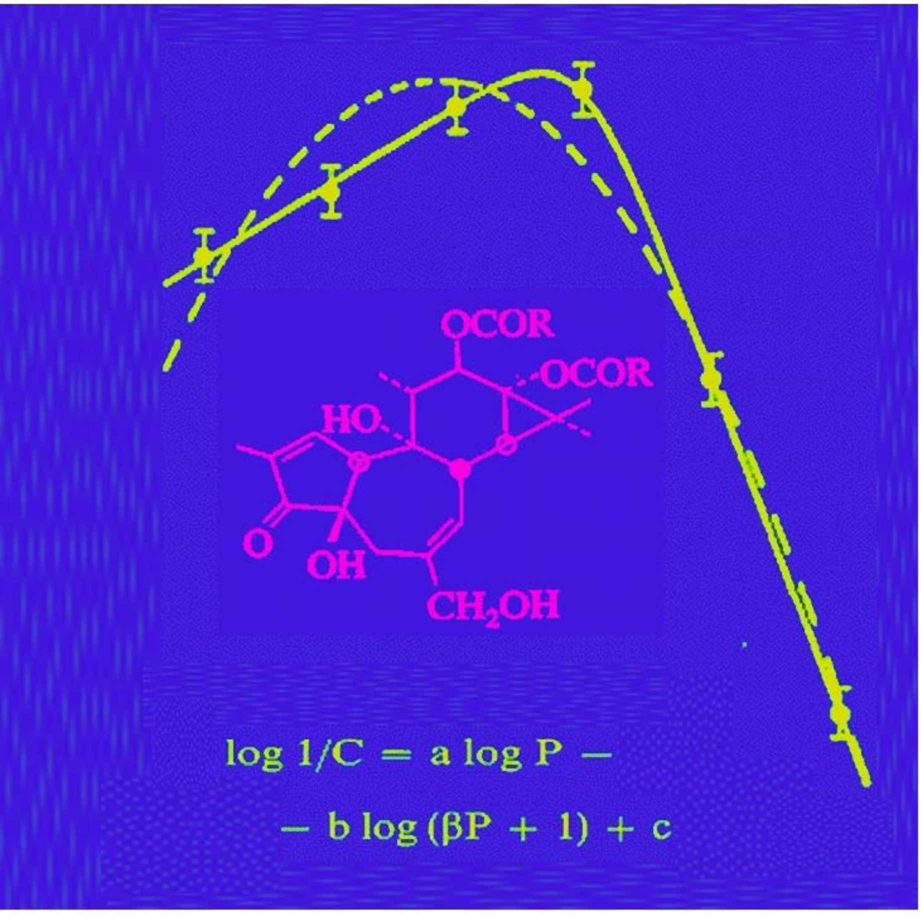
Edited by
R. Mannhold P. Krogsgaard-Larsen H. Timmerman

QSAR: Hansch Analysis and Related Approaches

by Hugo Kubinyi


OCOR
OCOR
HO
O
OH
CH₂OH
log 1/C = a log P −
− b log (βP + 1) + c

VCH

Hugo Kubinyi

QSAR: Hansch Analysis and Related Approaches

VCH

Methods and Principles in Medicinal Chemistry

Edited by
R. Mannhold
P. Krogsgaard-Larsen
H. Timmerman

QSAR: Hansch Analysis and Related Approaches

by Hugo Kubinyi

VCH

Weinheim · New York
Basel · Cambridge · Tokyo

Author:
Prof. Dr. Hugo Kubinyi
ZHV/W, A30
BASF AG
D-67056 Ludwigshafen

Editors:
Prof. Raimund Mannhold
Biomedical Research Center
Molecular Drug Research Group
Heinrich-Heine-Universität
Universitätsstraße 1
D-40225 Düsseldorf

Prof. Povl Krogsgaard Larsen
Dept. of Organic Chemistry
Royal Danish School of Pharmacy
DK-2100 Copenhagen
Danmark

Prof. Hendrik Timmerman
Faculty of Chemistry
Dept. of Pharmacochemistry
Free University of Amsterdam
De Boelelaan 1083
NL-1081 HV Amsterdam
The Netherlands

This book was carefully produced. Nevertheless, author, editors and publisher do not warrant the information contained therein to be free of errors. Readers are advised to keep in mind that statements, data, illustrations, procedural details or other items may inadvertently be inaccurate.

Published jointly by
VCH Verlagsgesellschaft mbH, Weinheim (Federal Republic of Germany)
VCH Publishers, New York, NY (USA)

Editorial Director: Dr. Hans F. Ebel
Production Manager: Dipl. Wirt.-Ing. (FH) Hans-Jochen Schmitt

Library of Congress Card No. applied for

British Library Cataloguing-in-Publication Data:
A catalogue record for this book is available from the British Library

Die Deutsche Bibliothek – CIP Einheitsaufnahme
Kubinyi, Hugo:
QSAR: Hansch analysis and related approaches / by Hugo
Kubinyi. – Weinheim ; New York ; Basel ; Cambridge ; Tokyo :
VCH, 1993
(Methods and principles in medicinal chemistry ; Vol. 1)
ISBN 3-527-30035-X (Weinheim ...)
ISBN 1-56081-768-2 (New York)
NE: GT

©VCH Verlagsgesellschaft mbH, D-69451 Weinheim (Federal Republic of Germany), 1993

Printed on acid-free and chlorine-free paper.

All rights reserved (including those of translation into other languages). No part of this book may be reproduced in any form – by photoprinting, microfilm, or any other means – nor transmitted or translated into a machine language without written permission from the publishers. Registered names, trademarks, etc. used in this book, even when not specifically marked as such, are not to be considered unprotected by law.

Printing and Bookbinding: Druckhaus "Thomas Müntzer" GmbH, D-99947 Bad Langensalza
Printed in the Federal Republic of Germany

© VCH Verlagsgesellschaft mbH, D-69451 Weinheim (Federal Republic of Germany), 1993

Distribution:

VCH, P. O. Box 101161, D-69451 Weinheim (Federal Republic of Germany)

Switzerland: VCH, P. O. Box, CH-4020 Basel (Switzerland)

United Kingdom and Ireland: VCH (UK) Ltd., 8 Wellington Court,
Cambridge CB1 1HZ (England)

USA and Canada: VCH, 220 East 23rd Street, New York, NY 10010–4606 (USA)

Japan: VCH, Eikow Building, 10-9 Hongo 1-chome, Bunkyo-ku, Tokyo 113 (Japan)

ISBN 3-527-30035-X (VCH, Weinheim)
ISBN 1-56081-768-2 (VCH, New York)

Dedicated to Corwin Hansch

Preface

The present monograph is the first volume in a new series of handbooks entitled "Methods and Principles in Medicinal Chemistry". The prime focus of this series is an educational introduction into the current knowledge of methodological aspects and basic principles in the rapidly developing field of Medicinal Chemistry.

Potentials and limitations of techniques will be critically and comparatively discussed and comprehensively exemplified. It is intended to provide the reader with the appropriate information for applying the adequate techniques to a given problem and to avoid misleading interpretations due to the improper use of methodology. Main topics under the scope of this new publication are:

- The determination of chemical properties of biologically relevant molecules.
- Innovative approaches in the characterization of biological activity.
- Methodological aspects in deriving SAR and QSAR analyses.
- Current developments in the physiological and biochemical understanding of diseases.
- Future perspectives in the development of Medicinal Chemistry.

The first volume in the series deals with Hansch analysis and related approaches. Publication of the Hansch model in the early sixties represents the starting point of modern QSAR methodology and correspondingly the present monograph focuses on these aspects of Medicinal Chemistry. But not the historical reasons have primarily led the editors to start the series with this topic. The "classical" QSAR methods also nowadays play an important role in Medicinal Chemistry. Despite the advances in protein crystallography, molecular modeling, and structure-derived molecular design, Hansch analysis and related approaches are continuously useful tools to quantitatively derive and prove hypotheses on structure-activity relationships. In addition, the quantitative treatise of kinetic aspects of drug action remains an exclusive domain of these methods.

According to the aim of this new series Hugo Kubinyi gives a practice-oriented introduction into Hansch analysis and related approaches which familiarizes the reader with the proper application of these methodologies. The comprehensive list of references gives an excellent access to current literature and comfortably introduces the reader to fields of his special interest.

Düsseldorf
Kopenhagen
Amsterdam
Summer 1993

Raimund Mannhold
Povl Krogsgaard-Larsen
Hendrik Timmerman

A Personal Foreword

The first lipophilicity-activity relationship was published by Charles Richet in 1893, exactly 100 years ago. From his quantitative investigations of the toxicities of ethanol, diethyl ether, urethane, paraldehyde, amyl alcohol, acetophenone, and essence of absinthe (!) he concluded *"plus ils sont solubles, moins ils sont toxiques"* (the more they are soluble, the less toxic they are). One year later Emil Fischer derived the lock and key model of ligand-enzyme interactions from his results on the stereospecificity of the enzymatic cleavage of anomeric glycosides.

In the following decades the receptor concept evolved from investigations of Paul Ehrlich; a continuous development of medicinal chemistry began, leading to better and better drugs against many diseases. However, despite important contributions by Meyer, Overton, Traube, Moore, Warburg, Fühner, and Ferguson to the dependence of nonspecific biological activities of drugs on their lipophilicity (most often expressed by oil/water partitioning), the field of quantitative relationships between chemical structures and their biological activities lay dormant for about 70 years.

The discipline of quantitative structure-activity relationships (QSAR), as we define it nowadays, was initiated by the pioneering work of Corwin Hansch on growth-regulating phenoxyacetic acids. In 1962–1964 he laid the foundations of QSAR by three important contributions: the combination of several physicochemical parameters in one regression equation, the definition of the lipophilicity parameter π, and the formulation of the parabolic model for nonlinear lipophilicity-activity relationships.

This was the time when I started my Ph. D. thesis on irritant and tumor-promoting phorbol esters, their isolation, partial synthesis, and structure-activity relationships at the Max Planck Institute of Biochemistry in Munich. Indeed, one diagram in this book (Figure 43, chapter 7.4) refers to these compounds. Although I recognized a nonlinear relationship between the biological activities and the chain length of the ester groups (I even measured partition coefficients and found a nice linear dependence on the lipophilicity of the compounds), the small step from drawing a diagram to formulating a mathematical model, *i.e.* deriving a parabolic equation, was too large for me at that time. Shortly afterwards, then doing research in pharmaceutical industry, I became aware of the work of Corwin Hansch, Toshio Fujita, William Purcell, and others on quantitative structure-activity relationships. Like some of my colleagues in pharmaceutical industry I noticed this new approach but did not consider to apply it to practical drug design. For years I lived with the prejudice that QSAR is a tool to describe only more or less nonspecific biological effects, like antibacterial, antifungal, hemolytic, narcotic, and toxic activities.

My conversion from Saulus to Paulus happened after a discussion with Rudolf Gompper in Munich in 1974. In his seminar on theoretical chemistry he also mentioned the pioneering contributions of Corwin Hansch to medicinal chemistry. I presented my scepticism but, at the same time, felt ashamed of my ongoing ignorance and decided to read some more papers. Three fortunate circumstances worked hand in hand: William Purcell's book *"Strategy of Drug Design. A Molecular Guide to Biological Activity"* had just arrived in our library and I read it in one day, fascinated by its content and style. An experienced technician helped me with his statistics programs (some months later I had discussions with a professional statistician who insisted that everything we QSAR people do is forbidden for this or that reason. I never would have started QSAR work if I had spoken to him first; now it was too late, I already was infected). A colleague provided a data set on antihistaminic compounds for which, another day later, a beautiful $\pi-\sigma$ relationship could be derived. A compound of this series came to preclinical and clinical development, but unfortunately it turned out to be only a drug for guinea pigs; it had almost no activity in humans.

After this big start I tried to understand the underlying theories and recalculated many published equations. My knowledge and experience increased, but I found a lot of numerical and also logical errors in the early QSAR literature. The consequence was to refine old models, to develop new ones, and to write scientific papers. My attempts to publish them were a difficult task. The comments of the reviewers ranged from *"much ado about nothing"* to *"wrong"* and it took a lot of patience, insistence, and several rebuttal letters to place them in the Journal of Medicinal Chemistry.

The publications of Corwin Hansch helped me to proceed. A two-month sabbatical in his group at the Pomona College followed in 1978. This visit led to a deeper understanding of quantitative structure-activity relationships and their physicochemical and biological foundations on my side. On the other hand, it stimulated Corwin Hansch to apply the bilinear model to the QSAR of enzyme inhibitors; the most interesting applications of this new model resulted from his work, from 1980 onwards.

Nowadays drug development is much too expensive to be guided by trial and error. QSAR, molecular modeling, and protein crystallography are important and valuable tools in computer-assisted drug design. The aim of this book is to give an introduction to QSAR methodology for beginners and practitioners and to present selected examples of typical applications. Comments are derived from about 20 years of practical applications, from thousands of calculated and recalculated QSAR equations. It still is my attitude to check other people's equations, especially when reviewing manuscripts. Some warnings are given and the limitations of QSAR methods will be discussed. As the commonly used methods are Hansch analysis, the Free Wilson model, and, recently coming up, comparative molecular field analysis (CoMFA), the focus is on these approaches.

Corwin Hansch initiated QSAR and he contributed the most to its development. Correspondingly, this book is dedicated to him on the occasion of his 75th anniversary in October 1993. He taught us how to apply QSAR in a proper manner to gain more insight into structure-activity relationships and biological mechanisms. The one and only way to thank him is to feel responsible to use and to develop the QSAR discipline in his sense. Thus, the book shall also be understood as a stimulus to further research on the real relationships between chemical structures and biological activities.

Heidelberg and Ludwigshafen Hugo Kubinyi

Content

1. Introduction 1

1.1. History and Development of QSAR 4
1.2. Drug-Receptor Interactions 7

2. Biological Data. The Additivity of Group Contributions 15

3. Parameters 21

3.1. Lipophilicity Parameters 27
3.2. The Measurement of Partition Coefficients and Related Lipophilicity Parameters 32
3.3. Lipophilicity Contributions and the Calculation of Partition Coefficients 34
3.4. Polarizability Parameters 40
3.5. Electronic Parameters 42
3.6. Steric Parameters 47
3.7. Other Parameters 49
3.8. Indicator Variables 54

4. Quantitative Models 57

4.1. The Extrathermodynamic Approach (Hansch Analysis) 57
4.2. The Additivity Model (Free Wilson Analysis) 62
4.3. The Relationships between Hansch and Free Wilson Analysis (The Mixed Approach) 65
4.4. Nonlinear Relationships 68
4.5. Dissociation and Ionization of Acids and Bases 77
4.6. Other QSAR Approaches 85

5. Statistical Methods 91

5.1. Regression Analysis 91
5.2. The Significance and Validity of QSAR Regression Equations 95
5.3. Partial Least Squares (PLS) Analysis and Other Multivariate Statistical Methods 100

6. Design of Test Series in QSAR 109

7. Applications of Hansch Analysis 115

- 7.1. Enzyme Inhibition 116
- 7.2. Other *in vitro* Data 121
- 7.3. Pharmacokinetic Data 126
- 7.4. Other Biological Data 129
- 7.5. Activity-Activity Relationships 133

8. Applications of Free Wilson Analysis and Related Models 139

9. 3D QSAR Approaches 149

- 9.1. Stereochemistry and Drug Action 149
- 9.2. Active Site Interaction Models 151
- 9.3. Comparative Molecular Field Analysis (CoMFA) 159
- 9.4. Molecular Similarity QSAR Analyses 172

10. Summary and Conclusions 179

References 183

Index 211

QSAR: Hansch Analysis and Related Approaches
by Hugo Kubinyi
© VCH Verlagsgesellschaft mbH, 1993

1. Introduction

The interactions of drugs with their biological counterparts are determined by intermolecular forces, *i.e.* by hydrophobic, polar, electrostatic, and steric interactions. Quantitative structure-activity relationships (QSAR) derive models which describe the structural dependence of biological activities either by physicochemical parameters (Hansch analysis), by indicator variables encoding different structural features (Free Wilson analysis), or by three-dimensional molecular property profiles of the compounds (comparative molecular field analysis, CoMFA).

Drugs, which exert their biological effects by interaction with a specific target, be it an enzyme, a receptor, an ion channel, a nucleic acid, or any other biological macromolecule, must have a three-dimensional structure, which in the arrangement of its functional groups and in its surface properties is more or less complementary to a binding site. As a first approximation the following can be concluded: the better the steric fit and the complementarity of the surface properties of a drug to its binding site are, the higher its affinity will be and the higher may be its biological activity.

A complication arises from the functionalities of the biological macromolecules typically involved in ligand-protein interactions: certain structural features of the ligand determine whether a compound is

a substrate (having a functional group which is hydrolyzed, acylated, oxidized, *etc.*, by an enzyme),
an inhibitor (exhibiting affinity to the binding site of an enzyme, but containing no such group),
a competitive receptor antagonist (having affinity to an agonist binding site, but mediating no receptor response),
an allosteric receptor antagonist (binding to a different site, see below),
a functional receptor antagonist (having no affinity to the receptor molecule, but inhibiting the receptor response *via* a different mechanism of action),
a receptor agonist (displaying intrinsic activity in addition to affinity, *i.e.* containing certain structural features which cause the receptor to respond in a certain manner), or
an allosteric effector molecule (binding at a different site of a protein and changing its 3D structure in such a way that a certain property of the protein, *e.g.* conformational flexibility or affinity to a substrate, an agonist, a cofactor, or any other small or large ligand is significantly changed).

The fit of the three-dimensional structure and the complementarity of the surface properties of a drug to its binding site are conditions for its biological activity. Another one, at least equally important, is that the drug has to reach this binding site. Even in simple *in vitro* systems, *e.g.* in enzyme inhibition, the surrounding water molecules compete to form hydrogen bonds to the binding site and to the functional groups of the ligand. The balance of hydrogen bonds in solution and in the bound state increases or reduces affinity. In more complex biological systems, like in cells, isolated organs, or whole animals, a certain range of lipophilicity enables the drug to walk its random way from the site of application to the site of action, *i.e.* to cross several lipophilic and hydrophilic barriers, lipid membranes as well as aqueous phases. In the case of nonspecific biological activities caused by membrane perturbation, only the distribution of the drug and its local concentration in a certain membrane compartment is responsible for its biological activity.

While the affinity of a ligand to its binding site results from the sum of all hydrophobic, polar, electrostatic, and steric interactions, the influence of lipophilicity and ionization on the distribution of a drug in a biological system is much more complex.

As long as the biological system is kept constant, the interaction of two different drugs with the binding site as well as their distribution in the system only depend on the chemical structures of the compounds. If these structures are closely related, *e.g.* having a chlorine atom instead of a hydrogen atom in a certain position, the differences in their physicochemical properties and thus the differences in the interaction forces can easily be described in a quantitative manner; the corresponding difference in biological activities should directly be related to the differences in these properties. This is indeed the case and all quantitative models of structure-activity relationships are based on the assumption of a more or less strict additivity of group contributions to biological activity values. In many cases nonlinear models are needed to describe, in addition to binding and intrinsic activity, the dependence of drug transport and distribution on lipophilicity and ionization.

While the classical models of quantitative structure-activity analyses do not consider the three-dimensional arrangement of functional groups, some recent approaches deal with this problem and describe biological activities in terms of favorable and unfavorable interaction spheres, derived from the hydrophobic, electrostatic, and steric interaction fields of the ligands.

The methods of quantitative structure-activity relationships which have developed during the past 30 years nowadays are widely applied to describe the relationships between chemical structures of molecules and their biological activities. Many attempts have been made to understand structure-activity relationships in physicochemical terms (or in terms of structural features, using indicator variables for individual substituents and groups) and to design new drugs on a more rational basis. However, the quantitative description of structure-activity relationships is no easy task and will remain difficult at least in the near future.

Most often QSAR analyses are retrospective studies, whether they follow a rational design of investigated structures or not. Only after performing syntheses and biological testing, a quantitative relationship is derived. Often the optimization of a lead compound is step by step accompanied by QSAR analyses.

The dispute, whether QSAR really aids to find the optimum within a series of biologically active molecules cannot generally be decided. Obviously, the QSAR results depend on the validity of the underlying hypotheses, on the complexity of the test model, and on the precision of the biological data. For new compounds within a congeneric series the quality of prediction of the biological activity values is related to the spanned parameter space and to the distance of the physicochemical properties of the new analogs to those of the other compounds. To mention only a few other effects, it also depends on the conformational flexibility of the ligand and its binding site, on multiple binding modes, and on differences in transport and metabolism.

Although sometimes taken as a criterion, prediction is not the primary goal of QSAR analyses. If it results from interpolation, it is often trivial; if extrapolation goes too far outside the included parameter space, it usually fails. QSAR helps to understand structure-activity relationships in a quantitative manner and to find the borders of certain properties, *e.g.* the optimum of lipophilicity within a series of analogs or the maximum size of a certain group in a stepwise procedure. The strategy and philosophy of QSAR enables medicinal chemists to look at their structures in terms of physicochemical properties instead of only considering certain pharmacophoric groups in it.

Many published structure-activity relationships do not meet generally accepted standards in scientific research and statistics. Most often hypotheses are not justified by the experimental data and, even worse, in some cases the results only reflect the patience of the authors to investigate many different variables to describe the biological activity values of a small number of compounds, until a certain combination of these variables gives a delusively good result.

Rational drug design did not start with QSAR. Chemists and biologists always followed rational guidelines, depending on the state of knowledge at their time. However, in the past 30 years several important qualitative concepts evolved from QSAR studies:

The role of different physicochemical properties being responsible for the drug-receptor interaction.

The understanding of the influence of lipophilicity and ionization on drug transport and distribution within a biological system.

The concept of optimum lipophilicity of a drug for passive transport, *e.g.* gastrointestinal absorption or transfer through the blood-brain barrier.

