COMPUTATIONAL DRUG DESIGN

A Guide for Computational and Medicinal Chemists

DAVID C. YOUNG
Computer Sciences Corporation

A JOHN WILEY & SONS, INC., PUBLICATION

COMPUTATIONAL DRUG DESIGN

COMPUTATIONAL DRUG DESIGN

A Guide for Computational and Medicinal Chemists

DAVID C. YOUNG
Computer Sciences Corporation

A JOHN WILEY & SONS, INC., PUBLICATION

Published by John Wiley & Sons, Inc., Hoboken, New Jersey
Published simultaneously in Canada

Library of Congress Cataloging-in-Publication Data:

Young, David C., 1964–
Computational drug design / David C. Young.
p.; cm.
Includes bibliographical references and index.
ISBN 978-0-470-12685-1 (cloth/CD)

1. Drugs—Design—Mathematical models. 2. Drugs—Design—Data processing.
I. Title.
[DNLM: 1. Computational Biology—methods. 2. Drug Design. 3. Biochemical Phenomena.
4. Chemistry, Pharmaceutical—methods. 5. Drug Delivery Systems.
6. Models, Chemical. QV 744 Y69c 2009]
RS420. Y68 2009
615′.190285--dc22 2008041828

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

This book is dedicated to my grandfathers, Harvey Turner and Ray Young.

Harvey Turner had the intelligence to work his way up from a draftsman to Chief Engineer at Donaldsons. Then he had the wisdom to leave that high pressure career behind and spend the next two decades teaching art.

Ray Young dropped out of high school to help make ends meet during the great depression. He never returned to school, but was the most widely read and knowledgeable person I have every met.

CONTENTS

PREFACE

A pharmaceutical company utilizing computational drug design is like an organic chemist utilizing an NMR. It won't solve all of your problems, but you are much better off with it than without it.

The design of a new drug is an incredibly difficult and frustrating task. If it weren't for the potential to earn equally incredible profits, the massive costs and aggravation over failed experiments would dissuade any reasonable person from undertaking such a career. There is no one scientific technique used to design a new pharmaceutical product. It is instead a collaborative process in which every available technique, and a few more invented on the spur of the moment, are utilized in order to achieve the desired results.

There are books that talk about drug design tools, algorithms, and mathematical functions, and books that give some results showing that one compound worked better than another for inhibiting a particular enzyme. However, these books spend surprisingly little time discussing the process that the chemist goes through to actually design a new drug molecule. This book is oriented around the way that computational techniques are utilized in the drug design process.

Typical design processes for a number of drug development scenarios are presented in the first part of the book. Multiple drug design processes are presented, because the process itself changes depending upon whether the drug target is a protein, DNA, a target within the central nervous system, etc. The design processes presented in this text do not reflect the process at any one specific pharmaceutical company, but are rather typical work flows incorporating the elements that are used in one way or another at almost all

pharmaceutical research campuses. The chapters on the drug design process are intended to show how each of the computational techniques are typically utilized. The comparison of different drug design processes illustrates where specific computational tools would and would not be appropriate. The text presents many rules of thumb for choosing which tools are best utilized under certain situations.

The second part of the book has a series of chapters, each focusing on one computational technique. The chapters on each of the computational techniques are intended to give a solid understanding of the strengths and weaknesses of the method. The underlying theory is discussed in concept, but with little if any mathematical derivation. The processes for using the software and important issues that tend to arise are described. Where there are significant differences between available software packages, those issues are discussed. However, the text is not specific to one manufacturer's software. The relative merits of various methods are discussed, and, where possible, a table with quantitative comparisons is presented.

The third part of the book gives a few chapters discussing related topics. These are topics that drug design chemists should have some familiarity with, but are not usually engaged in on a daily basis. Fields of research so new that they are still being defined at the time this book was written are also introduced here. Since any detailed information on such subjects would be obsolete before the ink on this book is dry, some of these introductions are kept intentionally broad and conceptual.

In a book that covers a broad subject area, it is always difficult to choose which references to include for each chapter. For this text, I have taken a two-fold approach. Key references are listed at the end of each chapter in an annotated bibliography. These references tend to be the next place that readers should look for additional information on the topics discussed in the chapter. This is supplemented by a longer reference list included on the accompanying CD. Readers wishing to delve very deeply into a particular subject will find this larger list of references valuable.

