**The IMA Volumes
in Mathematics
and its Applications**

Volume 108

Series Editors

Willard Miller, Jr. Robert Gulliver

Springer Science+Business Media, LLC

Institute for Mathematics and its Applications IMA

The **Institute for Mathematics and its Applications** was established by a grant from the National Science Foundation to the University of Minnesota in 1982. The IMA seeks to encourage the development and study of fresh mathematical concepts and questions of concern to the other sciences by bringing together mathematicians and scientists from diverse fields in an atmosphere that will stimulate discussion and collaboration.

The IMA Volumes are intended to involve the broader scientific community in this process.

Willard Miller, Jr., Professor and Director

* * * * *

IMA ANNUAL PROGRAMS

1982–1983	Statistical and Continuum Approaches to Phase Transition
1983–1984	Mathematical Models for the Economics of Decentralized Resource Allocation
1984–1985	Continuum Physics and Partial Differential Equations
1985–1986	Stochastic Differential Equations and Their Applications
1986–1987	Scientific Computation
1987–1988	Applied Combinatorics
1988–1989	Nonlinear Waves
1989–1990	Dynamical Systems and Their Applications
1990–1991	Phase Transitions and Free Boundaries
1991–1992	Applied Linear Algebra
1992–1993	Control Theory and its Applications
1993–1994	Emerging Applications of Probability
1994–1995	Waves and Scattering
1995–1996	Mathematical Methods in Material Science
1996–1997	Mathematics of High Performance Computing
1997–1998	Emerging Applications of Dynamical Systems
1998–1999	Mathematics in Biology
1999–2000	Reactive Flows and Transport Phenomena
2000–2001	Mathematics in Multi-Media

Continued at the back

Donald G. Truhlar W. Jeffrey Howe
Anthony J. Hopfinger Jeff Blaney
Richard A. Dammkoehler
Editors

Rational Drug Design

With 55 Illustrations

Springer

Donald G. Truhlar
Department of Chemistry
University of Minnesota
Minneapolis, MN 55455-0431, USA
Anthony J. Hopfinger
Department of Medicinal Chemistry
and Pharmacognocny
College of Pharmacy
University of Illinois at Chicago
Chicago, IL 60612-7231, USA
Richard A. Dammkoehler
Department of Computer Science
Washington University
St. Louis, MO 63130, USA

W. Jeffrey Howe
Unit 7247-267-1
Pharmacia & Upjohn Inc.
Kalamazoo, MI 49001, USA
Jeff Blaney
Metaphorics LLC
130 Alta Avenue
Piedmont, CA 94611, USA

Series Editors:
Willard Miller, Jr.
Robert Gulliver
Institute for Mathematics and its
Applications
University of Minnesota
Minneapolis, MN 55455, USA

Mathematics Subject Classifications (1991): 92B99, 92C40, 92B10, 82B80

Library of Congress Cataloging-in-Publication Data

Rational drug design / [edited by] Donald G. Truhlar . . . [et al.].

p. cm. — (The IMA volumes in mathematics and its applications ; 108)

Includes bibliographical referenes.

ISBN 978-1-4612-7159-8 ISBN 978-1-4612-1480-9 (eBook)

DOI 10.1007/978-1-4612-1480-9

1. Drugs—Design—Computer simulation—Congresses. 2. Drugs—Design—Mathematical models—Congresses. 3. Computer-aided design—Congresses. 4. QSAR (Biochemistry)—Congresses. I. Truhlar, Donald G., 1944— . II. Series.

[DNLM: 1. Drug Design congresses. 2. Models, Molecular congresses. 3. Computer Simulation congresses. QV 744 R236 1999] RS420.R36 1999

615'.19—dc21

DNLM/DLC

98-55465

Printed on acid-free paper.

© 1999 Springer Science+Business Media New York

Originally published by Springer-Verlag New York Berlin Heidelberg in 1999

Softcover reprint of the hardcover 1st edition 1999

All rights reserved. This work may not be translated or copied in whole or in part without the written permission of the publisher (Springer Science+Business Media, LLC), except for brief excerpts in connection with reviews or scholarly analysis. Use in connection with any form of information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed is forbidden. The use of general descriptive names, trade names, trademarks, etc., in this publication, even if the former are not especially identified, is not to be taken as a sign that such names, as understood by the Trade Marks and Merchandise Marks Act, may accordingly be used freely by anyone.

Authorization to photocopy items for internal or personal use, or the internal or personal use of specific clients, is granted by Springer Science+Business Media, LLC provided that the appropriate fee is paid directly to Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923, USA (Telephone: (508) 750-8400), stating the ISBN number, the title of the book, and the first and last page numbers of each article copied. The copyright owner's consent does not include copying for general distribution, promotion, new works, or resale. In these cases, specific written permission must first be obtained from the publisher.

