput structure determination by X-ray crystallography [4, 5]. The process from gene to crystal structure is clearly multidisciplinary and advances in molecular biology, biochemistry, crystallization, X-ray data collection and computational analysis underpin high-throughput X-ray crystallography. Many of these advances are being made in the public-funded initiatives focused on structural genomics. These programs began several years ago when Japan began building the RIKEN Genomic Sciences Center in Yokohama and the USA began funding structural genomics pilot projects under the NIGMS (National Institute of General Medical Sciences) Protein Structure initiative [6]. Similar programs are underway in other countries, for example, the Protein Structure Factory in Germany is focusing on solving structures of human proteins in collaboration with the German Human Genome Project (DHGP) and the European Union has recently funded the Structural Proteomics in Europe (SPINE) program, which is focused on both eukaryotic and prokaryotic proteins (www.spineurope.org).

The main focus of these structural genomics initiatives is to automate all steps of structural biology and to determine structures of proteins, using either NMR or X-ray crystallography, for which no three-dimensional information exists [7]. In addition to these public-funded centers, some specialist biotechnology companies were also formed in 1999 to pursue structural genomics programs. These include Structural GenomiX and Syrrx (San Diego, CA, USA), who are making significant developments in the automation of streamlining the gene to crystal structure process [8]. Other companies, such as Astex Technology (Cambridge, UK) and Abbott Laboratories (Abbott Park, IL, USA) have also developed high-throughput crystallography as a novel approach for fragment-based lead discovery (see below).

10.2.1

Clone to Crystal

The process of going from the gene of interest to a suitable crystal of the protein can be broadly separated into two stages. Firstly, significant amounts of chemically (and conformationally) homogenous protein need to be generated in a robust and reproducible process. Secondly, the protein then needs to be crystallized in order for the X-ray diffraction experiments to be performed. These two stages will now be briefly discussed.

10.2.1.1 Protein Production

Expression, purification and characterization of a novel protein in a quantity and form that is suitable for crystallization and X-ray analysis probably occupies 70–80% of the time in most academic structural biology groups. Consequently, methods for high-throughput parallel expression and purification are now being devel-

oped in many laboratories [9, 10]. Typically, 10–50 mg of protein are required to screen sufficient numbers of crystallization conditions to obtain the initial crystals. Traditionally, a handful of different DNA constructs would be generated, after analysis of the protein sequence, in an attempt to remove flexible regions of the protein that may hinder crystallization. Each construct would then be tested for expression in the host cell, usually *Escherichia coli* or insect cells, and the level of functional protein analyzed using bioassay and polyacrylamide gel electrophoresis (PAGE). In the past these different constructs would be analyzed sequentially, but recent developments in molecular biology, based on DNA recombination, now enable high-throughput approaches for cloning and expression where tens to hundreds of DNA constructs can easily be generated to test in parallel for high expression. Protein purification has also seen significant improvements owing to the development of affinity tags that allow proteins to be purified significantly faster and more efficiently [11]. Automated methods based on affinity chromatography, such as a nickel-nitrilotriacetic acid (Ni-NTA) column, are now available which can process samples in parallel using a 96-well format [10].

Most of these developments in protein expression have been designed to improve the throughput and efficiency of current *in vivo* methods. However, recent developments using *in vitro* systems have been reported and could potentially offer a new method for routine protein production. For several years there has been interest in bacterial and eukaryotic cell-free translation systems in which cellular components contained in artificial cells are used to generate protein but these systems have suffered from problems with stability and efficiency [12]. If they could be made to work, these systems could express proteins which would otherwise interfere with cell physiology and so offer advantages over traditional *in vivo* methods. For example, proteases that are toxic to the cell could be inhibited by small molecules during expression, and post-translational modifications such as phosphorylation could be better controlled using these cell-free systems. Furthermore, these systems can be used to produce stable-isotope labelling of proteins for NMR spectroscopy [12]. Recently, workers have overcome some of the stability problems and have developed a eukaryotic cell-free expression system, using components of the wheat seed, which was able to maintain productive translation for 14 days, resulting in several milligrams of protein [13]. These systems also lend themselves to robotic automation, which means the amounts of protein typically produced using this approach could be significantly scaled-up.

10.2.1.2 Crystallization

Crystallization is often regarded as a slow, resource-intensive step with low success rates in obtaining good X-ray quality crystals. However, much of the failure during this step can be attributed to poor quality protein samples that often have some level of chemical or conformational heterogeneity. The use of biophysical methods, such as fluorescence, circular dichroism, dynamic light scattering and mass spectrometry, to characterize the protein sample rigorously is a key step before performing crystallization experiments [14, 15].

However, significant advances in automation have also improved the process of crystallization with the new generation of robots able to sample the multidimensional space efficiently by varying precipitant concentration, buffers and pH, all variables known to affect crystallization. For example, a crystallization robot called Agincourt, jointly developed by the Genomics Institute of the Novartis Research Foundation (San Diego, CA, USA) and Syrrx, is able to set-up and observe 60 000 samples per day. Several other groups, such as scientists at Emerald Biostructures (Bainbridge Island, WA, USA), the Protein Structure Factory (Berlin, Germany) as well as commercial vendors such as Gilson-Cyberlab (Middleton, USA), have developed similar robots. Each of these crystallization robots has been designed to optimize and speed-up the experimental process and can prepare and monitor many thousands of experiments per day.

One key advantage of some of these robotic systems has been the capacity to perform nanocrystallization. For example, the Agincourt robot has been designed to dispense volumes in the 20–100 nL range, which significantly reduces the amount of protein required for screening of crystallization conditions [16]. Using this robot, crystals of Aurora-A, FAK and EphA2 protein kinases were grown in vapor diffusion experiments with 50 nL sitting drops, with 1–3 mg of protein being sufficient to set up over 1200 crystallization conditions [17]. The small drop size volume was also reported to accelerate crystal growth with maximal crystal size being reached within 48 hours rather than many days or weeks. These crystals were suitable for X-ray diffraction studies and were used to determine the structures of the protein kinases to high resolution [17]. Other groups have reported developing batch nanocrystallization methods where total crystallization volumes as low as 1 nL are being used [18].

One major technical challenge that nanocrystallization, in particular, has produced is that of automated crystal detection. Traditionally, a crystallographer would need to check each crystallization drop manually for signs of crystals or crystalline precipitant using a microscope. This clearly becomes impractical if the drops are 10–50 nL in size. Automated methods are being developed that allow the user to monitor and analyze the crystallization experiment [19]. Automatic crystal identification has remained a challenge as the size and morphology of crystals can vary greatly. Two approaches that are being developed for crystal detection involve edge-detection and birefringence as protein crystals often have sharp boundaries and are capable of altering polarized light [18, 20]. An ingenious method for analyzing the X-ray diffraction quality of a crystal while still in the crystallization drop has also been reported [21].

10.2.2

Crystal to Structure

Once X-ray quality crystals have been grown data collection using several wavelengths or derivatives is required in order to obtain the protein structure. X-ray data collection has been revolutionized in the last decade by both better X-ray sources and detectors. Third generation synchrotrons are now available across the

world which provide high intensity X-ray beams allowing the data collection time to be significantly reduced [22]. Synchrotron radiation coupled with charged-coupled device (CCD) detectors have allowed complete X-ray datasets for a crystal to be collected and processed within hours instead of days. Similarly, laboratory X-ray sources have also been significantly improved due largely to better optics [23].

Phase determination has also become dramatically easier by the application of synchrotron radiation to single and multi-wavelength anomalous diffraction techniques, known as SAD and MAD, respectively [24]. In fact, recent reports indicate that these anomalous diffraction methods can also be carried out on laboratory X-ray sources [25]. Elegant molecular biology methods for inserting heavy atoms such as selenium into the protein have been developed using auxotrophic cell expression systems [26]. This has replaced the approach of heavy atom derivatization of native protein crystals, which was often a major bottleneck in the novel structure determination of a protein. The percentage of protein structures solved using these methods, such as selenomethionine incorporation, has been increasing over the past decade [27].

Once a suitable electron density map has been generated many new methods of electron density interpretation and model-building are then available, which allow rapid and automated construction of protein models without the need for significant manual intervention [28]. In some cases, protein models have been automatically built into electron density maps in a matter of hours, rather than manually which often takes days or weeks. However, as much of this development has been in academic groups, most of these software tools focus only on the protein electron density and are unable to perform automated interpretation of ligand electron density. This has remained a significant problem for industrial groups who may be performing protein-ligand crystallography for as much as 80% of their time. To address this problem, specialized software has been developed such as the x-ligand module in Quanta from Accelrys Inc. (San Diego, CA, USA) and AutoSolve® from Astex. AutoSolve® allows automated analysis of the differences in electron density to detect binding of small molecules to the protein [29]. In the event of ligand binding, the software generates a correctly parameterized model for the ligand, which is then fitted into the electron density and the protein-ligand complex refined. AutoSolve® also identifies and models any solvent molecules and protein conformational movement that may have occurred on ligand binding.

10.2.3

Progress in Structural Genomics

From the beginning the structural biology community has been divided over the structural genomics programs, with many leading scientists critical of investing so heavily in these centers, and claiming these exercises are analogous to “stamp-collecting”. The funding is truly unprecedented with Japan promising to fund its RIKEN programs and eight new centers with \$100 million per year over the next five years and the NIGMS funding its centers at level of \$50 million a year [6]. The critics also voiced concern over the stated goals of the programs, for example,

the claim from NIGMS was to generate 10000 structures for unique proteins in 10 years and was considered by some to be unrealistic. Many felt that solving protein crystal structures was very different from sequencing genomes, and that the gain in productivity seen from “industrialization” in the human genome project would not extend to structural biology.

A recent conference to review the progress in structural genomics indicated that some of these initial goals were indeed over-ambitious [30]. Furthermore, the productivity of many of these centers has been disappointing with many reporting similar problems in the different parts of the “gene-to-structure process”. For example, the Northeast Structural Genomics Consortium (NESGC) led by Gaetano Montelione (Rutgers University, USA), reported that from 5187 DNA targets, 917 proteins have been purified from which only 50 structures have been solved using NMR or X-ray crystallography. Many other structural genomics centers reported similar attrition rates at this conference, which indicates that the new high-throughput technologies are not delivering, or at least not yet. The supporters of these public initiatives claim it is too early to pass judgement and point to 2004 when they believe the technology will be mature enough to be truly evaluated.

The privately based biotechnology companies have, however, reported better progress leading some observers to conclude that these “industrial-scale” programs are not best performed in academic centers. Syrrx, for example, claim to have determined 56 structures of unique proteins in the last 8 months. Although most of these structures are thought to be of bacterial origin, the company has also determined the structures of some important human protein kinases such as Aurora-A, FAK and EphA2 [17]. Structural GenomiX also claim to have solved more than 100 crystal structures using their high-throughput crystallography platform, again most are believed to be of bacterial proteins [29]. Astex has focused its high throughput crystallography methods solely on human proteins that are key to drug discovery and recently announced that it had solved the structures of human cytochrome P450 isozymes 2C9 and 3A4. The human cytochrome P450 enzymes play a vital role in drug metabolism and so have been of key interest to pharmaceutical companies for many years. However, despite many attempts to determine their structures, these proteins have remained intractable yielding only to high throughput crystallography methods.

Even if the original goals of the structural genomics initiatives prove to be over-ambitious, there is little doubt that the momentum, and technologies, generated by these programs will filter into the wider structural biology community and result in a significant increase in the number of protein structures. Currently, the Protein Data Bank (PDB) holds around 19000 protein structures, most of which have been determined using X-ray crystallography, and this number is expected to rise exponentially in the coming years (Fig. 10.1) [31]. Owing to this growing wealth of protein structure data, it is increasingly possible that a therapeutic target of interest to drug discovery scientists will have had its three-dimensional structure determined. Therefore the focus is likely to shift from solving the structure to how protein crystal structures can be exploited in drug discovery.

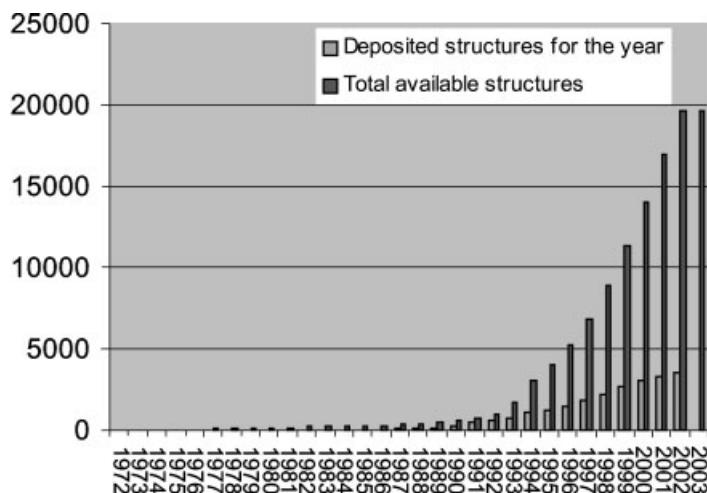


Fig. 10.1 Growth in the Protein Data Bank. For many years the number of protein structures being determined and deposited into the PDB was linear, however, with the advent of major technology advances over the last decade the deposition rate has become exponential. (Reproduced from the Protein Data Bank [31]).

10.3

High-throughput Crystallography in Lead Discovery

With the expectation that the structure of a drug target is probably going to be known, some groups have been developing methods for high-throughput crystallography for a very different purpose. Groups at Abbott and Astex have developed methods that focus on significantly improving the throughput of protein-ligand crystallography. This can be a bottleneck in drug discovery programs that utilize structure-based drug design methods. In many of these programs the structural information on the binding mode of compounds often lags behind the chemical synthesis, which reduces its impact. By obtaining crystal structures of protein-ligand complexes in a timelier manner, lead optimization can be performed more effectively.

Furthermore, although the structure of the native target protein is a useful start to guide a lead discovery program, the maximum value is only derived from structures of the protein in complexes with potential lead compounds. This is due to the fact that many proteins undergo some level of conformational movement on ligand binding which has proved very difficult to predict from the native structure alone. Also, water molecules often play a key role in the interactions between small molecules and proteins and their positions need to be established experimentally. The ability to determine crystal structures of protein-ligand complexes rapidly can not only effectively guide the lead optimization phase but also allows a new application of X-ray crystallography in drug discovery, that of a screening technology for fragment-based lead discovery [32].

10.3.1

Protein-Ligand Crystal Structures

The most reliable approach to determining the structure of a protein–ligand complex is either by co-crystallization or by soaking the ligand into the preformed crystal. However, when X-ray crystallography is used as a method for ligand screening, the soaking option is much preferred. After collecting the X-ray data from a protein crystal exposed to a ligand, the next step is to analyze and interpret the resulting electron density. This step is often time consuming and requires a crystallographer to spend several days assessing the data from a single protein-ligand experiment. This is a key bottleneck to using X-ray crystallography as a method for screening compounds. Technology advances have now been made to automate and accelerate this step. As mentioned earlier, software tools such as Quanta from Accelrys Inc. (San Diego, CA, USA) and AutoSolve® from Astex (Cambridge, UK) can assist the crystallographer in the analysis and interpretation steps.

