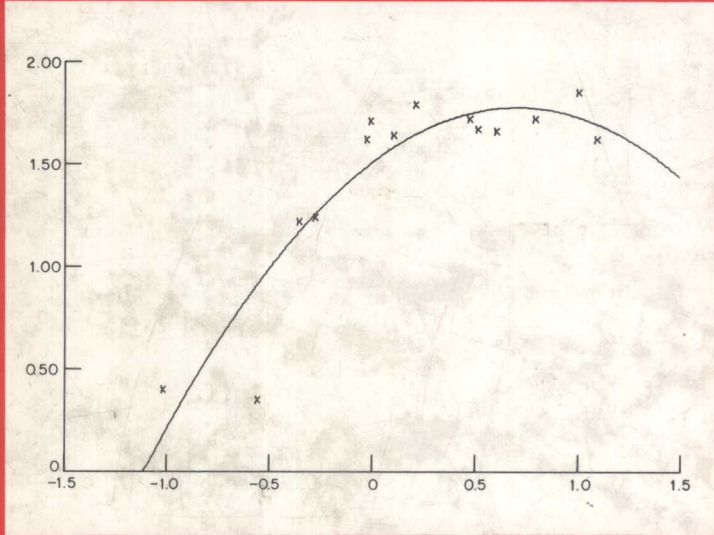


medicinal research series

volume 8



Yvonne
Connolly
Martin

QUANTITATIVE DRUG DESIGN

A Critical Introduction

Quantitative Drug Design

A Critical Introduction

Yvonne Connolly Martin

Abbott Laboratories

North Chicago, Illinois



MARCEL DEKKER, INC. New York and Basel

Library of Congress Cataloging in Publication Data
Martin, Yvonne Connolly, 1936-

Quantitative Drug Design.

(Medicinal research ; v. 8)

Includes bibliographical references and index.

1. Structure-activity relationship (Pharmacology)
2. Structure-activity relationship (Pharmacology)--
Mathematical models. 3. Drugs. I. Title. II. Se-
ries.

QP906.S75M37

615'.19

77-26709

ISBN 0-8247-6574-5

COPYRIGHT © 1978 by MARCEL DEKKER, INC. ALL RIGHTS RESERVED

Neither this book nor any part may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, microfilming, and recording, or by any information storage and retrieval system, without permission in writing from the publisher.

MARCEL DEKKER, INC.

270 Madison Avenue, New York, New York 10016

Current printing (last digit):

10 9 8 7 6 5 4 3

PRINTED IN THE UNITED STATES OF AMERICA

INTRODUCTION TO THE SERIES

This series of monographs was conceived by the late Dr. Fred Schueler of Tulane University. His untimely death came prior to the completion of the first volumes of the series, and his friend and very able colleague, Dr. Alfred Burger of the University of Virginia, took on the task of editing the first three volumes of the series. Dr. Burger's heavy responsibilities forced him to withdraw from further active participation with the series, and the next three volumes were selected from books already under contract to Marcel Dekker, Inc., with Dr. Gary Grunewald of the University of Kansas as Consulting Editor. At this point an Editorial Advisory Board was selected and Grunewald became the Series Editor. The traditions established under the leadership of Schueler and Burger will be continued.

It is hoped that books for the series will serve a useful role in the areas of medicinal chemistry, chemical pharmacology, and biochemistry. It is the intent of the Editor and the Editorial Advisory Board that books selected for the series should be timely and fill a definite need in the general areas mentioned. We welcome suggestions for future monographs in the series.

Gary L. Grunewald

PREFACE

My enthusiasm for writing this book arose from my experiences as a referee of manuscripts for the Journal of Medicinal Chemistry as well as from the unreliability of predictions of drug potency which I have made. Examination of the faults made in logic, statistics, or interpretation of quantitative structure-activity relationships by myself or others has made me aware of some of the traps into which one can fall in the execution of such studies.

The objective of this book is to provide a self-study guide for workers unfamiliar with the various methods used in quantitative drug design. Since particular attention will be paid for the standards of quality work, the reader should learn enough to be able to do publishable analyses. I fully realize that in setting such guidelines my own shortcomings become obvious. I hope that readers will inform me of their criticisms.