Nowadays many medicinal chemists are familiar with these relationships and do not any longer realize that much of our knowledge came from such analyses.

With the progress in protein crystallography and, derived from the resulting 3D structures, in molecular modeling, the interactions between a ligand and its binding site can be "seen" in three dimensions. Nevertheless, QSAR methods are still used to prove and to quantify the underlying hypotheses regarding the dependence of biological activities on physicochemical interactions. Protein crystallography-derived drug design only concerns ligand design. It does neither contribute to the design of optimum transport and distribution properties nor to the selection of metabolically stable analogs. These areas still remain in the field of classical QSAR studies.

1.1. History and Development of QSAR

In 1868 Crum-Brown and Fraser [1] published an equation which is considered to be the first general formulation of a quantitative structure-activity relationship. In their investigations of different alkaloids they recognized that alkylation of the basic nitrogen atoms produced significantly different biological effects of the resulting permanently charged quaternary ammonium compounds, as compared to the basic amines. Therefore they assumed that the "physiological activity" Φ must be a function of the chemical structure C (eq. 1).

$$\Phi = f(C) \tag{1}$$

Richet [2] discovered that the toxicity of organic compounds inversely follows their water solubility. Such a relationship corresponds to eq. 2, where $\Delta\Phi$ are the differences in biological activity values, caused by corresponding changes in the chemical and especially the physicochemical properties, ΔC.

$$\Delta\Phi = f(\Delta C) \tag{2}$$

Strictly speaking, still today there is no way to apply eq. 1 to biological data. All QSAR equations correspond to eq. 2, because only the differences in biological activities are quantitatively correlated with changes in lipophilicity and/or other physicochemical properties of the compounds under investigation.

At the turn of the last century Meyer [3] and Overton [4] independently of each other observed linear relationships between lipophilicity, expressed as oil-water partition coefficients, and narcotic activities [5]. Fühner [6] realized that within homologous series narcotic activities increase in a geometric progression, *i.e.* $1:3:3^2:3^3$, *etc.*, which gave the first evidence of an additivity of group contributions to biological activity values. This result was confirmed by many other studies, which used different lipophilicity parameters to describe various kinds of nonspecific biological activities. Ferguson gave a thermodynamic interpretation of such non-specific structure-activity relationships which also explained the often observed "cut-off" of biological activity values beyond a certain range of lipophilicity [7].

QSAR methodology rapidly developed from the mid fifties on: Bruice, Kharasch, and Winzler [8] formulated group contributions to biological activity values in a series of thyroid hormone analogs, which may be considered as a first Free Wilson-type analysis. Zahradnik [9–11] tried to apply the concept of the Hammett equation (eq. 3) [12], which at that time was used for three decades to describe the reactivity of organic compounds in a quantitative manner, also to biological data (eq. 4).

$$\log k_{R-X} - \log k_{R-H} = \varrho\sigma \tag{3}$$

$$\log \tau_i - \log \tau_{Et} = \alpha\beta \tag{4}$$

τ_i in this "biological Hammett equation" stands for the activity value of the ith member of a series, τ_{Et} is the biological activity value of the ethyl compound of the same series, β is a substituent constant (corresponding to the electronic σ parameter in the Hammett equation), and α is a constant characterizing the biological system, which corresponds to the Hammett reaction constant ϱ. Unfortunately, eq. 4 only holds true for nonspecific biological activities, most often within homologous series and within a certain lipophilicity range.

In 1962 Hansen [13] derived the first (and for a long time the only one) real Hammett-type relationship between the toxicities of substituted benzoic acids and the electronic σ constants of their substituents (eq. 38, chapter 3.5). In the same year the first QSAR publication of Corwin Hansch on *"The correlation of the biological activity of phenoxyacetic acids with Hammett substituent constants and partition coefficients"* [14] appeared.

1964 may be considered as the year of birth of modern QSAR methodology. Time was ready for more general formulations, how to treat structure-activity relationships in a quantitative manner. Independently, two papers were published, one by Hansch and Fujita on *"ϱ-σ-π Analysis. A method for the correlation of biological activity and chemical structure"* [15], the other by Free and Wilson on *"A mathematical contribution to structure activity studies"* [16]. Both contributions started the development of two new methods of quantitative structure-activity relationships, later called Hansch analysis (linear free energy-related approach, extrathermodynamic approach) and Free Wilson analysis, respectively. The real breakthrough in QSAR resulted from the combination of different physicochemical parameters in a linear additive manner (eq. 5; log 1/C is the logarithm of the inverse molar dose that produces or prevents a certain biological response, log P is the logarithm of the *n*-octanol/water partition coefficient P), as done earlier in theoretical organic chemistry. Further contributions were the definition of a calculated lipophilicity parameter π (eq. 6), to be used instead of measured log P values (like Hammett σ values are used instead of equilibrium constants of organic reactions), and the formulation of a parabolic equation for the quantitative description of nonlinear lipophilicity-activity relationships (eq. 7) [17–19].

$$\log 1/C = a \log P + b\sigma + \ldots + \text{const.} \tag{5}$$

$$\pi_X = \log P_{R-X} - \log P_{R-H} \tag{6}$$

$$\log 1/C = a\,(\log P)^2 + b \log P + c\sigma + \ldots + \text{const.} \tag{7}$$

Considering a significant contribution by Fujita and Ban [20], the Free Wilson model is defined by eq. 8, where a_{ij} is the group contribution of the substituent X_i in the position j and μ is the (theoretical) biological activity value of a reference compound within the series; all group contributions a_{ij} of the different substituents X_i refer to the corresponding substituents (most often being hydrogen) of this reference compound.

$$\log 1/C = \sum a_{ij} + \mu \tag{8}$$

Both models remained unchanged over the past three decades. Some improvements resulted from the combination of Hansch equations with indicator variables [21], which may be considered as a mixed Hansch/Free Wilson model (chapter 4.3) [22], and from the formulation of theoretically derived nonlinear models for transport and distribution of drugs in a biological system, *e.g.* the bilinear model (eq. 9; chapter 4.4) [23].

$$\log 1/C = a \log P - b \log(\beta P + 1) + c \qquad (9)$$

Various attempts have been made to use pattern recognition [24, 25] in QSAR studies and successful applications have been reported. Soft modeling techniques, *e.g.* the partial least squares (PLS) method [26, 27], now offer better opportunities. With the help of this principal component-like method the explanatory power of many, even hundreds or thousands of variables can be used for a limited number of objects, a task being absolutely impossible in regression analysis in which the number of objects must always be larger than the number of variables.

Three-dimensional quantitative structure-activity relationships (3D QSAR) were developed from the first attempts to map a receptor surface by analyzing a QSAR equation for noncovalent interactions of the ligands in the different positions of substitution (*e.g.* [28]). Höltje [29, 30] extended this approach. He postulated certain amino acid side chains as binding partners, calculated interaction energies in standard geometries, and correlated these energies with biological activity values. Several other attempts were made to map hypothetical interaction sites of a receptor, *e.g.* the distance geometry method of Crippen [31, 32]. Goodford's program GRID calculates interaction energies of certain probe atoms with the surface of a protein whose three-dimensional structure is known from crystallographic analysis [33].

If the three-dimensional structure of the protein is unknown, different fields of the ligands can be compared in 3D space. The molecules of a chemically related series are superimposed, following certain alignment hypotheses (the pharmacophore hypotheses), a grid is laid over the molecules, and values of the steric and electrostatic fields (and optionally other fields) are calculated in every grid point for each molecule of the series. An appropriate multivariate statistical method correlates thousands of such energy values in the different grid points (each one representing a column in the X block) with biological activities. In the first version, called DYLOMMS, principal component analysis was used [34]; later, PLS analysis turned out to be more suitable [35]. Comparative molecular field analysis (CoMFA), as it is used nowadays, was formulated in 1988 [36, 37]. The method, which still is under active development, has found many successful applications in a short time [38].

An excellent, recently published monograph which covers the whole field of QSAR, is the book *Quantitative Drug Design*, volume IV of the six-volume set *Comprehensive Medicinal Chemistry* [39]. In addition, numerous other monographs, either directly related to QSAR methodology and applications [40–47], on physicochemical parameters [48–56], or on related topics [57–64], have been published.

As it is impossible to cite all the relevant work and as the selection of original contributions is always more or less ambiguous, the reader is referred to ref. [39], to several monograph series [65–67], to proceedings of QSAR and QSAR-related symposia [68–84], to the journal *Quantitative Structure-Activity Relationships*, especially to the abstracts section of this journal [85], which year by year contains about 400–500 excellently prepared abstracts of QSAR-related publications, to other abstracts services [86–88], as well as to some other journals [89], including QSAR publications as their regular content.

The history of QSAR has been reviewed in books (*e.g.* [40]) and in dedicated articles [5, 90–93]; the development of 3D QSAR methods is commented in refs. [36, 38].

1.2. Drug-Receptor Interactions

"Corpora non agunt nisi fixata" (Ehrlich, 1913) [94] was an early formulation of the fact that drugs must interact with certain biological macromolecules to exert their biological activity.

The concept of the interaction of drugs with certain *"substances y with which they are capable of forming compounds, ... according to their chemical affinity to y"* goes back to the work of Langley in 1873–1878 [95]. The stereospecificity of such interactions was recognized by Fischer in 1894. In his investigations of the enzymatic cleavage of anomeric glycosides by invertin and emulsin (α-glucosidase and β-glucosidase, respectively), he formulated *„um ein Bild zu gebrauchen, will ich sagen, dass Enzym und Glukosid wie Schloss und Schlüssel zu einander passen müssen, um eine chemische Wirkung aufeinander ausüben zu können“* (to illustrate, I would like to say that enzyme and glucoside must fit together like lock and key, in order to exert a chemical effect on each other) [96]. The term receptor was first used by Ehrlich in his studies on dyestuffs and their interactions with biological tissues. In the following "receptor" sometimes is used as a synonym for any biological target, *e.g.* any specific binding site of a macromolecule; strictly speaking, this broad meaning is not correct from our today's definition of receptors as being soluble, membrane-anchored or membrane-embedded proteins that are able to produce a certain biological response *via* a series of mostly unknown events (for reviews see refs. [59, 97, 98]).

It should be mentioned that the work of Ehrlich also contains a prominent (and most probably the very first) example of a fortuitous success based on "rational" drug design, which later turned out to be based on a wrong hypothesis. *Prontosil rubrum*, *p*-[(2,4-diaminophenyl)-azo]benzenesulfonamide, was designed to stain and kill infectious microorganisms. However, the metabolite sulfanilamide is the active agent, not the dyestuff itself (cited from [97]). One of the most famous examples of serendipity (a term coined by Horace Walpole in 1754 from Serendip, a former name of Ceylon, in an old Persian fairy tale called *"The Three Princes of Serendip,"* in which the princes are described as making happy or interesting discoveries unexpectedly or by accident) was Fleming's finding of the antibacterial activities of certain fungi. The fortunate circumstance that he did not clean his dishes immediately after an unsuccessful experiment resulted in a fungal infection of a bacterial cell culture. This observation directly led to the discovery of penicillin. Less well-known is that the same Fleming, having a cold one day, just for fun "tested" his nasal mucus for antibacterial activity. This unplanned experiment led to the discovery of the enzyme lysozyme, which is also found in egg white, milk, blood serum, tears, saliva, some other secretions and tissues of animals, and in plant latices (cited from [99]).

Many drugs have been discovered by fortune; serendipity always played an important role in drug research [100] and, despite all our efforts in rational design, this will continue in the future. The consequences of short-term planning on the probability of success and the effect of a too bureaucratic management in drug research have been critically commented [100]. On the other hand, design and development of a new drug need the combined effort of a large team of specialists who can only work together in some form of organization; in addition, drug development is a costly and time-consuming multistage process which must be planned and controlled in a proper manner.

During the past decades the originally static lock and key model of ligand-receptor interaction was modified to a more realistic picture, with flexible drug molecules and dynamic receptors [101, 102]. Whenever a ligand approaches its binding site, both partners may change their shape (induced fit, flexible fit). The three-dimensional structures of only a few membrane-bound proteins and receptor-type protein complexes have been resolved at atomic resolution; amongst them are the photosynthetic reaction center [103], the light-driven proton pump bacteriorhodopsin [104], and the bacterial membrane-channel porin [105]. No three-dimensional structures of mammalian receptors are available at atomic resolution. Most of our knowledge regarding the geometry of ligand-binding site interactions resulted from 3D structures of soluble proteins, especially of enzymes and their inhibitor complexes [106 – 110]. Some common objectives against protein 3D structures from crystallographic analyses can easily be dissipated. The contacts between the individual protein molecules in the crystal are relatively weak (which makes it so difficult to crystallize proteins). Thus, it is very unlikely that they will have an effect upon the native conformation of the protein, with the possible exception of some outer loops. In principle protein crystals are ordered aqueous solutions, because they may contain up to 70% water. They still show some flexibility of individual amino acid side chains and of even larger domains. Some proteins (*e.g.* hemoglobin, as well as many enzymes) retain their functional properties in the crystal, although rate constants may be very different to those in aqueous solution due to less favorable diffusion conditions in the crystal; cofactors and inhibitors can be cocrystallized or soaked into the protein crystals.

Attempts have been made to model G protein-coupled receptors [111 – 113] because of their similarity in the number of trans-membrane domains to bacteriorhodopsin. Such models prove to be useful for gaining further insight into the structure and function of receptors. However, their value for ligand design is limited; at atomic resolution such models are far from reality.

An important contribution to the receptor concept resulted from recent investigations of Herbette [114, 115] of the partitioning into and the distribution of drugs in biological membranes. The correct spatial arrangement of the drug and its proper orientation in the membrane with respect to the binding site at the surface of the membrane-embedded receptor are considered to be of utmost importance for the drug-receptor interaction (Figure 1). In addition, the model of a drug being

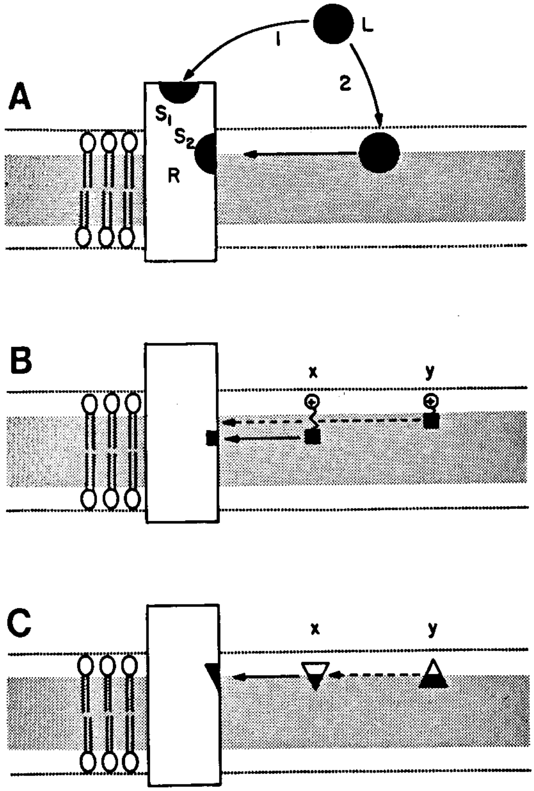


Figure 1: Drug-receptor interactions.
A) A ligand L may reach its binding site S_1 or S_2 at the receptor R by direct diffusion in the aqueous medium or (in the case of site S_2) by partitioning into the membrane and then diffusing to the binding site.
B) The highly ordered structure of the lipid bilayer may restrict lipophilic and especially amphiphilic drugs to a particular depth of penetration; drug **x** will fit the binding site because it is positioned at a proper depth for optimal interaction with the binding site, whereas drug **y** will be less active or inactive.
C) The orientation of the ligand relative to the binding site might also be optimized by the membrane by limiting the rotational degrees of freedom of the drug; drug **x** will be active, drug **y** not. In addition, the membrane may stabilize conformations of a drug which are different from those present in the liquid phase, thus enabling or disabling interaction with the receptor site in the membrane [114, 115].
(reproduced from Figure 2 of ref. [114] with permission from the Biophysical Society, Bethesda, MD, USA).

transferred from the aqueous phase to the membrane, finding its way inside the lipid bilayer, reduces the problem of a ligand approaching its binding site from three dimensions to only two dimensions.

Which forces are responsible for ligand binding to a receptor, be it an enzyme, a binding site at a receptor surface, a nucleic acid, or any other biological macromolecule? The affinity of a drug D to its binding site at the receptor R is determined by the free energy difference ΔG between the free states of both partners and the drug-receptor complex [DR], which is made up from the enthalpy change ΔH and the entropy change ΔS (eq. 10). The free energy ΔG is related to the equilibrium constant K for the reaction $D + R = [DR]$ by eq. 11.

$$\Delta G = \Delta H - T\,\Delta S \tag{10}$$

$$\Delta G = -2.303\,RT \log K \tag{11}$$

A short overview of the intermolecular interactions between drugs and their binding sites is given below.

Covalent bonds have energy values in the range of about 170–600 $\text{kJ} \cdot \text{mol}^{-1}$. As they are irreversible, they are not important for most therapeutically relevant drugs. Only alkylating agents (*e.g.* antitumor drugs like cyclophosphamide) as well as active site-directed and mechanism-based irreversible (suicide) enzyme inhibitors (*e.g.* the penicillins and cephalosporins as bacterial cell wall synthesis inhibitors, chloromethyl ketones as serine and cysteine protease inhibitors, or pargylin, deprenyl, and related analogs as monoamine oxidase inhibitors) form covalent bonds.

Electrostatic interactions are considered to be important attractive forces, due to their relative strength, [59, 116–119]. The molecular electrostatic field which surrounds a binding site guides the correct orientation of the drug [59] and is responsible for the first contact. However, the role of electrostatic interactions as being mainly responsible for high affinity has been questioned due to an often unfavorable solvation-desolvation energy balance. It is difficult to express their contribution in a quantitative manner, due to a number of reasons:

The dielectric constant inside a binding pocket may be significantly different from its value in water.

The strength of some interaction forces, *e.g.* of hydrogen bonds, depends on the interaction geometry [59, 117, 120].

In the case of additional dispersive interactions the resulting energy values heavily depend on small differences in the distances between the atoms participating in the interaction.

Even minor changes in the binding mode, *i.e.* in the geometry of the drug-receptor complex, in going from one analog to another, may increase or reduce the binding enthalpy ΔH considerably, a fact which is much too often neglected in quantitative structure-activity analyses.

Most interactions between charged groups include hydrogen bonds, *e.g.* between a positive ammonium group and a negative carboxylate, phenolate, phosphate, phosphonate, or sulfate group. Different energy values are given in literature (energies calculated *in vacuo* must not be compared with these values because they do not consider other intermolecular interactions, *e.g.* with the surrounding and competing water molecules). The values of charged hydrogen bonds have been estimated, mainly from the investigation of muteins, to be in the range of 15–19 $\text{kJ} \cdot \text{mol}^{-1}$, while those of neutral hydrogen bonds were estimated to be 2–6 $\text{kJ} \cdot \text{mol}^{-1}$ [121–123]. Correspondingly, the introduction of a neutral hydrogen bond increases the binding affinity by a factor of about 2–20, while the introduction of a charged hydrogen bond increases it by a factor of 400–2,000. Differences in free energy values, derived from reaction rates of ligands containing a hydroxyl group and ligands having a hydrogen atom instead, have been compiled for different enzymes [124]. From a recent comparison of the binding energies of amide-amide hydrogen bonds in aqueous solution and in nonpolar solvents it was concluded that earlier values of neutral hydrogen bond energies may be too small [125].

Dispersion forces are attractive forces between atoms at close distances. Even molecules with no permanent dipole moment have, due to the movement of their electrons, local dipole moments which induce dipoles in the opposite molecule, leading to fluctuating electrostatic attractions. At a closer distance repulsive forces develop due to an unfavorable overlap of the van der Waals spheres of both molecules. These relationships are typically described by the Lennard Jones potential, with an r^6 attractive term and an r^{12} repulsive term (Figure 2) [59, 116]. Dipole-dipole interactions and dispersion forces are much weaker than other electrostatic interactions. Nevertheless, if there is a close contact between both molecules over a relatively large surface area, they may sum up to large values of overall interaction energies.

Hydrophobic interactions are the most important single factor providing the driving force for noncovalent interactions in aqueous solution, especially in the case of large hydrophobic areas. They are merely entropic interactions. The molecules which surround the hydrophobic surfaces are loosely associated; they have a certain degree of order and are therefore in an unfavorable entropic state. The association of the hydrophobic areas of a ligand and its binding site displaces and releases the ordered water molecules into solution, which leads to a gain in entropy. The corresponding contribution of a methylene group (which is not in the neighborhood

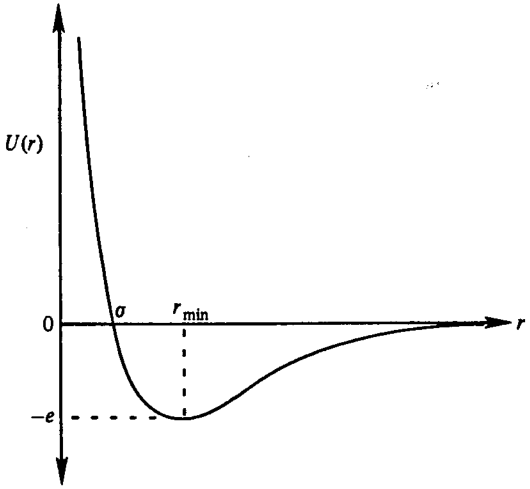


Figure 2: Dependence of the potential energy U of two atoms on their distance r (Lennard Jones potential). Coming from an infinite distance r, the energy decreases (attraction due to electrostatic interactions) until a minimum distance r_{min} is reached; from thereon repulsion due to increasing van der Waals overlap of the atoms results; σ is the distance for which the interaction energy is zero (reproduced from Figure 3.1 of ref. [59] with permission from Cambridge University Press, Cambridge, UK).

