

WOODHEAD PUBLISHING SERIES IN BIOMEDICINE

# A PRACTICAL GUIDE TO RATIONAL DRUG DESIGN

HONGMAO SUN



# **A Practical Guide to Rational Drug Design**

## **Related titles**

*Orphan Drugs: Understanding the Rare Disease Market and its Dynamics*  
(ISBN 978 1 907568 09 1)

*Concepts and Models for Drug Permeability Studies*  
(ISBN 978 0 08 100094 6)

Woodhead Publishing Series in Biomedicine:  
Number 81

# A Practical Guide to Rational Drug Design

*Sun Hongmao*



ELSEVIER

AMSTERDAM • BOSTON • CAMBRIDGE • HEIDELBERG  
LONDON • NEW YORK • OXFORD • PARIS • SAN DIEGO  
SAN FRANCISCO • SINGAPORE • SYDNEY • TOKYO

Woodhead Publishing is an imprint of Elsevier



Woodhead Publishing is an imprint of Elsevier  
80 High Street, Sawston, Cambridge, CB22 3HJ, UK  
225 Wyman Street, Waltham, MA 02451, USA  
Langford Lane, Kidlington, OX5 1GB, UK

Copyright © 2016 Sun Hongmao. Published by Elsevier Ltd. All rights reserved.

No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval system, without permission in writing from the publisher. Details on how to seek permission, further information about the Publisher's permissions policies and our arrangements with organizations such as the Copyright Clearance Center and the Copyright Licensing Agency, can be found at our website: [www.elsevier.com/permissions](http://www.elsevier.com/permissions).

This book and the individual contributions contained in it are protected under copyright by the Publisher (other than as may be noted herein).

### Notices

Knowledge and best practice in this field are constantly changing. As new research and experience broaden our understanding, changes in research methods, professional practices, or medical treatment may become necessary.

Practitioners and researchers must always rely on their own experience and knowledge in evaluating and using any information, methods, compounds, or experiments described herein. In using such information or methods they should be mindful of their own safety and the safety of others, including parties for whom they have a professional responsibility.

To the fullest extent of the law, neither the Publisher nor the authors, contributors, or editors, assume any liability for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions, or ideas contained in the material herein.

ISBN: 978-0-08-100098-4 (print)

ISBN: 978-0-08-100105-9 (online)

### British Library Cataloguing-in-Publication Data

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

**Library of Congress Control Number:** 2015941459

For information on all Woodhead Publishing publications  
visit our website at <http://store.elsevier.com/>



Working together  
to grow libraries in  
developing countries

[www.elsevier.com](http://www.elsevier.com) • [www.bookaid.org](http://www.bookaid.org)

# Dedication

To my mother, Jinzhu Xu, a wise woman.

This page intentionally left blank

# Contents

<b>Introduction to the Book</b>	<b>ix</b>
<b>Foreword</b>	<b>xi</b>
<b>Acknowledgements</b>	<b>xiii</b>
<b>About the Author</b>	<b>xv</b>
<b>Part One Structure-Based Ligand Design</b>	<b>1</b>
<b>1 Structures, Limitations, and Pitfalls</b>	<b>3</b>
1.1 Introduction	3
1.2 The limitations of experimentally determined structures	3
1.3 The pitfalls of misusing structural information	7
1.4 Protein structural change upon activation	10
References	12
<b>2 Structure-Based Ligand Design I: With Structures of Protein/Lead Compound Complex Available</b>	<b>15</b>
2.1 Introduction	15
2.2 Case study 1: BACE1 – Fill the pocket by growing a molecule	16
2.3 Case study 2: heat shock protein 90 – Restore the electrostatic complementarity	36
2.4 Case study 3: estrogen receptor $\alpha$ agonists recognized by optimized pharmacophore models	43
2.5 Summary	57
References	58
<b>3 Structure-Based Ligand Design II: With Structure of Protein/Lead Compound Complex Unavailable</b>	<b>61</b>
3.1 Introduction	61
3.2 Case study 1: Plk1 kinase domain inhibitors as antitumor drugs	62
3.3 Case study 2: XIAP inhibitors to trigger apoptosis as an antitumor therapy	71
3.4 Case study 3: Bcl-xl inhibitors as anticancer drugs	82
3.5 Case study 4: design of kinase/bromodomain-containing 4 dual inhibitors	89
References	104
<b>4 Homology Modeling and Ligand-Based Molecule Design</b>	<b>109</b>
4.1 Introduction	109
4.2 Case study 1: prediction of human Yes1 kinase structure	110