Another area of automation that has assisted in the high throughput crystallography of proteins and protein–ligand complexes involves hardware for mounting and removing crystals from X-ray detectors, tasks currently performed by scientists. Robotic systems have been developed which can remove cryogenically frozen crystals from a storage dewar, place and correctly align the crystal on the X-ray detector and remove after the X-ray data have been collected, ready for the next crystal. Several of these systems have been developed by groups at synchrotron beamlines, where the data collection is often completed in a matter of minutes and the time-consuming step has been scientists having to change crystals [33]. However, with the advent of more sensitive X-ray detectors and more powerful laboratory X-ray sources, in-house data collection times have also been significantly reduced, so that a crystal-changing robot can have a significant impact on productivity within a laboratory setting. In fact, the Automated Crystal Transport, Orientation and Retrieval (ACTOR) robot, which was developed by the Abbott group, was one of the first systems and was designed for use on a laboratory X-ray source [34]. ACTOR has subsequently been licensed to Rigaku-MSC (Woodlands, TX, USA), a leading X-ray equipment manufacturer, who has made further improvements and now markets this robot for both laboratory and synchrotron X-ray sources (Fig. 10.2).

The integration of these software and hardware technological developments can allow protein-ligand crystallography to be performed with greater automation and at a significantly faster pace in the laboratory setting. For example, the Abbott group have reported collecting X-ray data from 18 crystals of dihydronoepterin al-dolase in 42 hours using a Rigaku rotating anode source with a MarCCD detector (MarUSA, Evanston, IL, USA) and the original ACTOR system [34]. Similarly, the Astex group have a hardware configuration based on a Jupiter 140 CCD X-ray detector on a Rigaku Ru-300 HR 5KW rotating anode laboratory X-ray source with an ACTOR robot (all supplied by Rigaku-MSC). X-ray data were recently collected using this set-up from crystals of protein tyrosine phosphatase 1B (PTP1B), a key therapeutic target for type 2 diabetes, as part of a fragment screening campaign.

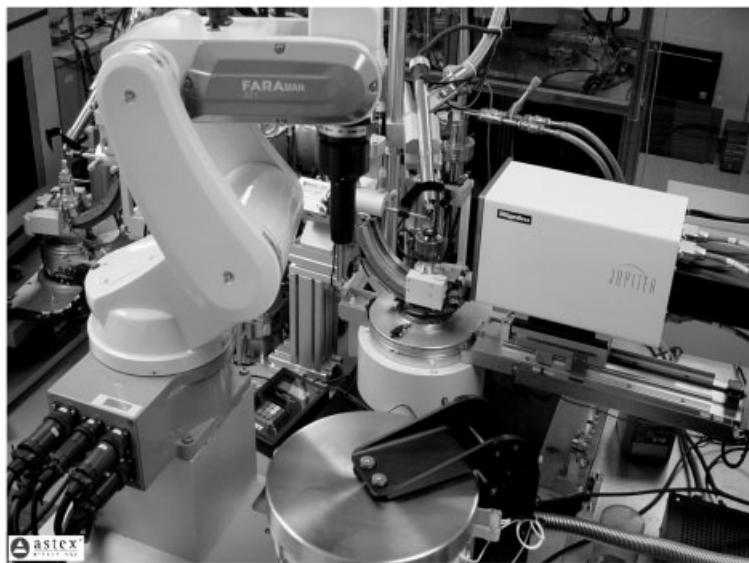


Fig. 10.2 Automation for X-ray data collection. The ACTOR robot coupled with the new generation Jupiter CCD detectors from Rigaku-MSC (Woodlands, TX, USA) allows rapid, unmanned data collection from laboratory x-ray sources [33].

X-ray data from 52 PTP1B crystals were collected and processed in 48 hours without manual intervention (personal communication). These data were analyzed simultaneously for ligand binding using AutoSolve®, which in the event of binding is able to generate a refined protein-ligand crystal structure automatically. To further optimize the impact of protein-ligand data, crystal structures should then be automatically presented to the medicinal chemists using a web-based molecular viewer such as AstexViewer™, which is capable of displaying key protein-ligand interactions and also electron density [35]. This level of automation and throughput is required if X-ray crystallography is to be used effectively as a screening technology for fragment-based lead discovery.

10.4 Fragment-Based Lead Discovery

There is growing interest in the use of molecular fragments for lead discovery. One reason for this interest is due to a problem that is evident in the nature of “hits” identified from traditional bioassay-based high throughput screens (HTS). The average MW of successful drugs in the World Drug Index is in the low 300s, which is similar to the average MW in current corporate compound collections [36]. This implies that corporate compound collections have evolved to be broadly “drug-like” instead of “lead-like” with respect to MW and other features. Therefore, an HTS hit from a compound collection with μM affinity towards the target

may well already have an average drug MW, however, it is likely that the MW will increase significantly during the lead optimization process, leading to poor drug-like properties with respect to solubility, absorption and clearance [37].

In order to address this issue several groups have been developing methods to identify low MW fragments (MW 100–250) that could be efficiently optimized into novel lead compounds possessing good drug-like properties. These molecular fragments would by definition have limited functionality and will therefore exhibit weaker affinity (typically in the 100 μm – mM range). This level of affinity is outside of the normal HTS sensitivity range and as such cannot routinely be identified in standard bioassays due to the high concentration of compound that would be required, interfering with the assay and leading to significant false positives. However, biophysical methods such as NMR and X-ray crystallography are ideal for detecting weakly binding molecules. Different libraries of molecular fragments can be used to target a particular protein. For example, computational analysis of the proteininactive site using pharmacophoric mapping or protein-ligand docking methods can be used to generate a focused-fragment library for a target or target class. Alternatively, diversity-based fragment libraries can be used that have been generated by sampling scaffolds in known drug molecules [38]. As in conventional drug discovery, the use of different fragment libraries in a screening campaign is influenced by the varying importance of novelty in the initial lead compound.

A key advantage of using molecular fragments for screening is the significant amount of chemical space that is sampled using a relatively small library of compounds, typically ranging from 500–1000 compounds. For example, if the binding of several heterocycles is probed against specific binding pockets in a protein, the discrimination between a binding and non-binding event is dependent solely on the molecular complementarity and not constrained or modulated by the heterocycle being part of a larger molecule. This is a far more comprehensive and elegant way to probe for new interactions than having the fragments attached to a rigid template, as might derive from a conventional combinatorial chemistry approach.

One of the earliest and most successful applications of fragment-based lead discovery using biophysical methods is that of the NMR methods pioneered by Fesik and colleagues at Abbott and coined “SAR-by-NMR” [39, 40]. In determining structure-activity relationships (SAR) by NMR, perturbations to the NMR spectra of a protein are used to indicate that ligand binding is taking place and to give some indication of the location of the binding site. Once molecular fragments bound to the target protein have been identified they can be linked together using structure-based chemical synthesis to improve the affinity for the target. Compounds with nanomolar affinities for FK506-binding protein were discovered using SAR-by-NMR where two ligands with micromolar affinities were tethered together [39].

10.4.1

Fragment-Based Lead Discovery Using X-ray Crystallography

X-ray crystallography has the advantage of defining the ligand-binding sites with more certainty than NMR and the binding orientations of the molecular frag-

ments play a critical role in guiding efficient lead optimization programs. Furthermore, ligands that bind away from the target active site can be ignored, which provides another advantage over some NMR methods. Fragment libraries can be screened as singlets or in cocktails by soaking pre-formed crystals of unliganded proteins using X-ray crystallography as the primary method of detection. As the output from an X-ray experiment is a visual description of the bound compound (its electron density) it is possible to screen cocktails of compounds without the need to deconvolute. An optimum cocktail size is typically between 4 and 8 and is defined by the tolerance of the protein crystals to organic solvents and the concentration at which you wish to screen each fragment. Given the affinity of many of the fragments is expected to be within the high micromolar to low millimolar range, the final concentration of each individual fragment is often within the 10–100 micromolar range.

Some of the first experiments in which X-ray crystallography was used as a “screening tool” were reported by Verlinde and colleagues who exposed crystals of trypanosomal triosephosphate isomerase to cocktails of compounds in their search for inhibitors [41]. More recently, Nienaber and colleagues have described a method for screening using X-ray crystallography that focuses on soaking the tar-

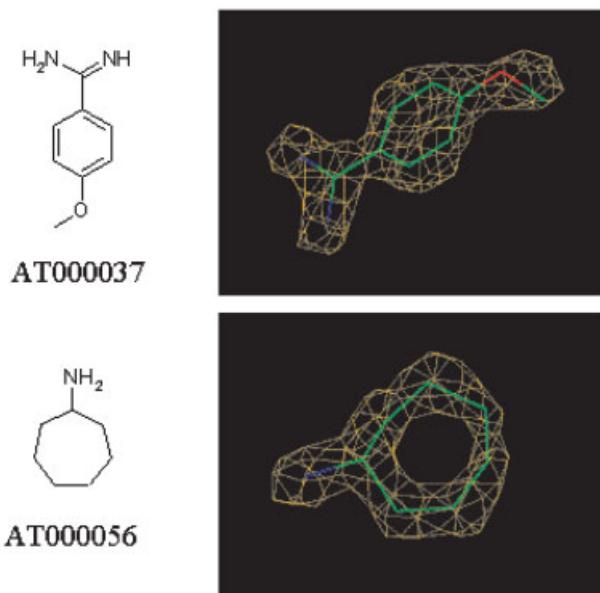


Fig. 10.3 Automatic analysis of ligand electron density. The electron density was interpreted and models of compounds automatically fitted using AutoSolve®. Although the binding affinity is weak (AT000037 $IC_{50}=46\ \mu M$; AT000056 $IC_{50}=1\ mM$) the fragments bound into the pocket of Trypsin adopt a clearly ordered conformation. Electron density maps are contoured at 3σ and density due to protein and solvent has been removed for clarity.

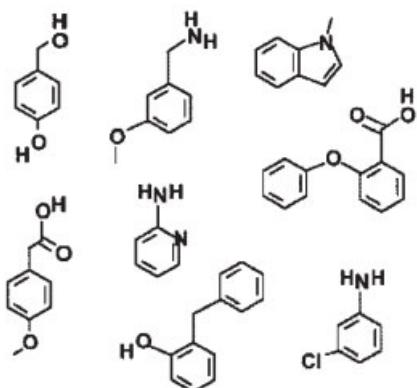
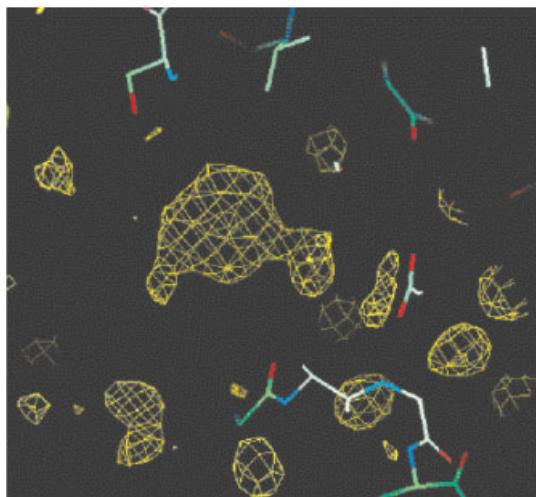


Fig. 10.4 Analyzing fragment cocktails using AutoSolve®. A protein crystal was exposed to a cocktail of 8 fragments and the resultant electron density is shown below. AutoSolve attempts to fit each of the 8 molecules into the electron density and ranks the goodness of fit for each one. In this way the need for deconvolution of an active cocktail is reduced.



get crystals with cocktails of compounds having differing shapes that can easily be distinguished by visual inspection of the electron density [42]. In these experiments crystals of the anticancer target urokinase were used to screen cocktails of 6–8 compounds in an attempt to discover novel inhibitors with improved pharmacokinetic properties. The active compounds were identified visually from the electron density map and the experiment repeated with the actives removed from the cocktail. In this way, multiple binders in a cocktail could be identified. Three different compound classes were discovered using this approach, one of which, the 2-aminoquinoline, provided the basis of a new optimized lead compound [42]. Fragment screening using X-ray crystallography can also be used in conjunction with other biophysical methods such as surface plasmon resonance (SPR). For ex-

ample, SPR was used by Lesuisse and colleagues to identify novel phosphotyrosine mimetics to incorporate into inhibitors of the Src SH2 domain [43]. Over 150 fragments were evaluated for their binding affinity for Src SH2 using SPR and then soaked into crystals. From these experiments, 20 fragments were found to bind at the active site and were used to design low nanomolar range Src SH2 inhibitors devoid of phosphate groups.

However, to fully exploit X-ray crystallography as a screening approach it is desirable to implement an objective and automated process to address the key bottleneck of data interpretation and analysis [29]. AutoSolve[®], from Astex, allows rapid and automated analysis of electron density from fragment-soaking experiments that use singlets and/or cocktails of compounds. Examples of electron density that were unambiguously interpreted by AutoSolve[®] are shown in Fig. 10.3. In each case the binding mode of the small-molecule fragment is clearly defined by the electron density, which means that although the affinity may be in the millimolar range, the binding is ordered, with key interactions being made between the compound and the protein. In fact, AutoSolve[®] requires no human intervention if the quality of electron density is high, and can identify the correct compound bound at the active site from an experiment where the crystal has been exposed to a cocktail of compounds, therefore reducing the need to deconvolute (Fig. 10.4).

10.4.2

Structure-Based Optimization of Fragment Hits

When the binding of one or more molecular fragments has been determined in the proteininactive site, this provides a starting point for medicinal chemistry to optimize the interactions using a structure-based approach. The fragments can be combined onto a template or used as the starting point for “growing out” an inhibitor into other pockets of the protein (Fig. 10.5). The potency of the original weakly binding fragment can be rapidly improved using iterative structure-based chemical synthesis. For example, in one of our lead discovery programs targeted against the cancer target cyclin-dependent kinase 2, we identified an initial fragment, AT381 (MW=134), which exhibited an IC₅₀ of 1 mM in an enzyme assay. Using the crystal structure of AT381 bound to CDK2 we were able to improve the potency more than 100 000-fold by synthesizing only 30 analogues (Fig. 10.6). The resulting compound, AT3851, had an IC₅₀=30 nM and was generated using an iterative cycle of design and synthesis (unpublished results). Compounds from this novel lead series are being further optimized to improve pharmacokinetic properties and show cellular activity against a range of normal and cancer cell lines including MRC5, HCT116 and HT29.