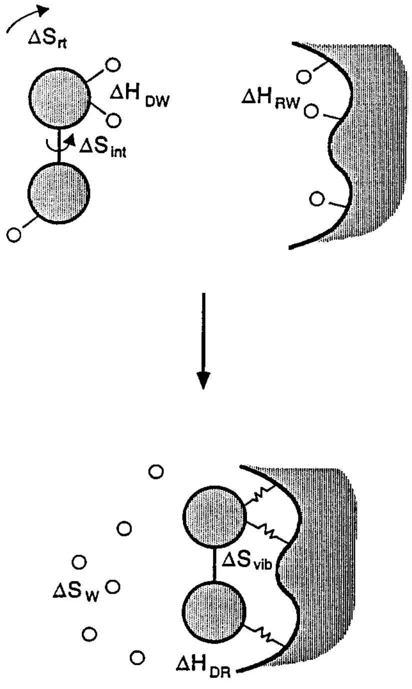


Figure 3: Drug-receptor interactions. ΔH_{DW} and ΔH_{RW} are the enthalpies of hydration of the drug and the receptor, respectively, ΔH_{DR} is the enthalpic contribution of the drug-receptor interaction, ΔS_{rt} is the overall rotational and translational entropy in solution, and ΔS_{int} the internal rotational entropy of the free drug. ΔS_W is the increase in entropy due to the release of bound water molecules (small circles) and ΔS_{vib} the entropy gain due to low-frequency vibrational modes associated with the drug-receptor noncovalent interaction (reproduced from Figure 1 of ref. [128] with permission from the American Chemical Society, Washington, DC, USA).

of a polar group, *i.e.* shielded by a bound water molecule) is estimated to be about 2 kJ · mol^{-1}, the contribution of a phenyl ring is about 8 kJ · mol^{-1} [116]. However, there still is a considerable discussion about the actual strength of hydrophobic interactions [118, 126].

Negative contributions to drug-receptor binding result from the loss of translational and rotational energies of the ligand in going from the free to the bound state, the loss of internal rotational degrees of freedom (conformational entropy) in the case of flexible molecules, and the enthalpy that is needed to remove water molecules associated to polar groups of both partners, *i.e.* from desolvation. After the drug-receptor complex has formed, a positive contribution results from the increase in entropy due to a low frequency vibration associated with the drug-receptor noncovalent bonds.

The net balance of favorable (enthalpic and entropic) and unfavorable (entropic) contributions shows the influence of the flexibility of a drug molecule as well as the importance of the quality of fit. As a first approximation, the loss of translational and rotational entropy does not increase proportionally to the size of a molecule, while the loss of internal conformational degrees of freedom depends on the number of rotatable bonds. This explains why rigid analogs (if they contain the correct conformation of the pharmacophore) are often much more active and show a higher degree of selectivity than the more flexible ones.

The contribution of polar and electrostatic interactions is often overemphasized because the transfer of the ligand from the aqueous medium to the binding site and especially the negative influence of desolvation are neglected. But hydrophobic interactions also have negative consequences; limited solubility renders transport and distribution in the biological system more difficult or even impossible if a drug molecule becomes too lipophilic. The most important favorable and unfavorable enthalpic and entropic contributions to drug-receptor interactions are summarized in Figure 3.

Whether a specific ligand-receptor interaction is enthalpy- or entropy-driven can be determined by thermodynamic analysis. According to Hitzemann [127], no general conclusions can be drawn whether the binding of agonists and antagonists is either stabilized by enthalpy or by entropy. Even the change of the receptor conformation, caused by the binding of an agonist and producing the receptor response, need not be the result of an enthalpic contribution. Furthermore, the ionic strength of the medium, *e.g.* the presence or absence of NaCl in the case of opiate receptors, may lead to opposite results.

The different contributions to the overall free energy of the binding of a ligand to a biological macromolecule [38, 116, 118, 119, 128, 129] and the important role of the surrounding water molecules [59, 130] have been reviewed.

QSAR: Hansch Analysis and Related Approaches
by Hugo Kubinyi
© VCH Verlagsgesellschaft mbH, 1993

2. Biological Data.
The Additivity of Group Contributions

Biological data [41, 131, 132] of all different kinds can be (and have been) used in quantitative structure-activity relationships; it should be noted that (especially in the case of pharmacokinetic parameters) only the dependence of these values on structural variation is considered in the context of QSAR studies:

Affinity data, like substrate or receptor binding constants,
rate constants, like association/dissociation, and Michealis Menten constants,
inhibition constants, especially K_i and IC_{50} values of different enzymes,
pharmacokinetic parameters, like absorption rate constants, distribution parameters, clearance, rate constants of metabolic degradation, and elimination rate constants,
in vitro biological activity values, using bacterial, fungal, and other cell cultures, as well as isolated and *in situ* perfused organs,
in vivo biological activity values, *i.e.* various pharmacodynamic and toxic activities of drugs.

All these data can be used in QSAR studies, provided they are in the right scale. Each single step of drug distribution and even the binding of the drug to its receptor site correspond to a partitioning of the drug between an aqueous and a nonaqueous phase, the latter being a membrane or another lipid phase, a serum protein, the active site of an enzyme, or the binding site of a receptor.

Equilibrium as well as rate constants are related to free energy values ΔG by relationships of the type of eq. 11 (chapter 1.2). Thus, only equilibrium constants (*e.g.* K_i values or at least IC_{50} values, not % inhibition at a certain concentration) and rate constants (*e.g.* log k values, not % absorption or % concentration in a certain compartment) are suited for QSAR studies, which means that all biological data have to be transformed in an appropriate manner before being used in quantitative analyses. In the case of complex biological data resulting from a sequence of several independent processes (in the worst case whole animal data), sometimes one effect predominates: *e.g.* the bioavailability, the penetration of the blood-brain barrier, or the affinity to the receptor site. In other cases several effects overlap, which makes the QSAR analysis much more difficult. Due to the nonlinear characteristics of dose-response relationships, % effect values at a certain dose must not be used in QSAR equations. In each case they have to be transformed to equieffective molar doses (*i.e.* dose levels which produce or prevent a certain pharmacodynamic effect; dose levels that increase the life span of animals to a certain extent; dose levels which kill a certain percentage of the animals).

According to eq. 11, all values have to be converted into a logarithmic scale. Decadic logarithms are used in classical QSAR analyses; only in CoMFA studies (chapter 9.3) are natural logarithms (which are related to decadic logarithms by a factor of 2.303) occasionally applied instead. As a convention negative logarithms, *i.e.* logarithms of reciprocal molar concentrations (*e.g.* log 1/C or pC values), are preferred to obtain larger values for more active compounds (values in this inverse scale are directly comparable to pH and pK_a values).

There is another reason to use the logarithmic scale for biological activity values in quantitative structure-activity analyses. A condition for the application of regression analysis is a normal distribution of the experimental error in the dependent variable (not of the data themselves as sometimes stated in the literature; the biological activity values should be reasonably distributed over the whole range, without clustering of data, but they need not fulfill any other conditions). For biological data this holds true for the logarithmic scale, not for the linear scale. Thus, log 1/C (C being a molar concentration that produces a certain effect, *e.g.* an ED_{50} value), log $1/K_i$, pI_{50}, log k values, *etc.* are appropriate biological parameters for QSAR studies. Other data, like mg/kg values, sometimes give satisfactory results because they are highly correlated with molar doses, at least within a narrow molecular weight range of the compounds included in the analysis. The same holds true for some % effect or % concentration data, ranging from a few percent values to values near 100% (measured at a single dose); also these values are not too far from being linearly related to the theoretically correct parameters; however, in good QSAR practice they should not be applied. Linear values (instead of logarithmic values) are sometimes found in the QSAR literature. If they are within a small range, they are closely correlated with logarithmic values. Any conclusions drawn from a comparison of the results as to whether one or the other scale might be "better" are meaningless. They only reflect an excessively narrow range of data; fortuitous errors may lead to wrong interpretations.

Due to the relationship between biological activity and the free energies of binding (or partitioning) also the terms "extrathermodynamic relationships" or "linear free energy-related approach" are used for quantitative structure-activity relationships, especially Hansch analysis.

A medicinal chemist, whether being familiar with the principles of quantitative structure-activity relationships or not, implicitly applies the additivity concept of group contributions to biological activity values. From general experience it is assumed that the activity contribution of the parent system of a lead structure will remain constant if a certain group of the molecule is exchanged for another group. Certain frustrations arising from exceptions to this rule (*e.g.* the older medicinal chemist's joke: methyl, ethyl, propyl, futile; how about phenylbutazone and suloctodil?) nowadays may easily be explained by nonlinear lipophilicity-activity relationships. In general, the assumption of a more or less strict additivity of group contributions to biological activity values has been proven to be correct by thousands of QSAR equations and by a few dedicated investigations which will be discussed below.

While the success of QSAR analyses may be taken as sufficient evidence for the additivity of group contributions to biological activity values, the following question arises: are these group contributions more or less constant from one system to the other or do they depend on the choice of the compounds and/or the biological system?

To answer this question, they are definitely not constant for different systems. The "biological Hammett equation" (eq. 4, section 1.1) only applies to certain groups of compounds, covering a narrow range of lipophilicity, and to nonspecific biological data. On the other hand, the work of Andrews [116, 128] gives evidence that certain functional groups contribute to receptor affinity within a more or less constant range. He used eq. 12 (ΔS_{rt} = overall translational and rotational entropy of the drug in solution; n_{DOF} and E_{DOF} = numbers and energies of internal degrees of freedom; n_X and E_X = numbers and energies of interaction of different functional groups) to calculate "mean binding energies" AVERAGE ΔG and derived eq. 13 from affinity constants which were converted to free energy values ΔG by eq. 11 (all E_X values are given in $\mathrm{kJ \cdot mol^{-1}}$, in contrast to the values in refs. [116, 128]). The 200 different ligands which were included in the investigation covered a wide variety of different chemical structures. The only weak point in eq. 13 is that the $T\,\Delta S_{rt}$ term (first term at the right side of eq. 13) had to be fixed at 59 $\mathrm{kJ \cdot mol^{-1}}$ before the regression was performed.

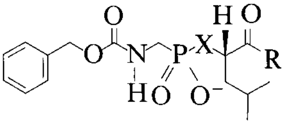
$$\text{AVERAGE } \Delta G = T\,\Delta S_{rt} + n_{DOF} \cdot E_{DOF} + \Sigma n_X \cdot E_X \tag{12}$$

$$\begin{aligned}\text{AVERAGE } \Delta G = {} & -59 - 3.0 n_{DOF} + 3.0 n_{C(sp^2)} + 3.4 n_{C(sp^3)} + 48 n_{N^+} + \\ & + 5.0 n_N + 34 n_{CO_2^-} + 42 n_{PO_4^{2-}} + 10.5 n_{OH} + 14.2 n_{C=O} + \\ & + 4.6 n_{O,S} + 5.4 n_{Hal}\end{aligned} \tag{13}$$

Eq. 13 is by no means predictive; if the intrinsic binding energy contributions are calculated from different subsets, they cover a wide range of values. AVERAGE ΔG means that there are "poor fit compounds", like methotrexate ($-74\ \mathrm{kJ \cdot mol^{-1}}$), ouabain (−64), penfluridol (−52), carboxyribitol 1,5-diphosphate (−40), thyroxine (−39), buprenorphine (−38), and ketanserin (−38). They all bind to the receptor site much worse than predicted by eq. 13, which indicates that not all functional groups interact with the binding site or that some conformational energy is needed to adopt a conformation which allows the interaction of these groups with the binding site. On the other hand, there are "exceptional fit compounds", like biotin (+69), camphor (+62), the phosphate^{3-} ion (+49), 2,4-diamino-5-methyl-6-pentylquinazoline (DMPQ; +48), phenol (+45), valproate (+41), estradiol (+41), diazepam (+36), and oxalate (+26). They bind much better than predicted by eq. 13, which indicates that all possible interactions contribute and that the values given in eq. 13 should be interpreted as the lower limits of interaction energies in a geometrically favorable arrangement of the interacting groups [128].

General rules of bioisosterism [133–135] are reflected by eq. 13. However, significant exceptions may be obtained, depending on the biological system. Bartlett [136, 137] described thermolysin inhibitors (**1**), where the exchange of the −NH− group of phosphonamidate analogs for oxygen (phosphonates) reduces affinities to the enzyme by a factor of about 1,000, while in the case of the $-CH_2-$ analogs (phosphinates) affinities are retained (Table 1). Protein crystallography shows that the −NH− group of the phosphonamidates interacts as a hydrogen bond donor with the backbone carbonyl group of an alanine (Figure 4) [138].

Table 1. Inhibition constants (K_i values in nmol) of thermolysin inhibitors (**1**) [137]



(1)

Series	X		
(R)	-NH-	-O-	$-CH_2-$
OH	760	660,000	1,400
Gly-OH	270	230,000	300
Phe-OH	78	53,000	66
Ala-OH	16.5	13,000	18.4
Leu-OH	9.1	9,000	10.6

Neither the oxygen nor the methylene analogs are capable of forming such an interaction. The affinity differences can be explained by a consideration of the hydrogen bond balance: the phosphonamidates and phosphonates are hydrated in solution, which is unfavorable for binding. Only in the phosphonamidates is this negative effect counterbalanced by a hydrogen bond to the binding site. Therefore, the situation is balanced for the phosphonamidates (hydrogen bonds in water and at the binding site) and for the phosphinates (no hydrogen bonds at all); it is unfavorable for the phosphonates (hydrogen bonds only in water), thus explaining the large differences to the other analogs. The differences in affinities between the phosphonamidates and the phosphonates were also theoretically explained by a thermodynamic perturbation method, implemented with molecular dynamics [139].

There is another important limitation of the concept of bioisosterism, which arises from different conformational preferences of closely related analogs. A systematic investigation of the influence of X on allowed and preferred conformations of

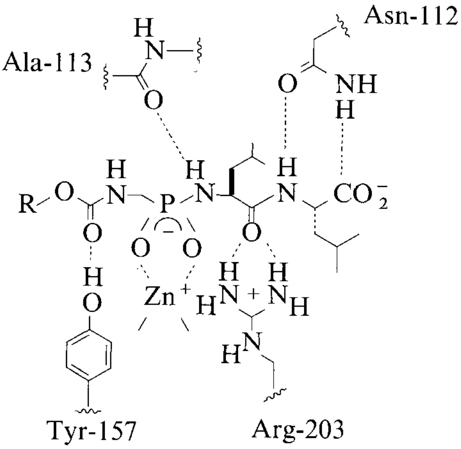


Figure 4: Electrostatic interactions between inhibitors (**1**) and thermolysin in the vicinity of the $-PO_2NH-$ linkage (Table 1) (reproduced from Figure 1 of ref. [137] with permission from the American Chemical Society, Washington, DC, USA).

Phe-X-Phe (with and without *ortho*-substitution of the phenyl rings) demonstrates that even small structural variations of X (*e.g.* in going from X = —O— to X = —NH—) drastically change the conformation maps [140]. The results of semiempirical calculations [140] are identical with the experimental data for corresponding structures as contained in the Cambridge database of crystal structures [141].

To answer the question whether, within a certain group of compounds and for a certain type of biological interaction, the group contributions to biological data are strictly additive or not, the problem has to be defined in the correct manner. Evidence for additivity of binding energies has been derived in cases where a small molecule A and a larger molecule B interact with different parts of a binding site. If such molecules are combined to A—B, the group A may now be considered as a substituent of the larger molecule B (the "anchor principle") [142—144]. Most of the loss of entropy due to freezing translational and rotational degrees of freedom can be attributed to the binding of B. Thus, the difference of the free energy values between B and AB reflects the true binding energy of A. The difference $\Delta G_{AB} - \Delta G_B$ is much larger than the binding energy ΔG_A, since it does no longer include the unfavorable entropy term $T\,\Delta S_A$ (eqs. 14—16).

$$\Delta G_A = \Delta H_A - T\,\Delta S_A \tag{14}$$

$$\Delta G_{AB} - \Delta G_B = \Delta H_{AB} - T\,\Delta S_{AB} - \Delta H_B + T\,\Delta S_B \approx \Delta H_{AB} - \Delta H_B \tag{15}$$

$$\Delta G_{AB} - \Delta G_B \approx \Delta H_A \tag{16}$$

Consequently, the binding energy of a certain functional group can be estimated by comparing the binding energies of a pair of compounds, one bearing this group and the other lacking it. The same principle applies to mutagenesis experiments, to peptide or protein ligands and to the binding sites. Additivity in this sense has elegantly been proven by a series of single and double/multiple point mutation experiments [123, 145], which demonstrate a strict additivity of multiple exchanges as compared to single exchanges, for transition-state stabilization energies as well as for free energies of interaction at protein-protein surfaces. Protein crystallography helps to understand some slight deviations from this additivity rule, resulting from unfavorable side-chain interactions, *e.g.* a difference of 1.7 $\text{kJ} \cdot \text{mol}^{-1}$ in the case of a Cys-35, Thr-51 double mutant of tyrosyl-tRNA synthetase, as compared to the single site mutations [146]. It should be mentioned that such mutation experiments also contribute to the understanding of enzymatic reactions [147] and to the function of distant groups, *e.g.* mobile loops, which participate in the enzymatic process [148]. The ongoing progress in the genetic expression of receptor proteins will lead to a better understanding of receptor function, regulation, and selectivity, even before the first experimental 3D structures become available.

In only a few investigations has the number of various kinds of interactions of ligands with their binding sites been correlated with affinities or some other type of biological activity. The two most prominent ones shall be mentioned here: first, an investigation of hemoglobins from different species, where the shift of the oxygen binding curves has been explained by the different interactions of the allosteric effector molecules with the protein (eq. 17; n_I, n_C = numbers of ionic and covalent interactions) [149]; oxygen affinities were first converted to K_D values and then to free energies of binding, ΔG (values in $kJ \cdot mol^{-1}$). In eq. 17, as in all following examples of QSAR equations, the values given in parentheses after the regression coefficients are 95% confidence intervals (not standard deviations, which are usually smaller by a factor of two), n is the number of compounds, r is the multiple correlation coefficient, and s is the standard deviation; sometimes F, the Fisher significance ratio, is given in addition to r and s values (chapter 5.1).

$$\Delta G = -3.14(\pm 0.62)\, n_I - 6.78(\pm 1.39)\, n_C - 8.29(\pm 2.87) \qquad (17)$$

$$(n = 29;\ r = 0.928;\ s = 3.34;\ F = 81.15)$$

Eq. 17 indicates that the binding energies of the different ligands to the various hemoglobins are closely related to the numbers of ionic and covalent interactions. A reversible covalent bond contributes about twice as much ($-6.78\ kJ \cdot mol^{-1}$) as an ionic interaction ($-3.14\ kJ \cdot mol^{-1}$).

The second example is a recent investigation of thermolysin inhibitors [150], in which the hydrophobic and polar contacts between the inhibitors and the binding site were correlated with their inhibitory potencies (eq. 18, recalculated from the data given in ref. [150]; NPHO = total complementary nonpolar carbon contacts, *i.e.* hydrophobic interactions; NHBOND = buried complementary contacts, where the ligand provides a hydrogen donor or a hydrogen acceptor, *i.e.* hydrogen bonds are formed between the ligand and the binding site). One compound had to be excluded because of an extremely unfavorable contact between two hydrogen bond acceptor atoms (see above, discussion on phosphonates as thermolysin inhibitors).

$$\log 1/K_i = 0.624(\pm 0.10)\, \mathrm{NPHO} + 0.217(\pm 0.08)\, \mathrm{NHBOND} - 3.623(\pm 0.59) \qquad (18)$$

$$(n = 9;\ r = 0.993;\ s = 0.228;\ F = 202.51)$$

Also the affinity constants of the phosphonamidate analogs (**1**, Table 1) confirm the additivity concept of group contributions to biological activity values. The different residues R are separated from the bridge atom X by two carbon atoms; all affinity values of the three series with X = $-NH-$, $-O-$, and $-CH_2-$ are closely correlated [137].

There is overwhelming evidence that the additivity of group contributions to biological data is not just a working hypothesis, but an intrinsic feature resulting from the thermodynamic relationships between the free energies of binding (and partitioning, respectively) and biological activity values.

Deviations only arise from:

nonlinear dependences of transport and distribution on lipophilicity (the most important reason for nonadditivity of group contributions; see chapter 4.4),
different metabolic pathways,
steric crowding of ligand groups, leading to conformational distortions,
multiple binding modes, caused by steric interference or other repulsive interactions at the binding site (see chapter 9.3), and
changes in the mechanism of action, *e.g.* substrates and inhibitors or agonists and antagonists (normally not being responsible for nonadditivity).

QSAR: Hansch Analysis and Related Approaches
by Hugo Kubinyi
© VCH Verlagsgesellschaft mbH, 1993

3. Parameters

Parameters which encode certain structural features and properties are needed to correlate biological activities with chemical structures in a quantitative manner. Of special value are physicochemical properties which are directly related to the intermolecular forces involved in the drug-receptor interaction as well as to the transport and distribution properties of drugs. In this respect hydrophobic, polar, electronic, and steric properties are most important; most often, polarizability parameters are considered as being closely related to steric parameters and are discussed together with these parameters, although they are definitely different (if structural variation is appropriate).

In this chapter physicochemical and other parameters will be differentiated according to:

Lipophilicity parameters, *e.g.* partition coefficients and chromatographic parameters.

Polarizability parameters, *e.g.* molar refractivity, molar volume, and parachor.

Electronic parameters, *e.g.* Hammett σ constants, field and resonance parameters, parameters derived from spectroscopic data, charge transfer constants, dipole moments, and quantum-chemical parameters.