Production managed by A. Orrantia; manufacturing supervised by Nancy Wu.

Camera-ready copy prepared by the IMA.

SPIN 10709698

FOREWORD

This IMA Volume in Mathematics and its Applications

RATIONAL DRUG DESIGN

is based on the proceedings of a workshop on “Mathematical and Computational Issues in Drug Design” which was an integral part of the 1996–97 IMA program on “MATHEMATICS IN HIGH-PERFORMANCE COMPUTING.” The workshop brought together top researchers in computer-aided drug discovery, computational chemistry, mathematics, and computer science to present state-of-the-art research in both the science and the underlying mathematics.

We thank Donald G. Truhlar of University of Minnesota (Chemistry and Supercomputer Institute), W. Jeffrey Howe of Pharmacia and Upjohn, Anthony J. Hopfinger of University of Illinois at Chicago (Pharmacy), Jeff Blaney of Metaphorics, and Richard A. Dammkoehler of Washington University (Computer Science) for their excellent work in organizing the workshop and editing the proceedings.

We also take this opportunity to thank the National Science Foundation (NSF), Minnesota Supercomputer Institute (MSI), and the Army Research Office (ARO), whose financial support made the workshop possible.

Willard Miller, Jr., Professor and Director

Robert Gulliver, Associate Director

PREFACE

Drug research and discovery are of critical importance in human health care and are becoming increasingly expensive, while the need for new drugs is also increasing. Computational approaches for discovery and optimization of drug leads have proven successful in many recent research programs. (A “lead” compound is one with sought for bioactivity but which requires further optimization, for example to improve its bioavailability or reduce certain side reactions, in order to become a useful drug.) Methods for drug lead discovery and optimization have grown in their effectiveness not only because of improved understanding of the basic science — the biological events and molecular interactions that define a target for therapeutic intervention — but also because of advances in algorithms, representations, and mathematical procedures for studying drug processes. In order to promote the interaction of mathematicians, computer scientists, and chemists to further the progress in the field and alert researchers to the opportunities for interdisciplinary research, the University of Minnesota’s Institute for Mathematics and Its Applications and the University of Minnesota Supercomputer Institute sponsored a Workshop on Rational Drug Design on the Minneapolis campus, April 7–11, 1997. The workshop was devoted primarily to mathematical and computational issues in drug design. This volume contains the proceedings of that Workshop.

The workshop brought together top researchers in computer-aided drug discovery, computational chemistry, mathematics, and computer science to present state-of-the-art research in both the science and the underlying mathematics and to identify new problems for possible collaborations. General subject areas of the workshop included receptor-based applications such as binding energy approximations, molecular docking, and de novo design; non-receptor-based applications such as molecular similarity, conformational analysis, and structural diversity; molecular dynamics simulations; and solvation issues related to partitioning of a solute between aqueous and nonpolar media. The workshop focused on the mathematical procedures and algorithms upon which the scientific applications are based. These include graph theory and topology, non-linear multidimensional optimization, the processing and representation of information obtained from simulation studies, global optimization and search strategies, plus performance enhancement through parallel computing architectures. In addition to the oral presentations, the workshop included two panel discussions, one examining the most important current problems in drug design that may be computationally tractable, and the second focusing on emerging areas of study in which improvements in scientific knowledge over the next few years may enable the fruitful application of computational methods. The overall goal of this workshop was to bring together scientists and mathematicians

to examine the current state of this very broad and interdisciplinary field of research, and to identify the areas where cross-fertilization of ideas and collaborative research might most effectively advance the field.

A broad landscape of high-profile topics in computer-assisted molecular design (CAMD) directed to drug design were covered over the course of the Workshop. Several of these topics involve finding working solutions to problems where mathematicians and mathematically oriented physical scientists might provide new direction and insight. Among the problems presented were two which permeate many fields of research today. One of these two problems is sampling high-dimensional spaces. In CAMD applications, sampling problems arise in performing conformational analysis, searching for molecular alignments, and carrying out molecular simulations. The other of these two problems is determining all extrema of a high-dimension functional interrelationship. Identification of stable states (conformations, crystal structures, and ligand-receptor binding modes) corresponds to finding minima in potential energy functions; barriers to reaction, conformational interconversion, and melt transitions correspond to saddle points. Optimum structure-activity relations correspond to extrema of penalty functions or of average deviations of prediction from experiment in correlation models.