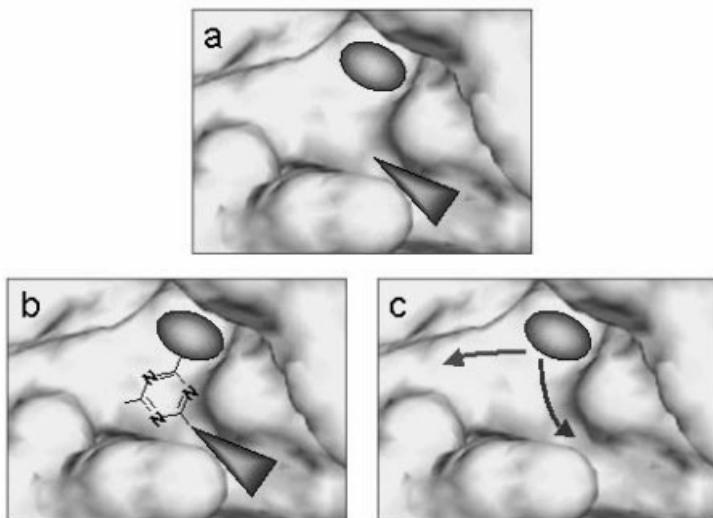


Fig. 10.5 Fragment-based lead generation. Once fragments have been identified bound into the active site they can be used as a start point for iterative structure-driven chemistry resulting in a drug-size lead compound. If two fragments are bound in 2 different pockets (a) then they could be used to decorate an appropriate scaffold (b). Alternatively, a single fragment can be rationally modified to occupy other neighbouring pockets (c).

10.5 Conclusions

The recent increase in available protein crystal structures is driving a new wave of interest in structure-based drug design. This trend is expected to continue as crystal structures of more therapeutic drug targets are determined in the coming years. Technology advances in all aspects of structural biology are fuelling this surge of protein structure data. Many of these advances in protein production, crystallization and structure determination have been developed as part of the current structural genomics initiatives. High throughput crystallography methods are now being used to solve novel protein crystal structures and protein-ligand complexes. The ability to generate protein-ligand crystal structures rapidly allows not only accelerated lead optimization but a new application of X-ray crystallography, as a screening technology. This may have particular value for fragment-based lead discovery where the initial molecular fragments are likely to have an affinity too weak to enable detection using traditional bioassay-based methods. Initial data generated using X-ray crystallographic screening of molecular fragment libraries indicates that novel scaffolds can be identified and subsequently optimized using rapid structure-based synthesis to generate useful lead compounds. The potential of this fragment-based screening approach using X-ray crystallography may be significant, in particular against targets which have remained intractable using conventional screening methods.

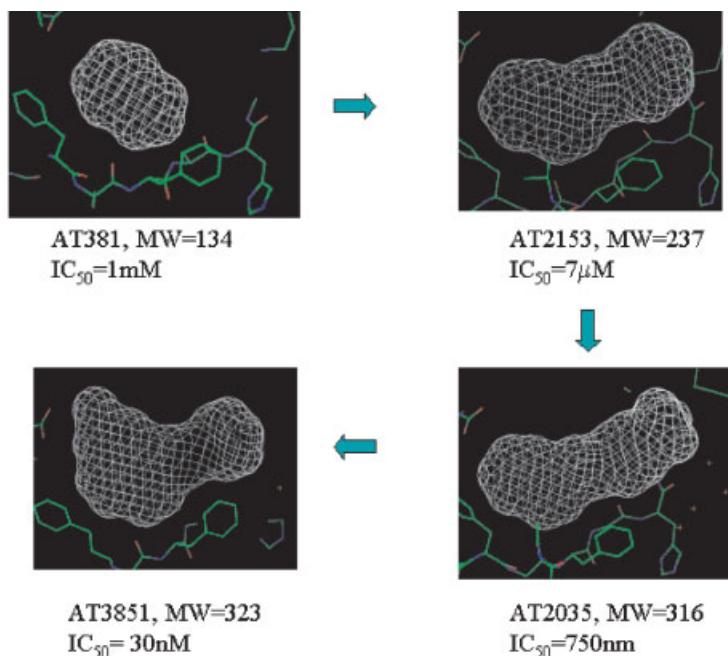


Fig. 10.6 Structure-based lead optimization. Structure-based optimization of the initial fragment hit (AT381) allows affinity to be rapidly and efficiently built-up using focused chemical synthesis. The resulting compound (AT3851) is a potent inhibitor of CDK2 and has good physicochemical properties and shows cell-based activity against a variety of cancer cell lines.

10.6

Acknowledgements

I wish to thank Mike Hartshorn and Ian Tickle who developed AutoSolve and Jeff Yon, Marc O'Reilly, Dominic Tisi and Andrew Sharff for contributions towards this chapter. I also appreciate the assistance of Emma Southern and Michelle Jones in the production of this text.

10.7

References

- 1 International Human Genome Sequencing Consortium, Initial sequencing and analysis of the human genome. *Nature* **2000**, *409*, 860–921.
- 2 CAMPBELL, S.F., Science, art and drug discovery: a personal perspective. *Clin. Sci.* **2000**, *99*, 255–260.
- 3 BURLEY, S. K., An overview of structural genomics. *Nat. Struct. Biol. Suppl.* **2000**, Nov., 932–934.
- 4 HEINEMANN, U., ILLING, G., OSCHKINAT, H., High-throughput three-dimensional protein structure determination. *Curr. Opin. Biotech.* **2001**, *12*, 348–354.

- 5 BLUNDELL, T. L., JHOTI, H., ABELL, C., High-throughput crystallography for lead discovery in drug design. *Nat. Rev. Drug Disc.* **2002**, *1*, 45–54.
- 6 NORVELL, J. C., MACHALEK, A. Z., Structural genomics programs at the US National Institute of General Medical Sciences. *Nat. Struct. Biol.* **2000**, *7*, 931.
- 7 VITKUP, D., MELAMUD, E., MOULT, J., SANDER, C., Completeness in structural genomics. *Nat. Struct. Biol.* **2001**, *8*, 559–566.
- 8 DRY, S., McCARTHY, S., HARRIS, T., Structural genomics in the biotechnology sector. *Nat. Struct. Biol.* **2000**, *7*, 946–949.
- 9 STEVENS, R. C., Design of high-throughput methods of protein production for structural biology. *Structure* **2000**, *8*, R177–R185.
- 10 LESLEY, S. A., High throughput proteomics: protein expression and purification in the post-genomic world. *Protein Exp. Purif.* **2001**, *22*, 159–164.
- 11 CROWE, J., DOBELI, H., GENTZ, R., HOCHULI, E., STUBER, D., HENCO, K., 6xHis-Ni-NTA chromatography as a superior technique in recombinant protein expression/purification. *Methods Mol. Biol.* **1994**, *31*, 371–387.
- 12 KIGAWA, YABUKI, T., YOSHIDA, Y., TSUTSUI, M., ITO, Y., SHIBATA, T., YOKOYAMA, S., Cell-free production and stable-isotope labeling of milligram quantities of proteins. *FEBS Lett.* **1999**, *442*, 15–19.
- 13 SAWASAKI, et al., A cell-free protein synthesis system for high-throughput proteomics. *PNAS* **2002**, *99*, 14652–14657.
- 14 VAN MIERLO, C. P. M., STEENSMA, E., Protein folding and stability investigated by fluorescence, circular dichroism and nuclear magnetic resonance spectroscopy: the flavodoxin story. *J. Biotech.* **2000**, *79*, 281–298.
- 15 D'ARCY, A., Crystallising proteins – a rational approach? *Acta Crystallogr.* **1994**, *D50*, 469–471.
- 16 STEVENS, R. C., High-throughput protein crystallisation. *Curr. Opin. Struct. Biol.* **2000**, *10*, 558–563.
- 17 NOWAKOWSKI, J., CRONIN, C. A., MCREE, D. A., KNUTH, M. W., et al., Structures of the cancer-related aurora-A, FAK, and EphA2 protein kinases from nanovolume crystallography. *Structure* **2002**, *10*, 1659–1667.
- 18 BODENSTAFF, E. R., HOEDEMAEKER, F. J., KUIL, M. E., DE VRIND, H. P. M., ABRAHAMS, J. P., The prospects of protein nanocrystallography. *Acta Crystallogr.* **2002**, *D58*, 1901–1906.
- 19 STEWART, L., CLARK, R., BEHNKE, C., High-throughput crystallisation and structure determination in drug discovery. *Drug Disc. Today* **2002**, *7*, 187–196.
- 20 WILSON, J., Towards the automated evaluation of crystallisation trials. *Acta Crystallogr.* **2002**, *D58*, 1907–1914.
- 21 WATANABE, N., Semi-automatic protein crystallisation system that allows in situ observation of X-ray diffraction from crystals in the drop. *Acta Crystallogr.* **2002**, *D58*, 1527–1530.
- 22 HENDRICKSON, W., Synchrotron crystallography. *Trends Biochem. Sci.* **2000**, *25*, 637–643.
- 23 ARNDT, U. W., BLOOMER, A. C., New developments in X-ray optics for macromolecular crystallography using laboratory X-ray sources, *Curr. Opin. Struct. Biol.* **1999**, *9*, 609–614.
- 24 DE LA FORTELLE, E., BRICOGNE, G., Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. *Methods Enzymol.* **1997**, *276*, 472–494.
- 25 LEMKE, C. T., S-SAD, Se-SAD and S-Se-SIRAS using Cu *K_a* radiation: why wait for synchrotron time. *Acta Crystallogr.* **2002**, *D58*, 2096–2101.
- 26 DOUBLIE, S., Preparation of selenomethionyl proteins for phase determination. *Methods Enzymol.* **1997**, *276*, 523–530.
- 27 OGATA, C. R., MAD phasing grows up. *Nat. Struct. Biol., Synchrotron Supplm.* **1998**, *1997*, 638–640.
- 28 PERRAKIS, A., MORRIS, R., LAMZIN, V. S., Automated protein model building combined with iterative structure refinement. *Nat. Struct. Biol.* **1999**, *6*, 458–463.
- 29 BLUNDELL, T. L., ABELL, C., CLEASBY, A., HARTSHORN, M. J., TICKLE, I. J., PARASINI, E., High-throughput crystallography for drug discovery. Proceedings of the Royal Society of Chemistry meeting, Cutting

- Edge Approaches to Drug Design, March 2001 (FLOWER, D., ed.) RSC Publications, Cambridge. 2001, 53–59.
- 30 SERVICE, R., Tapping DNA for structures produces a trickle. *Science* 2002, 298, 948–950.
- 31 BERMAN, H. M., The Protein Data Bank. *Acta Crystallogr.* 2002, D58, 899–907.
- 32 CARR, R., JHOTI, H., Structure-based screening of low-affinity compounds. *DDT*, 2002, 7, 522–527.
- 33 KARAIN, W. I., Automated mounting, centering and screening of crystals for high-throughput protein crystallography. *Acta Crystallogr.* 2002, D58, 1519–1522.
- 34 MUCHMORE, S. W. et al., Automated crystal mounting and data collection in protein crystallography. *Structure* 2000, 8, R243–R246.
- 35 HARTSHORN, M. J., AstexViewerTM: a visualisation aid for structure-based drug design. *J. Comp. Aided Mol. Design*, 2003, 16, 871–881.
- 36 OPREA, T. I., Is there a difference between leads and drugs? A historical perspective. *J. Chem. Inf. Comp. Sci.* 2001, 41, 1308–1315.
- 37 LIPINSKI, C. A., LOMBARDI, F., DOMINY, B. W., FEENEY, P. J., Experimental and computational approaches to estimate solubility and permeability in drug discovery and development. *Adv. Drug Delivery Rev.* 2001, 46, 3–26.
- 38 FEJZO, J., LEPRE, C. A., PENG, J. W., BEVIS, G. W., MURCKO, M. A., MOORE, J. M., The SHAPES strategy: an NMR-based approach for lead generation in drug discovery. *Chem. Biol.* 1999, 6, 755–769.
- 39 SHUKER, S. B., HAJDUK, R. P., MEADOWS, R. P., FESIK, S. W., Discovering high-affinity ligands for proteins: SAR by NMR. *Science* 1996, 274, 1531–1534.
- 40 HAJDUK, P. J., MEADOWS, R. P., FESIK, S. W., NMR-based screening in drug discovery. *Quart. Rev. Biophys.* 1999, 32, 211–240.
- 41 VERLINDE, C., KIM, H., BERNSTEIN, B. E., MANDE, S. C., HOL, W. G. J., Antitrypanosomiasis drug development based on structures of glycolytic enzymes. Structure-based drug design (VEERAPANDIAN, P., ed.), Marcel Dekker, New York, 1997, 365–394.
- 42 NIENABER, V. L., RICHARDSON, P. R., KLIGHOFER, V., BOUSKA, J. L., GIRANDA, V. L., GREER, J., Discovering novel ligands for macromolecules using X-ray crystallographic screening. *Nat. Biotech.* 2000, 18, 1105–1108.
- 43 LESUISSE, D., LANGE, G., DEPREZ, P., BERNARD, D., SCHOOT, B., DELETTRE, G., MARQUETTE, J.-P., BROTO, P. et al., SAR and X-ray. A new approach combining fragment-based screening and rational drug design: application to the discovery of nanomolar inhibitors of Src SH2. *J. Med. Chem.* 2002, 45, 2379–2387.

11

Micro-Crystallization

CARL L. HANSEN, MORTEN SOMMER, KYLE SELF, JAMES M. BERGER,
and STEPHEN R. QUAKE

11.1

Introduction

With the completion of the various genome projects including the human genome, X-ray crystallography is poised to be the workhorse in the next great endeavor: illuminating the cellular proteomes. Just as the invention of the microscope helped to bring about an understanding of life at the cellular scale [1], X-ray crystallography has allowed scientists to observe proteins at the atomic level, driving a structural revolution in the biological and medical sciences. Unprecedented technological advances in synchrotron X-ray sources, phasing techniques and computing power have further amplified the power of this technique [2, 3].

These innovations in diffraction methods and model generation have not however been matched by the development of techniques for reliably and rapidly crystallizing macromolecules. Obtaining diffraction-quality crystals remains a major bottleneck in structure determination [4], and has prevented X-ray crystallography from realizing its true potential. The development of generally applicable technologies that enable the robust routine crystallization of biological macromolecules will bring about a paradigm shift in biology and drug discovery, allowing researchers not only to picture life at the molecular level, but also to open areas of investigation that were previously inaccessible. Emergent technologies utilizing microfluidics now have the potential to solve these problems on several levels, both by allowing researchers to conduct efficient assays on the nanoliter scale, and by pushing the development of novel crystallization techniques.

The crystallization of a biological macromolecule is realized by manipulation of one or more chemical and thermodynamic variables, such that the solubility of a target molecule in a concentrated solution is reduced, thereby promoting a transition to the solid phase in the form of a well-ordered crystal. In principle, any thermodynamic variable that may directly, or indirectly, affect protein solubility may be used to induce crystallization. Variables that are most often manipulated include macromolecule concentration, ionic strength, identity and concentration of precipitating agents, pH, temperature and small-molecule additives. Together, these variables comprise a vast multi-dimensional chemical phase space that must be systematically explored to discover crystallization conditions.