Steric parameters, derived from linear free energy relationships or from geometric considerations.

Parameters like molecular weight, geometric parameters, conformational entropies, connectivity indices, and other topological parameters.

Indicator (Free Wilson-type, dummy) variables.

Hundreds of different parameters have been used in QSAR studies in the past 30 years. Some scientists are really creative in steadily inventing new parameters, arguing that all problems in QSAR will disappear with their use. While some parameters may indeed turn out to be useful, most of them have no general significance. A certain saturation effect is seen in this field of QSAR research (like in the case of mathematical models for nonlinear lipophilicity-activity relationships; chapter 4.4). On the other hand, there still is a lack of adequate parameters to describe some important interactions like the membrane partitioning of drugs, the strength of hydrogen bonds, the influence of desolvation energies on drug-receptor affinity, and steric interactions with a (most often unknown) binding site.

Early parameter collections, published by Hansch *et al.* [151, 152] included π, σ_m, σ_p, $\mathcal{F}$, $\mathcal{R}$, and MR values of 284 aromatic substituents. Another publication contained π, π^-, $\mathcal{F}$, $\mathcal{R}$ (separated according to *ortho*-, *meta*-, and *para*-substitution) and MR values of 34 aromatic substituents [153]. In the book *"Substituent Constants for Correlation Analysis in Chemistry and Biology"*, published by Hansch and Leo in 1979, π, MR, E_s, and all known σ values were compiled for about 2,000 different substituents [50]. Two tables included π, MR, $\mathcal{F}$, $\mathcal{R}$, σ_m, σ_p, and indicator variables for hydrogen bond acceptor and donor properties of 166 aromatic substituents as well as hydrophobic fragmental constants, hydrogen bond acceptor and donor properties, MR, and $\mathcal{F}$ values of 103 aliphatic substituents. By cluster analysis the substituents were separated into 20, 10, and 5 different groups, respectively, to allow a rational selection of substituents in the design of new analogs according to their distances in multidimensional parameter space [50, 154].

Physicochemical parameters, the underlying theories, and their use in QSAR studies have been reviewed in several other monographs [39–43, 48, 49, 51, 52, 54–56] and in dedicated articles (*e.g.* [155–158]).

The most comprehensive compilation of physicochemical parameters mentions more than 220 different parameters and contains a table of 58 parameters of 59 different substituents, groups, and fragments, together with a correlation matrix of all 58 parameters [158]. Later, another 16 parameters of the same set of substituents were added; about 18% of missing values had to be estimated [159, 160]. All parameter values are included in a commercially available database DESBASE [161]. An in-house compilation of Eli Lilly, mentioned in the literature but not available (!) to the scientific community, contains more than 17,000 physicochemical parameters of 3,000 different substituents [162], organized in a relational database for fast and efficient retrieval.

Principal component analysis (PCA; chapter 5.3) reduces the data matrices of physicochemical properties to fewer, orthogonal dimensions [159, 160, 163–169]. Six properties (aqueous solvation energy, partition coefficient, boiling point, molar refractivity, volume, and vaporization enthalpy) of 114 liquid compounds are nearly quantitatively correlated with the first two principal components, termed B (bulk) and C (cohesiveness), which were derived from a large number of physicochemical properties of these compounds [164–166]. Some other properties can be explained by B and C and additional minor components D, E, and F; only a few properties, *e.g.* dipole moment, melting point, and molecular weight, are not correlated with these parameters. The BC(DEF) parameters were used to "predict" some known properties of 139 further compounds with a relatively high degree of reliability [165].

Principal component analysis of several electronic parameters showed that more than 90% of the information is contained in their first principal component [167]. PCA of seven different chemical descriptors (lipophilic, electronic, and steric properties of 28 aromatic substituents) led to a clear grouping of substituents in a plot of the loadings of the first two components according to hydrogen bond acceptors, hydrogen bond donors, alkyl groups, and halogens [168]. Nine descriptor variables, *i.e.* π, MR, σ_m, σ_p, and the STERIMOL parameters L, B_i, B_{ii}, B_{iii}, and B_{iv}, of a set of 100 aromatic substituents (Table 2) were investigated by PCA [169]. The first four components contained 39%, 21%, 9%, and 7% (together explaining 76%) of the total variance; from the weights and the loadings of the components it could be concluded that the first component is mainly related to steric bulk and hydrophobicity, the second component to electronic parameters, and the third one again to hydrophobicity and shape.

Principal component analysis was also used to investigate 74 parameters of 59 different substituents [159, 160]. Five significant components contained about 84%

Table 2. Representative parameters of aromatic substituents (reproduced from Table 1 of ref. [169] with permission from the copyright owner)

Substituents	π	MR$^{a)}$	σ_m	σ_p	L	**B$_i$$^{b)}$**	**B$_{ii}$$^{b)}$**	**B$_{iii}$$^{b)}$**	**B$_{iv}$$^{b)}$**
Br	0.86	0.888	0.39	0.23	3.83	1.95	1.95	1.95	1.95
Cl	0.71	0.603	0.37	0.23	3.52	1.80	1.80	1.80	1.80
F	0.14	0.092	0.34	0.06	2.65	1.35	1.35	1.35	1.35
SO_2F	0.05	0.865	0.80	0.91	3.50	2.03	2.70	2.45	2.51
SF_5	1.23	0.989	0.61	0.68	4.65	2.49	2.49	2.49	2.49
I	1.12	1.394	0.35	0.18	4.23	2.15	2.15	2.15	2.15
NO	-1.20	0.520	0.62	0.91	3.44	1.70	2.44	1.70	1.70
NO_2	-0.28	0.736	0.71	0.78	3.44	1.70	1.70	2.44	2.44
N_3	0.46	1.020	0.27	0.15	4.62	1.50	4.18	2.34	2.57
H	0.00	0.103	0.00	0.00	2.06	1.00	1.00	1.00	1.00
OH	-0.67	0.285	0.12	-0.37	2.74	1.35	1.93	1.35	1.35
SH	0.39	0.922	0.25	0.15	3.47	1.70	2.33	1.70	1.70
NH_2	-1.23	0.542	-0.16	-0.66	2.93	1.50	1.50	1.84	1.84
SO_2NH_2	-1.82	1.228	0.46	0.57	3.82	2.11	3.07	2.67	2.67
$NHNH_2$	-0.88	0.844	-0.02	-0.55	3.40	1.50	2.82	1.84	1.84
$N{=}CCl_2$	0.41	1.835	0.21	0.13	5.65	1.70	1.80	1.84	4.54
CF_3	0.88	0.502	0.43	0.54	3.30	1.98	2.61	2.44	2.44
OCF_3	1.04	0.786	0.38	0.35	4.57	1.35	3.33	2.44	2.44
SCF_3	1.44	1.381	0.40	0.50	4.89	1.70	3.69	2.44	2.44
CN	-0.57	0.633	0.56	0.66	4.23	1.60	1.60	1.60	1.60
NCS	1.15	1.724	0.48	0.38	4.29	1.50	2.24	1.64	1.76
SCN	0.41	1.340	0.41	0.52	4.08	1.70	4.45	1.70	1.70
SO_2CF_3	0.55	1.286	0.79	0.93	4.11	2.11	3.64	2.67	2.67
NHCN	-0.26	1.014	0.21	0.06	3.53	1.50	3.08	1.90	1.90
CHO	-0.65	0.688	0.35	0.42	3.53	1.60	1.60	2.00	2.36
CO_2H	-0.32	0.693	0.37	0.45	3.91	1.60	1.60	2.36	2.66

Table 2. (continued)

OCH_2CH_3	0.38	1.247	0.10	-0.24	4.92	1.35	3.36	1.35	1.90
CH_2OCH_3	-0.78	1.207	0.02	0.03	4.91	1.52	2.88	1.90	1.90
SC_2H_5	1.07	1.842	0.18	0.03	5.24	1.70	3.97	1.90	1.90
NHC_2H_5	0.08	1.498	-0.24	-0.61	4.96	1.50	3.42	1.90	1.90
$SO_2C_2H_5$	-1.09	1.814	0.60	0.72	5.31	2.11	3.67	2.67	2.67
NMe_2	0.18	1.555	-0.15	-0.83	3.53	1.50	2.56	2.80	2.80
PMe_2	0.44	2.119	0.03	0.31	3.88	2.00	2.97	2.84	3.29
cyclo-PROPYL	1.14	1.353	-0.07	-0.21	4.14	1.98	2.24	2.29	2.88
$CO_2C_2H_5$	0.51	1.747	0.37	0.45	5.96	1.90	1.90	2.36	4.29
$(CH_2)_2CO_2H$	-0.29	1.652	-0.03	-0.07	5.96	1.52	3.05	2.35	2.67
$NHCO_2C_2H_5$	0.17	2.118	0.07	-0.15	4.45	1.50	4.97	1.90	5.57
$CHMe_2$	1.53	1.496	-0.07	-0.15	4.11	2.04	2.76	3.16	3.16
n-C_3H_7	1.55	1.496	-0.07	-0.13	5.05	1.52	3.49	1.90	1.90
$OCHMe_2$	0.85	1.706	0.10	-0.45	4.59	1.35	3.61	1.90	3.16
OC_3H_7	0.85	1.706	0.10	-0.25	6.05	1.35	4.30	1.90	1.90
SC_3H_7	1.61	2.307	0.15	0.00	6.21	1.70	4.90	1.90	1.90
NHC_3H_7	0.62	1.963	-0.24	-0.61	6.07	1.50	4.36	1.90	1.90
2-Thienyl	1.61	2.404	0.09	0.05	5.97	1.65	1.77	3.13	3.16
$CH{=}CHCOCH_3$	-0.06	2.110	0.21	-0.01	5.80	1.60	3.24	1.83	3.73
COC_3H_7	0.53	2.048	0.38	0.50	4.67	2.36	3.69	3.16	3.16
$CO_2C_3H_7$	1.07	2.217	0.37	0.45	6.90	1.90	1.90	2.36	4.83
sec-C_4H_9	2.04	1.959	-	-0.12	5.02	1.90	3.16	2.76	3.49
iso-C_4H_9	-	1.959	-	-0.12	5.05	1.52	4.21	1.90	3.16
n-C_4H_9	2.13	1.969	-0.08	-0.16	6.17	1.52	4.42	1.90	1.90
tert-C_4H_9	1.98	1.962	-0.10	-0.20	4.11	2.59	2.97	2.86	2.86
OC_4H_9	1.55	2.166	0.10	-0.32	6.99	1.35	4.79	1.90	1.90
NHC_4H_9	1.16	2.426	-0.34	-0.51	7.01	1.50	4.97	1.90	1.90
C_5H_{11}	2.67	2.426	-0.08	-0.16	7.11	1.52	4.94	1.90	1.90
C_6H_5	1.96	2.536	0.06	-0.01	6.28	1.70	1.70	3.11	3.11
$N{=}NC_6H_5$	1.69	3.131	0.32	0.39	8.43	1.70	1.70	1.92	4.31

Table 2. (continued)

CH_2Br	0.79	1.339	0.12	0.14	4.09	1.52	3.75	1.95	1.95
CH_2Cl	0.17	1.049	0.11	0.12	3.89	1.52	3.46	1.90	1.90
CH_2I	1.50	1.886	0.10	0.11	4.36	1.52	4.15	2.15	2.15
NHCHO	-0.98	1.031	0.19	0.00	4.22	1.50	1.50	1.94	3.61
$CONH_2$	-1.49	0.981	0.28	0.36	4.06	1.60	1.60	2.42	3.07
CH=NOH	-0.38	1.028	0.22	0.10	4.88	1.60	1.60	1.92	3.11
CH_3	0.56	0.565	-0.07	-0.17	3.00	1.52	2.04	1.90	1.90
$NHCONH_2$	-1.30	1.372	-0.03	-0.24	5.09	1.84	1.84	1.94	3.61
$NHCSNH_2$	-1.40	2.219	0.22	0.16	4.62	1.50	4.18	2.34	2.57
OCH_3	-0.02	0.787	0.12	-0.27	3.98	1.35	2.87	1.90	1.90
CH_2OH	-1.03	0.719	0.00	0.00	3.97	1.52	2.70	1.90	1.90
$SOCH_3$	-1.58	1.370	0.52	0.49	4.03	1.60	2.93	2.49	3.36
OSO_2CH_3	-0.88	1.699	0.39	0.36	4.03	1.35	3.86	1.90	3.57
SO_2CH_3	-1.63	1.349	0.60	0.72	4.37	2.11	3.15	2.67	2.67
SCH_3	0.61	1.382	0.15	0.00	4.30	1.70	3.26	1.90	1.90
$NHCH_3$	-0.47	1.033	-0.30	-0.84	3.53	1.50	3.08	1.90	1.90
$NHSO_2CH_3$	-1.18	1.817	0.20	0.03	4.06	1.50	1.90	3.59	3.88
CF_2CF_3	1.68	0.923	0.47	0.52	4.11	1.98	3.64	2.44	2.44
$C\equiv CH$	0.40	0.955	0.21	0.23	4.66	1.60	1.60	1.60	1.60
$NHCOCF_3$	0.08	1.430	0.30	0.12	-	-	-	-	-
CH_2CN	-0.57	1.011	0.16	0.01	3.99	1.52	4.12	1.90	1.90
$CH=CHNO_2$	0.11	1.642	0.32	0.26	4.29	1.60	3.24	1.83	4.21
$CH=CH_2$	0.82	1.099	0.05	-0.02	4.29	1.60	1.60	2.00	3.09
$COCH_3$	-0.55	1.118	0.38	0.50	4.06	1.90	1.90	2.36	2.93
$SCOCH_3$	0.10	1.842	0.39	0.44	5.19	1.70	4.01	1.90	1.90
$OCOCH_3$	-0.64	1.247	0.39	0.31	4.87	1.35	3.68	1.90	1.90
CO_2CH_3	-0.01	1.289	0.37	0.45	4.85	1.90	1.90	2.36	3.36
$NHCOCH_3$	-0.97	1.493	0.21	0.00	5.15	1.50	3.61	1.90	1.94
$CONHCH_3$	-1.27	1.457	0.35	0.36	5.00	1.60	2.23	2.42	3.07
CH_2CH_3	1.02	1.030	-0.07	-0.15	4.11	1.52	2.97	1.90	1.90

Table 2. (continued)

OC_6H_5	2.08	2.768	0.25	-0.03	4.51	1.35	5.89	3.11	3.11
$SO_2C_6H_5$	0.27	3.320	0.61	0.70	5.82	2.11	6.01	2.67	2.67
$OSO_2C_6H_5$	0.93	3.670	0.36	0.33	8.20	1.61	3.64	1.80	3.57
NHC_6H_5	1.37	3.004	-0.12	-0.40	4.53	1.50	5.95	3.11	3.11
$NHSO_2C_6H_5$	0.45	3.788	0.16	0.01	-	-	-	-	-
cyclo-HEXYL	2.51	2.669	-0.15	-0.22	6.17	2.04	3.49	3.16	3.16
COC_6H_5	1.05	3.033	0.34	0.43	4.57	2.36	5.98	3.11	3.11
$OCOC_6H_5$	1.46	3.233	0.21	0.13	8.15	1.70	4.40	1.70	1.84
$N{=}CHC_6H_5$	-0.29	3.301	-0.08	-0.55	8.40	1.70	3.55	1.80	3.66
$CH{=}NC_6H_5$	-0.29	3.301	0.35	0.42	8.50	1.70	1.70	2.36	4.07
$CH_2C_6H_5$	2.01	3.001	-0.08	-0.09	3.63	1.52	6.02	3.11	3.11
$CH_2OC_6H_5$	1.66	3.219	0.03	0.04	8.19	1.52	3.09	3.11	3.11
$C{\equiv}CC_6H_5$	2.65	3.321	0.14	0.16	8.88	1.70	1.70	3.11	3.11
$NHCOC_6H_5$	0.49	3.464	0.02	-0.19	8.40	1.94	3.61	3.11	3.11

[a] MR is scaled by a factor of 0.1, as usual

[b] B_i – B_{iv} are defined as in [169]; B_i is the smallest value orthogonal to L, B_{ii} is opposite to B_i; B_{iii} and B_{iv} are orthogonal to B_i and B_{ii} and arranged in such a manner that $B_{iii} < B_{iv}$.

of the total variance; lipophilic, steric, and electronic parameters were nicely separated in a plot of the first two components (containing 33% and 25% of the information), while several less well defined properties as well as a random variable were close to the origin of this diagram. After discarding 12 irrelevant descriptors, three significant components explained 77% of the total variance (38%, 28% and 11%). Lipophilicity was interpreted as being related to bulk (*i.e.* the size of the substituents) and polarity, which was also confirmed by corresponding combinations of such parameters in regression equations [160].

Amino acids were characterized by a principal component analysis of their side-chain properties [170, 171]; first, for 20 coded amino acids three principal components z_1, z_2 and z_3 were derived (interpreted as being related to hydrophilicity, side-chain bulk, and electronic properties) [170] and afterwards new z scales resulted [171] from a partial least squares (PLS) analysis of the side-chain properties of these amino acids and additional 35 noncoded (unnatural) amino acids. The use of these scales (instead of the original variables) was recommended for structure-activity analyses.

Principal component scores have several advantages as compared to the original variables:

The individual error of a single parameter value does not significantly distort the values of the principal component scores.

The problem of missing data is greatly reduced.

Correlation of biological activity values with the first few components (which can be interpreted in terms of lipophilicity, bulk, and electronic properties) gives a clear picture of the inherent properties which might be responsible for the variation in the activity values.

On the other hand, there are serious shortcomings. The principal component scores of substituents clearly depend on a proper selection of the original variables. In one of the cases cited above, five out of seven variables were electronic parameters [168]. In another study five out of nine variables were related to the size of the substituents [169]. A further disadvantage is that higher components can no longer be interpreted in physicochemical terms. Nowadays the use of the PLS method (chapter 5.3) [26, 27], including all physicochemical parameters in the block of X variables, seems to be a better alternative.

3.1. Lipophilicity Parameters

No other physicochemical property has attracted as much interest in QSAR studies as lipophilicity (synonymously called hydrophobicity; any differentiation between both terms is only a semantic nicety; the opposite to lipophilicity is hydrophilicity) [172, 173], due to its direct relationship to solubility in aqueous phases, to membrane permeation, and to its (merely entropic) contribution to ligand binding at the receptor site. Several monographs [49, 50, 54, 56] and numerous reviews (*e.g.* [158, 173 – 179]) have been published, the recent article by Taylor [173] providing the most comprehensive overview (the reading of this excellent review, being a masterpiece in elegant style and critical comments, is highly recommended; also the theoretical background of partitioning and solubility is discussed there in detail).

Lipophilicity is defined by the partitioning of a compound between an aqueous and a nonaqueous phase. While early definitions of the partition coefficient P referred to "light" and "heavy" phases (leading to complications in the case of organic solvents having higher density than water), nowadays P is defined as the ratio of substance concentrations in the organic and aqueous phases of a two-compartment system under equilibrium conditions (eq. 19); due to possible association of the solute in the organic phase (*e.g.* dimers of carboxylic acids), partition coefficients should be measured at low concentrations or P values must be extrapolated to infinite dilution of the solute in the system (the partitioning of acids and bases is discussed in chapter 4.5).

$$P = \frac{c_{org}}{c_{aq}} \qquad (19)$$

While partition coefficients from many different organic solvent/water systems have been used in early structure-activity relationships, *n*-octanol became the organic solvent of choice after the pioneering work of Hansch on *n*-octanol/water partition coefficients of substituted phenoxyacetic acids and the lipophilicity parameter π derived from these partition coefficients [14, 15, 17, 18, 173, 180].

n-Octanol/water has many important advantages as compared to other systems [173, 180]:

It is a suitable model of the lipid constituents of biological membranes, due to its long alkyl chain and the polar hydroxyl group.

Its hydroxyl group is a hydrogen bond donor and a hydrogen bond acceptor, interacting with a large variety of polar groups of different solutes.

Despite its lipophilic character it dissolves many more organic compounds than alkanes, cycloalkanes, or aromatic solvents do.

While the aqueous phase of the *n*-octanol/water system contains nearly no octanol at equilibrium, the octanol phase dissolves an appreciable amount of water (2.3 $\text{mol} \cdot \text{l}^{-1}$, corresponding to a molar ratio of *n*-octanol/water $\approx 4/1$); therefore, polar groups need not be dehydrated on their transfer from the aqueous phase to the organic phase.

n-Octanol has a low vapor pressure, allowing reproducible measurements; on the other hand, its vapor pressure is high enough to allow its removal under mild conditions.

n-Octanol is UV-transparent over a large range, making the quantitative determination of many compounds relatively easy.

By far the most partition coefficients have been measured in *n*-octanol/water [50, 181]; also calculated partition coefficients (chapter 3.3) refer to this system.

Last but not least, many lipophilicity-activity relationships, using *n*-octanol/water partition coefficients or lipophilicity parameters derived therefrom, prove the relevance of this system (*e.g.* [18, 19, 182]).

The use of a single standard system for drug partitioning is justified by the Collander equation (eq. 20) [183], which relates partition coefficients from different solvent systems.

$$\log P_2 = a \log P_1 + c \qquad (20)$$

To give an example, slopes a = 1.04 to 1.06 and correlation coefficients r = 1.00 were obtained for the partitioning of a homologous series of 4-alkylpyridines in different organic solvent/water systems. Values c = −1.20, −1.14, −0.78, −0.50, and 0.46 resulted for the organic phases hexadecane, octane, dibutyl ether, carbon tetrachloride, and chloroform (based on *n*-octanol as reference), corresponding to a quantitative scale of solvent polarity [184]. Numerous other examples (*e.g.* Figure 5) [174, 185, 186] confirm the validity of Collander-type relationships. For some solvents different equations have been obtained for hydrogen bond donor and acceptor solutes [174]. Inherent limitations of the Collander equation have been discussed by Taylor [173].