This book is very industry-centric. The discussions of when and how tools are used is based on a typical pharmaceutical industry drug design process. As such, I have intentionally avoided using cartoon-like illustrations of geometric figures fitting together. The majority of the figures in this book are screen shots of actual software packages that drug designers might use on a daily basis. This is the environment that a drug designer in the pharmaceutical industry must learn to work in.

For students interested in pursuing a career in the drug design field, this text is intended to give an ideal starting point for their studies. The text assumes a solid background in chemistry, a basic understanding of biochemistry, and only minimal previous exposure to computational chemistry.

Researchers already employed in the drug design field will be particularly interested in the tables comparing accuracies of docking methods. There is also a fairly large table of bioisosteric substitutions. Providing an overview of the whole field may turn out to be this book's greatest contribution.

I wish you the best of success in pursuing your drug design activities.

DAVID C. YOUNG

ACKNOWLEDGMENTS

There is a popular myth that books are written by solitary people typing away in a lonely, deserted house. Indeed, there are many hours spent in front of a keyboard. However, a book would never come into being without the help, support, and hard work of the author's family, colleagues, co-workers, editors, graphic artists, and random other people saying, "Wow, that sounds complicated."

My family has been exceptionally tolerant of my ever-present laptop in the car, during swimming lessons, in front of the TV, and at this very minute sitting off to the side as my wife Natalie displays her stained glass work at an art show. My oldest son Gregory is a man of very few words, but the occasional "cool dad" speaks volumes. My daughter Ariel thinks it is neat that her dad is a scientist, but still won't ask for help with her college freshman chemistry homework. My youngest, Isaac, has little interest in anything that doesn't involve video games or reptiles, but he seems to consider docking calculations with solvation and entropy corrections to be dad's form of video game.

My current job at the Alabama Supercomputer Center allows me the chance to interact with faculty and students of many different disciplines throughout the state of Alabama. Randy Fulmer and the Alabama Supercomputer Authority staff are always interested to hear about the scientific research utilizing the supercomputers here. I've had bosses both good and bad, and David Ivey at CSC is definitely the best. Charles Wright and Derek Gottlieb always think of me as the software guy. They won't forget the day they asked about quantum chemistry software and got way more than they bargained for (the rest of the staff is afraid to ask).

This is my second book with John Wiley & Sons. I wouldn't consider working with any other publisher as long as Wiley will have me. Anita Lekhwani and Rebecka Ramos have been wonderful to work with. There are many others at Wiley who contribute to creating a high quality book as they format the tables, integrate the artwork, and lovingly cover the manuscript in red ink.

Within the pharmaceutical field, I have had the pleasure of working with some excellent people. Andy Peek, now at Integrated DNA Technologies Inc., manages to be top notch at bioinformatics without succumbing to the high pressure of the drug design world. It has also been my privilege to work with Brad Poland, now at Pfizer, who is a wonderful crystallographer and co-worker. Working with Stephan Reiling, now at Aventis, and the entire SARNavigator development team was the most enjoyment I ever got from my job. Mitch Polley, who left Tripos to return home to Australia, has become a good friend as well as teaching me much about drug design.

I wanted the majority of the figures in the book to show commercial drug design software, which is the environment that drug designers must learn to work in. I greatly appreciate getting demo copies of software from ACD, Accelrys, Cambridge Crystallographic Data Centre, Chemical Computing Group, Conflex, COSMOlogic, SimBioSys, Simulations Plus, Tripos, and Wavefunction for this purpose. Those same companies were invited to contribute product literature, white papers, and demo software to the accompanying CD.

DAVID C. YOUNG

ABOUT THE AUTHOR

David Young's career has taken him to the far corners of computational chemistry. He was assistant director of drug design for a now nonexistent startup company, eXegenics. David once taught introductory science and graduate computer programming courses at Auburn University. He has also written quite a bit of software for Tripos and others. Dr Young is currently employed by Computer Sciences Corporation (CSC) as a chemistry software expert, under contract to the Alabama Supercomputer Authority. Much earlier in his life, he ran the nuclear reactor aboard a ballistic missile submarine.

