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

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

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

Dedication

To my mother, Jinzhu Xu, a wise woman.

This page intentionally left blank

Contents

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 physicochemical, 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.[1]

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.[1,2] Not only did the size of the Protein Data Bank (PDB) experience an exponential growth in the past 40 years,[3,4] but many therapeutically important targets, such as G protein–coupled receptors (GPCRs),[5–11] also became available in the past decade. In 2013 alone, more than 10,000 new structures were deposited in the PDB worldwide (http://www.wwwpdb.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.[12] 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.[2] The ever-increasing availability of the structures of the therapeutically relevant proteins, including the difficult targets such as GPCRs,[5–11] greatly facilitated the research on structure-based ligand

A Practical Guide to Rational Drug Design. http://dx.doi.org/10.1016/B978-0-08-100098-4.00001-6