Natural macromolecule targets for crystallography are both large and extremely varied. Proteins, for example, are complicated polymers of amino acids with polar, non-polar, charged and aromatic residues that interact with each other and with the external environment. These complicated interactions result in a highly varied and complicated phase space diagram that cannot be determined *a priori*. Consequently, macromolecular crystallization requires a brute force approach in which as large a volume of chemical phase space as possible must be explored. For macromolecules possessing a large volume of phase space that is conducive to crystallization, there is a good chance that one of the randomly screened conditions will result in crystal formation. Since the first “hit” is not likely to be optimal, initial successes are usually of poor quality, and may consist of spherulites (Fig. 11.1 A), phase separation (Fig. 11.1 B), micro-crystals (Fig. 11.1 B), needles, needle clusters, thin plates (Fig. 11.1 C), plate stacks (Fig. 11.1 D), or small single crystals (Fig. 11.1 E). These starting conditions can be used to initiate focused and refined screening however, eventually producing diffraction quality crystals (Fig. 11.1 F). For macromolecules that are more difficult to crystallize, initial screens may be too coarse to uncover promising conditions and further screening may be required. If the conditions required for crystallization are very specific, it may ultimately require many thousands of experiments before a hit is detected, if at all.

The large number of experiments required to uncover successful crystallization conditions therefore is what represents the most formidable obstacle to determining the structure of many important biological macromolecules. For example, membrane proteins play a central role in cell signaling and are often excellent targets for small molecule therapeutics. Unfortunately, the structures of only a very few have been determined to date, primarily because of the difficulties associated

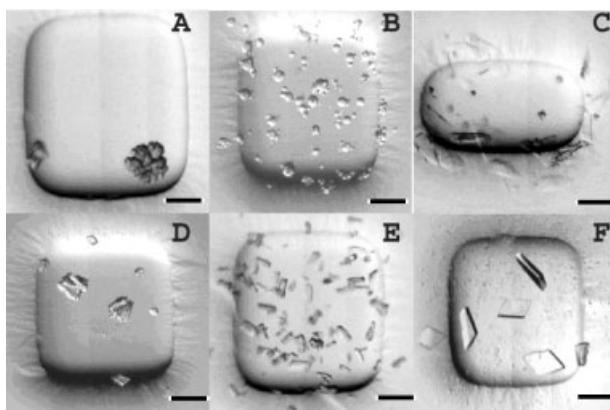


Fig. 11.1 Examples of initial crystallization hits from a single on-chip screening experiment of a type II topoisomerase ATPase domain/ADP, 12 mg mL⁻¹. (A) Irregular spherulite. (B) Phase separation and spherulites with nucleating microcrystals. (C) Thin plate clusters. (D) Thick plate stacks. (E) Well-formed microcrystals. (F) Large single crystals. All scale bars are 100 μ m.

with expressing, solubilizing and stabilizing these molecules in large quantities (>1 mg) [5–7]. Similarly, many proteins in the cell work as large complexes. Structural information of these targets provides invaluable information regarding complex reactions and protein–protein interactions but such assemblies are exceedingly difficult to purify in large amounts, and at times must even be purified by processing kilogram quantities of native sources [8, 9].

Traditional techniques for the crystallization of macromolecules include concentration through slow dehydration (vapor diffusion), batch and dialysis methods and both liquid-liquid and liquid-gel diffusion experiments (for a review see [10]). Practical limitations of traditional fluid handling approaches require that the minimum volume per assay for these techniques ranges from 0.2 to 1 μL for vapor diffusion and micro-batch methods, and up to 50 μL for micro-dialysis. Since protein samples may only be available in milligram quantities, and target molecule concentrations are generally required in excess of 10 mg mL^{-1} , the ability to perform nanoliter scale assays routinely would not only help to promote a more comprehensive screening strategy, but would also allow for experimentation with ultra-low-abundance macromolecules.

To date robotic automation has been the dominant strategy both for miniaturizing standard crystallization protocols and for fostering high-throughput crystallization [11, 12]. This allows more trials to be conducted with a given sample volume. State of the art liquid handling robots are capable of accurately dispensing low viscosity solutions in small volumes. These machines have been shown to increase throughput, reduce sample consumption, and improve reproducibility [4, 13–15]. Nonetheless, there are however several drawbacks that limit robotic success and applicability. Since dispensing robots are sensitive to solution viscosity and surface tension, they must be properly calibrated for each working fluid, and frequently have difficulty dispensing the highly viscous solutions (such as MPD or PEG) used in crystallography. Furthermore, these systems require significant dedicated space and are expensive to both purchase and maintain, making them unavailable to the vast majority of researchers, particularly academic labs. From a progress standpoint, these systems essentially only scale down the volumes needed to conduct screening by traditional methods employing low viscosity solutions. Robotics have not necessarily changed or improved upon the efficiency of the crystallization method itself.

11.2

Microfluidics – Method and Design

In the same way that miniaturization has impacted the electronics industry, microfluidic technologies promise to spark a revolution in the biological sciences by integrating ultra small-volume sample processing within a chip format. The use of nanoliter reaction volumes and highly scaleable parallel sample processing make microfluidic technologies ideally suited to protein crystallography, where the screening and processing of precious reagents is required. Beyond reduction in

sample consumption, the unique properties of mixing and fluid flow at the micron scale also allow for the implementation of assays that are highly efficient at detecting crystallization conditions. Despite this enormous potential, several technical problems have prevented the realization of the full potential of microfluidic devices for protein crystallization.

In order to perform assays on the nanoliter scale, a system for the accurate metering and dispensing of fluids is required. For such a system to be scaleable and to have general applicability, it must be insensitive to both the surrounding fluidic architecture and to the properties of the working fluids. In the case of protein crystallography, this latter requirement is particularly important since the solutions used in crystal trials cover a large range of chemistries, viscosities, ionic strengths and pH. Previous work on microfluidic metering has resulted in the development of electrokinetic metering systems that manipulate fluids by the application of an electric potential to the terminals of the device [16–19]. Using this technique, once the devices are calibrated, fluids may be accurately metered and mixed on the picoliter scale [16–21]. This method is not scaleable however, and there are additional fundamental limitations that prevent its universal application. For example, the electrokinetic force used in these devices depends strongly on both the properties of the working fluid and on its interaction with the channel walls. Small changes in pH or salt concentration that result from ion drift over time can lead to more than a ten-fold variation in injected volume. These systems are also dependent on the viscosity and on the fluidic resistance due the surrounding channel architecture. As a result, electrokinetic devices must be recalibrated for every fluid, and are not suitable for high-throughput screening applications with a diverse ensemble of unrelated fluids. Moreover, since electrokinetic systems have no active valves, reagents diffuse through junctions and channels over time. This leaking dilutes and contaminates samples over time, restricting the maximally achievable incubation times and the density at which assays may be integrated on chip. The problem of reagent storage is particularly acute for crystallization applications where assays may be required to incubate for several days or weeks.

All of these problems may be addressed by using MEMS (microelectromechanical systems) fabrication techniques to incorporate active mechanical valves on chip. Traditional MEMS techniques, using “hard” materials such as glass or silicon, may be used to fabricate true, leak-proof mechanical valves. However, MEMS fabrication techniques are expensive and require many processing steps, rendering the integration of many valves on a chip a difficult and expensive process. Furthermore, since these valves are fabricated from hard materials, a large valve actuator is needed to achieve valve closure at attainable actuation forces. The large actuators of each valve and the low yield of the fabrication process impose practical constraints on the degree of integration that is possible in microfluidic devices made from traditional MEMS techniques.

As an alternative to traditional MEMS techniques, multilayer soft lithography (MSL) enables the facile and inexpensive large-scale integration of valves on chip [22, 23]. MSL is an extension of the technique of soft lithography [24], which uses

replica molding of non-traditional polymer materials from a micromachined master. Soft lithography may be used to quickly reproduce features as small as 80 nm, creating a soft and inexpensive negative of the original master. This technique has found diverse applications from patterned surface deposition methods to the fabrication of passive microfluidic devices [24–34]. MSL uses consecutive soft lithography molding and bonding steps to generate complex multilayer fluidic devices with active mechanical valves, pumps, mixers and flow control logic.

Owing to their excellent optical properties, low surface energy, and low Young's modulus, silicone rubbers have become the most popular material for multi-layer soft lithography. Silicone rubbers are typically formed from the combination of two liquid components that cross-link into a flexible solid upon curing. One such example polymer is General Electric RTV 615. Part A of this compound consists of a platinum catalyst with polydimethylsiloxane polymers that have been functionalized with vinyl groups. Part B contains silicon hydride (Si–H) groups that covalently bond to the vinyl groups of part A, cross-linking the polymers. When mixed at the stoichiometric ratio of 10:1 (A:B) there is an equal number of vinyl and silicon hydride groups, allowing all groups to become incorporated into a covalent bond. When combined at a different ratio, however, there remain active groups that do not participate in a covalent bond, and which instead may be used to bond two surfaces together permanently, creating a monolithic device.

A process flow diagram illustrating the steps in MSL is shown in Fig. 11.2. Two negative molds, one defining the flow structure, and the other defining the control structure of valves, are first patterned on a silicon wafer using conventional photolithography, leaving 10 µm raised features of photo-resist. The flow layer master is then annealed so that the photo-resist is allowed to re-flow, creating rounded flow channels. Masters are reproduced by replica molding in silicone rubber. A 30:1 ratio of (A:B) silicone rubber, having excess vinyl groups, is spun onto the flow mold to a final thickness of 30 µm. A 3:1 silicone rubber layer, containing excess silicon hydride groups, is cast over the control mold to a thickness of approximately 7 mm. Once both layers are heated and allowed to partially cure, the structures solidify. The control layer is then peeled from the mold, punched to create valve access ports, aligned to the flow structure and the entire device is then heated once again, causing the excess vinyl and silicon hydride groups to

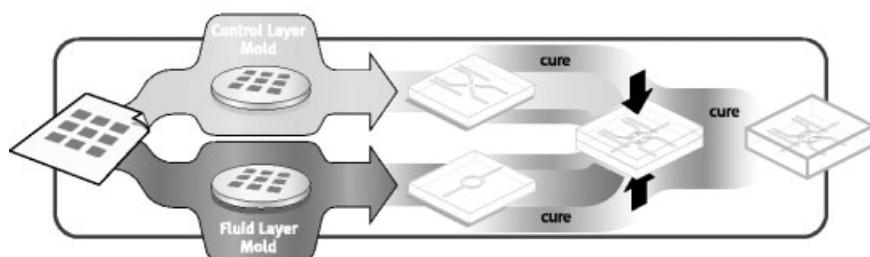


Fig. 11.2 Flow diagram illustration of the technique of multilayer soft lithography.

covalently bond. The resulting monolithic device is peeled from the flow mold and channel access ports are punched into it prior to sealing the multilayer polymer to a glass substrate. The substrate itself may additionally have structural features, such as microwells of defined volumes that may be accessed through the molded channels, thus providing larger localized volumes for reactions to occur.

When a control channel crosses over a flow channel, only a thin square membrane of elastomer separates the two, forming a valve (Fig. 11.3 A). By pneumatic or hydraulic pressurization of the control channel, the membrane may be deflected down into the flow channel, causing it to seal against the glass substrate (Fig. 11.3 B). Because of the compliance of the membrane, a hermetic seal may be easily achieved at moderate actuation pressures even in the presence of particulates or imperfections. The low Young's modulus of the elastomer ($\approx 1 \text{ MPa}$ versus $\approx 100 \text{ GPa}$ for crystalline silicon) allows for the closing of valves having areas as small as $100 \mu\text{m}^2$, so that many thousands of active valves may be integrated on a chip smaller than a credit card [23].

The actuation pressure required to close a valve is dependent on the geometry of the valve junction. For a standard geometry of a $10 \mu\text{m}$ high flow channel, a

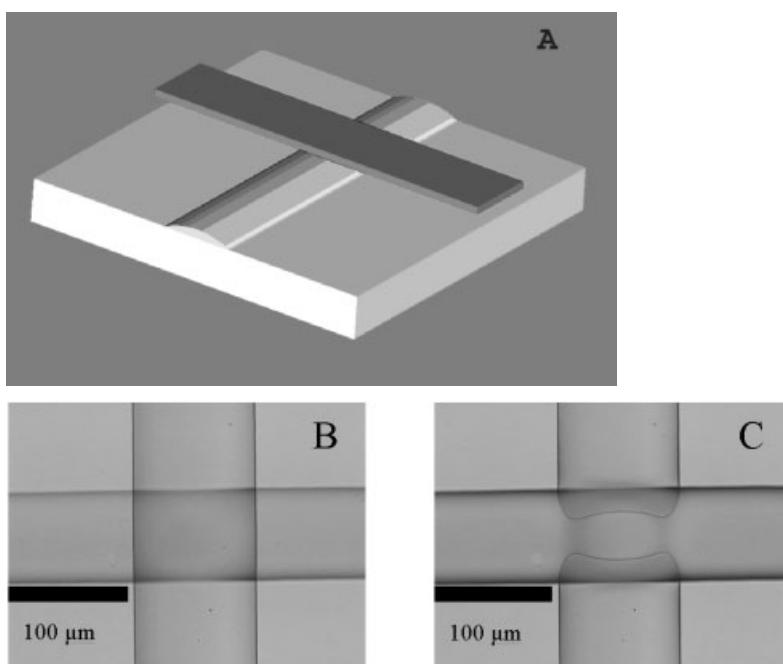


Fig. 11.3 Active valves fabricated by multilayer soft lithography. (A) Rendering of valve geometry; control channel (narrow diagonal) crosses flow channel (wider diagonal) forming a valve at the intersection. (B) Micrograph of an open valve showing control (horizontal) and flow (vertical) structures. (C) Closed valve; application of pressure to control structure deflects the membrane and pinches off the flow structure, creating a fluidic seal.

20 μm thick membrane, and a 100 $\mu\text{m} \times 100 \mu\text{m}$ valve, the required actuation pressure is approximately 6 psi. The strong dependence of actuation pressure on the channel and valve geometry may be exploited to engineer bridges (valves with very high actuation pressure) by simply tapering the thickness of a control channel in the vicinity of a flow channel that is not to be closed. By applying an intermediate actuation pressure, a single control channel may be used to close a plurality of selected flow channels without significantly impeding flow in adjacent lines.

It is, however, not enough simply to create a device with many integrated valves. The realization of highly integrated and complex fluidic devices also requires that the problem of priming, or initially filling the device with fluid, be addressed. For microfluidic devices made from conventional hard materials such as silicon or glass, this requirement may preclude the use of multiply crossing, highly complex fluidic architectures. Such devices may therefore need to be primed using a flow-through method, which requires an outlet through which displaced gas may be vented. The introduction of the priming fluid through a complex fluidic structure may trap air bubbles at junctions and other channel features. Surface tension effects at the liquid–gas interfaces of these bubbles result in a high effective viscosity such that they may not be easily removed and can adversely affect the performance of the device. Furthermore, since a single priming fluid must pass through the entire device, it may subsequently contaminate or dilute sample solutions.