Dr Young received his PhD in chemistry from Michigan State University, under the direction of James Harrison. He also has degrees in computational mathematics and business. His previous book, *Computational Chemistry: A Practical Guide for Applying Techniques to Real World Problems*, has been on the John Wiley & Sons bestseller list.

David currently lives in Huntsville, Alabama, where he provides software technical support to users of the Alabama Supercomputer Center.

SYMBOLS AND ACRONYMS USED IN THIS BOOK

1D	one-dimensional
2D	two-dimensional
2.5D	two-and-a-half-dimensional
3D	three-dimensional
3D-QSAR	three-dimensional quantitative structure–activity relationship
Å	ångström
ACD	Advanced Chemistry Development
ACE	angiotensin-converting enzyme
AcrB	multidrug efflux pump protein
ADAM	a docking method from the Institute of Medicinal Molecular Design
ADME	absorption, distribution, metabolization, excretion
ADMET	absorption, distribution, metabolization, excretion, and toxicity
AI	artificial intelligence
AMBER	Assisted Model Building and Energy Refinement
AMOEBA	a force field for proteins
API	applications programming interface
ASP	atomic solvation parameter
AZT	azidothymidine
BBB	blood–brain barrier
BCI	Barnard Chemical Information
C++	a computer programming language

CAESA	Computer Assisted Estimation of Synthetic Accessibility
CAMEO	Computer Assisted Mechanistic Evaluation of Organic Reactions
CAOS	computer-aided organic synthesis
CASINO	Computer-Aided Synthesis Inference for Organic Compounds
CASP	Critical Assessment of Techniques for Protein Structure Prediction
CATH	Class, Architecture, Topology, Homology
CC	coupled cluster
CFD	computational fluid dynamics
CFF	Consistent Force Field
CFF93	Consistent Force Field 1993
CFR	Code of Federal Regulations
CHARMM	Chemistry at Harvard Macromolecular Mechanics
CHEAT	Carbohydrate Hydroxyls represented by Extended Atoms
CHIRON	Chiral Synthon
CI	configuration interaction
CLIX	a docking method
ClogP	method for predicting log *P*
CMC	Comprehensive Medicinal Chemistry
CNS	central nervous system
CoMFA	Comparative Molecular Field Analysis
CoMSIA	Comparative Molecular Shape Indices Analysis
CPE	Chemical Potential Equalization
CPU	central processing unit
CSI	Carbó Similarity Index
CVFF	Consistent Valence Force Field
DEREK	Deductive Estimation of Risk from Existing Knowledge
DFT	density functional theory
DRF90	Direct Reaction Field 90
DNA	deoxyribonucleic acid
EFF	Electron Force Field
EROS	Elaboration of Reactions for Organic Synthesis
EVB	Empirical Valence Bond
FDA	Food and Drug Administration
FEP	free energy perturbation
FEP-MD	Free Energy Perturbation Molecular Dynamics
FLOG	Flexible Ligands Oriented on Grid
FRED	Fast Rigid Exhaustive Docking

FSSP	Fold Classification based on Structure–Structure Alignment of Proteins/Families of Structurally Similar Proteins
GA	genetic algorithm
GB/SA	Generalized Born Solvent Accessible
GLUT2	glucose transporter 2
GPCR	G-protein-coupled receptor
GROMACS	Groningen Machine for Chemical Simulations
HADDOCK	High Ambiguity Driven Biomolecular Docking
HASL	Hypothetical Active Site Lattice
hERG	human ether-a-go-go related gene
HF	Hartree–Fock
HOLOWin	Holosynthon and Windows
HOMO	highest occupied molecular orbital
hPEPT1	human intestinal small peptide carrier
HQSAR	hologram quantitative structure–activity relationship
HTVS	high throughput virtual screening
HUPO	Human Proteome Organisation
IC_{50}	concentration at which activity is decreased by 50%
ICM	a docking program from MOLSOFT
IGOR	Interactive Generation of Organic Reactions
InChI	IUPAC International Chemical Identifier
IRC	intrinsic reaction coordinate
K_d	dissociation constant
K_I	inhibition constant
K_M	Michaelis constant
LBDD	ligand-based drug design
LD_{50}	lethal dose for 50% of test subjects
LHASA	Logical and Heuristics Applied to Synthetic Analysis
LIE	Linear Interaction Energy
LIGIN	a docking program
log *D*	log *P* for ionization state at a specific pH
log *P*	octanol–water partition coefficient
log *S*	aqueous solubility
log S_w	intrinsic water solubility
LOO	leave one out
LR	linear regression
LUDI	a scoring method for docking and *de novo* design
LUMO	lowest unoccupied molecular orbital
MAb	monoclonal antibody
MCASE	Multi-Computer Automated Structure Evaluation
MCS	maximal common subgraph