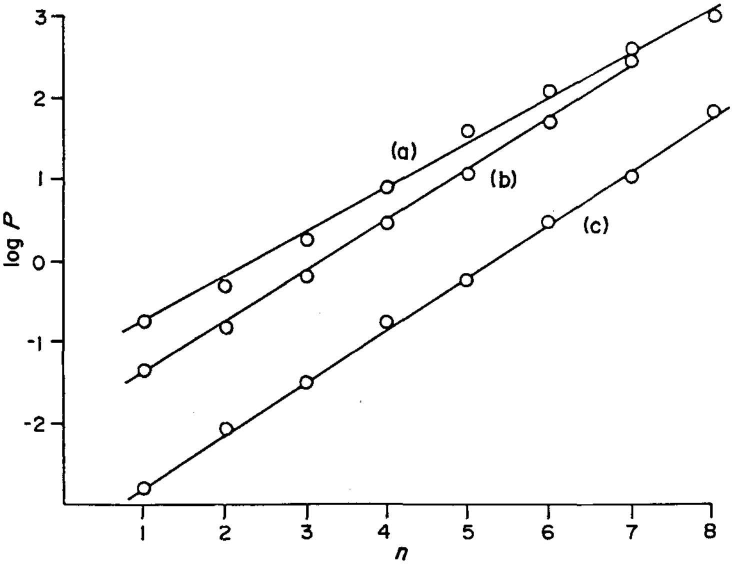


Figure 5: Log P of homologous alcohols as a function of n (number of alkyl carbon atoms): (a) n-octanol/water, slope 0.54 (lipophilicity contribution of the methylene group); (b) chloroform/water, slope 0.62; (c) cyclohexane/water, slope 0.65 (reproduced from Figure 11 of ref. [173] with permission from Pergamon Press Ltd., Headington Hill Hall, Oxford OX3 0BW, UK).

Eq. 21 correlates the partitioning of some alcohols between human erythrocyte membranes and aqueous buffer with *n*-octanol/water partition coefficients [187].

$$\log P_{\text{membrane}} = 1.003(\pm 0.13) \log P_{\text{oct}} - 0.883(\pm 0.39) \tag{21}$$

$$(n = 5;\ r = 0.998;\ s = 0.082)$$

Several hundreds of linear relationships between various kinds of (mostly nonspecific) biological data and *n*-octanol/water partition coefficients have been published (*e.g.* [18, 182]). However, the choice of *n*-octanol/water as the standard system for drug partitioning must be reconsidered in the light of some recent results. Principal component analysis of partition coefficients from different solvent systems [188 – 190] shows that lipophilicity depends on solute bulk, polar, and hydrogen-bonding effects [189]; isotropic surface areas, *i.e.* areas where no water molecules bind and hydrated surface areas, were correlated with the first and the second principal components of such an analysis [190].

The hydrogen-bonding ability I_H of different functional groups was first defined by Seiler (Table 3) [191] as the difference between cyclohexane/water and *n*-octanol/water partition coefficients (eq. 22).

$$\Delta \log P = \log P_{\text{oct}} - \log P_{\text{cyclohexane}} = \sum I_H - 0.16 \tag{22}$$

$$(n = 195;\ r = 0.967;\ s = 0.333;\ F = 107)$$

Corresponding scales can be derived, *e.g.* from log P values measured in *n*-octanol/water, heptane/water, and other systems [192], from the "water dragging

Table 3. Hydrogen-bonding ability values I_H of various substituents and substructures (eq. 22) [191]

Substituent / substructure	I_H
Aromatic -COOH	2.87
Aromatic -OH	2.60
-CONH-	2.56
$-SO_2NH-$	1.93
Aliphatic -OH	1.82
Aliphatic $-NH_2$	1.33
Aromatic $-NH_2$	1.18
$-NR_1R_2$ ($R_1, R_2 <> H$)	0.55
$-NO_2$	0.45
>C=O	0.31
-C≡N	0.23
-O-	0.11
ortho-Substitution to -OH, -COOH, $-NR_1R_2$	-0.62

effect" (the ability of a solute to carry water molecules from the aqueous phase into a polar organic solvent like *e.g.* dibutyl ether) [193, 194], and from water/gas phase equilibrium constants of different solutes [195]. The significance of hydrogen-bonding parameters in QSAR studies has been discussed [196–198] and examples for their application have been given [196, 199–203], the most interesting ones being relationships between the blood-brain barrier penetration of H_2-antihistaminic drugs and Δ log P [199, 200], the difference between *n*-octanol/water and cyclohexane/water partition coefficients (eqs. 23 and 24).

$$\log(C_{\text{brain}}/C_{\text{blood}}) = -0.604(\pm 0.17)\,\Delta\log P + 1.23(\pm 0.56) \tag{23}$$

$(n = 6;\ r = 0.980;\ s = 0.249;\ F = 98.0)$

$$\log(C_{\text{brain}}/C_{\text{blood}}) = -0.485(\pm 0.16)\,\Delta\log P + 0.889(\pm 0.50) \tag{24}$$

$(n = 20;\ r = 0.831;\ s = 0.439;\ F = 40.2)$

The use of different solvents, alkane (inert), *n*-octanol (amphiprotic), chloroform (hydrogen bond donor), and propylene glycol dipelargonate (PGDP; hydrogen bond acceptor), was proposed to model different membranes and tissues [204, 205]; a large number of PGDP/water partition coefficients have been measured. While *n*-octanol seems to be a suitable model of amphiprotic regions of a membrane, PGDP resembles more its lipid part. Therefore, PGDP should be considered as a supplementary solvent, not as a substitute for *n*-octanol. "*The proper choice of solvent system is more open now than at any time in the last 50 years*" [173].

Systematic investigations by Herbette [114, 115, 206–211], based on membrane/water partition coefficient measurements and on small-angle neutron diffrac-

Table 4. Membrane and *n*-octanol/buffer (pH = 7.2) partition coefficients [114]

Drug	Biological Membranes	*n*-Octanol/Buffer
Bay P 8857	125,000	42
Nisoldipine	6,000	38
Nimodipine	5,000	730
Propranolol	200	18
Acetylcholine	32	0.003
Timolol	16	0.7
Ethanol	3	0.6

tion experiments, show that an isotropic two-phase solvent system cannot be an appropriate model for the lipid bilayer of membranes, with their inner hydrophobic part and the outer, polar and negatively charged surface of the phospholipids. Indeed, biological membrane/buffer partition coefficients of drugs are significantly different from those measured in *n*-octanol/water systems (Table 4). Not only the partitioning of the drugs into the membrane has to be considered to understand the structure-activity relationships of receptor ligands in a quantitative manner but also their orientation relative to the binding site at the membrane-bound receptor or channel protein (Figure 1, chapter 1.2) [115].

The important role of membranes in drug action was also recognized by Seydel [212–214]. In the drug/membrane interaction, the membrane acts on the drug molecules:

The diffusion through the membrane may become the rate-limiting step.
The membrane may prevent diffusion to the active site.
The membrane may bind or accumulate drugs.
The solvation of the drug in the membrane may lead to a conformational change of its structure.

Vice versa, the drug acts on the membrane properties:

The drug may change the conformation of acyl groups.
The drug may increase the membrane surface.
The drug may change the thickness of the membrane.
The drug may change the fluidity of the membrane.
The drug may change the membrane potential and the hydration of the polar head groups.

All these effects have an important, hitherto mostly neglected influence on drug activity [214]. The differences in inhibitory activities of trimethoprim (TMP) analogs in cell-free and whole cell systems of *Escherichia coli* strains being sensitive and resistant to TMP, the interaction of amphiphilic benzylamines with artificial bilayers, the interaction of neuroleptics with bilayer membranes (measured by NMR techniques), and the reversal of multidrug resistance by amphiphilic agents could quantitatively be described in relation to these effects [214].

The role of biomembranes in mediating the receptor subtype selectivities of peptides was investigated by Schwyzer [215, 216]; eq. 25 defines the different contributions to the overall drug-membrane interaction.

$$\Delta G_{\text{total}} = \Delta G_{\text{electrostatic}} - \Delta G_{\text{transfer}} + \Delta G_{\text{immobilization}} + \Delta G_{\text{amphiphilic}} + + \Delta G_{\text{dipole}} \tag{25}$$

3.2. The Measurement of Partition Coefficients and Related Lipophilicity Parameters

The measurement of partition coefficients [40, 41, 49, 156, 173, 174, 217, 218] is not as easy as one would expect from their simple definition (eq. 19, chapter 3.1):

Practical problems arise for polar and highly lipophilic ($\log P > 4$) compounds.

Even small impurities may drastically distort the experimental values, especially if the analytical method is more sensitive to impurities.

Radioactive material is only suited if it has a high degree of radiochemical purity.

Phases must be equilibrated in advance without the solute.

After the addition of the solute and the attainment of equilibrium (after extensive shaking and waiting for 24 hours), centrifugation is necessary to separate the phases quantitatively.

Partition coefficients should be measured with as small amounts of solute as possible to avoid association phenomena in either phase; in ambiguous cases the concentration dependence of the partition coefficient should be determined.

Aqueous buffers should not contain extractable ions.

Solute concentrations should be determined in both phases and at least two independent measurements should be made.

It is desirable to dissolve the solute independently in both phases in order to prove that equilibrium is attained from both sides (surprising deviations may be obtained).

To sum up, a lot of experience is needed to determine log P values by the classical shake flask method. Alternatives have been developed and compared with each other, *e.g.* filter probe methods [219, 220], the AKUFVE method [221], and different centrifugal partition chromatographic techniques (which correspond to true partitioning because only two immiscible liquid phases and no solid support are involved) [222–225]. As the scope and limitations of most of these techniques have been reviewed [173, 217, 218, 225], they shall not be discussed here in detail.

Chromatographic parameters obtained from reversed-phase thin-layer chromatography are occasionally used as substitutes for partition coefficients [226–229]. Silica gel plates, being coated with hydrophobic phases (nowadays being commercially available), are eluated with aqueous/organic solvent systems of increasing water content. Eq. 26 converts the resulting R_f values to R_M values, which are true measures of lipophilicity.

$$R_M = \log(1/R_f - 1) \tag{26}$$

Chromatographic R_M values are at least useful as a "quick and rough" estimation of lipophilicity. If they are determined in a proper manner, *i.e.* by using several aqueous/organic solvent systems with increasing water content and by extrapolating all R_M values to 100% water (provided the extrapolated values are not too far outside the range of solvent concentrations used for their experimental determination), they are closely related to log P values [227–229]. This relationship has been confirmed by several investigations, *e.g.* of a large, chemically heterogeneous group of compounds ($n = 415$; $r = 0.933$; $s = 0.377$; $F = 2803$) [228] and by systematic studies of the correlations between R_M values and experimental log P values as well as log P values calculated by two different methods (compare chapter 3.3) [229].

The determination of R_M values offers many important advantages, as compared to the measurement of log P values:

Compounds need not be pure.
Only traces of material are needed.
A wide range of hydrophilic and lipophilic congeners can be investigated.
The measurement of practically insoluble analogs poses no problems.
No quantitative method for concentration determination is needed (the spots can be detected by any specific or nonspecific method).
Several compounds can be investigated simultaneously.

The main disadvantage is a certain lack of precision and reproducibility and the use of different organic solvent systems, which renders the derivation of π- or f-related scales (chapter 3.3) impossible. In addition, a chromatographic method is no equilibrium process (even if it results from distribution), leading to deviations, *e.g.* in the case of charged compounds [229].

Nowadays high-performance liquid chromatography (HPLC) [230–235] is the method of choice in many (especially industrial) laboratories. Log k′ values, which are calculated from t_r, the retention time of the compound, and t_0, the retention time of the solvent front (eq. 27), are closely correlated with *n*-octanol/water partition coefficients, *e.g.* by eq. 28 [231].

$$k' = (t_r - t_0)/t_0 \tag{27}$$

$$\log P = 1.025(\pm 0.06)\log k' + 0.797 \tag{28}$$

$$(n = 33;\ r = 0.987;\ s = 0.127)$$

Technical details, including solid support, coating and column filling techniques, eluents, and factors affecting the reproducibility have been reviewed [173, 235, 236]. Lipophilicity values from HPLC measurements are not on a unique scale, but the log k′ values can be converted to *n*-octanol/water partition coefficients with the help of a set of HPLC calibration standard compounds for which classical shake-flask partition coefficients are known. Experience shows: once the HPLC method is successfully established in a laboratory, it will remain the method of choice for lipophilicity determinations.

Aqueous solubility, although closely related to lipophilicity (*e.g.* for neutral liquids, log 1/S *vs.* log P: $n = 140$; $r = 0.955$; $s = 0.344$) [18], is not and should not be used as a parameter in QSAR studies for two reasons: first, for solids it is no pure measure of lipophilicity due to the intermolecular forces in the crystal (equations relating partition coefficients to solubilities indeed contain the melting point as an additional parameter), and second, its experimental measurement is even more difficult than in the case of partition coefficients. Wrong values result if equilibrium is not attained due to low dissolution rate constants. On the other hand, solubilities increase considerably with even small amounts of impurities and hypersaturation phenomena are very common.

3.3. Lipophilicity Contributions and the Calculation of Partition Coefficients

One of the most important contributions of Hansch was the discovery that partition coefficients are additive constitutive parameters, like some other molecular properties. He defined a lipophilicity parameter π (eq. 29) [15, 17, 18, 174, 218, 237, 238] in the manner that Hammett σ constants (eq. 3, chapter 1.1) were defined.

$$\pi_X = \log P_{R-X} - \log P_{R-H} \tag{29}$$

The only difference to the Hammett equation is the absence of a term like ϱ, because π values only refer to aromatic substituents and to *n*-octanol/water partition coefficients, if not stated otherwise. Slightly different π values were obtained for *meta*- and *para*-substituents and for electron donor, acceptor, and neutrally substituted benzenes. π Values from different scales can be interrelated if electronic effects are considered (*e.g.* eq. 30) [237].

$$\Delta\pi = \pi_{\text{Phenol}} - \pi_{\text{Benzene}} = 0.823\sigma + 0.061 \tag{30}$$

$(n = 24;\ r = 0.954;\ s = 0.097)$

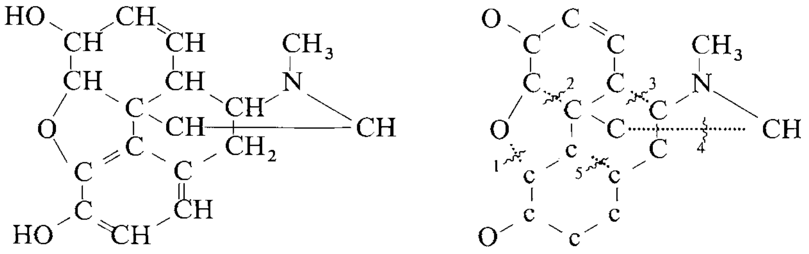
However, the differences between most π scales are not significant; nowadays the π_{benzene} values which were derived from monosubstituted benzenes are most often used.

The application of π values for lipophilicity calculation of aliphatic compounds led to significant deviations between observed and calculated values (back folding of aliphatic chains to an aromatic ring was one of the desperate attempts to support a wrong concept by a wrong hypothesis). The problem could easily be solved by Rekker, who recognized that the use of π values is inappropriate in the case of aliphatic compounds. He defined a hydrophobic fragmental constant f (eq. 31; a_i = number of occurrences of the fragment with the lipophilicity contribution f_i) [49, 56, 239–241], which is a measure of the absolute lipophilicity contribution of the corresponding substituent or group and is no longer based on the exchange of H for X, as π values are.

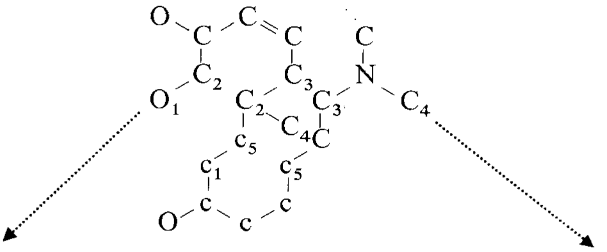
$$\log P = \sum a_i f_i \tag{31}$$

f Scales were not only derived for *n*-octanol/water but also for other solvent systems, *e.g.* alkane/water, chloroform/water, and PGDP/water systems [173, 204, 205].

Morphine: Break and number five ring closures:



Generate SMILES for resulting noncyclic structure:



O1C2C(O)C=CC3C2(C4)c5c1c(O)ccc5CC3N(C)C4

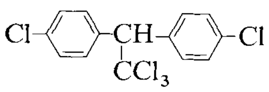
Figure 6: Generation of the SMILES code of morphine (reproduced from Figure 6 of ref. [256] with permission from Pergamon Press Ltd., Headington Hill Hall, Oxford OX3 0BW, UK).

Methylene group fragment values are even known for 24 different organic solvent/water systems [242, 243]. They regularly decrease with increasing water content of the organic phase, *i.e.* with increasing polarity of the organic solvent.

The new f system, later modified by Leo and Hansch [50, 244, 245] and others [246, 247], was a real advance because it allowed the *de novo* calculation of partition coefficients. The computer program CLOGP, developed by Chou and Jurs [177, 248] from the hydrophobic fragmental constant approach of Leo and Hansch, was later largely extended and fully computerized (*e.g.* [179, 218, 249 – 252]) to its current version [253]. The input of structures is being done in SMILES notation [253 – 256], an easy and powerful language for converting chemical structures into a computer-readable form (*e.g.* CCCC = butane, Cc1ccccc1 = toluene, c1ncccc1C(=O)O = nicotinic acid) (Figure 6); no human interaction, like the definition of certain groups or correction factors, is needed after the input of the structures.

Calculations based on the two different f scales, the one by Rekker, derived from a set of thousand compounds by statistical methods [241], and the other by Leo and Hansch, derived from a few highly accurate measurements of appropriate standard compounds [244], have been compared in their predictive ability [56, 158, 257, 258]; some step by step calculations (Tables 5 and 6) [56, 158, 257, 259] demonstrate how difficult it is to calculate log P values by hand, without the help of a computer program.

Table 5. Log P calculation of DDT (**2**); comparison of the Leo/Hansch and the Rekker method (reproduced from ref. [259] with permission from the copyright owner)

DDT, $C_{14}H_9Cl_5$ **log P_{obs} = 5.98**[a]					**(2)**
Leo/Hansch			**Rekker**		
$2\,f_{C_6H_5}$	$= 2\cdot(1.90)$	$= 3.80$	$f_{C_6H_4}$	$= 2\cdot(1.658)$	$= 3.316$
$-\,2\,f_H$	$= -2\cdot(0.23)$	$= -0.46$	f_{CH}		$= 0.337$
$2\,f_C$	$= 2\cdot(0.20)$	$= 0.40$	$f^{al}_{CCl_3}$		$= 2.060$
f_H		$= 0.23$	$2\,f^{ar}_{Cl}$	$= 2\cdot(0.924)$	$= 1.848$
$3\,f_{Cl}$	$= 3\cdot(0.06)$	$= 0.18$			
$2\,f^{\varnothing}_{Cl}$	$= 2\cdot(0.94)$	$= 1.88$		**log P_{calc}**	$=$ **7.561**
$(6 - 1)\cdot F_b$	$= 5\cdot(-0.12)$	$= -0.60$			
F_{cBr}		$= -0.13$			
$3F_{mhG3}$	$= 3\cdot(0.53)$	$= 1.59$			
	log P_{calc}	$=$ **6.89**[b]			

[a] new value: log P_{obs} = 6.36 (HPLC determination) [253]

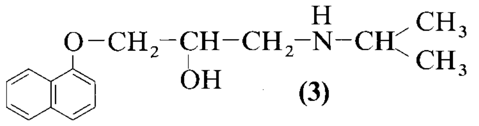
[b] new value: log P_{calc} = 6.61 (CLOGP 3.54) [253]

No general conclusions can be drawn from such comparisons. Both approaches give reliable results and both have their limitations. Although partially computerized versions of the Rekker method have been developed [259–261], nowadays CLOGP is by far the most convenient, advanced, and accurate computer program for the calculation of *n*-octanol/water partition coefficients (Table 7); together with the POMONA database (Table 8) it is part of the MEDCHEM software [253]. The database contains experimental *n*-octanol/water and other partition coefficients, experimental pK_a values, and additional valuable information on about 11,000 drugs and other standard compounds (version 3.54; 25,400 compounds in the current UNIX version 4.2) [253]. Compounds can be retrieved by various criteria; in addition, substructure, superstructure, similarity, and even 3D searches (3D structures generated by the program CONCORD, see chapter 9.3) can be performed on the structures which are included in this database.

Some complications arise in the calculation of partition coefficients of highly polar compounds and in the case of fragments not known to the program CLOGP. Possible improvements of CLOGP have been discussed [262].