These difficulties can be surmounted in silicone devices by exploiting the gas permeability of a soft silicone polymer [35, 36]. Arbitrarily complex, connected fluidic structures may be filled in minutes by a technique called pressurized out-gas priming (POP). Using the POP technique, a fluid is injected into a closed channel structure, causing the gas ahead of it to be pressurized. Owing to the permeability of the elastomer, the pressurized gas quickly diffuses into the bulk material, allowing the priming fluid to completely fill the flow structure. Despite the low surface energy of PDMS (22 mN m^{-1}), aqueous solutions may be easily introduced, at moderate pressures (1–8 psi), into channels having a minimum dimension of 1 μm , eliminating the need for surface modification protocols. Since no outlet is needed for venting the gas, dead-end reaction chambers and channels may be used. By incorporating integrated valves into the flow structure, the device may be selectively primed with different solutions, allowing for greater design flexibility.

The ability to fill a device with many different fluids in different channels and chambers allows for the implementation of a simple and robust geometric method of metering solutions. The principle behind this scheme is to set up a geometry in which two microfluidic chambers may be primed with different solutions and connected via a fluidic interface that is controlled by a microfabricated valve (Fig. 11.4). With a central interface valve closed, the two chambers are first dead-end filled with two different solutions using the POP technique. Once both chambers have been completely filled, containment valves are actuated, isolating the chambers from the rest of the chip. The interface valve is then opened, creating a

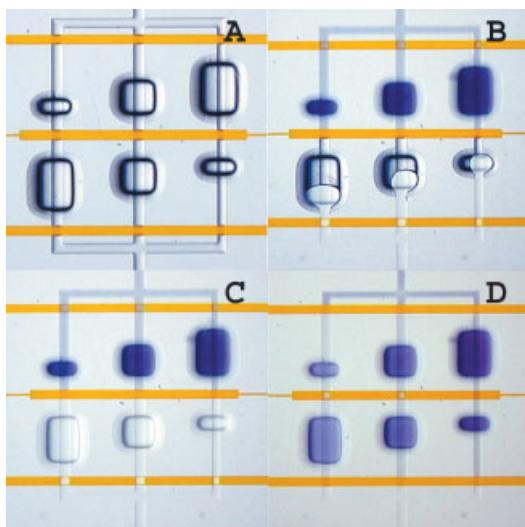


Fig. 11.4 Barrier interface metering. (A) A set of microfluidic reaction chambers designed to mix two fluid samples at three distinct mixing ratios. Reaction chambers having approximate volumes of 5 nL, 12.5 nL and 20 nL are paired to produce final mixing ratios of 4:1, 1:1 and 1:4 while maintaining a constant reaction volume of 25 nL. (B) Dead-end filling of reaction chambers by the pressurized outgas priming (POP) method. Top wells have been filled

with dye (13 mM bromophenol blue sodium salt) and bottom wells are being filled with water. (C) Chambers are sealed by actuation of the containment valves and the interface valve is released, allowing the solutions to slowly mix by diffusion. (D) The complete diffusive mixing of an organic dye with water, resulting in three distinct concentrations in the different mixing chambers.

constricted fluidic connection between the chambers, and allowing them to mix by diffusion. In this way, the chambers effectively act as microfluidic measuring cups, precisely determining the mixing ratio by the relative chamber volumes. Since the metering depends only on the volume of the chambers, it is an inherently robust technique that is insensitive to resident fluid properties.

11.3

Utility of Microfluidics for Crystallization

In the simple configuration shown in Fig. 11.4, mixing is completely diffusive and occurs on the order of an hour for small molecules in an aqueous solution. Although rapid mixing of solutions at the low Reynolds numbers presents a challenge, several schemes for fast and efficient mixing have been developed. In cases where rapid mixing is required, it is straightforward to implement the BIM scheme into a geometry such as a ring, and then enhance mixing by active pumping around the ring using a kneading or chaotic mixing configuration [37–39]. Nonetheless, for some applications it is desirable to suppress the confounding ef-

fects of convection and instead to allow mixing that is completely diffusive. One situation where this scheme is particularly useful lies in the crystallization of macromolecules.

The simple geometric metering scheme shown in Fig. 11.4 has been used to develop a highly efficient microfluidic device for protein crystallization in ultra-small volume reactions. The crystallization chip implements 144 simultaneous metering and mixing reactions while consuming only 3.0 μL of protein solution. A layout of the chip (Fig. 11.5) shows 48 reaction centers (Fig. 11.4), each consisting of three pairs of microfluidic reaction chambers with relative volumes of 1:4, 1:1 and 4:1. Each pair of chambers is connected to the protein sample and one of 48 crystallization solution reservoirs. The chip has 480 integrated valves that are actuated through three separately addressable control lines. As a precaution, 48 safety valves are included at the solution inlets to avoid the unwanted loss of protein sample in the unlikely event of an interface valve failure. The remaining two lines simultaneously control all interface and containment valves. By virtue of this parallel architecture and the robustness of the BIM scheme, solutions of varying viscosity, surface tension, pH and ionic strength may be simultaneously metered and mixed at three different mixing ratios using only two hydraulic control lines.

The chip is contained within a carrier device (Fig. 11.6), which facilitates loading, storage and interfacing with the control lines. The chip is secured in the base of the carrier with the safety and interface lines directly connected to two of three

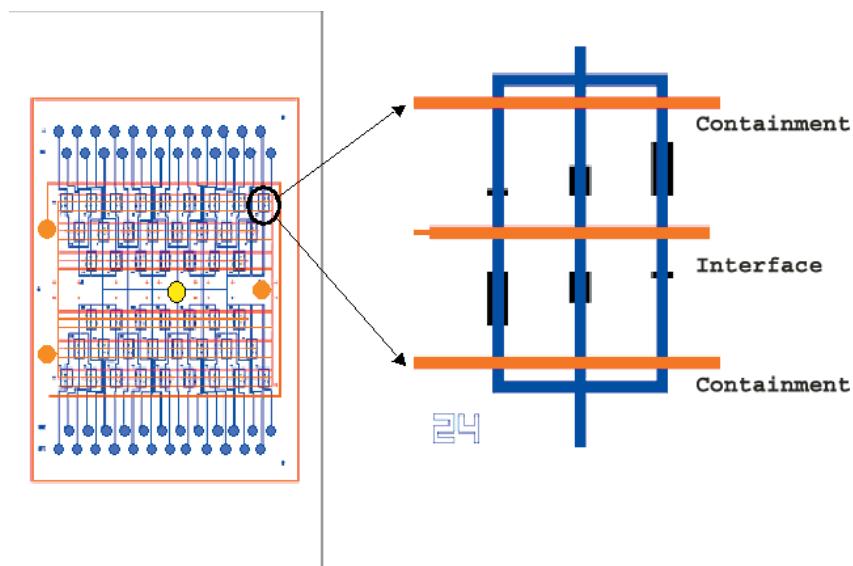


Fig. 11.5 Parallel architecture of crystallization chip and blow-up of a single unit cell. All unit cells are connected to a central sample port (yellow) and one of 48 unique reagent ports (blue). All interface and containment

valves are actuated in parallel through two control ports (red). An additional control port (red; top right) actuates safety valves in the unlikely contingency of a valve failure.

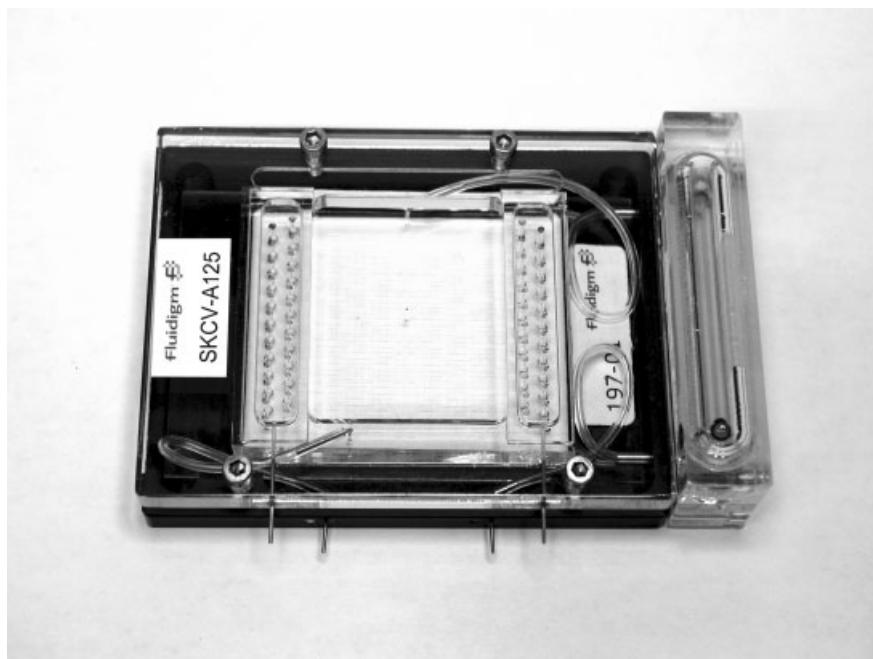


Fig. 11.6 Crystallization chip inside carrier device. Pressure reservoir (bottom) allows for free transport and storage of chip. Interface pins (right) allow for facile loading and control of chip valves.

carrier interface pins that are externally accessible. The containment line interface pin is connected to the chip through a pressure accumulator. Once charged, this accumulator acts as an on-board pressure source to maintain actuation of the containment valves for several weeks, thereby allowing the chip to be stored and transported without the need for any external connections. The top of the carrier has two cavities with a raised lip around the periphery and stainless steel input ports for pressurization. The cavities mate with the 48 reagent wells, creating a seal against the compliant elastomer chip when the plates are pressed together. Once the carrier is assembled the cavities are pressurized, simultaneously injecting the 48 crystallizing agents into the chip. The protein sample is then loaded through a single port located in the center of the chip. A 3 μL sample of protein is aspirated into a pipette tip, and the tip is inserted into the protein port. The tip is then pressurized through an adapter, injecting the protein into the 144 reaction chambers. Since the sample is introduced through a single port, there is very little lost at the interface, and a true economy of scale is realized with more than half the sample being used in the crystallization assays.

While the most obvious advantage of microfluidics is the enormous reduction in sample consumption, the growth of protein crystals can also be fundamentally improved by taking advantage of the physical properties of fluid flow at the mi-

cron scale [36]. The successful crystallization of a macromolecule is determined both by thermodynamic and kinetic considerations. A concentrated solution of the target molecule must first be brought into a state of supersaturation in which the crystal phase is energetically favorable, and then kept in this state to allow crystal nucleation and growth to occur. Super-saturation is induced through the addition of a precipitating agent. Traditional precipitating agents include salts, polymers (e.g., polyethylene glycol), organic solvents, buffers and various additives, all of which are chosen specifically to manipulate thermodynamic variables such as solution pH, dielectric constant, salt concentration and effective protein concentration. Since there is currently no way to predict *a priori* which conditions will be favorable to crystallization, determination of conditions is a purely empirical process of trial and error. Moreover, because a thorough investigation of phase space by conventional techniques is impractical, the initial search is directed towards a sparse matrix or incomplete factorial sampling of likely crystallization agents [40, 41]. Once preliminary conditions have been identified, a refined search is used for optimizing crystal size and quality.

The chance of success in this search depends both on the number of assays and the effectiveness of each assay. While there is no general strategy for predicting what conditions will give rise to crystallization, an understanding of the physics of crystal nucleation and growth may be used to design experiments that increase the chances of success. A hypothetical two-dimensional phase space illustrating the protein behavior as a function of protein concentration and precipitating agent concentration is shown in Fig. 11.7 [36]. This phase space represents the interaction of a given protein with a specific precipitating solution, and is therefore unique to that solution. The shape and the extent of the phase space region will vary dramatically depending on the precipitating agent and target molecule characteristics. For some precipitating agents there may exist a region of supersaturation bounded by precipitation and solubility curves. In this region the protein solution is out of equilibrium, and given sufficient time undergoes a phase transition to a crystalline solid. The time required for such a transition depends strongly on the degree of supersaturation defined as the ratio of the protein concentration to the maximum soluble concentration in equilibrium. This region of supersaturation may further be divided into a metastable and a labile region. Within the labile region, near the precipitation curve, rapid nucleation occurs, leading to the growth of a large number of small low quality crystals. In many cases these microcrystals may be too small to be distinguished from the amorphous precipitate, a problem that can cause promising conditions to be overlooked. In the metastable region, near the solubility curve, the growth of large, high quality crystals is supported, but nucleation is a rare event requiring impractically long incubation times. Experiments that stay in this region will likely appear as clear drops and will be unremarkable.

The three-dimensional aggregation of target molecules into a critical nucleus from which crystal growth may proceed, is a process that requires a higher activation energy, and hence higher supersaturation, than the subsequent one- or two-dimensional nucleation needed for crystal facet growth. For this reason, an opti-

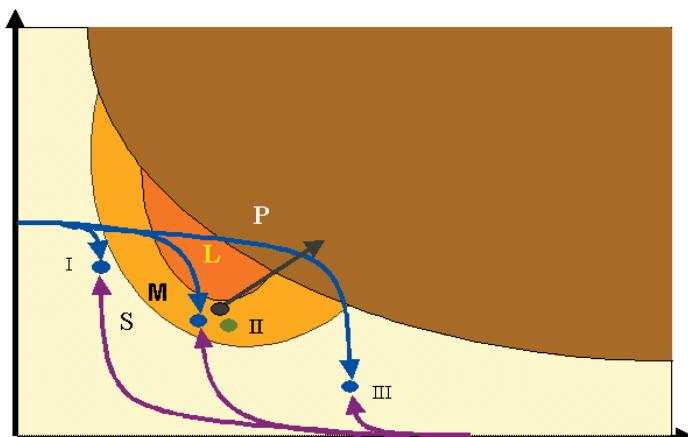


Fig. 11.7 Schematic diagram of the evolution of vapor diffusion, microbatch and μ FID experiments through a two-dimensional projection of phase space having protein concentration and precipitating agent concentration as variables. The phase space is divided into soluble (S), metastable (M), labile (L) and precipitation (P) regions. Microbatch and hanging-drop experiments start at point II where the target molecule is combined 1:1 with the precipitating agent. Microbatch experiments sample only a point in phase space since incubation under immiscible oil prevents subsequent concentration of reagents (green). In

vapor diffusion experiments, equilibration with a reservoir of precipitating agent slowly concentrates the reagents, driving the sample monotonically into the super-saturation region (black). The evolution of a μ FID reaction site having 3 different mixing ratios is shown. Curves represent the average state of both the sample side (blue) and precipitating agent side (purple) of each compound well. The final states (I, II, III) are determined by the mixing ratio. A decrease in protein concentration due to precipitation or crystal growth is not included in the figure.

mal crystal growth scheme should allow for initial nucleation by transiently high levels of supersaturation, followed by passage into lower supersaturation levels that support high quality crystal growth [42, 43].