The references in any book of this type provide the reader with an entrance to the literature on the various topics. In order that the reader becomes familiar with the individuals involved, I have used the complete first name of authors when that is provided in the original article. The title and inclusive page numbers of the article are included for convenience. The majority of the text for this book was completed in early 1976; important new references up to the first six months of 1977 were added where possible.

There are many people who have helped me in my study of structure-activity relationships. Corwin Hansch enthusiastically helped me to learn the methods which he introduced to medicinal chemistry; Julius Taylor in 1960 taught me the importance of physical properties to drug activity and in 1966 badgered me into trying multiple regression analysis on some pargyline data we had; Ronald Wiegand had such faith in the potential usefulness of the extrathermodynamic approach that he supported its application at Abbott when others were highly critical. I am also grateful to those who read part or all of the manuscript and who offered helpful criticism: Paul Craig, William Dunn III, Rainer Franke, Gary Grunewald, James Hackbarth, Corwin Hansch, Al Leo, and Julius Taylor. Willard Bass encouraged and helped me compose the manuscript on the Abbott DEC-10 computer. Arlene Davidson was a great help in this task and also in spotting typographical errors. My most staunch supporter is my husband; it is he who convinced me to write this book and supported the effort with labor of his own at home.

Yvonne Connolly Martin

CONTENTS

Introduction to the Series	iii
Preface	v
Chapter 1	
INTRODUCTION AND OVERVIEW OF DRUG DESIGN PROBLEMS	1
I. Scientific Foundations of Modern Drug Design Methods	1
II. Topics Covered	3
III. Steps in Drug Design	6
References	9
Chapter 2	
THE EXTRATHERMODYNAMIC AND <u>de novo</u> APPROACHES: OVERVIEW	11
I. Classic Concepts of Relationships between Physical Properties and Biological Potency	12
II. Quantitative Linear Free-Energy Relationships for Chemical Reactions of Organic Compounds	15
III. Description of the Extrathermodynamic Method	19
IV. <u>de novo</u> or Free-Wilson Approach	25
V. Strengths and Weaknesses of the Extrathermodynamic Method	26
References	29
Chapter 3	
NONCOVALENT INTERACTIONS OF IMPORTANCE TO BIOLOGICAL SYSTEMS	33
I. Review of Pertinent Thermodynamic Relationships	35
II. The Importance of the Solvent	36
III. Electrostatic Interactions	41
IV. Hydrogen Bonds	44
V. Dispersion Bonds	46
VI. Hydrophobic Bonds	49
VII. Charge-Transfer Bonds	55
VIII. Steric Repulsion	56
IX. Summary of Physical Properties Correlated with Different Types of Noncovalent Interactions	56
References	58

Chapter 4	
CALCULATION OF THE PHYSICAL PROPERTIES OF COMPOUNDS	61
I. Partition Coefficients	62
II. Parameters for Dispersion Bonding	81
III. Constants for the Steric Effects of Substituents	85
IV. Constants for the Electronic Effects of Substituents	87
V. Indicator (Dummy) Variables	109
References	110
Chapter 5	
BIOLOGICAL DATA	115
I. Consequences of Drug-Receptor Interaction	116
II. Selection of Data for Analysis: Characteristics of an Ideal Biological Test	118
III. Calculation of Relative Potency	126
IV. Choice of Classification Boundaries for Discriminant Analysis and Pattern Recognition	136
References	137
Chapter 6	
THE FORM OF THE EQUATIONS WHICH RELATE POTENCY AND PHYSICAL PROPERTIES	139
I. Empirical versus Theoretical Equations	140
II. A Simple Equilibrium Model for an Ionizable Substance	142
III. General Equations for Equilibrium Models	147
IV. Model-based Equations for the Observed Rates of Biological Processes	158
V. Whole Animal Tests for Which no Model can be Postulated	160
VI. Empirical Equations	161
VII. Perspective	163
References	163
Chapter 7	
PRINCIPLES OF REGRESSION ANALYSIS	167
I. Fundamental Concepts of Statistics	168
II. Calculation and Statistical Evaluation of a Simple Least Squares Line	176
III. Multiple Regression: Description and Statistical Evaluation	184
References	187
Chapter 8	
STRATEGY IN THE STATISTICAL EVALUATION OF A REAL DATA SET	189
I. Desirable Characteristics of a Multiple Regression Computer Program	191