---

4.3	Case study 2: homology modeling of human melanocortin-4 receptor, a G protein-coupled receptor target	121
4.4	Case study 3: ligand-based approaches to human MC4R	148
4.5	Summary	155
4.6	Summary of part I	155
	References	156
<b>Part Two QSAR and ADMET Predictions</b>		<b>161</b>
<b>5</b>	<b>Quantitative Structure–Activity Relationships: Promise, Validations, and Pitfalls</b>	<b>163</b>
5.1	Introduction	163
5.2	QSAR and its role in drug discovery	168
5.3	Preparation of datasets	171
5.4	Geometrical description of PLS and SVM	173
5.5	Support vector machine	178
5.6	Roles of molecular descriptors	181
5.7	Validation of QSAR models	182
5.8	Pitfalls in QSAR modeling	183
5.9	Summary	189
	References	189
<b>6</b>	<b>Quantitative Structure–Property Relationships Models for Lipophilicity and Aqueous Solubility</b>	<b>193</b>
6.1	Introduction	193
6.2	Lipophilicity as estimated by $\log P$	193
6.3	Atom type-based molecular descriptors and their optimization	196
6.4	Less complex PLS linear regression model and highly accurate SVR model of $\log P$	205
6.5	Is $\log D$ more relevant for drug discovery?	209
6.6	Aqueous solubility is a key property of drug molecules	211
6.7	QSPR modeling of aqueous solubility	213
6.8	Summary	220
	References	222
<b>7</b>	<b><i>In Silico</i> ADMET Profiling: Predictive Models for CYP450, P-gp, PAMPA, and hERG</b>	<b>225</b>
7.1	Introduction	225
7.2	CYP450 for drug metabolism	225
7.3	Intestinal absorption and PAMPA models	240
7.4	Prediction of P-gp activities	250
7.5	Toxicity in drug discovery	255
7.6	Predictive models for hERG	256
7.7	Delivery of modeling results	263
7.8	Summary	265
	References	266
<b>Index</b>		<b>269</b>

# Introduction to the Book

This book is intended for junior computational chemists and graduate students who are interested in learning tactics to deal with real-life problems in rational drug design. It will also be useful to experienced medicinal chemists who have already been exposed to molecular modeling and desire to gain in-depth knowledge on the capabilities and boundaries of computer-aided drug design and property predictions.

Today we have more protein structures of therapeutic interest, we have more sophisticated molecular modeling software, and we have accumulated a huge amount of high-quality biologic data; naturally, high expectation has been placed on molecular modeling to play a more important role in drug discovery. However, the reality lags behind expectation. This book is intended to uncover the great potentials of molecular modeling and to have computational chemists' voices heard.

This book is divided to two parts. In the first part, 10 case studies are used to elaborate how to use the right tools to solve the right problems in structure-based ligand design. Different strategies are explored for designing novel drugs when protein–ligand complex structures are determined and when only limited structural information of target proteins or active ligands is available. Many common diseases—various cancers, Alzheimer's disease, endocrine disorder, and obesity—are covered in the case studies. The coverage of protein targets is maximized, including kinases, proteases, G protein–coupled receptors (GPCRs), nuclear receptors, chaperone proteins, epigenetic proteins, protein–protein interactions, and so on. Structure features and structural changes upon activation are reviewed for kinases and GPCRs.

The second part of the book focuses on the key properties of druglike molecules and how to make these properties immediately available for decision making in drug discovery. Critical questions, such as what might cause the high attrition rate of clinical trials, are addressed from the standpoint of a computational chemist. The three determinant components for a predictive and robust quantitative structure–activity relationships (QSAR) model—data, molecular descriptors, and statistics—are discussed in detail. Implementation and optimization of an atom type–based molecular descriptor system are highlighted. Highly accurate models are constructed for logP, solubility, CYP450 isoforms, PAMPA, P-gp, and hERG on the basis of exactly the same descriptors. The performance of different methods to rebalance biased training data is analyzed. The potential pitfalls of QSAR modeling, such as overfitting, are also covered. (The datasets and atom type calculator are available online.)

This page intentionally left blank

# Foreword

By targeting, in a systematic and pragmatic fashion, the many challenges facing medicinal chemists who are focused on drug discovery and development, Dr. Hongmao Sun has provided an excellent introduction to both the power and the limitations of molecular modeling, quantitative structure–activity relations, and structure-based design. Dr. Sun brings expertise in the area derived from both industry as well as academia, providing a synergistic interface between theory and logistics. Exposure to students and their enthusiastic, but naïve, attitudes forces simplifications that are balanced by real examples. Convey the concept; then add essential complications by example.

This book is a gift for academics whose exposure to drug discovery is limited but who want to expose their students to the realities and excitement of the hunt for new therapeutics. In addition, synthetic chemists who are new to the pharmaceutical industry will find much that they need to understand regarding the strengths and weaknesses of computational approaches. Knowledge protects the naïve from exploitation as well as guiding those with expertise.