The microfluidic reactor we designed provides exactly this property by implementing a microfluidic free-interface diffusion (μ FID) between the precipitant and the protein solutions on chip. The opening of the microreactor interface valve that separates the protein from the precipitating solution creates a well-defined fluidic interface that allows for diffusive equilibration of the coupled micro-wells. The phase-space trajectory taken by each assay on the chip therefore depends on both the diffusion constants of the species involved, and the relative volume of the wells. Shortly after the interface valves are opened, the target macromolecule concentration in the target well changes very little, while that of the precipitating agent, which typically has a much larger diffusion constant, increases towards one of three final values determined by the mixing ratios. Subsequently, over a time of approximately 8–24 hours the target macromolecule concentration equilibrates, increasing on the precipitating agent side and decreasing on target macromolecule side, towards the final protein concentration that is again determined by the mix-

ing ratios. The macromolecule targets in the μ FID assay therefore travel a curved path through phase space, sampling efficient crystal nucleation conditions in the labile region prior to settling into a high quality growth regime in the metastable region. This trajectory can be contrasted with that of microbatch and hanging drop methods, which are the two most popular assays for crystallization screening, but have phase trajectories that are stagnant or monotonically increasing in supersaturation (Fig. 11.7).

The kinetics of the chip μ FID assays have many similarities to conventional free interface diffusion, a technique that has long been recognized as an efficient method for exploring crystallization space [43–45]. Despite the favorable kinetics of conventional free interface diffusion, this method has not achieved success in the protein crystallography community for a number of reasons. For example, conventional FID set-ups create a liquid interface in a capillary tube, so that solutions must be delicately introduced into opposite ends of the capillary by inserting a thin needle, making the technique labor intensive and ill-suited to high throughput screening. Furthermore, the diameter of the capillary must be of a substantial size (typically 100 μm), necessitating the use of relatively large sample volumes (generally $>5 \mu\text{L}$). In addition, the introduction of the second fluid causes transient convection, which results in a poorly defined interface that may only be reduced by cumbersome procedures such as the introduction of hydro-gels or the freezing of the first solution. The drawbacks of the large volume that is required, delicate dispensing and the intrinsic poorly defined interface, have thus prohibited the use of FID as either a routine or large scale automated screening technique.

Even when a well-defined fluidic interface can be created, buoyancy-driven convection due to density differences in the solutions causes complex mixing at the interface. To avoid this unwanted mixing, capillaries must be stored with the long axis parallel to gravity and the more dense solution on the bottom. This configuration creates a stable fluidic interface, but often causes nucleated crystals to fall away from the interface and out of the optimal growth conditions [10]. It has been proposed that free interface diffusion would therefore only realize its practical advantages in microgravity environments where gravity-induced convection is eliminated [10, 43]. However, the unusual properties of fluid flow in microfluidic devices make it both possible and practical to implement nearly ideal free interface diffusion conditions in terrestrial devices. At the length scales that are relevant in microfluidic devices, the Grashoff number (which measures the ratio of buoyant to viscous forces) is small so that convection is suppressed and mass transport is dominated by diffusion [46, 47]. Furthermore, the BIM method allows many near ideal fluidic interfaces to be established with negligible transient mixing. Lastly, since there are no significant concentration gradients parallel to gravity, nucleated crystals do not fall out of the ideal growth region.

Conventional free interface diffusion achieves high transient levels of supersaturation, but has a complicated spatial/temporal gradient due to the constant cross-section of the capillary. This gradient couples the kinetics and thermodynamics of traditional free interface diffusion assays in a way that μ FID does not.

In the μ FID assay, the fluidic interface is established between the two wells in a constricted channel where the cross-section is $10\text{ }\mu\text{m} \times 100\text{ }\mu\text{m}$. In contrast, the cross-section of a well is approximately $300\text{ }\mu\text{m} \times 100\text{ }\mu\text{m}$. This constriction acts as a high-impedance connection between the two channels, localizing the concentration gradient only to the length of the connecting channel. Fig. 11.8 shows a finite difference time domain simulation of the diffusive equilibration of a low molecular weight dye in two micro-wells coupled by a constricted channel (PDE Toolbox, MATLAB[®]; The MathWorks Inc. of Natick, MA, USA). With the exception of the region in close proximity to the inlet, no appreciable concentration gradient forms in the micro-well itself. In a protein crystallization experiment, this implies that as the wells equilibrate, the vast majority of the sample evolves simultaneously through a continuum of thermodynamic conditions. By monitoring the experiment over time it is therefore possible to determine the exact transient conditions that resulted in crystal nucleation.

In the optimization of crystallization conditions, it is often desirable to slow down the equilibration process so that favorable conditions are approached more slowly to produce fewer nucleation events and larger crystals. In vapor diffusion experiments, this regime can be achieved by methods such as placing a thin layer

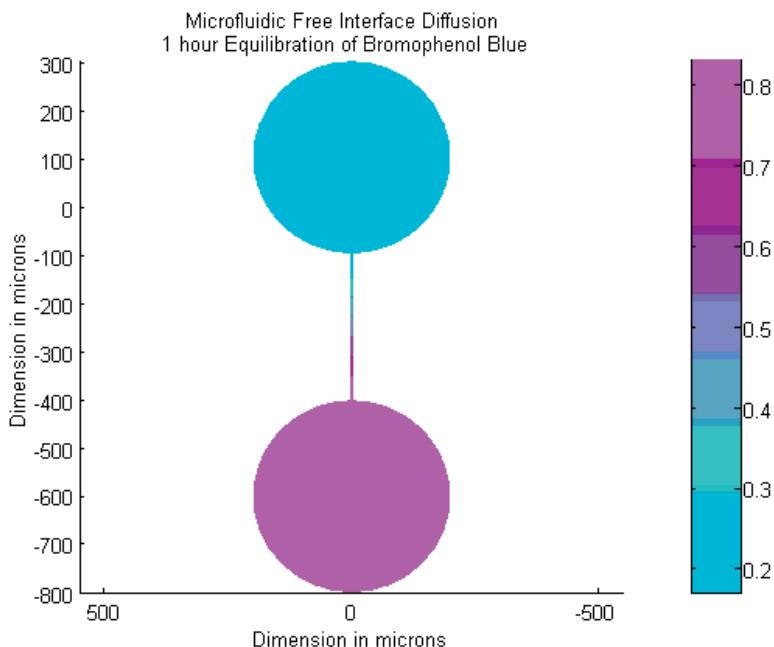


Fig. 11.8 Two-dimensional finite element modeling of diffusion of an organic dye in aqueous solution between two micro-wells connected by a constricted channel. Simulation shows the bulk of the concentration drop

occurring along the channel with no appreciable gradient within the micro-wells. For modeling purposes, constriction of the channel in the vertical dimension has been represented by a lateral constriction.

of semi-permeable oil over the precipitant, increasing the size of the crystallization drop to slow equilibration, or by inclusion of a chemical additive such as glycerol to the crystallization drop in order to reduce the vapor pressure. While these techniques are effective in slowing down the equilibration of the drop with the reservoir, they allow for only coarse control of the equilibration kinetics. Conversely, microfluidic free interface diffusion allows for precise and straightforward control of the equilibration rate while decoupling the kinetics and thermodynamics of crystallization.

The absence of any spatial gradient within the micro-wells allows for a simple analytical solution of the time-dependent evolution of the concentration in each chamber. The net transport of a diffusing species along the channel is equal to the product of the diffusive current density and the channel cross-sectional area. Since the channel is constricted in both height and width, the problem is one-dimensional and the gradient along the channel is given by the difference in concentration divided by the channel length. The equation governing the change of concentration in a well of volume V_1 and concentration $C_1(t)$ that is coupled to a second well of volume V_2 at concentration $C_2(t)$ is therefore given by:

$$\frac{dC_1(t)}{dt} = -DA \frac{C_1(t) - C_2(t)}{V_1 L} \quad (1)$$

Integration of this equation and application of the initial conditions $C_1(0)=C_0$, $C_2(0)=0$, gives

$$C_1 = \frac{C_0}{1 + \frac{V_2}{V_1}} + \frac{C_0}{1 + \frac{V_1}{V_2}} \times \exp\left[-D \frac{V_1 + V_2}{V_2} \frac{1}{L} \frac{A}{V_1} t\right] \quad (2)$$

From this equation, it can be seen that for a given diffusion constant and relative well volume, the rate of equilibration depends only on two characteristic lengths: the length of the connecting channel and the ratio of the well volume to the channel cross-sectional area. Thus, modifying the channel geometry allows for intuitive and accurate control over the kinetics of equilibration without changing the chemistry of the solutions. For example, by making the channel twice as long, the reaction proceeds at one half the speed, whereas reducing the channel cross-sectional area by a factor of two increases the equilibration time by the same factor. Furthermore, since these length scales only scale time in the exponent, the locus of concentrations (path through phase space) achieved during a complete equilibration depends only on the relative volume of the wells, and is independent of the channel geometry. The decoupling of the kinetics and thermodynamics of diffusive equilibration has important implications in crystal optimization where it is often desirable to approach crystallization conditions slowly while conserving the successful thermodynamic variables.

A further advantage of the constricted channel between the two wells is that only a very small fraction of the protein sample is exposed to large transient gradi-

ents that occur shortly after the interface is established. In micro-batch or hanging-drop experiments, the sudden addition of the precipitating agent to the protein sample induces rapid convective mixing, causing large transient concentration gradients throughout the drop, and may often result in the immediate precipitation of the protein. In the μ FID experiments only a very small volume of the protein is exposed to this gradient, so that more concentrated precipitating solutions should be used with negligible immediate precipitation, and hence ultimately higher levels of supersaturation may be achieved.

Experiments comparing the effectiveness of the chip to two conventional macroscopic techniques, vapor diffusion and microbatch, show that the chip may be used to conduct large numbers of nanoliter-scale assays while equaling or increasing crystallization success rates of these methods. In a recent study [36], in-chip crystallization experiments were conducted on 11 model macromolecules including 7 commercially available crystallization standards (Lysosyme, Glucose Isomerase, Xylanase, Thaumatin, Protease K, Bovine Trypsin and Beef Liver Catalase), 3 proteins with unpublished structures (bacterial primase catalytic core domain, type II topoisomerase ATPase domain and a mycobacterial RNase) and a bacterial 70S ribosome. Each target was screened against two or more standard sparse matrices of precipitants. To compare crystallization in chip against standard crystallization methods, crystallization experiments were repeated for 9 of the model macromolecules using the conventional micro-batch and hanging-drop techniques, thereby keeping the precipitant chemistries constant while varying the kinetic scheme for crystal growth. Successful crystallization conditions were identified in the μ FID format for all model macromolecules tested, and crystal growth was observed at incubation times ranging from 3 hours to 7 days. High quality crystals of 4 different proteins grown in chips are shown in Fig. 11.9.

A histogram comparing the number of successful experiments for each method (Fig. 11.10) shows that identical sparse matrix screens led to crystal growth more often in the chip than by conventional techniques in all but two cases [36]. For the bacterial primase catalytic core domain, 11 conditions producing needle crystals of dimensions greater than 100 μ m were detected on chip; no hits were initially observed in either macroscopic method. Further on-chip experimentation, optimizing around the crystallization conditions, identified in initial screens crystals produced whose largest dimension exceeded 400 μ m. These conditions were subsequently used to reproduce crystallization in microbatch and vapor diffusion formats, demonstrating that successful on-chip conditions may be exported to more conventional macroscopic techniques.

During these trials, the chip also produced crystals of two targets that had not been seen by conventional screening. A previously unidentified crystal form of the bacterial 70S ribosome was obtained in three conditions of a sparse matrix of precipitants (Hampton Crystal Screen I), demonstrating that large protein-nucleic acid complexes may be crystallized in chip (C. Hansen, A. Vila-Sanjurjo and J. Cate, personal communication). Crystals of a previously uncrystallized mycobacterial RNase were also obtained from a single experimental condition on chip, whereas no crystals had been observed for this sample despite prior extensive

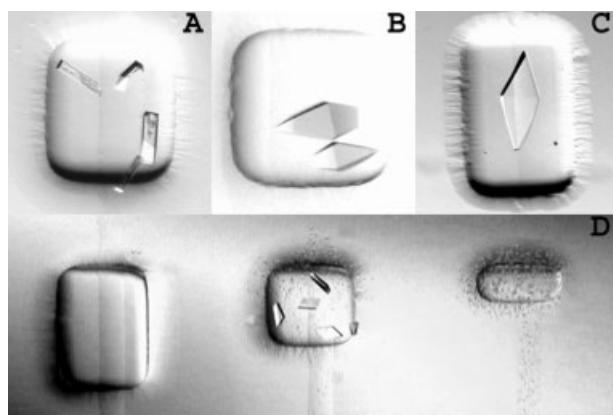


Fig. 11.9 High-quality protein crystals grown in the crystallization chip. (A) Glucose isomerase, 31 mg mL^{-1} in deionized water. Mixing ratio of 1:1 with 0.2 M magnesium acetate tetrahydrate, 20% w/v polyethylene glycol 8000, 0.1 M sodium cacodylate, pH 6.5. (B) Thaumatin, 50 mg mL^{-1} in 0.1 M ADA (Sigma Aldrich) pH 6.5. Mixing ratio of 1:1 with 0.8 M potassium sodium tartrate tetrahydrate, 0.1 M HEPES pH 7.5. (C) Chicken egg white lysosome, 50 mg mL^{-1} in 0.05 M potassium dihydrogen phosphate, 20% w/v polyethylene glycol 8000. (D) Crystals of type II topoisomerase ATPase domain/ADP showing the importance of mixing ratio as a screening parameter; 12 mg mL^{-1} in 100 mM sodium chloride, 20 mM TRIS pH 7.0. Mixing ratios of 4:1, 1:1 and 1:4 with 0.2 M ammonium fluoride, 20% w/v polyethylene glycol 3350, pH 6.2.

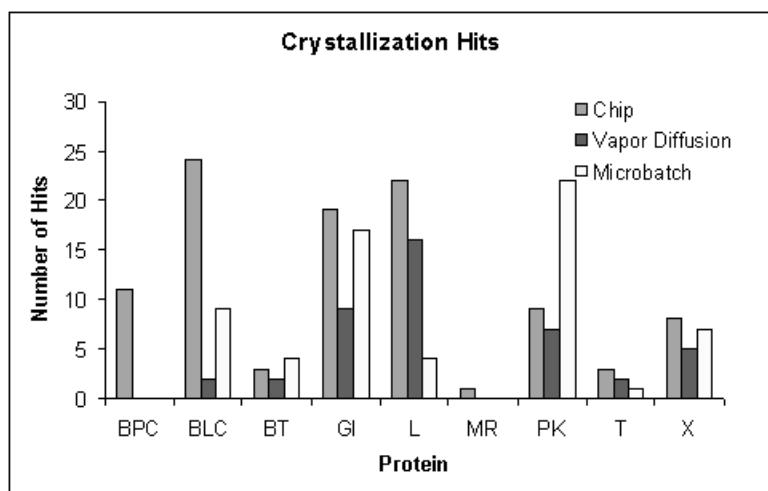


Fig. 11.10 Histogram of crystallization hits for sparse matrix screens of model proteins. Number of screens tested on each protein are lysosome (L)=2, glucose isomerase (GI)=2, protease K (PK)=1, bovine liver catalase (BLC)=1, xylanase (X)=2, bacterial primase catalytic core domain (BPC)=3, bovine pancreas trypsin (BT)=1, thaumatin (T)=1, mycobacterial RNase (MR)=3.

trials using traditional methods. Subsequent broad-based screening efforts for the RNase around this condition using the hanging-drop vapor diffusion set-ups proved successful, but only after the protein concentration was increased to $>40 \text{ mg mL}^{-1}$.