II. Discussion of Characteristics of Two Multiple Regression Programs: HANSCH and SAS	193
III. How to Find the Important Linear Multiple Regression Equations for a Set of Data	194
IV. Nonlinear Regression Analysis	198
V. Special Techniques and Pitfalls in the Examination of a Data Set	204
References	213

Chapter 9

THE EXTRATHERMODYNAMIC APPROACH: A DETAILED EXAMPLE OF CALCULATIONS	215
I. Calculation of the Physical Properties	215
II. Statistical Analysis	221
III. Interpretations	229
References	232

Chapter 10

OTHER MATHEMATICAL METHODS OF USE IN QUANTITATIVE STRUCTURE-ACTIVITY STUDIES	233
I. Measures of Distance in Multidimensional Space	234
II. Eigenvalue, Principal Component, and Factor Analysis	236
III. Discriminant Analysis	242
IV. Cluster Analysis	246
V. Pattern Recognition	247
VI. Simplex Methods	248
References	251

Chapter 11

SYNTHESIS TO FOLLOW UP A LEAD: HOW TO START AND WHEN TO STOP	253
I. The Original Series	254
II. Synthesis Based on a Tentative Equation	267
III. Termination of Synthesis	269
References	269

Chapter 12

CASE STUDIES IN THE APPLICATION OF THE EXTRATHERMODYNAMIC APPROACH	271
I. Inhibition of Dopamine- <u>beta</u> -oxidase by 5-Substituted Picolinic Analogs	272
II. The Rate of Hydrolysis of Dopamine Amino Acid Amides	278
III. Analgetic Potency of Carbolines	285
IV. Erythromycin Analogs	290
References	325

Chapter 13	
FREE-WILSON OR <u>de novo</u> METHOD	329
I. Description of the Calculations	329
II. Design of the Series	332
III. Evaluation of the Results of a <u>de novo</u> Calculation	334
IV. Example: Erythromycin Esters	335
V. Overall Evaluation of the Method	343
References	346
 Chapter 14	
OVERVIEW OF OTHER METHODS USED IN THE DESIGN OF BIOLOGICALLY ACTIVE MOLECULES	349
I. Modification of Known Substances	350
II. Enzyme Inhibitors	350
III. Conformational Methods	361
References	366
 Appendix I	
IMPORTANT EQUATIONS AND DEFINITIONS	371
 Appendix II	
SUBSTITUENT CONSTANTS	377
 Appendix III	
WISWESSER LINE NOTATION	395
 Appendix IV	
SOURCES OF COMPUTER PROGRAMS AND DATA BASES	401
 AUTHOR INDEX	403
 SUBJECT INDEX	411

Chapter 1

INTRODUCTION AND OVERVIEW OF DRUG DESIGN PROBLEMS

I. SCIENTIFIC FOUNDATIONS OF MODERN DRUG DESIGN METHODS	1
II. TOPICS COVERED	3
III. STEPS IN DRUG DESIGN	6
References	9

I. SCIENTIFIC FOUNDATIONS OF MODERN DRUG DESIGN METHODS

The techniques involved in the rational design of biologically active molecules have been broadened by the availability of many new instruments and techniques. For example, computers can now do the calculations which could only be contemplated as recently as twenty years ago. This book will discuss primarily those techniques of drug design on which the impact of computers has been largest. That is, the Hansch, extrathermodynamic, or linear free energy and the Free-Wilson or *de novo* methods are treated in most detail because these methods have been widely used. For example, reviews of these methods have been written by Hansch (1973), Neely (1973), and Purcell et al., (1973) among others. There is also a book of 15 chapters entitled "Biological Correlations--The Hansch Approach", edited by R. F. Gould (1972). Although Purcell et al., treat the subject briefly, no one has described in detail how one

actually applies these methods. In addition, some space is devoted to pattern recognition which is in the beginning stages as far as drug design is concerned. Conformational methods are covered briefly. The design of enzyme inhibitors -- active-site-directed irreversible inhibitors, transition state analogs, and k_{cat} inhibitors -- is also discussed. These methods compliment the extrathermodynamic approach and offer much promise for the future.