The construction of correlation models actually presents a whole family of mathematical problems identified at the Workshop. The most direct CAMD application area in correlation model construction is quantitative structure-activity relationship (QSAR) analysis. Central to the construction of QSARs is the maximum extraction of information from data sets which are highly oversubscribed, that is, for which the number of independent variables is much greater than the number of dependent variables, as is the case for applications of comparative molecular field analysis (CoMFA). The opposite case, the number of independent variables being much less than the number of dependent variables, an undersubscribed problem, is also of concern. Here the issue is to get the most representative, robust, and reliable model from the data set.

The multiple extrema problem also arises in constructing statistical models. Data sets can contain multiple stable correlation models. There is a need to know how many distinct models are inherent to the data set and to be able to rank those models with respect to measures of significance, reliability, and robustness.

Theory can also contribute to expanding the data set by producing theoretical data to characterize potential drug molecules that have never been made. Use of force fields to determine the likely steric and electrostatic fit of a proposed molecule to a binding site is the most prominent example. An example of a frontier in this area was a paper on the use of quantum mechanical modeling to calculate charge distributions, solvation energies, and partition coefficients for small molecules.

Many of the opportunities for new mathematical contributions to this

field are occasioned by the recent prominence of combinatorial libraries. Such libraries can be synthesized and assayed by robots, but mathematical modeling can play a very important role in prioritizing the targets to be made and making sense of the results. Because very large numbers of molecules can be involved, there is a new emphasis on rapidity and efficiency of the computational tools.

In a stimulating after-dinner speech at the Workshop banquet, Dr. Ralph Hirschmann of the University of Pennsylvania drew on his long industrial experience to present another perspective on drug design, focusing on many non-computational issues. For example, he discussed a 1997 paper in the *Journal of the American Chemical Society* where researchers found that the shape of a base inserted in DNA, rather than its hydrogen-bonding ability, may be the key to the polymerase recognition process that leads to faithful copying of DNA. Although this is an experimental result, by underscoring the role of 3-D shape it further dramatizes the role that computation can play in designing biotechnological molecules that mimic one or another capability of natural biological molecules. Dr. Hirschmann, however, took exception to the use of “rational drug design” and “computer-aided drug design” as near synonyms; he claims that pharmaceutical researchers were not totally irrational before they had computers! While this may be true, we hope that the Workshop and these proceedings will continue the trend of more and more productive uses of mathematical and computational techniques for the eminently humanistic goal of the design of new and better drug molecules.

Donald G. Truhlar

W. Jeffrey Howe

Anthony J. Hopfinger

Jeff Blaney

Richard A. Dammkoehler

CONTENTS

Foreword	v
Preface	vii

PANEL DISCUSSIONS

Important current problems in drug design that may be computationally tractable	1
<i>Panelists: Dave Doherty, Bill Dunn, Graham Richards, and Doug Rohrer</i>	
<i>Moderators: A.J. Hopfinger and W.J. Howe</i>	
New problems that should be addressed in the next ten years	7
<i>Panelists: Gordon Crippen, Simon Kearsley, Garland Marshall, and Phil Portoghese</i>	
<i>Moderator: Jeff Blaney</i>	

MOLECULAR PROPERTIES

Matching of chemical and biological structures using subgraph and maximal common subgraph isomorphism algorithms	11
<i>Peter Willett</i>	
Molecular similarity	39
<i>W. Graham Richards and Daniel D. Robinson</i>	
Modeling the effect of solvation on structure, reactivity, and partitioning of organic solutes: Utility in drug design	51
<i>Candee C. Chambers, David J. Giesen, Gregory D. Hawkins, Christopher J. Cramer, Donald G. Truhlar, and Wouter H.J. Vaes</i>	
Cell-based methods for sampling in high-dimensional spaces	73
<i>Jason A. Rush</i>	
A preliminary multiobjective analysis of the chemotherapeutic benefits of hairpin-linked polyamides	81
<i>Wynn L. Walker, David S. Goodsell, Richard E. Dickerson, and Elliot M. Landaw</i>	

RECEPTOR-BASED MODELING

Using structural information for the creative design of molecules	93
---	----

Colin McMartin and Regine S. Bohacek

GrowMol, a de novo computer program, and its application to thermolysin and pepsin: Results of the design and synthesis of a novel inhibitor	103
--	-----

Regine Bohacek, Colin McMartin, Peter Glunz, and Daniel H. Rich

Deducing objective site models by mixed integer programming	115
---	-----

Gordon M. Crippen

NUMERICAL METHODS

Molecular dynamics information extraction	127
---	-----

Dennis Sprous, Weidong Wang, Ganesan Ravishanker, Matthew A. Young, and David L. Beveridge

Ionic charging free energies using Ewald summation	149
--	-----

Tom Darden

Genetic function approximation: Evolutionary construction of novel, interpretable, nonlinear models of experimental data	163
--	-----

David Rogers

Applications of distributed computing to conformational searches	191
--	-----

Brian T. Luke