Consistent with the localization of large initial concentration gradients to the relatively small volume of the connecting channel, μ FID-based chip experiments result in reduced protein precipitation. It was observed that the mixing ratios and concentration of crystallization agents that lead to crystallization on chip often caused the protein to precipitate immediately in hanging drop and microbatch experiments. In the case of a type II topoisomerase ATPase domain, the final concentration of precipitating agent achievable in chip was 4 times greater than that possible for the microbatch method.

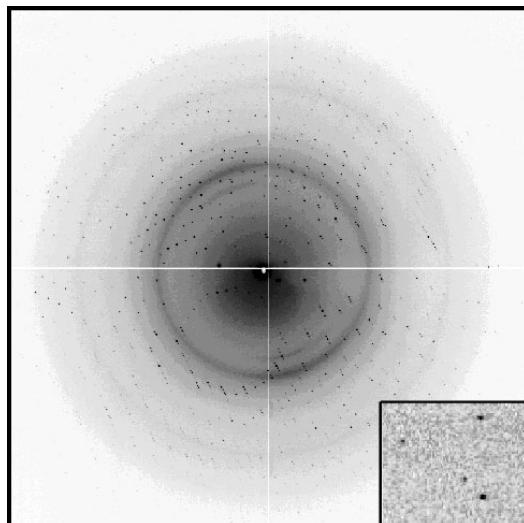
Finally, crystal growth in μ FID experiments was generally observed to be faster than in microbatch or hanging-drop. For the type II topoisomerase ATPase domain, crystal growth in microbatch required 2 weeks while crystals grown on chip with the same conditions appeared after only 4 hours of incubation. When crystals grew on chip in less than 12 hours, they were always observed on the protein side of the compound well, suggesting that the short crystallization times are due to the high degree of supersaturation achieved in the initial phase of diffusive equilibration.

Despite the small volumes, crystals of sufficient size for crystallographic structure determination may be grown and harvested directly from the chip. Fig. 11.11 shows a high-resolution diffraction pattern for a single thaumatin crystal grown from only 5 nL of protein solution [36]. The high-resolution data obtained from chip-grown crystals exceeds that of crystals grown by conventional ground-based techniques, and is comparable to that obtained from thaumatin crystals grown in space [48].

In general, crystallization experiments in space have suggested that microgravity environments may be well suited for the growth of large, high quality crystals [48–51]. The mechanism for the increased order of space-grown crystals is not completely known but is largely attributed to the lack of buoyancy-driven convection during crystal growth, an effect also present in terrestrial microfluidic devices [52–54].

Microfluidic technologies have the potential to play a dominant role in the development of high-throughput crystallography. While the simple parallel architecture of the present device is ideally suited to preliminary screening experiments, more sophisticated control and flow architectures will allow for increased functionality and applicability. Using the simple valve configuration as a building block, multi-valve elements such as peristaltic pumps, rotary mixers and addressable arrays have already been created on chip [22, 23, 38, 55–57]. These components have further been incorporated into higher-level microfluidic mixing and metering tools that can process samples in the sub-nanoliter range with picoliter accuracy. Given the current rate of progress in the field these components will no doubt lead to new technologies with applications both in and beyond protein crystallization.

Fig. 11.11 X-ray diffraction pattern from a single thaumatin crystal grown and harvested from a 5 nL chip micro-well. The inset shows a clean reflection at 1.35 Å resolution.



11.4 References

- 1 HOOKE, R., *Micrographia* **1665**.
- 2 OGATA, C. M., MAD phasing grows up. *Nat. Struct. Biol.* **1998**, *5* (*S*), 638–640.
- 3 CUSACK, S., et al., Small is beautiful: protein micro-crystallography. *Nat. Struct. Biol.* **1998**, *5* (*S*), 634–637.
- 4 CHAYEN, N. E., SARIDAKIS, E., Protein crystallization for genomics: towards high-throughput optimization techniques. *Acta Crystallogr., Sect. D, Biol. Crystallogr.* **2002**, *58* (*6*), 921–927.
- 5 AI, X., CAFFREY, M., Membrane protein crystallization in lipidic mesophases: Detergent effects. *Biophys. J.* **2000**, *79* (*1*), 394–405.
- 6 OSTERMEIER, C., MICHEL, H., Crystallization of membrane proteins. *Curr. Opin. Struct. Biol.* **1997**, *7* (*5*), 697–701.
- 7 ROSENBUSCH, J. P., Stability of membrane proteins: Relevance for the selection of appropriate methods for high-resolution structure determinations. *J. Struct. Biol.* **2001**, *136* (*2*), 144–157.
- 8 BAN, N., et al., A 9 angstrom resolution X-ray crystallographic map of the large ribosomal subunit. *Cell* **1998**, *93* (*7*), 1105–1115.
- 9 CRAMER, P., et al., Architecture of RNA polymerase II and implications for the transcription mechanism. *Science* **2000**, *288* (*5466*), 640–649.
- 10 MCPHERSON, A., *Crystallization of Biological Macromolecules*. 1st edn. Cold Spring Harbor Laboratory Press, **1999**.
- 11 RC, S., Development of high-throughput technologies for protein crystallography and structure-based drug design. *Abstracts of the Papers of the American Chemical Society* **2001**, *221* (*2*), U38.
- 12 RC, S., High-throughput protein crystallization. *Curr. Opin. Struct. Biol.* **2000**, *10* (*5*), 558–563.
- 13 ABOLA, E., et al., Automation of X-ray crystallography. *Nat. Struct. Biol.* **2000**, *7*, 973–977.
- 14 NYARSIK, M. U., et al., Development of a technology for automation and miniaturization of protein crystallization. *J. Biotechnol.* **2001**, *85*, 7–14.
- 15 SERVICE, R. F., Robots enter the race to analyse proteins. *Science* **2001**, *292* (*5515*), 187–188.
- 16 KANHOLZ, A. E., et al., Quantitative analysis of molecular interaction in a micro-

- fluidic channel: The T-sensor. *Anal. Chem.* **1999**, *71*, 5340–5347.
- 17** JENSON, D., Smaller, faster chemistry. *Nature* **1998**, *393*, 735–737.
- 18** JACOBSON, S.C., MCKNIGHT, T.E., RAMSEY, J.M., Microfluidic devices for electrokinetically driven parallel and serial mixing. *Anal. Chem.* **1999**, *71*, 4455–4459.
- 19** ZHANG, C., MANZ, A., Narrow sample channel injectors for capillary electrophoresis on microchips. *Anal. Chem.* **2001**, *73*, 2656–2662.
- 20** DERTINGER, S.K.W., et al., Generation of gradients having complex shapes using microfluidic networks. *Anal. Chem.* **2001**, *73*, 1240–1246.
- 21** QUI, C.X., HARRISON, D.J., Integrated self-calibration via electrokinetic solvent proportioning for microfluidic immunoassays. *Electrophoresis* **2001**, *22*, 3949–3958.
- 22** UNGER, M., et al., Monolithic microfabricated valves and pumps by multilayer soft lithography. *Science* **2000**, *288* (5463), 113–116.
- 23** THORSEN, T., MAERKL, S.J., QUAKE, S.R., Microfluidic large-scale integration. *Science* **2002**, *298*.
- 24** XIA, Y., WHITESIDES, G.M., Soft lithography. *Angew. Chem.* **1998**, *37*, 550–575.
- 25** KIM, Y., SUH, K., LEE, H., Fabrication of three-dimensional microstructures by soft molding. *Appl. Phys. Lett.* **2001**, *79* (14), 2285–2287.
- 26** CHIU, D.T., et al., Patterned deposition of cells and proteins onto surfaces by using three-dimensional microfluidic systems. *Proc. Natl. Acad. Sci. USA* **2000**, *97* (6), 2408–2413.
- 27** DELAMARCHE, E., Patterned delivery of immunoglobulins to surfaces using microfluidic networks. *Science* **1997**, *276*, 779–781.
- 28** LI, H.W., HUCK, W.T.S., Polymers in nanotechnology. *Curr. Opin. Solid State Mater. Sci.* **2002**, *6* (1), 3–8.
- 29** BREHMER, M., et al., Soft lithography on block copolymer films. *Abstracts of Papers of the American Chemical Society* **2002**, *224* (2), U502–U503.
- 30** TIEN, J., NELSON, C., CHEN, C., Fabrication of aligned microstructures with a single elastomeric stamp. *Proc. Natl. Acad. Sci. USA* **2002**, *99* (4), 1758–1762.
- 31** SHAH, H., et al., Direct generation of optical diffractive elements in perfluorocyclobutane (PFCB) polymers by soft lithography. *IEEE Photon. Technol. Lett.* **2000**, *12* (12), 1650–1652.
- 32** KANE, R., et al., Patterning proteins and cells using soft lithography. *Biomaterials* **1999**, *20* (23/24), 2363–2376.
- 33** XIA, Y., GM, W., Soft lithography. *Annu. Rev. Mater. Sci.* **1998**, *28*, 153–184.
- 34** BRITAIN, S., et al., Soft lithography and microfabrication. *Phys. World* **1998**, *11* (5), 31–36.
- 35** MONAHAN, J., GEWIRTH, A.A., NUZZO, R.G., A method for filling complex polymeric microfluidic devices and arrays. *Anal. Chem.* **2001**, *73*, 3193–3197.
- 36** HANSEN, C.L., et al., A robust and scalable microfluidic metering method that allows protein crystal growth by free interface diffusion. *PNAS* **2002**, *99*, 16531–16536.
- 37** HE, B., et al., A picoliter-volume mixer for microfluidic analytical systems. *Anal. Chem.* **2001**, *73* (9), 1942–1947.
- 38** CHU, H., UNGER, M.A., QUAKE, S.R., A microfabricated rotary pump. *Biomed. Microdev.* **2001**, *3*, 323–330.
- 39** STROOCK, A., et al., Chaotic mixer for microchannels. *Science* **2002**, *295* (5555), 647–651.
- 40** CARTER, C.W., BALDWIN, E.T., FRICK, L., Statistical design of experiments for protein crystal-growth and the use of a pre-crystallization assay. *J. Cryst. Growth* **1988**, *90* (1–3), 60–73.
- 41** JANCARIK, J., KIM, S.H., Sparse-matrix sampling – A screening method for crystallization of proteins. *J. Appl. Crystallogr.* **1991**, *24*, 409–411.
- 42** LUFT, J.R., DETITTA, G.T., Kinetic aspects of macromolecular crystallization. *Methods Enzymol.* **1997**, *276*, 110–130.
- 43** SALEMME, F.R., A free interface diffusion technique for the crystallization of proteins for X-ray crystallography. *Arch. Biochem. Biophys.* **1972**, *151*.
- 44** SYGUSCH, J., et al., Protein crystallization in low gravity by step gradient diffusion method. *J. Cryst. Growth* **1996**, *162* (3/4), 167–172.

- 45 GARCIA-RUIZ, J. M., et al., A supersaturation wave of protein crystallization. *J. Cryst. Growth* **2001**, *232* (1–4), 149–155.
- 46 NERAD, B. A., SHLICHTA, P. J., Ground-based experiments on the minimization of convection during the growth of crystals from solution. *J. Cryst. Growth* **1986**, *75*, 591–608.
- 47 GARCIA-RUIZ, J. M., et al., Agarose as crystallization media for proteins I: Transport processes. *J. Cryst. Growth* **2001**, *232* (1–4), 165–172.
- 48 BARNES, C. L., SNELL, E. H., KUNDROT, C. E., Thaumatin crystallization aboard the International Space Station using liquid-liquid diffusion in the Enhanced Gaseous Nitrogen Dewar (EGN). *Acta Crystallogr., Sect D, Biol. Crystallogr.* **2002**, *58* (5), 751–760.
- 49 WANG, Y. P., et al., Protein crystal growth in microgravity using a liquid/liquid diffusion method. *Micrograv. Sci. Technol.* **1996**, *9* (4), 281–283.
- 50 MILLER, T. Y., HE, X. M., CARTER, D. C., A comparison between protein crystals grown with vapor diffusion methods in microgravity and protein crystals using a gel liquid liquid diffusion ground-based method. *J. Cryst. Growth* **1992**, *122* (1–4), 306–309.
- 51 NG, J. D., et al., Comparative analysis of space-grown and earth-grown crystals of an aminoacyl-tRNA synthetase: space-grown crystals are more useful for structural determination. *Acta Crystallogr., Sect. D, Biol. Crystallogr.* **2002**, *58* (4), 645–652.
- 52 THOMAS, B. R., et al., Distribution coefficients of protein impurities in ferritin and lysozyme crystals – Self-purification in microgravity. *J. Cryst. Growth* **2000**, *211* (1–4), 149–156.
- 53 LIN, H., et al., Lower incorporation of impurities in ferritin crystals by suppression of convection: Modeling results. *Cryst. Growth Design* **2001**, *1* (1), 73–79.
- 54 CHERNOV, A. A., GARCIA-RUIZ, J. M., THOMAS, B. R., Visualization of the impurity depletion zone surrounding apoferritin crystals growing in gel with holoferitin dimer impurity. *J. Cryst. Growth* **2001**, *232* (1–4), 184–187.
- 55 CHOU, H., et al., A microfabricated device for sizing and sorting DNA molecules. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 11–13.
- 56 LIU, J., ENZELBERGER, M., QUAKE, S., A nanoliter rotary device for polymerase chain reaction. *Electrophoresis* **2002**, *23* (10), 1531–1536.
- 57 FU, A., et al., An integrated microfabricated cell sorter. *Anal. Chem.* **2002**, *74* (11), 2451–2457.