Table 6. Log P calculation of propranolol (**3**); comparison of the Leo/Hansch and the Rekker method (reproduced from ref. [259] with permission from the copyright owner)

Propranolol, $C_{16}H_{21}NO_2$
log P_{obs} = 3.56, 3.18a)



(3)

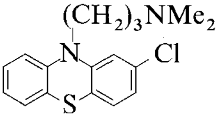
Leo/Hansch			**Rekker**		
$7\,f^{\varnothing}_{\underline{CH}}$	= 7·(0.355)	= 2.49	$f_{C_{10}H_7}$		= 3.113
$2\,f^{\varnothing}_{\underline{C}}$	= -2·(0.225)	= 0.45	$2\,f_{CH_2}$	= 2·(0.519)	= 1.038
$f^{\varnothing}_{\underline{C}}$		= 0.13	$2\,f_{CH_3}$	= 2·(0.701)	= 1.402
$f^{\varnothing}_{O}$		= -0.61	$2\,f_{CH}$	= 2·(0.337)	= 0.674
$2\,f_{CH2}$	= 2·(0.66)	= 1.32	f^{ar}_{O}		= -0.439
$2\,f_{CH}$	= 2·(0.43)	= 0.86	f^{al}_{OH}		= -1.470
f_{OH}		= -1.64	f^{al}_{NH}		= -1.814
f_{NH}		= -2.15	$2 \times 2C_M$	= 4·(0.289)	= 1.156
$2\,f_{CH3}$	= 2·(0.89)	= 1.78			
$(9 - 1)\cdot F_b$	= 8·(-0.12)	= -0.96		**log P_{calc}**	= **3.660**c)
$F_{P2}(O/OH)$	= -0.26·(-0.61 - 1.64)	= 0.59			
$F_{P2}(OH/NH)$	= -0.26·(-1.64 - 2.15)	= 0.99			
F_{gBr}		= -0.22			
F_{cBr}		= -0.13			
	log P_{calc}	= **2.90**b)			

a) actual value: log P_{obs} = 3.56 [253]
b) new value: log P_{calc} = 2.75 (CLOGP 3.54) [253]
c) revised value: P_{calc} = 3.46 [56]

Nonadditivities of *n*-octanol/water partition coefficients [173] limit the applicability of the CLOGP program. There is experimental evidence that some "polar" compounds may not be as polar as expected from their chemical structures, due to a network of intramolecular hydrogen bonds [263].

A modified version of the Rekker method, which has to be further evaluated, was recently published [56]. Some other atom-, bond- and group-based calculation procedures (*e.g.* [264–271]) have been developed; most of them are reviewed and critically commented in ref. [173]. The calculation of hydrophobic fields from lipophilicity contributions is discussed in chapter 9.2.

Table 7. Log P calculation of chlorpromazine (**4**) by the program CLOGP [253]. Input of the structure in SMILES code: c12cc(Cl)ccc2Sc3ccccc3N1CCCN(C)C; unique SMILES code, generated by the program CN(C)CCCN2c1ccccc1Sc3ccc(Cl)cc23 (reproduced from ref. [253] with permission from the copyright owner).



(4)

```
NAME: CHLORPROMAZINE

SMILES:      CN(C)CCCN2c1ccccc1Sc3ccc(C1)cc23
ISOC-ID:     A--A-AAA--a-aaaaa--a-aaa----aa--
FRAG-ID:     _1_______2__________3______4______
H-COUNT:     3__3_222_____1111_____11_____1___
RING 1:      ___________________________a_aaa____aa__
RING 2:      ___________A_A____A_AA__________A_
RING 3:      _______________a_aaaaa____________________
```

Class	Type	Log(P) Contribution Description	Comment	Value
FRAGMENT	# 1	Tertiary Amine (ZW+)	MEASURED	−2.180
FRAGMENT	# 2	Tertiary Amine (ZW+)	MEASURED	−0.390
FRAGMENT	# 3	Sulfide	MEASURED	−0.770
FRAGMENT	# 4	Chloride	MEASURED	−0.940
ISOLATING	CARBON	5 Aliphatic isolating carbon(s)		0.975
ISOLATING	CARBON	12 Aromatic isolating carbon(s)		1.560
EXFRAGMENT	HYDROG	19 Hydrogen(s) on isolating carbons		4.313
EXFRAGMENT	BONDS	5 chain and 0 alicyclic (net)		−0.600
FRAGBRANCH	FRAG 1	5 net bonds (out of 6) count.		−0.400
ELECTRONIC	SIGRHO	2 Potential interactions; 1.50 used	Within Ring	0.213
RESULT	v3.4	All fragments measured	CLOGP	5.201

React: list of electronically active fragments

Frag No.	Ring No.	type	sigma	rho
4	1	attached	0.280	0.000
3	1	attached	0.000	0.300
2	1	attached	0.000	0.610
2	3	attached	0.000	0.610
3	3	attached	0.000	0.300

Pact: list potential electronic activity

frag	ring	sigma	frag	ring	rho	dist	value
4	1	0.280	3	1	0.300	0	1.000
4	1	0.280	2	1	0.610	0	1.000
4	1	0.280	2	3	0.610	−1	0.000
4	1	0.280	3	3	0.300	−1	0.000

Elect: Details of sigma-rho corrections used. 2 possible interactions allocated as follows:

Sigma frag	drop	Rho frag	drop	Potential Corr.	Net Corr.
4	1.0000	2	1.0000	0.1708	0.1708
4	0.5000	3	1.0000	0.0840	0.0420

Table 8. The POMONA database [253]. Part of the information given on chlorpromazine (**4**); entries # 1–11 and 29–33, out of 67 entries on chlorpromazine and additional 14 entries on different salts of chlorpromazine (reproduced from ref. [253] with permission from the copyright owner)

```
  >>>> THOR "POMONA89" database TYPE ALL <<<<
+------------------------------------------------------------+
| SMILES      CN(C)CCCN2c1ccccc1Sc3ccc(Cl)cc23              1|
+------------------------------------------------------------+
| CMR         9.3765                                        2|
| ERROR LEV   High confidence CMR estimate                   |
| VERSION     3.54                                           |
+------------------------------------------------------------+
| CLOGP       5.201                                         3|
| ERROR LEV   All fragments measured                         |
| VERSION     3.54                                           |
+------------------------------------------------------------+
| TIMESTAMP   1988 Aug 25 14:28:51                          4|
+------------------------------------------------------------+
| MOLFORM     C17H19CL1N2S1                                 5|
+------------------------------------------------------------+
| WLN         T C666 BN ISJ B3N1&1                          6|
|   +--------------------------------------------------------+
|   | ACTIVITY TYPE  ANTI-EMETIC, ANTIPSYCHOTIC             7|
|   +--------------------------------------------------------+
|   | LOCAL NAME  CHLORPROMAZINE                            8|
|   +--------------------------------------------------------+
|   | LOGP        2.04                                      9|
|   | SOLV PAIR   Octanol                                    |
|   | REFERENCE   Ahmed,A., et.al., Pharm.Res., 119 (1985)   |
|   | pH          6.0                                        |
|   +--------------------------------------------------------+
|   | LOGP        5.25                                     10|
|   | SOLV PAIR   Octanol                                    |
|   | REFERENCE   Clarke, F., J. Pharm. Sci., 73, 226 (1984) |
|   +--------------------------------------------------------+
|   | PKA         9.20                                     11|
|   | REFERENCE   Clarke, F., J. Pharm. Sci., 73, 226 (1984) |
|   +--------------------------------------------------------+
      ...         ...
      ...         ...
|   +--------------------------------------------------------+
|   | LOGP        5.35                                     29|
|   | SOLV PAIR   Octanol                                    |
|   | REFERENCE   Azzaro, M., Lab. Of Physical Org. Chem.,   |
|   |             U. Of Nice, France, Private Comm.          |
|   | FOOTNOTE    Ion-corrected                              |
|   | ...  2      Phosphate buffer                           |
|   | pH          7.4                                        |
|   +--------------------------------------------------------+
|   | PKA         9.30                                     30|
|   | REFERENCE   Azzaro, M., Lab. Of Physical Org. Chem.,   |
|   |             U. Of Nice, France, Private Comm.          |
|   +--------------------------------------------------------+
|   | LOGP        5.32                                     31|
|   | SOLV PAIR   Octanol                                    |
|   | REFERENCE   Church, C. & Hansch, C., Pomona College,   |
|   |             Unpublished Analysis                       |
|   +--------------------------------------------------------+
|   | LOGP        5.35                                     32|
|   | SOLV PAIR   Octanol                                    |
|   | REFERENCE   Anderson, S. & Hansch, C., Pomona College, |
|   |             Unpublished Analysis                       |
|   | SELECTED    *                                          |
|   +--------------------------------------------------------+
|   | LOGPSTAR    5.35                                     33|
|   +--------------------------------------------------------+
      ...         ...
      ...         ...
```

From a theoretical point of view it might be worthwhile to correlate lipophilicity with principal components of other molecular properties, *e.g.* the BC(DEF) parameters (eq. 32) [164, 165]. However, these parameters are available only for a very limited number of organic compounds.

$$\log P = 3.65(\pm 0.08)\,B - 7.66(\pm 0.11)\,C - 5.74(\pm 0.37)\,D - 0.31(\pm 0.71)\,E + 5.09(\pm 0.90)\,F + 1.604(\pm 0.014) \quad (32)$$

$(n = 114;\ r = 0.998;\ s = 0.08)$

Many attempts were made to explain lipophilicity by related properties, *e.g.* by solubility, solvent-accessible surface, and charge distributions calculated from semiempirical methods ([190, 272–284] and references cited therein); while some of the results allow a better understanding of the intrinsic nature of lipophilicity, none of these alternative approaches have led to a reliable log P prediction system so far.

3.4. Polarizability Parameters

Molar volume MV, molar refractivity MR, and parachor PA are theoretically and practically closely interrelated parameters (eqs. 33–35; MW = molecular weight, ϱ = density, n = refractive index; γ = surface tension) [285]. Another related parameter is the molar polarization P_M (dielectric constant ε instead of the n^2 term in eq. 34) [286].

$$MV = MW/\varrho \quad (33)$$

$$MR = MV \cdot \frac{n^2 - 1}{n^2 + 2} \quad (34)$$

$$PA = MV \cdot \gamma^{1/4} \quad (35)$$

Molar volume itself is not strictly additive, but the corrected volume parameters MR and PA are additive constitutive molecular properties, like log P and the Hammett σ parameter. While molar refractivity has attracted much attention [50, 286–288], molar volume [55, 289] and parachor [50, 290] have only rarely been used in QSAR studies.

MR still is the chameleon [91] amongst the physicochemical parameters, despite its broad application in QSAR studies. It has been correlated with lipophilicity, molar volume, and steric bulk. Of course, due to its MW/ϱ component it is indeed related to volume and size of a substituent. But two recent statements that MR is only based on these properties [91, 291] cannot be accepted. The refractive index-related correction term in MR accounts for the polarizability and thus for the size and the polarity of a certain group [158, 173, 286]. The larger the polar part of a molecule is, the larger its MR value will be. Even hydrophobic substituents have a weak enthalpic interaction due to dispersion forces, in addition to their entropic contribution to ΔG, coming from the displacement of ordered water molecules. For hydrophobic substituents a close interrelation between volume, surface, lipophilicity, and MR is to be expected, which breaks if polar substituents are included. No correlation could be obtained between MR and the steric parameter E_s^c ($n = 142$; $r = 0.67$; $s = 12.82$) [286]. Since the refractive index n varies only slightly for most organic compounds, molar volume (MV) is usually highly interrelated with MR.

In the above-mentioned parameter collection of 59 different substituents [158], MR is correlated with

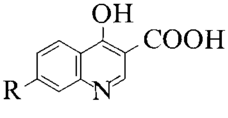
van der Waals volume ($r = 0.97$),
parachor ($r = 0.92$),
the bulk parameter B of the BC(DEF) components ($r = 0.84$),
aromatic ($r = 0.90$) and aliphatic ($r = 0.91$) fragmental volume constants, and
the connectivity parameters ${}^0\chi_{ar}^v$ ($r = 0.95$) and ${}^1\chi_{ar}^v$ ($r = 0.95$) (chapter 3.7),

but not with

different lipophilicity parameters ($r = 0.53-0.62$),
the steric parameters E_s ($r = 0.56$) and E_s^c ($r = 0.36$), and
Charton's υ parameter ($r = 0.58$) (chapter 3.6) [158].

The significance of molar refractivity terms in QSAR equations of some ligand-enzyme interactions could be interpreted with the help of 3D structures. These investigations show that substituents modeled by MR bind in polar areas, while substituents modeled by π bind in hydrophobic space [288, 292]. Correspondingly, a positive sign of MR in a QSAR equation can be explained by binding of the substituents to a polar surface, while a negative sign or a nonlinear relationship indicates a limited area or steric hindrance at this binding site. Most often the MR values are scaled by a factor of 0.1 to achieve reasonable values of the regression coefficients of the resulting QSAR equations.

The different nature of MR, as compared to hydrophobic and steric properties, can only be detected in cases where a proper selection of substituents allows this. One such example is the inhibition of malate dehydrogenase by 4-hydroxyquinoline-3-carboxylic acids (**5**), where the interaction of the ligands with the enzyme is described better by MR (eq. 36) than by π (pI_{50} vs. π: $n = 13$; $r = 0.604$; $s = 0.716$), and the respiration inhibition of ascites tumor cells, where the transport into or the accumulation in the cells is more appropriately described by π (eq. 37; same set of compounds) than by MR (pI_{50} vs. MR: $n = 14$; $r = 0.699$; $s = 0.554$) (Table 9) [293].



(**5**)

$$pI_{50}(\mathrm{MDH}) = 0.70(\pm 0.17)\,\mathrm{MR} + 2.29(\pm 0.30) \qquad (36)$$
$$(n = 13;\ r = 0.939;\ s = 0.315)$$

$$pI_{50}(\mathrm{ascites}) = 0.46(\pm 0.11)\,\pi + 3.22(\pm 0.16) \qquad (37)$$
$$(n = 14;\ r = 0.933;\ s = 0.280)$$

Table 9. 7-Substituted 4-hydroxyquinoline-3-carboxylic acids (**5**) as inhibitors of cellular respiration [293]

Substituent R	**pI_{50}, Malate Dehydogenase**	**pI_{50}, Ascites**	π[a]	**MR**[b]
H	-	2.98	0.00	0.103
Cl	2.44	3.84	0.55	0.603
F	1.98	3.30	0.06	0.092
OCH_3	-	3.28	0.49	0.787
$COCH_3$	3.04	3.10	-0.39	1.118
$N(CH_3)_2$	3.32	3.33	1.10	1.555
$OCH_2C_6H_5$	4.49	4.41	1.81	3.174
$OCH_2C_6H_3(3,4\text{-}Cl_2)$	5.32	4.82	3.23	4.174
NO_2	2.72	3.24	-0.40	0.736
$CONH_2$	3.13	2.24	-1.18	0.981
COOH	2.97	2.24	-2.80	0.605
SO_2CH_3	3.18	2.75	-1.39	1.349
OH	3.31	3.04	0.06	0.285
SO_2NH_2	3.02	2.47	-1.36	1.228
SO_3^-	2.67	2.88[c]	-4.76	0.971

[a] calculated from measured log P values
[b] MR values are scaled by a factor of 0.1
[c] not included in the calculation

Atomic molar refractivity contributions have been defined [267–269]. The MEDCHEM software [253] also contains a routine for the calculation of MR values.

3.5. Electronic Parameters

Electronic properties of molecules [12, 40–43, 53, 57, 158, 294–297] can be described by a wide variety of different parameters, *e.g.* by

Hammett σ constants,
field and resonance parameters $\mathscr{F}$ and $\mathscr{R}$,
pK_a values,
parameters derived from molecular spectroscopy,
charge transfer constants,
dipole moments,
hydrogen-bonding parameters, and
parameters derived from quantum-chemical calculations, *e.g.* orbital energies and partial charges.

As all these parameters describe the influence of a certain group or substituent on electron density distribution, all have been used in QSAR studies. In contrast to global molecular properties, like lipophilicity and molar refractivity, they normally refer to a certain atom or group (the only exception being dipole moments). To give an example: if σ_p is used to describe the influence of *para*-substituents on the acidity of a phenolic group, this term models the ability of the phenolic group to interact as a hydrogen bond donor or acceptor with a (most often unknown) acceptor or donor atom at the binding site.

The use of σ constants in QSAR studies started with the early finding of Hansen [13] that the toxicities of substituted benzoic acids against mosquito larvae are correlated with Hammett σ values (Table 10); eq. 38 was derived after the *p*-nitro analog had been excluded (all analogs: n = 14; r = 0.711; s = 0.427).

$$\log 1/C = 1.454(\pm 0.42)\,\sigma + 1.787(\pm 0.17) \tag{38}$$

$$(n = 13;\ r = 0.918;\ s = 0.243;\ F = 58.91)$$

Indeed, all analogs can be described much better by using π instead (eq. 39) [15]. The log 1/C *vs.* σ relationship in eq. 38 only exists by fortune. It results from a close relationship between π and σ, if the nitro analog is excluded (n = 13; π *vs.* σ: r = 0.91); all equations were recalculated by using the π_{benzene}, σ_m, and σ_p values presented in Table 10.

Table 10. Toxicity of substituted benzoic acids *vs.* mosquito larvae [13, 15]

Substituent R	π[a]	σ	**log 1/C obsd.**	**log 1/C calc.**[b]
3,4,5-Tri-I	3.36	0.88	3.54	3.40
3,5-Di-I	2.24	0.70	2.85	2.80
4-I	1.12	0.18	2.31	2.20
3,4-Di-Cl	1.42	0.60	2.28	2.36
4-Cl	0.71	0.23	2.06	1.98
4-Br	0.86	0.23	2.03	2.06
3-Cl	0.71	0.37	2.00	1.98
3,4-CH=CH-CH=CH-	1.27	0.17	1.92	2.28
4-F	0.15	0.06	1.85	1.68
4-CH_3	0.56	-0.17	1.66	1.90
H	0.00	0.00	1.64	1.60
4-OCH_3	-0.02	-0.27	1.60	1.59
4-NO_2	-0.28	0.78	1.52	1.45
4-OH	-0.67	-0.37	1.29	1.24

[a] π_{benzene} values
[b] eq. 39

$$\log 1/C = 0.535(\pm 0.09)\,\pi + 1.602(\pm 0.11) \tag{39}$$

$$(n = 14;\ r = 0.969;\ s = 0.151;\ F = 181.88)$$

In considering electronic effects, one has to differentiate between field (inductive) effects and resonance effects. Due to the characteristic features of a benzene ring, σ_m mainly describes the inductive effect while σ_p stands for a combination of both effects, with the resonance effect predominating. Over the past decades many different σ scales were developed in organic chemistry, besides σ_m and σ_p also σ^+ (to account for substituents which donate electrons to the aromatic ring system by direct resonance interaction), σ^- (for corresponding acceptor substituents), σ^0 and σ^n (normal or unexalted σ constants), σ_I and σ_R (inductive and resonance contributions), *etc.*

In 1968 Swain and Lupton [298] tried to stop the proliferation of σ scales. They defined field and resonance components $\mathcal{F}$ and $\mathcal{R}$, by assuming that any set of σ values can be expressed by a weighted combination $a\mathcal{F} + b\mathcal{R}$, that there is no resonance contribution in the case of 4-substituted bicyclo[2.2.2]octane-carboxylic acids ($b = 0$), and that there is no resonance contribution of a $N^+(CH_3)_3$ substituent ($\mathcal{R} = 0$). They were able to correlate 43 different electronic parameters with linear combinations of these two parameters (many r values being larger than 0.98), *e.g.* eqs. 40 and 41.

$$\sigma_m = 0.60(\pm 0.00)\,\mathcal{F} + 0.27(\pm 0.00)\,\mathcal{R} + 0.00(\pm 0.00) \tag{40}$$

$$(n = 42;\ r = 1.00;\ s = 0.00)$$

$$\sigma_p = 0.56(\pm 0.00)\,\mathcal{F} + 1.00(\pm 0.00)\,\mathcal{R} + 0.00(\pm 0.00) \tag{41}$$

$$(n = 42;\ r = 1.00;\ s = 0.00)$$

Hansch redefined $\mathcal{F}$ and $\mathcal{R}$ values in a more consistent manner (eqs. 42 and 43); the $\mathcal{F}$ values in eq. 42 are now in the right scale, as compared to the $\mathcal{R}$ values [50, 151].

$$\mathcal{F} = 1.369(\pm 0.19)\,\sigma_m - 0.373(\pm 0.14)\,\sigma_p - 0.009(\pm 0.04) \tag{42}$$

$$(n = 14;\ r = 0.992;\ s = 0.042)$$

$$\mathcal{R} = \sigma_p - 0.921\mathcal{F} \tag{43}$$

There has been considerable discussion *pro* [299, 300] and *contra* [301–305] the validity of the underlying assumptions. Despite this (more philosophical) question, the separation of σ values into inductive and resonance effects seems to be justified, at least from a practical point of view.

A recent compilation [306] contains σ_m, σ_p, and the redefined $\mathcal{F}$ and $\mathcal{R}$ values of 530 substituents, together with σ_p^+, σ_p^-, $\mathcal{R}^+$ values (eq. 44, $\mathcal{R}^+$ values of π-electron donor substituents; eq. 45, $\mathcal{R}^+$ values of π-electron acceptor substituents), and $\mathcal{R}^-$ values (eq. 46, $\mathcal{R}^-$ values of π-electron donor substituents) of 223 substituents as well as some other electronic parameters derived from spectroscopic data; as compared to normal $\mathcal{R}$ values, the $\mathcal{R}^+$ and $\mathcal{R}^-$ values of π-electron donor substituents are significantly enhanced (eqs. 44 and 46).

$$\mathcal{R}^{+} = 1.90(\pm 0.26)\,\mathcal{R} - 0.07(\pm 0.11) \tag{44}$$
$$(n = 29;\ r = 0.945;\ s = 0.170)$$

$$\mathcal{R}^{+} = 1.16(\pm 0.45)\,\mathcal{R} - 0.01(\pm 0.07) \tag{45}$$
$$(n = 16;\ r = 0.829;\ s = 0.060)$$

$$\mathcal{R}^{-} = 1.93(\pm 0.59)\,\mathcal{R} + 0.19(\pm 0.11) \tag{46}$$
$$(n = 34;\ r = 0.758;\ s = 0.173)$$

In addition, optimized and normalized S values (field-inductive σ bond perturbation) and P values (resonance π bond perturbation; orthogonal to S) have been defined and used in QSAR studies [307]. The factorization of σ values into different field and resonance contributions is reviewed in ref. [306].