MD	molecular dynamics
MEP	Molecular Electrostatic Potential
MFA	Molecular Field Analysis
MlogP	a method for predicting log P
MLP	molecular lipophilic potential
MM	molecular mechanics
MM+	a molecular mechanics force field
MM1	a molecular mechanics force field
MM2	a molecular mechanics force field
MM2X	a molecular mechanics force field
MM3	a molecular mechanics force field
MM4	a molecular mechanics force field
MMFF	Merck Molecular Force Field
MMX	a molecular mechanics force field
MOGA	Multiobjective Genetic Algorithm
MOMEC	Molecular Mechanics
MPn	Møller–Plesset Perturbation Theory ($n = 2, 3, \ldots$)
MRSA	methicillin-resistant *Staphylococcus aureus*
MSA	molecular shape analysis
MVP	Molecular Visualization and Processing Environment
NBTI	Non-Boltzmann Thermodynamic Integration
NLM	nonlinear map
NMR	nuclear magnetic resonance
NOE	nuclear Overhauser effect
OCSS	Organic Chemistry Synthesis Simulator
OPLS	Optimized Potential for Liquid Simulations
OPLS-2001	Optimized Potential for Liquid Simulations 2001
OPLS-2005	Optimized Potential for Liquid Simulations 2005
OPLS-AA	Optimized Potential for Liquid Simulations All Atom
OPLS-UA	Optimized Potential for Liquid Simulations United Atom
OSET	Organic Synthesis Exploration Tool
OWFEG	One Window Free Energy Grid
PAMPA	parallel artificial membrane permeability assay
PBE	Poisson–Boltzmann Equation
PB/SA	Poisson Boltzmann Solvent Accessible
PCA	principal components analysis
PFF	Polarizable Force Field
π (pi)	electron orbitals or bonds perpendicular to the sigma bond
pK_a	acidity equilibrium constant
PLP	piecewise linear potential
PLS	partial least squares

PMF	potential of mean force
PTMs	posttranslational modifications
QCFF/PI	Quantum Consistent Force Field for Pi electrons
QM	quantum mechanics
QMFF	Quantum Mechanical Force Field
QM/MM	a method combining quantum mechanics and molecular mechanics
QPLD	Quantum-Polarized Ligand Docking
QSAR	quantitative structure–activity relationship
QSM	Quantum Similarity Measure
QXP	a force field-based docking program
ReaxFF	Reactive Force Field
RFF	Reaction Force Field
RMSD	root mean square deviation
ROCS	Rapid Overlay of Chemical Structures
ROSDAL	Representation of Organic Structure Descriptions Arranged Linearly
RNA	ribonucleic acid
SAR	structure–activity relationship
SBDD	structure-based drug design
SCR	structurally conserved region
SDS	synthesis design systems
SCOP	Structural Classification of Proteins
SECS	Simulation and Evaluation of Chemical Synthesis
SESAM	Search for Starting Materials
SIBFA	Sum of Interactions Between Fragments *Ab Initio* Computed
SIE	Solvated Interaction Energy
SLN	SYBYL Line Notation
SMILES	Simplified Molecular Input Line Entry Specification
SMoG	Small Molecule Growth
SP	standard precision
SST	Starting Material Selection Strategies
SUA	structural unit analysis
SVL	Scientific Vector Language
SYBYL	the Greek word for oracle names a force field and software from Tripos
SYNGEN	Synthesis Generation
TPSA	topological polar surface area
UBCFF	Urey–Bradley Consistent Force Field
UFF	Universal Force Field
VALIDATE	a docking scoring function

VR	variable region
WLN	Wiswesser Line Notation
WODCA	Workbench for the Organization of Data for Chemical Applications
XP	extra precision
YETI	a force field