Subject Index

a

Abl 51, 64ff., 74
 Abl-gleevec 51
 ACE 211
 activation loop 65
 adenine 49, 61, 66
 adenine pocket 56
 AGC family of kinases 52, 61
 AKT 61
 AKT2 68
 aldehydes 93, 95, 111, 133
 Alk5 72
 alkaline phosphatase 211
 alkoxyketone 141
 alkoxymethyl ketones 140
 amicetin 118
 aminoacyl-4-aminohexosyl-cytosine antibiotics 118
 aminoacyl-tRNA 101
 androgen receptor (AR) 17
 angiotensin-converting enzyme (ACE) 211
 anilinopyrimidines 68
 anilinoquinazoline 61, 67, 70f.
 anisomycin 116, 120
 antibiotics 99, 103ff., 109, 113f., 116f., 120
 – chemical structures 106, 116
 – nucleoside analogues 117
 – resistance 99, 105, 110ff.
 anti-tumor activity 94, 119
 AP21967 201
 apoptosis 84f., 96
 aromatic-aromatic interactions 135
 aryl pyridopyrimidines 70
 aryl sulfonamide 63
 aryl-pyridopyrimidones 71
 A-site 100f., 104f., 114, 117, 119
 aspartate 2-oxoglutarate aminotransferase (AspAT) 197

ATP 49f., 52, 60, 65
 – analogues 202
 – binding site 50, 52f., 55
 – binding pocket 152, 157, 159, 161f.
 – binding set 74
 – binding site 56f., 63, 67
 – non-hydrolyzable 49
 Aurora A 68
 automated analysis of electron density interpretation 221
 – AutoSolve 221, 224f., 227ff.
 – x-ligand 221
 automation 220, 224
 AZ-242 36
 azepanone 138ff.
 azithromycin 106, 108

b

baculovirus expression 4
 balanol 58f., 63, 68
 bamicetin 118
 Bcl-2 85
 BIRB-796 74
 bisindolyl maleimides 69
 blasticidin 116
 blasticidin S 118f.
 boronates 93
 bovine pancreatic ribonuclease 213
 Bower model 57

c

calpain inhibitor I 95
 calpains 93
 cancer 47, 68, 83, 85
 carbinolamine 112
 carbinolamine bond 111
 carbomycin A 106, 108, 114f.
 carbomycin M 108, 111

- carrier device 244
 - casein kinase 2 58
 - caspases 85
 - cathepsin B 128
 - cathepsin K inhibitors 127
 - cathepsins 93, 128
 - Cdc28 204
 - cell-cycle progression 96
 - CEP1612 93f.
 - C-helix 48, 51ff.
 - chemical inducers of dimerization (CID) 200
 - Chk1 66
 - chloramphenicol 115ff.
 - chymotrypsin 181
 - C-H \cdots O=C hydrogen bonds 31, 32
 - CI-1033 70, 72
 - CID 202
 - c-Kit 65
 - clarithromycin 112
 - clindamycin 106, 114
 - clinical trials 68
 - collagenase 176, 179, 181
 - computational methods and programs 151ff.
 - Alpha Site Finder 151
 - CHARMM 151
 - de novo design 153
 - LeapFrog 153, 163
 - LEGEND 72, 153, 155ff., 167
 - LUDI 153, 163
 - SiteID 151
 - virtual screening 72, 152
 - SEEDS 72, 156f., 158, 167
 - C-Raf 73
 - crystal formation 236
 - micro-crystals 236
 - phase separation 236
 - plate stacks 236
 - single crystals 236
 - spherulites 236
 - thin plates 236
 - crystal growth 245
 - crystal nucleation 245, 247
 - crystallization chip 243f., 251
 - crystallization techniques 237, 247
 - batch methods 237
 - dialysis methods 237
 - free interface diffusion 247
 - hanging drop 247, 250
 - liquid-gel diffusion 237
 - liquid-liquid diffusion 237
 - microbatch 247, 250
 - micro-dialysis 237
 - vapor diffusion 237, 248
 - free-interface diffusion 246
 - c-Src 60, 66, 71, 202f.
 - crystallization
 - phase space 245f.
 - cyclic diaminoketone inhibitors 136
 - cyclin A 56, 65
 - cyclin B1 85
 - cyclin-dependent kinase 1 (cdk-1) 73
 - cyclin-dependent kinase 2 (cdk-2) 48f., 51, 53, 55f., 58, 61f., 64ff., 71ff., 148, 150, 165, 229
 - activation mechanism 150
 - mutants 165
 - cyclin-dependent kinase 4 (cdk-4) 55, 67, 71f., 74, 147ff., 151
 - ATP binding pocket 151
 - homology modeling 148
 - cyclin-dependent kinase 1 (cdk-5) 54f.
 - cyclin-dependent kinase 1 (cdk-6) 54
 - cyclins 52, 63, 85
 - cyclophilin A (CyP) 190
 - cyclosporin A (CsA) 190
 - cysteine protease inhibitors 131
 - cysteine proteases 128, 131
 - cytochrome P450 enzymes 222
- d**
- de novo ligand design 72, 73, 154f., 160, 167
 - deletion mutagenesis 210
 - des-chloroflavopiridol 58
 - dexamethasone 18
 - DFG motif 50, 53, 55, 74
 - DFLB 93f.
 - diabetes 47
 - diabetic retinopathy 68
 - diacylcarbohydrazides 142f.
 - diaminoketone 134, 139
 - diaminoketone-based inhibitors 133
 - dianilinophthalimides 69
 - diaryl urease 68
 - diethyl stilbene 8
 - 5,11-cis-diethyl-5,6,11,12-tetrahydrochrysene-2,8 diol 14
 - diffraction pattern 253
 - dihydroeponemycin 94
 - dihydrofolate reductase (DHFR) 202
 - dihydroxytestosterone 17
 - DNA gyrase B subunit 210
 - DNA-binding domain (DBD) 200

e

- E1 83
- E2 83
- E3 83
- E64 131ff.
- e.coli expression 54
- ecotin 171ff., 179
 - crystallization 175
 - expression 173
 - primary site 173
 - purification 74
 - secondary site 173, 176, 182
 - tool for drug design 183
- EGFR 68
- EGFR 61, 67, 69ff.
- eicosapentaenoic acid (EPA) 34
- elongation factor 101
- elongation factor Tu 189
- EphA2 68
- EphB2 54
- epoxomicin 94f.
- epoxyketones 93f.
- ERK 67
- ERK2 49, 53f., 60
- erythromycin 106, 112
- E-site 101
- estradiol 5, 7f., 42, 193
- estrogen receptor (ER) 5, 7, 8, 13, 193
 - ER- α 5, 9, 193
 - ER- β 5, 9, 193
 - ligands 6ff.
 - SERMs 15
 - agonists 13
- dexamethasone 202
- exit tunnel 107, 110ff., 116f.

f

- factor binding site 101
- factor Xa 178f., 181
- FAK 68
- farglitazar 39, 41f.
- FGFR 71
- FGFR1 55, 60ff., 71
- fibrate ligand 36, 41
- FK506-binding protein (FKBP) 201
- FKBP rapamycin-associated protein (FRAP) 201
- flavopiridol 67f.
- fragment-based lead discovery 153, 223, 225 ff., 230
- FRAT1 55
- furanone 141f.
- FxxF motif 52

g

- gatekeeper 70f., 73
- gatekeeper pocket 56, 59f., 62, 69ff.
- gatekeeper residue 49, 60f.
- genistein 12
- Gleevec 51, 74
- glucocorticoid receptor (GR) 18, 202
 - agonist 18
- GluR2 receptor 210
- glycine rich loop 51
- glycosidases 211
- glycosylation sites 212
- gougerotin 118
- granzyme B 177, 179, 181
- granzyme C 181
- GSK3 54
- guanine 110

h

- H-7 57f.
- H-chain ferritin 212
- Hck 52, 58, 60, 64, 204
- high-throughput screening 68
- high-throughput crystallization 237
- high-throughput crystallography 217ff., 251
- HIV-1 integrase 213
- homology modeling 66, 69ff., 74, 128, 167
- human genome 47
- hydrogen bond donor-acceptor 58
- hydrophobic collapse 63
- 4-hydroxytamoxifene 10
- hymenaldisene 68
- hymenialdisine 57, 58

i

- IFNy 86
- IGF-1 55
- immune surveillance 83
- immunoproteasome 86, 89, 91f.
- indirubin-5-sulfonate 57f., 65
- inhibitor design 67ff., 72
- inhibitor-induced binding 63
- interface pins 244
- IRK 49, 51, 53, 55
- isoquinoline sulfonamides 58f.

k

- KDR 68
- kinase domain 48, 54
- kinase inhibitors 56, 57f.
 - pharmacophore 56
 - structures 57

kinases 47 ff., 202
 – activation loop 50
 – catalysis 49, 53
 – conformational flexibility 50
 – constructs 54
 – crystallization 54
 – hinge region 48
 – serine/threonine 55, 60
 – solvent channel 53
 – substrate binding 49
 – tertiary structure 48
 – T-loop 50
 – tyrosine 66, 71

I

L 868276 57, 68
 lactacystin 93 ff.
 lactam amide 69
 large ribosomal subunit 99
 Lck 60, 64, 66
 leptin 213
 leupeptin 132 f.
 LexA 202
 limited proteolysis 54, 210
 lincomycin 106
 lincosamides 105, 114
 lipophilic plug 56, 59, 61 f., 71
 LY 333531 67 f.

m

macrolides 105 ff., 110 ff.
 – 14-membered 107, 110, 112 ff.
 – 15-membered 107 f., 111
 – 16-membered 107 f., 110 ff.
 MAPKAPK2 68
 MEK 67
 methotrexate 202
 MG132 93 f.
 MHC class I molecules 92
 microbatch 246
 micro-crystallization 235
 microcrystals 245
 microelectromechanical systems 238
 microfluidics 237
 miniaturization 237
 MSK1 52
 multilayer soft lithography 238
 muscle atrophy 83
 mutagenesis 55, 61
 mycaminoose 109
 mycinose 109 ff.

n

nanocrystallization 220
 NL VS 94
 (Ntn-) hydrolase 85, 88 f.
 NU2058 63, 65
 NU6094 57 f., 63
 NU6102 63, 65
 nuclear hormone receptor (NHR) 1 ff., 3, 5, 7, 191
 – activation function 1 (AF-1) 1
 – activation function 2 (AF-2) 1
 – agonist conformation 5
 – antagonist conformation 7
 – A-ring ketone 16 ff., 42
 – A-ring phenol 8 ff., 14, 16, 20 f., 42
 – co-repressor 1, 3, 34, 41
 – co-activators 1, 3, 5, 15, 33 f., 41
 – ligand-binding domain (LBD) 1 ff., 20, 191
 – partial agonist 12, 15
 – antagonist 3, 5, 9 f., 12, 14, 41
 – agonists 3, 6 ff., 15, 18, 21
 – nucleoside analogue antibiotics 117

o

orthogonal ligand-receptor pair
 (OLRP) 187 ff., 193, 195
 – estrogen receptor 193
 – RAR 195
 – RXR 195
 – near drugs 196
 OSI-774 61 f.
 osteoporosis 127
 oxindoles 68
 oxyanion hole 88

p

p27 83, 85
 p38 mitogen-activated protein kinase
 (p38) 202
 p53 83, 85
 PA28 86
 PA700 85 f.
 PAK 54
 papain 128, 131, 134, 143
 PD 173074 60
 PD 173955 65
 PD0153035 67
 PD-173955 67
 PDGFR 65, 71
 peptidyl transferase center 102 f., 104 f., 108, 110, 114, 118
 – secondary structure 104
 peptidyl transferase reaction 101 f., 104

- peroxisome proliferator-activated receptor (PPAR)
 – isotype selective ligands 34f.
 – PPAR- α 3, 34
 – PPAR- δ 34
 – PPAR- γ 2, 34
 PGFR-SU5402 51
 PGPH 86, 91f.
 pharmacophore 73
 phosphorylation 55
 piperidinones 137
 PKA 49, 52, 58f., 61, 63f., 66, 69
 PKB 52
 PKC 58f., 61
 polar surface site 61, 71
 PP1 60
 PP2 60
 pressure reservoir 244
 pressurized outgas priming 241
 PRK 52
 PRK2 52
 proenzymes 131
 progesterone receptor (PR) 16, 190
 proteasome 83ff.
 proteasome inhibitors 93, 95
 – irreversible 94
 – structure of proteasome inhibitor complexes 95
 protein engineering 214
 protein kinases 204
 protein synthesis 99
 PS-273 93
 PS-341 93f.
 PS-519 94
 PSI 94
 P-site 100f., 105, 114, 118f.
 PTK 787 67f., 68
 puromycin 115ff., 119f.
 Purvalanol B 62
 pyrazolopyrimidines 69f.
 pyridinylimidazoles 64
 pyridopyrimidines 71
 pyrrolidinone 137, 141f.
 pyrrolopyrimidines 70
- q**
 quercetin 57f., 68
 quinazolines 68
- r**
 Raf 67
 raloxifene 9
 raloxifene-core 8
- rapamycin 201
 rational drug design 99, 111ff., 120f.
 retinoic acid (RA) 24
 – all-trans RA 24
 – 9-cis RA 24
 retinoic acid receptor (RAR) 23, 27, 191,
 195
 – enantiomers binding to RAR γ 27ff., 43
 – isotype selectivity 30ff.
 – RAR α 30
 – RAR β 30
 – RAR γ 25, 195
 – recognition element for binding 30
 retinoic X receptor (RXR) 2, 23, 195
 – RXR- α 33, 195
 – RXR- β 33
 – selective ligands 33ff.
 rheumatoid arthritis 47
 rhinovirus 3C protease 210
 Rho-specific guanine nucleotide dissociation inhibitor (RhoGDI) 213
 – glutamate mutant 213
 – lysine mutant 213
 ribose rings 49
 ribosome 102
 – exit tunnel 104f.
 – structure 99, 102
 – structure in complex with antibiotic 100
 – 30S subunit 99
 – 50S subunit 99, 104, 119f.
 robotic automation 237
 rosiglitazone 37, 42
 roxithromycin 112
 rRNA 102f., 107f., 111, 115, 118
 – secondary structure 102f.
- s**
 SB 203580 59f.
 SB 238039 62
 SB 220025 60
 selectivity 47ff., 60, 73f.
 serine proteases 171ff., 179
 SGK1 52
 silicone rubbers 239
 small ribosomal subunit 99
 SNAAP 86, 89
 soft lithography 238f.
 sparse matrix screens 251
 sparsomycin 116, 119f.
 spiramycin 106, 108, 111
 20S proteasome 85ff., 96
 – active sites 88

– catalytic mechanism 91
– structure 86, 88
– substrate specificities 89
26S proteasome 85f.
Src 61
Src kinases 52, 71
70S ribosome 101
staurosporine 57f., 63f., 66, 68f.
steroid ligand selectivity 19, 21
STI-571 51f., 64f.
streptogramin A 106, 114f., 121
streptogramin B 105f., 114f.
streptogramins 114
structural genomics 217f., 221f.
structure-based design 65, 69, 72, 127,
133f.
structure-guided mutagenesis 191, 194,
196ff.
SU4984 61f.
SU5402 52
substrate-based design 72
sugar pocket 56, 59, 62, 70
surface site 56, 59, 62
surrogate 66f., 69
synchrotron radiation 221

t

tetracycline 121
thiomethyl ketones 140
thrombin 181
thymidylate synthase (TS) 209, 212
Tie2 51
TMC-95A 94f.
Traxler model 57
triphosphate group 49
triphosphate region 56, 59, 62
Twitchin kinase 51, 54
tylosin 106, 108f., 110f.

u

urokinase 211, 228

v

vapor diffusion 246
VEGFR2 49, 55, 61
vinyl sulfones 93f.
virginiamycin M 106
vitamin D receptor (VDR) 22, 210
VX-745 74

z

ZD1839 61f.
ZD6474 61f.
zinc finger domain (ZFD) 201