Ortho-substituents pose special problems because their σ values include a steric contribution; furthermore, as compared to *meta*- and *para*-substituents, many *ortho*-substituents cause conformational changes, sometimes being favorable for binding, sometimes being very unfavorable; although the problem of *ortho*-substituent parametrization has been discussed in detail (*e.g.* [41, 296]), it is (and will remain) a difficult problem in QSAR studies.

Aliphatic σ constants (σ^* values) are defined by eq. 47, where log k and log k_0 are the rate constants of acid- (subscript A) and base-catalyzed (subscript B) hydrolysis of RCOOR′ and CH_3COOR', respectively [308].

$$\sigma^* = \frac{\log(k/k_O)_B - \log(k/k_O)_A}{2.48} \tag{47}$$

pK_a Values [53, 295] reflect electronic properties in a direct manner, due to the definition of Hammett σ values relative to the ionization constants of benzoic acids ($\varrho = 1.00$); thus, they may be used as substitutes for σ values. However, their use is inadequate if they are taken to describe the relative amounts of the unionized form of different drugs (see chapter 4.5). One of the rare exceptions to this rule is *e.g.* eq. 48, which describes the antibacterial activities of sulfonamides *vs. Escherichia coli* [309].

$$\begin{aligned}\log 1/C = {} & 1.044(\pm 0.13)\,pK_a - 1.640(\pm 0.18)\log(\beta \cdot 10^{pK_a} + 1) + \\ & + 0.275(\pm 0.65)\end{aligned} \tag{48}$$
$$\log \beta = -5.96 \qquad pK_a \text{ optimum} = 6.22$$
$$(n = 39;\ r = 0.956;\ s = 0.275;\ F = 124.1)$$

Parameters derived from molecular spectroscopy, *e.g.* from infrared or NMR measurements, have relatively early been used in QSAR studies (*e.g.* [158, 310–312]). They are extremely valuable in describing the electronic influence of substituents in systems for which no σ values are known, which applies to most heterocyclic compounds.

Charge-transfer constants C_T [313] and κ [314, 315] were derived from association constants of different analogs with tetracyanoethylene and 1,3,5-trinitrobenzene, respectively. Although charge-transfer interactions are supposed to be important for the interaction of some drugs with their binding sites (*e.g.* in DNA intercalation) [116], they have not attracted much attention in QSAR studies. Values are known for only few substituents but they are definitely interrelated with other electronic parameters.

The use of dipole moments μ [296, 316] in QSAR studies was proposed by Lien [316–318]; group dipole moments of 311 aromatic [317] and 214 aliphatic substituents [318] are tabulated in literature. A certain problem of the use of dipole moments is their directionality. Vector addition has to be used to estimate the dipole moment of compounds having more than one site of substitution.

Hydrogen-bonding parameters have, in part, been discussed in chapter 3.1. In addition to Seiler's I_H values [191] several other scales were derived, *e.g.* by Hansch and Leo (discriminating donor, acceptor, and neutral substituents) [50, 319] and by Taft (pK_{HB} values, derived from the complex formation of bases with *p*-fluorophenol in CCl_4) [320, 321]; the subject has been reviewed in [296].

In contrast to the enormous importance of hydrogen bonds in drug-receptor interactions, only a few QSAR analyses have been performed by using such parameters. This may reflect the fact that hydrogen-bonding ability is related to some other electronic parameters, at least within structurally related series. A more pessimistic but, nevertheless, plausible explanation might be that the choice of parameters in QSAR studies mainly depends on their availability and ready-to-use tabulation in handbooks. It seems that a lot of further work has to be done with respect to hydrogen bonds in drug-receptor interactions and their proper, *i.e.* quantitative consideration in QSAR studies.

Quantum-chemical parameters [297] are a matter of their own. Their frequent abuse in the early, heroic times of QSAR discredited them for long time. Quantum-chemical calculations yield so many different values, *e.g.* net atomic charges, charge densities, superdelocalizabilities, electrostatic potentials, values for inductive, resonance, and polarizability effects, as well as HOMO and LUMO energies, *etc.*, that their uncritical combination, as done in these early days, often ended in chance correlations. Nowadays some more appropriate applications demonstrate their real value for QSAR studies [322–328], *e.g.* eq. 49 (correlating biodegradability, expressed as BOD = biological oxygen demand, with an atomic charge difference parameter, $\Delta\,|\delta|_{X-Y}$) [324] and eq. 50 (mutagenicity of various aromatic compounds; TA98 = Ames test results, using *Salmonella typhimurium*, strain TA98) [328]. The advantage of ε_{LUMO} (= lowest unoccupied molecular orbital) in eq. 50, as compared to classical σ values, results from the possibility to include various heteroaromatic compounds for which no σ values are known.

$$\%\,BOD = 1015\,\Delta\,|\delta|_{X-Y} + 1.523 \qquad (49)$$

$(n = 197;\ r = 0.991;\ s = 3.822)$

$$\log TA98 = 0.65(\pm 0.16)\log P - 2.90(\pm 0.59)\log(\beta P + 1) - 1.38(\pm 0.25)\,\varepsilon_{LUMO} + 1.88(\pm 0.39)\,I_1 - 2.89(\pm 0.81)\,I_a - 4.15(\pm 0.58) \quad (50)$$

$\log \beta = -5.48$ optimum $\log P = 4.93$

($n = 188$; $r = 0.900$; $s = 0.886$)

Classical and magnetic aromaticities of heterocycles have been defined and correlated with biological activities [329].

Rapid calculation procedures of electronic effects in organic molecules were proposed by Gasteiger and Marsili [330–332] as an alternative to *ab initio* and semiempirical quantum-chemical calculations, which are relatively time-consuming procedures.

3.6. Steric Parameters

Steric effects are difficult to describe, due to the fact that the 3D structures of the binding sites of drugs are most often unknown. Several reviews of steric effects and steric descriptors [50, 52, 287, 308, 333–336] contribute to this problem but a general solution seems to be inherently impossible.

E_s constants, as defined by eq. 51 (acid-catalyzed hydrolysis of RCOOR′ *vs.* CH_3COOR') [308], were the first parameters which have been used to describe steric effects in QSAR studies.

$$E_s = \log(k/k_0)_A \quad (51)$$

Van der Waals radii of symmetrical substituents followed, due to the relationship given in eq. 52 ($r_{v(av)}$ = average van der Waals radii of the substituents) [50].

$$E_s = -1.839 r_{v(av)} + 3.484 \quad (52)$$

($n = 6$; $r = 0.996$; $s = 0.132$)

Charton [337–339] defined a steric substituent parameter υ (r_v = minimum van der Waals radius of a substituent) (eq. 53), which is highly correlated with E_s values (eq. 54) [50].

$$\upsilon_X = r_{vX} - r_{vH} = r_{vX} - 1.20 \quad (53)$$

$$E_s = -2.062(\pm 0.86)\,\upsilon - 0.194(\pm 0.10) \quad (54)$$

($n = 104$; $r = 0.978$; $s = 0.250$)

Hancock [340] modified E_s to E_s^c values by correcting them for the number of α-hydrogen atoms, n_H (eq. 55).

$$E_s^c = E_s + 0.306(n_H - 3) \quad (55)$$

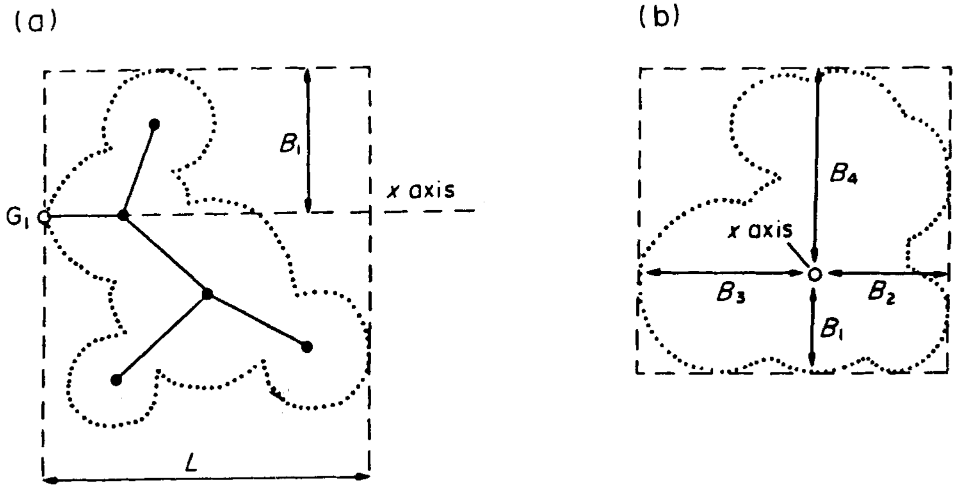


Figure 7: STERIMOL parameters L and $B_1 - B_4$. Projections of a substituent (a) along and (b) perpendicular to the axis of substitution (x axis) [335]; according to the definition of Skagerberg *et al.* [169], $B_i = B_1$, $B_{ii} = B_4$, $B_{iii} = B_2$, and $B_{iv} = B_3$ (reproduced from Figure 12 of ref. [287] with permission from Pergamon Press Ltd., Headington Hill Hall, Oxford OX3 0BW, UK).

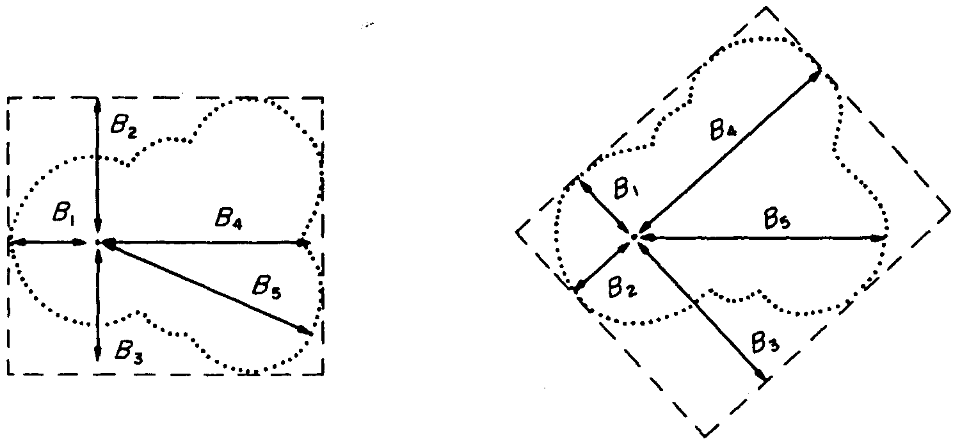


Figure 8: Revised STERIMOL parameters L, B_1, and B_5 of the OCH_3 group. B_1 is the minimum width, B_5 the maximum width of the substituent [336, 343]. Different possibilities result from the definition of the parameters $B_2 - B_4$, depending on the selection of B_1 (reproduced from Figure 13 of ref. [287] with permission from Pergamon Press Ltd., Headington Hill Hall, Oxford OX3 0BW, UK).

Fujita expressed E_s^c values of substituents of the type $CR^1R^2R^3$ as a weighted sum of the individual E_s^c values of R^1, R^2, and R^3 [341] to overcome problems arising from the use of steric parameters of unsymmetrical substituents. A simple steric parameter S_b was formulated on the basis of substituent branching [342].

Many other parameters related to size, *e.g.* van der Waals volumes, molar volumes, solvent-accessible surface, molar refractivity, *etc.* have been used to describe steric effects in QSAR equations.

A real progress resulted from the definition of the STERIMOL parameters L, B_1, B_2, B_3, and B_4 [335, 336]. L is defined as the length of a substituent along the axis of its substitution to the parent skeleton; the width parameters B are all orthogonal to L and have angles of 90 degrees to each other. They are arranged in such a manner that B_1 has the smallest value and B_4 the largest one; in most cases B_4 is close to the maximum width of the substituent (Figure 7). A slightly different definition was used by Skagerberg *et al.* (Table 2, chapter 3) [169]: B_i still is the smallest value orthogonal to L but B_{ii} is opposite to B_i; B_{iii} and B_{iv} are orthogonal to B_i and B_{ii} and arranged in such a manner that $B_{iii} < B_{iv}$.

Due to some criticism [50] and due to a number of chance correlations in the QSAR literature, arising from uncritical combinations of linear and squared terms (*i.e.* 10 possible values for one site of substitution and many more for several different sites), Verloop reduced the width parameters to B_1, being the smallest width, and B_5, now being the largest width orthogonal to L, but independent of the angle between B_1 and B_5 (Figure 8) [336, 343]. The use of the STERIMOL parameters in QSAR studies has been reviewed [287, 335, 336, 343].

However, as said before, even the most elegant description of steric bulk of a substituent cannot cope with the problem of the unknown geometry of the binding site and its steric constraints. The quantitative description of steric effects is a puzzling problem even in the case of known 3D structures of the protein, due to

different conformations of different ligands,
small (or even larger) differences in the binding modes of different analogs (see chapter 9.3), and
variations in the positions of side chains and even backbone atoms of the protein in different ligand-protein complexes.

Compilations of different steric parameters are given in refs. [158, 287, 335, 336].

3.7. Other Parameters

An important group of parameters in QSAR studies are terms derived from 2D and 3D structures of drugs, which are not directly related to lipophilic, electronic, or steric properties.

A molecular weight term was used by Lien [344] to improve the fit of parabolic Hansch equations, *e.g.* for the antifungal activities of homologous aliphatic amines *vs. Rhinocladium beurmanni* (log 1/C *vs.* $(\log P)^2$, log P, and log MW: $n = 15$; $r = 0.994$; $s = 0.161$). However, in this case the MW term only accounts for systematic deviations between the parabolic model and the experimental data, due to a close interrelation between log P and log MW ($r = 0.985$). A better and more consistent result is obtained if the bilinear model (chapter 4.4) is used instead, without the log MW term ($n = 15$; $r = 1.000$; $s = 0.031$) [345].

A more appropriate application of a molecular weight parameter has been demonstrated in a QSAR study of the multidrug resistance CR of tumor cells [346], where the MW term stands for the dependence of CR values on diffusion rate constants. The relationship between MW and volume implies that $\sqrt[3]{MW}$, corresponding to a linear dimension of size, *i.e.* being directly related to diffusion rate constants [347], should be better suited than log MW ($n = 29$; $r = 0.871$; $s = 0.394$) [346], which indeed is the case (eq. 56) [348].

$$\log CR = 0.70(\pm 0.24)\sqrt[3]{MW} - 1.01(\pm 0.32)\log(\beta \cdot 10^{\sqrt[3]{MW}} + 1) - $$
$$- 0.10(\pm 0.05)\log P + 0.38(\pm 0.28)\,I - 3.08(\pm 1.42) \qquad (56)$$

$\log \beta = -6.851$ optimum $\sqrt[3]{MW} = 7.21$

($n = 40$; $r = 0.891$; $s = 0.344$)

No general survey can be given of the use of geometric parameters in QSAR studies; such parameters are derived from known or hypothetical pharmacophores and therefore they only apply to certain sets of compounds. Much too often conformational preferences are unknown and cannot adequately be considered in QSAR studies. An impressive example, how this can be done, is given below. Eq. 57 describes the calcium channel affinities of a series of cyclic lactones (**6**) (Figure 9) which are structurally closely related to nifedipine. Due to their conformational restrictions these lactones have different torsion angles between the two rings which should adopt a 90° angle in the biologically active conformation (KI = displacement of ^{3}H-nimodipine, $\Delta\alpha$ = deviation from 90° angle) (Figure 10) [349].

$$\log KI = 0.067(\pm 0.017)\,\Delta\alpha + 0.19(\pm 0.34) \qquad (57)$$

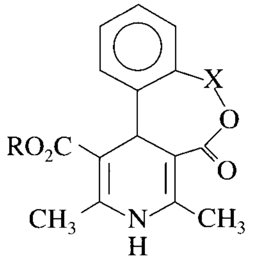
($n = 7$; $r = 0.88$; $F = 16.5$)

Conformational energy and entropy values have been proposed and used as parameters in QSAR studies [350–353].

The hyperstructure approaches MSD and MTD (minimal steric difference and minimal topological difference) and molecular shape analysis (MSA) are discussed in chapter 4.6, the use of similarity indices is discussed in chapter 9.4.

A large (unfortunately much too large) group of parameters in QSAR studies are topological indices, *e.g.* connectivity values χ (Figure 11), based on the characterization of chemical structures by graph theory. Since the first publications on the use of molecular connectivity values in correlation analysis and QSAR studies appeared [354–367], a huge number of related papers followed, most of them by Kier and Hall (for reviews see [48, 158, 175, 178, 287, 368–371]).

Molecular connectivity indices χ are calculated from molecular formulas in a unique manner and, due to their mathematical definition, some physicochemical properties of branched and unbranched isomers can be described with high accuracy



(6)

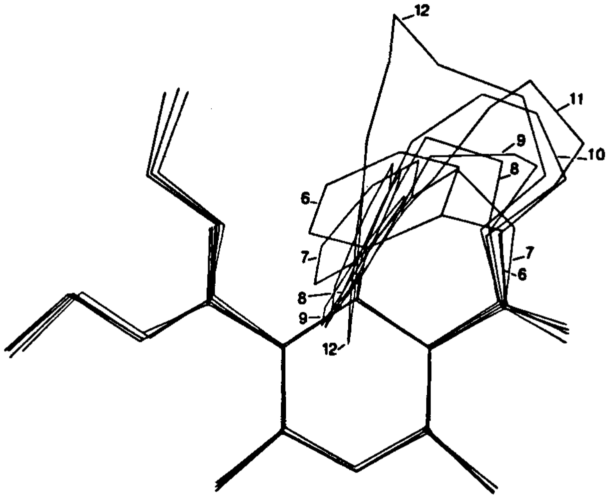


Figure 9: Superimposed X-ray structures of dihydropyridine lactones with six- to twelve-membered rings (**6**; $X = -(CH_2)_n-$, $n = 0-6$) (reproduced from Figure 5 of ref. [349] with permission from the copyright owner).

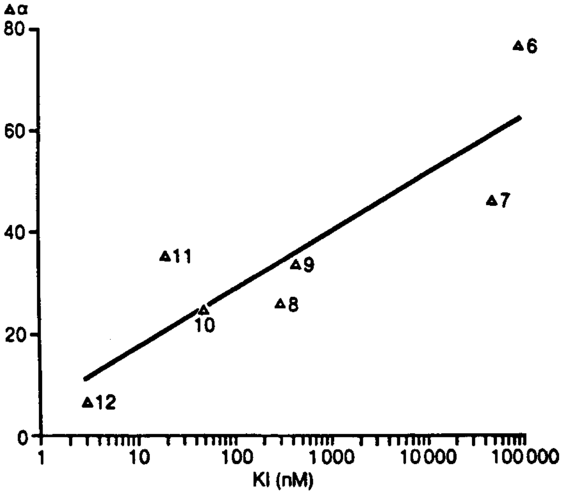
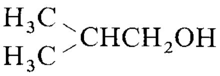
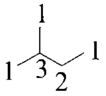
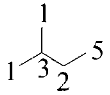


Figure 10: Calcium antagonism of the dihydropyridine lactones (**6**); plot of the angular deviation Δα *vs.* KI (reproduced from Figure 6 of ref. [349] with permission from the copyright owner).



$(H_3C)_2CHCH_2OH$ =



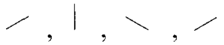


iso-butanol (Atom connectivity) (Valence connectivity)

0th order: $^0\chi = \sum (\delta_i)^{-1/2}$

$$^0\chi = 3(1)^{-1/2} + (3)^{-1/2} + (2)^{-1/2} = 4.284$$
$$^0\chi^v = 2(1)^{-1/2} + (3)^{-1/2} + (2)^{-1/2} + (5)^{-1/2} = 3.732$$

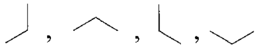
1st order:



$^1\chi = \sum (\delta_i\delta_j)^{-1/2}$

$$^1\chi = 2(1 \cdot 3)^{-1/2} + (2 \cdot 3)^{-1/2} + (2 \cdot 1)^{-1/2} = 2.270$$
$$^1\chi^v = 2(1 \cdot 3)^{-1/2} + (2 \cdot 3)^{-1/2} + (2 \cdot 5)^{-1/2} = 1.879$$

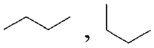
2nd order:



$^2\chi = \sum (\delta_i\delta_j\delta_k)^{-1/2}$

$$^2\chi = (1 \cdot 3 \cdot 1)^{-1/2} + 3(1 \cdot 3 \cdot 2)^{-1/2} = 1.802$$
$$^2\chi^v = (1 \cdot 3 \cdot 1)^{-1/2} + 2(1 \cdot 3 \cdot 2)^{-1/2} + (3 \cdot 2 \cdot 5)^{-1/2} = 1.576$$

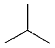
3rd order path:



$^3\chi_P = \sum (\delta_i\delta_j\delta_k\delta_l)_P^{-1/2}$

$$^3\chi_P = 2(1 \cdot 3 \cdot 2 \cdot 1)^{-1/2} = 0.816$$
$$^3\chi_P^v = 2(1 \cdot 3 \cdot 2 \cdot 5)^{-1/2} = 0.365$$

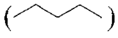
3rd order cluster:



$^3\chi_C = \sum (\delta_i\delta_j\delta_k\delta_l)_C^{-1/2}$

$$^3\chi_C = (1 \cdot 1 \cdot 3 \cdot 2)^{-1/2} = 0.408$$
$$^3\chi_C^v = (1 \cdot 1 \cdot 3 \cdot 2)^{-1/2} = 0.408$$

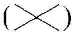
4th order path:



$^4\chi_P = \sum (\delta_i\delta_j\delta_k\delta_l\delta_m)_P^{-1/2}$

not present

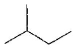
4th order cluster:



$^4\chi_C = \sum (\delta_i\delta_j\delta_k\delta_l\delta_m)_C^{-1/2}$

not present

4th order path/cluster:



$^4\chi_{PC} = \sum (\delta_i\delta_j\delta_k\delta_l\delta_m)_{PC}^{-1/2}$

$$^4\chi_{PC} = (1 \cdot 1 \cdot 3 \cdot 2 \cdot 1)^{-1/2} = 0.408$$
$$^4\chi_{PC}^v = (1 \cdot 1 \cdot 3 \cdot 2 \cdot 5)^{-1/2} = 0.183$$

Figure 11: Calculation of various molecular connectivity indices of *iso*-butanol (reproduced from ref. [175] with permission from Birkhäuser Verlag AG, Basel, Switzerland).