The practice of reliable drug design draws upon a knowledge of aspects of several disciplines; physical and organic chemistry, statistics, pharmacology, and biochemistry are perhaps the most important. The success of any one individual or team of individuals in this task depends upon how well they can integrate these diverse disciplines. For example, predictions based on careful experimental quantification of the physical properties of a series of drugs will be incorrect if the proper statistical evaluation of the relationship between these properties and potency have not been used. Alternatively, brilliant calculations on faulty data are likewise faulty. Thus serious workers will always be alert for gaps in their knowledge of the subjects on which drug design are based.

In view of the fact that quantitative drug design derives from many areas, this book will provide references to the key literature on these subjects. Thus physical organic chemistry works written by physical organic chemists will be cited, for statistics, work by statisticians, etc. In this manner, the new worker will be at once in touch with the knowledge basis of the drug design methods. Subjects which are central to the methods or which have not been reviewed elsewhere are treated in more detail than those for which adequate reviews are available.

The chapters in this book are designed to teach the reader how to apply modern drug design methods. It is a how-to-do-it book. Topics are treated in such a manner that they should be understandable by readers with a modern undergraduate training in chemistry. Each subject will be treated in enough detail that a careful student could apply the method, properly, to their own

data. Readers who do not intend to apply these methods, but merely want to be able to evaluate other workers' applications, should have developed a critical facility by the time they finish the study of this book.

Standards for quality work are an integral part of the discussions. For example, for which types of analogs can one reliably calculate a lipid-water partition coefficient? For which analogs are such calculations unreliable? What precautions apply to the accurate determination of partition coefficients or pK_a 's? How does one decide if a computer program is suitable for these purposes? What are the important statistical criteria for a successful regression analysis? Can data points be dropped from an analysis? How well should a structure activity equation fit a set of data, and conversely when is an equation an over fit of data? These same standards can be applied to the evaluation of the work of others.

The words "drug design" are used instead of "structure-activity relationships" to emphasize that the end result of the application of the various methods should not be a rationalization of existing data, but rather a prediction of new data. Conclusions which have predictive power are much more useful than those which merely summarize a set of observations.

"Drug" is used throughout in its broadest sense, that is as any biologically active compound which is not normally found in the organism, or which is administered in amounts which produce abnormally high levels in the body. By this definition, a pesticide is a drug; so is a natural hormone administered in large doses.

II. TOPICS COVERED

The book is arranged as follows: Chapter 1 includes an introduction to the discipline of drug design. What are its scientific origins? How do scientists go about trying to discover a drug? Chapter 2 is an overview of the extrathermodynamic and Free-Wilson methods. These chapters are intended to acquaint

readers unfamiliar with medicinal chemistry with the type of topics which will be discussed in later chapters.

Chapter 3 is a review of the noncovalent forces between molecules. It is these forces which determine the specificity and often the strength of the interaction between a drug and its receptor. The calculation of physical properties of molecules which can be related to these forces is discussed in Chapter 4. These topics are important to the whole issue of reliability.

Desirable characteristics of the biological data and methods of transformation of raw data into potency values are covered in Chapter 5. These topics are important to understand--more than one expert at linear free-energy calculations has made an error in the calculation of the relative potency of analogs.

The mathematical form of the relationship between potency and physical properties is the subject of Chapter 6. Particular emphasis is given to the complex relationships between potency, partition coefficient, and fraction ionized. Both the empirical Hansch equations and model-based equations are included. These subjects have only recently attracted attention (Martin and Hackbarth, 1976); they have not been reviewed elsewhere.

Chapter 7 contains an introduction to the statistics of multiple linear regression. The characteristics of an acceptable regression analysis computer program, the steps one should follow in the first examination of a set of data, pitfalls and special techniques in regression analysis, and distinctive characteristics of the application of nonlinear regression are discussed in Chapter 8.

Chapter 9 is a detailed illustration of the introductory material with a series of esters of erythromycin. This step-by-step explanation should be a useful guide for readers who have never attempted a regression analysis. Some of the tables are copies of the actual computer printouts from this data.