**Garland R. Marshall, PhD**

Professor of Biochemistry and Molecular Biophysics,  
Adjunct Professor of Chemistry  
Washington University in St. Louis

This page intentionally left blank

# Acknowledgements

I am greatly indebted to a gentleman who has influenced my career so much that I wouldn't be a computational chemist today without him. In 1997, I randomly wrote an email to Professor Garland Marshall to ask for any opening position in his group at Washington University in St. Louis. Subsequently, he invited me for a visit, but I had to reject his invitation because I had a tough schedule to finish my PhD thesis; to take care of my newborn baby; and most important, to begin a local job as a software developer. Professor Marshall strongly suggested that I visited his group before making a final decision. He did not want me to waste my training and scientific background. After learning my difficulties, he actually flew to the Boston area to interview me. I was touched by his unexpected decision, and our meeting changed my career path. I am deeply grateful to Professor Marshall for being so understanding, kind, and encouraging to the unknown graduate student that I was.

I am grateful to my wife, Ashley, for always placing her career and needs behind mine and for freeing me from household chores so that I could focus on writing this book.

I am indebted to Professor Jane Tao at Rice University and Mr. Alan Sun at University of Maryland for sharing their opinions on the manuscripts. Both readers and I appreciate Ms. Lynda Boomer's proofreading. Without her tremendous work and continuous encouragement, the book would not have been completed smoothly and in a timely manner.

Finally, I want to acknowledge my appreciation for the continuous support that Dr. Christopher Austin, Dr. Anton Simeonov, and Mr. Ajit Jadhav at the National Center for Advancing Translational Sciences, has provided.

**Hongmao Sun**

May 2015

This page intentionally left blank

# About the Author

Dr. Hongmao Sun received his first PhD degree from the University of Science and Technology of China by studying atomic structures of high-temperature superconductors using extended X-ray absorption fine structure. He was awarded a second PhD degree in medicinal and computational chemistry by Clark University. After postdoctoral research under the guidance of Prof. Garland Marshall at Washington University Medical School in St. Louis, he joined Hoffmann-La Roche as a computational chemist in 1999. Bringing 11 years of industrial experience, he switched to Biotechnology HPC Software Applications Institute in Frederick, Maryland, in 2010. He is currently a research scientist at the National Center for Advancing Translational Sciences of the National Institutes of Health.



This page intentionally left blank

Part One

# Structure-Based Ligand Design

This page intentionally left blank

# Structures, Limitations, and Pitfalls

# 1

## 1.1 Introduction

The terms *structure-based design* and *structure-based drug design* are sometimes used interchangeably. In this book, the concept of structure-based ligand design is confined to the processes of improving the potency of a ligand on the basis of structural information of proteins or protein–ligand complexes, which are the focus of the first part of the book. The term *structure-based drug design* (or *rational drug design*) is reserved for the multiple-dimensional optimization of properties such as physico-chemical, ADME (absorption, distribution, metabolism, and elimination), and toxicologic profile of a drug molecule to meet the predefined pharmacologic efficacy, which is addressed in the second part of this book.<sup>1</sup>

The adoption of the new technologies and the ever-increasing demand for protein structural information by the pharmaceutical industry keep boosting the output of experimentally determined three-dimensional (3D) structures of target proteins and protein–ligand complexes.<sup>1,2</sup> Not only did the size of the Protein Data Bank (PDB) experience an exponential growth in the past 40 years,<sup>3,4</sup> but many therapeutically important targets, such as G protein–coupled receptors (GPCRs),<sup>5–11</sup> also became available in the past decade. In 2013 alone, more than 10,000 new structures were deposited in the PDB worldwide (<http://www.wwpdb.org/stats.html>). The large amount of structural information lays a solid ground for structure-based ligand design. The role played by structure-based ligand design is becoming indispensable, and nowadays it has become an integral part of modern drug discovery.<sup>12</sup> As structural information plays an increasingly important role in drug discovery, it is critical for us to acknowledge the limitations of the experimentally determined structures so that the pitfalls of misusing structural information can be avoided. Choosing an appropriate structure from the beginning and bearing in mind the inherent flexibility of proteins are the two keys toward a successful structure-based ligand design campaign.

## 1.2 The limitations of experimentally determined structures

*Wisdom is fortified, not destroyed, by understanding its limitations.*

—Mortimer J. Adler

Single-crystal X-ray diffraction is the most widely used technique for determining 3D structures of proteins and protein–ligand complexes.<sup>2</sup> The ever-increasing availability of the structures of the therapeutically relevant proteins, including the difficult targets such as GPCRs,<sup>5–11</sup> greatly facilitated the research on structure-based ligand