Statistical methods other than regression analysis are reviewed in Chapter 10. These have not received wide attention, but they can be used in situations in which regression analysis is inappropriate.

Chapter 11 consists of considerations in the design and termination of synthesis in a series. These in general are nonmathematical applications of the concepts covered earlier. This chapter may be of special use to those readers who do not intend to become experts in the calculations of the extrathermodynamic approach, but who do want to apply some of the concepts to their own work.

Chapter 12 consists of several case studies of the actual application of the extrathermodynamic approach to problems in the drug industry. These are meant to illustrate by example points dealt with previously, and also to give the reader a feeling for the strengths and weakness of the extrathermodynamic method.

The Free-Wilson method is covered in Chapter 13. The data on the erythromycin esters is again used as an example to allow comparison between this and the linear free energy approach.

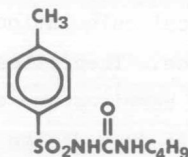
Finally, Chapter 14 provides a brief introduction to design of enzyme inhibitors and the consideration of conformation as a basis for the design of analogs. The design of enzyme inhibitors has been the focus of real advances in recent years. The methods discussed appear to be useful in appropriate cases. The consideration of the three-dimensional shape of the molecule has been one aspect of planning analogs for many years. Recent advances include the calculation of the conformation by potential energy and quantum chemical calculations and measurement of it by nuclear magnetic resonance. Theoretical considerations of the importance of low-energy barriers are even more recent. The conformational approach to drug design may also be evaluated by observations of small-molecule/large-molecule interactions in the crystal state.

Appendix I contains a summary of the important equations used or derived in the text. Appendix II is a list of values of substituent constants for a large number of substituents. Appendix III is a brief introduction to Wisswesser Line Notation. Appendix IV provides information on purchasing computer programs and data banks discussed in the text.

III. STEPS IN DRUG DESIGN

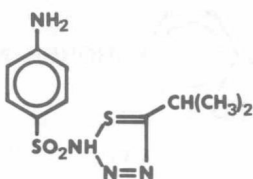
There are several separate steps in the design of a synthetic compound to be used as a drug. These are presented from a slightly different point of view by Biel and Martin (1971). Examples are found in the books edited by Ariens (1971-1976) and by Burger (1970). Each particular step uses a different type and quality of information. The various "rational" approaches to drug design are in general more applicable to one or another of these steps. The stages are briefly described for the newcomer to the field.

After the definition of the type of biological properties the sought-after compound should have, the first step in the discovery of a drug of novel structure is to find a "lead" compound. This is usually a molecule with demonstrable but weak activity of the desired type. An important source of lead compounds is "random" screening. This means simply that many compounds of varied structural types are tested at a high dose in an appropriate biological test. A second rich source for leads is in the observed side effects in animals or man of known drugs. The oral antidiabetic agents, of which tolbutamide (I) is an example, were



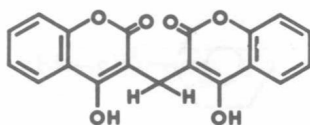
I

discovered by careful molecular modification of a sulfa drug (II) which showed lowering of blood glucose as a side effect (Grunwald,



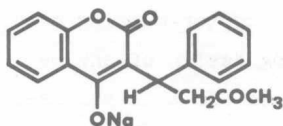
II

1970). A third source of leads for new drugs is found in natural products. For example, the structure of the natural anticoagulant, bishydroxycoumarin (III), led to the synthesis of a distant analog,



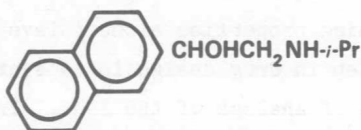
III

the widely used anticoagulant drug warfarin (IV) (Cutting, 1969).



IV

A fourth source of leads is the structure of natural body substances. Pronethalol (V) was designed to antagonize the effects



V

of norepinephrine (VI) (Barrett, 1972). Also, enzyme inhibitors are often designed from the structure of substrates. For example, the monoamine oxidase inhibitor pargyline (VII) is an analog of the substrate, benzylamine (VIII) (Martin et al., 1975).