[48]. It is this relationship to physicochemical properties like partition coefficients, molar refractivities, and steric properties that allows a quantitative description of (most often nonspecific) biological activities within chemically closely related series.

In addition to normal and valence connectivity [372, 373] values, many other topological indices have been defined (for reviews see [52, 158, 287, 369, 371]).

Some more topological indices have been created in the last few years:

A molecular shape index κ [371, 374–376].

A fragment approach to structure-activity correlations, based on the characterization of fragments by suitably weighted paths [377].

A flexibility index, calculated from molecular shape descriptors [378, 379].

Linearly transformed χ values [380].

The concept of topological state and the total topological index [381].

A redefinition of the Wiener index [382], according to the contributions of different bond types [383].

Topology-based electronic descriptors [384].

An electrotopological state index [385].

A differential molecular connectivity index [386].

Needless to say that every new parameter and new approach is praised to be superior to the older ones. However, nobody should be surprised if some biological activities can fortuitously be explained by any combination of so many different parameters.

The use of topological indices in QSAR studies has been criticized (*e.g.* [158, 175, 173, 287, 387]). "Good QSAR practice" is violated in many papers:

In contrast to general recommendations on the selection of biologically meaningful parameters [307], the physicochemical meaning of the topological parameters is never clear.

Chance correlations may arise from the uncritical combination of a large number of closely interrelated connectivity terms, including normal and valence connectivity values, higher order path and cluster connectivity terms, and squared terms.

In many cases standard deviations of regression coefficients are given instead of confidence intervals, suggesting that terms are significant which in reality are not significant.

Although Taylor's comments on topological indices are generally favorable, with respect to their use in QSAR studies he formulates [173]: *"Calculation is simple and their use has become widespread. There has been some tendency, therefore, to use molecular connectivity in contexts for which it is unsuitable, and unfortunately, this has been encouraged by Kier and Hall themselves"* and *"regard molecular connectivity as an irrelevance which has had the unfortunate effect of diverting attention from the real work that needs doing"*. Nothing can be added to this criticism. To whom it may concern.

3.8. Indicator Variables

Indicator variables (sometimes called dummy variables or *de novo* constants) are used in linear multiple regression analysis [388, 389] to account for certain features which cannot be described by continuous variables [21, 22, 41, 390, 391]. In QSAR equations they normally describe a certain structural element, be it a substituent or another molecular fragment; thus, Free Wilson analysis may be interpreted as a regression analysis approach using only indicator variables [21, 22, 390, 391]. At a first glance, indicator variables do not appear to be linear free energy-related parameters. Indeed, they are (chapter 4.3), but most often the relationships to linear free energy terms remain unknown. Numerical values of activity contributions to biological activity values can be derived from a linear free energy relationship (*i.e.* a linear Hansch equation) for each substituent of a group of compounds. The values are numerically equivalent to the group contributions of a corresponding Free Wilson analysis, if both analyses are based on hydrogen as reference substituent in every position of substitution (compare chapter 4.3) [390–394].

Indicator variables have also been used to account for other structural features, *e.g.* for intramolecular hydrogen bonds, hydrogen bond donor and acceptor properties, *ortho* effects, cis/trans isomerism, different parent skeletons, different test models, *etc.* [22, 390]. Some precautions are necessary: indicator variables should not describe a single compound (in this case the corresponding group contribution includes the experimental error of this one biological activity value) and the use of indicator variables should be justified from a theoretical point of view; otherwise, continuous variables will be mixed with meaningless dummy variables, just to fit the data.

The proper use of an indicator variable to combine a group of different but related analogs can be demonstrated with two subsets of papain ligands. Eq. 58 was derived for a series of N-mesylglycine phenyl esters (7, R = SO_2CH_3) and eq. 59 for a corresponding series of N-benzoylglycine phenyl esters (7, R = COC_6H_5) (Table 11) [395].

$$\log 1/K_m = 0.529(\pm 0.23)\,MR + 0.370(\pm 0.20)\,\sigma + 1.877(\pm 0.13) \tag{58}$$

(n = 13; r = 0.935; s = 0.105; F = 34.51)

$$\log 1/K_m = 0.771(\pm 0.67)\,MR + 0.728(\pm 0.37)\,\sigma + 3.623(\pm 0.34) \tag{59}$$

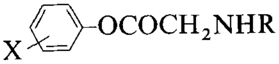
(n = 7; r = 0.971; s = 0.148; F = 32.85)

While the coefficients of the MR and σ terms are not significantly different (as indicated by an overlap of their confidence intervals), the constant terms are. The combination of eqs. 58 and 59 with the help of an indicator variable I (I = 1 for mesylamides; I = 0 for benzamides) [395] leads to eq. 60 (recalculated).

$$\log 1/K_m = 0.569(\pm 0.26)\,MR + 0.561(\pm 0.19)\,\sigma - 1.922(\pm 0.15)\,I + {} + 3.743(\pm 0.17) \tag{60}$$

(n = 20; r = 0.990; s = 0.148; F = 272.04)

Table 11. Binding of glycine X-phenyl ester amides (7) to papain [395]



(7)

X	R	log $1/K_m$	MR	σ	I
4-OH	SO_2Me	2.05	0.28	-0.37	1
4-OMe	SO_2Me	2.13	0.79	-0.27	1
4-Me	SO_2Me	2.08	0.56	-0.17	1
3-Me	SO_2Me	2.23	0.56	-0.07	1
H	SO_2Me	1.79	0.10	0.00	1
4-F	SO_2Me	1.95	0.09	0.06	1
3-OMe	SO_2Me	2.29	0.79	0.12	1
4-CHO	SO_2Me	2.33	0.69	0.42	1
4-Cl	SO_2Me	2.38	0.60	0.23	1
3-F	SO_2Me	1.98	0.09	0.34	1
4-COMe	SO_2Me	2.57	1.12	0.50	1
3-NO_2	SO_2Me	2.53	0.74	0.71	1
4-NO_2	SO_2Me	2.71	0.74	0.78	1
4-NH_2	COC_6H_5	3.58	0.54	-0.66	0
4-Me	COC_6H_5	4.02	0.56	-0.17	0
H	COC_6H_5	3.77	0.10	0.00	0
4-Cl	COC_6H_5	4.00	0.60	0.23	0
4-F	COC_6H_5	3.69	0.09	0.06	0
3-NO_2	COC_6H_5	4.74	0.74	0.71	0
4-NO_2	COC_6H_5	4.85	0.74	0.78	0

The value of the indicator variable I corresponds to the biological activity contribution of a mesylamido group, based on the benzamido group as the reference substituent. While not too much information can be derived from this value, there is no other way to combine eqs. 58 and 59 to one equation. The large increase in the value of the correlation coefficient r (from 0.935 and 0.971, respectively, to 0.990) results from the fact that the overall variance of the data increases by combining both subsets, while the unexplained variance remains constant, as can be seen from a comparison of the standard deviations s of all equations (compare chapter 5.1).

Indicator variables are especially useful in the early phases of a QSAR analysis and for large, complex data sets. Different subsets can be combined with their help, until the real dependence of biological activity values on some physicochemical parameters can be derived from a more extensive structural variation.

QSAR: Hansch Analysis and Related Approaches
by Hugo Kubinyi
© VCH Verlagsgesellschaft mbH, 1993

4. Quantitative Models

4.1. The Extrathermodynamic Approach (Hansch Analysis)

Hansch analysis [14, 15, 17, 18, 40–44, 396] correlates biological activity values with physicochemical properties by linear, linear multiple, or nonlinear regression analysis; thus, Hansch analysis is indeed a property-property relationship model. As practically all parameters used in Hansch analysis are linear free energy-related values (*i.e.* derived from rate or equilibrium constants), the terms "linear free energy-related approach" or "extrathermodynamic approach" [396] are sometimes used as synonyms for Hansch analysis. Also the biological activity values are, if they are properly defined, linear free energy-related values (*e.g.* binding or inhibition constants, absorption and distribution rate constants, or complex data which correspond to a weighted combination of several such unit processes).

Early attempts to correlate biological activity values with lipophilicity, expressed *e.g.* by solubility or partition coefficients [3–7], only explained nonspecific structure-activity relationships; the application of the concept of a general biological Hammett equation (eq. 4, chapter 1.1) failed. The methodological breakthrough came from a suggestion by Fujita, at that time working in Hansch's group [17], to apply an approach used earlier by Taft [308]. Like he, Hansch and Fujita combined different physicochemical parameters in one equation, *e.g.* eq. 61 (C = molar concentration that produces a certain biological effect; k_1, k_2, k_3 = coefficients determined by a least squares procedure, *e.g.* linear multiple regression analysis, to fit the biological data).

$$\log 1/C = k_1 \log P + k_2\sigma + k_3 \qquad (61)$$

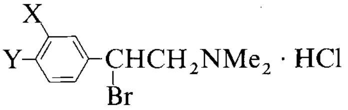
For *in vivo* data eq. 61 was extended to eq. 62 by including a parabolic lipophilicity term. The idea behind eq. 62 was that molecules which are too hydrophilic or too lipophilic will not be able to cross lipophilic or hydrophilic barriers, respectively. Therefore, they will have a lower probability to arrive at the receptor site than molecules with intermediate lipophilicity, being readily soluble in aqueous phases as well as in lipid phases.

$$\log 1/C = -k_1(\log P)^2 + k_2 \log P + k_3\sigma + k_4 \qquad (62)$$

Eqs. 61 and 62 are only two examples of an enormous wide variety of Hansch equations. Later steric parameters were added to this general model and even later molar refractivity values. With such multiparameter equations it was possible to describe much more complex dependences of biological activities on physicochemical properties than before. In the last three decades nearly all conceivable combinations of lipophilic, polarizability, electronic, and steric parameters, with and without additional indicator variables, have been used and correlated with biological activity values in linear, parabolic, and bilinear equations.

Only one illustrative example is given here to describe and explain the proper application of Hansch analysis (for further examples see chapter 7). Graham and Karrar [397] determined the antiadrenergic activities of a series of α-bromophenethylamines (**8**). Hansch and Lien [398] derived eq. 63, which was at this time considered to give the best quantitative description of the data (Table 12; only some

Table 12. Antiadrenergic activities of *meta*- and *para*-substituted N,N-dimethyl-α-bromo-phenethylamines (**8**) [307, 393, 397]



(8)

meta (X)	*para* (Y)	π	σ^+	E_s^{meta}	**log 1/C obsd.**	**log 1/C calc.**[a]	**log 1/C calc.**[b]
H	H	0.00	0.00	1.24	7.46	7.82	7.88
H	F	0.15	-0.07	1.24	8.16	8.09	8.17
H	Cl	0.70	0.11	1.24	8.68	8.46	8.60
H	Br	1.02	0.15	1.24	8.89	8.77	8.94
H	I	1.26	0.14	1.24	9.25	9.06	9.26
H	Me	0.52	-0.31	1.24	9.30	8.87	8.98
F	H	0.13	0.35	0.78	7.52	7.45	7.43
Cl	H	0.76	0.40	0.27	8.16	8.11	8.05
Br	H	0.94	0.41	0.08	8.30	8.30	8.22
I	H	1.15	0.36	-0.16	8.40	8.61	8.51
Me	H	0.51	-0.07	0.00	8.46	8.51	8.36
Cl	F	0.91	0.33	0.27	8.19	8.38	8.34
Br	F	1.09	0.34	0.08	8.57	8.57	8.51
Me	F	0.66	-0.14	0.00	8.82	8.78	8.65
Cl	Cl	1.46	0.51	0.27	8.89	8.75	8.77
Br	Cl	1.64	0.52	0.08	8.92	8.94	8.94
Me	Cl	1.21	0.04	0.00	8.96	9.15	9.08
Cl	Br	1.78	0.55	0.27	9.00	9.06	9.11
Br	Br	1.96	0.56	0.08	9.35	9.25	9.29
Me	Br	1.53	0.08	0.00	9.22	9.46	9.43
Me	Me	1.03	-0.38	0.00	9.30	9.56	9.47
Br	Me	1.46	0.10	0.08	9.52	9.35	9.33

[a] eq. 65 [b] eq. 68

selected physicochemical parameters are included); slightly different results are obtained if different scales, *e.g.* π values from the phenoxyacetic acid system, π_{benzene} values, or any other π values are used.

$$\log 1/C = 1.22\pi - 1.59\sigma + 7.89 \tag{63}$$
$$(n = 22;\ r = 0.918;\ s = 0.238)$$

Later Cammarata [399] presented eq. 64 (recalculated) which describes *meta*-substituents by their π and σ values and *para*-substituents by a steric parameter r_V^{para}. However, the steric parameter has the wrong sign; the positive value of its regression coefficient implicates that steric bulk increases biological activity, which cannot be true. In addition, it is difficult to understand how an electronic effect can only be obtained for the *meta*-substituents.

$$\log 1/C = 0.747(\pm 0.26)\,\pi_m - 0.911(\pm 0.52)\,\sigma_m + {} + 1.666(\pm 0.26)\,r_V^{para} + 5.769(\pm 0.45) \tag{64}$$
$$(n = 22;\ r = 0.962;\ s = 0.168;\ F = 74.01)$$

Correspondingly, eq. 64 was criticized by Unger and Hansch [307] in a noteworthy paper, which constitutes a milestone in the development of Hansch analysis. They formulated rules for the derivation of extrathermodynamic equations which are summarized here because of their general validity (supplementary comments are given in parentheses):

Selection of independent variables. A wide range of different parameters, like log P or π, σ, MR, and steric parameters, should be tried; MO parameters and indicator variables should not be overlooked (the problem of testing too many parameters will be discussed below). The parameters selected for the "best equation" should be essentially independent (*i.e.* the intercorrelation coefficients r should not be larger than 0.6–0.7; exceptions are combinations of linear and squared terms, like $(\log P)^2$ and log P, which are usually highly interrelated, with r values >0.9).

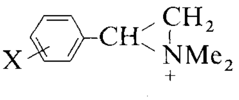
Justification of the choice of independent variables. All "reasonable" parameters must be validated by an appropriate statistical procedure, *e.g.* by stepwise regression analysis (nowadays sometimes cross-validation is recommended, see chapters 5.2 and 5.3). The "best equation" is normally the one with the lowest standard deviation, all terms being significant (indicated by the 95% confidence intervals or by a sequential F test). Alternatively, the equation with the highest overall F value may be selected as the "best" one (see chapter 5.1).

Principle of parsimony (Occam's Razor; William of Ockham, 1285–1349/50, English philosopher and logician). All things being (approximately) equal, one should accept the simplest model.

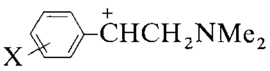
Number of terms. One should have at least five to six data points per variable to avoid chance correlations (this rule only applies to data sets of intermediate size; for small data sets more parameters may be allowed if they are based on a reasonable model; for large data sets, *e.g.* $n > 30$, this recommendation leads to equations which include too many variables).

Qualitative model. It is important to have a qualitative model which is consistent with the known physical-organic and biomedicinal chemistry of the process under consideration.

Following these recommendations, Unger and Hansch reconsidered the mechanism of action of the α-bromophenethylamines and argued that σ^+ instead of σ might be a better electronic descriptor, because the compounds are supposed to interact with the receptor site *via* the formation of an ethyleneiminium ion (**9**) and subsequently a benzyl cation (**10**). They derived eq. 65 (recalculated values) [307], which gave a much better description of the data than eq. 63.



(9)



(10)

$$\log 1/C = 1.151(\pm 0.19)\,\pi - 1.464(\pm 0.38)\,\sigma^+ + 7.817(\pm 0.19) \tag{65}$$

$(n = 22;\ r = 0.945;\ s = 0.196;\ F = 78.63)$

Out of many other equations derived by Unger and Hansch [307], separating hydrophobic and electronic effects for *meta*- and *para*-substituents and splitting electronic effects into inductive and resonance contributions S and P, only eqs. 66 and 67 shall be discussed here.

$$\log 1/C = 0.83(\pm 0.27)\,\pi_m + 1.33(\pm 0.20)\,\pi_p - 0.92(\pm 0.50)\,\sigma_m^+ - 1.89(\pm 0.57)\,\sigma_p^+ + 7.80(\pm 0.17) \tag{66}$$

$(n = 22;\ r = 0.966;\ s = 0.164)$

$$\log 1/C = 0.86(\pm 0.30)\,\pi + 0.47(\pm 0.26)\,S_m - 0.36(\pm 0.21)\,S_p - 0.92(\pm 0.61)\,P_p + 0.62(\pm 0.49)\,r_v^{para} + 7.08(\pm 0.62) \tag{67}$$

$(n = 22;\ r = 0.967;\ s = 0.167)$

Later eq. 68 was derived from the assumption that hydrophobic and electronic influences are identical in the *meta*- and *para*-positions, but that there might be a steric hindrance to the *meta*-substituents (the E_s scale is opposite in sign to the r_v scale, having smaller and even negative values for large substituents; thus, the positive sign of the coefficient in eq. 68 indeed indicates a negative steric influence of large substituents) [393].

$$\log 1/C = 1.259(\pm 0.19)\,\pi - 1.460(\pm 0.34)\,\sigma^+ + 0.208(\pm 0.17)\,E_s^{meta} + 7.619(\pm 0.24) \tag{68}$$

$(n = 22;\ r = 0.959;\ s = 0.173;\ F = 69.24)$

Eqs. 63–68 reveal a typical dilemma in Hansch analysis: while eqs. 65–68 are significantly better than eq. 63 and are based on more reasonable assumptions than eq. 64, which one of them is the "best" equation? On the basis of the correlation coefficients r (the crutches of a QSAR beginner), eq. 67 is to be preferred; eq. 66 seems to be the best one if the standard deviation s, a much better criterion, is considered. The differences in the correlation coefficients r and in the standard deviations s of eqs. 65–68 are rather small. However, if one applies the principle of parsimony, eqs. 66 and 67 should be omitted because too many parameters are included for such a small data set.

How to decide between eqs. 65 and 68? The confidence interval of the E_s term in eq. 68 is rather large and the overall F value of eq. 68 is smaller than the one of eq. 65, indicating that there may be too few degrees of freedom to favor eq. 68. On the other hand, the additional E_s term in eq. 68 is justified by a sequential F test (see chapter 5.1).

With the evidence on hand it is impossible to differentiate between both equations on a rational basis and to prefer either eq. 65 or 68. The equations are derived from a less well designed group of compounds; with the exception of hydrogen and the methyl group, all other substituents are halogens. A validation of all these equations can only be achieved after synthesis and testing of additional compounds with larger *meta*-substituents, hydrophilic, and/or electron donor substituents (the importance of a proper selection of substituents in Hansch analysis is discussed in chapter 6).

In addition to the ambiguity of the results there is another serious problem in Hansch analysis. One compound of the original series, the *para*-phenyl analog, was not included in all the analyses described above, because of its bad fit. This procedure is the only choice if a single compound represents a largely different substituent which could only be described by a separate parameter. But most often the elimination of data points is an arbitrary, subjective, and therefore dangerous procedure. An important effect may be overlooked or a false hypothesis may incorrectly be justified, starting from a wrong selection of so-called "outliers".

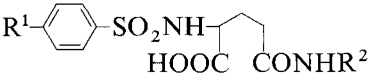
The biological activity values of other analogs, having different substituents in the positions for which the equations were derived, can be predicted from Hansch equations. While this is easy from a mathematical point of view (the parameter values of the new substituents have to be inserted in the corresponding Hansch equation), such predictions are indeed a rather difficult task. Of course, the predicted values depend on the underlying equation; different predictions result from different equations, *e.g.* for large *meta*-substituents from eqs. 65 and 68. If a point to be predicted is far outside the included parameter range (once compared by Hansch with predictions on Paris, derived from a map of London) there is not only a risk, but certainly a guarantee for failure. In the case of the phenethylamines it might happen that hydroxy analogs are no longer antagonists but agonists instead.

Often predictive ability is considered to be a criterion for the relevance of quantitative structure-activity analyses. While this is an obviously reasonable demand, it should be realized that the main purpose of Hansch analysis is a better understanding, not prediction. New hypotheses can be established from quantitative analyses, which are proven or disproven by synthesis and testing of new analogs. If the predicted values are close to the experimental ones, the model can be accepted.

Otherwise, the hypothesis was wrong or limited to a certain parameter range; new conclusions and new models must be derived.

Topliss was the first to investigate the risk of chance correlations in Hansch analyses in a systematic manner. Several simulations, using random numbers instead of real parameter values, revealed that for a given number of compounds the chance of obtaining correlation coefficients r larger than 0.9 not only drastically increases with the number of variables included in the equation, but also with the number of variables from which the different combinations are being selected [400, 401].

Probably the most terrific example of an overprediction of biological data, resulting from the use of too many variables, was published in 1986 (!). Several hundred equations were derived for the antitumor activities (a type of biological activity that usually is associated with standard deviations much larger than 0.3) of eight glutamic acid derivatives (**11**, $R^1 = H, CH_3$; $R^2 = H$, alkyl, cycloalkyl) [402]. Only one frightening example, eq. 69 (%ITW = % inhibition of tumor weight), is cited here; six variables for eight compounds leave one degree of freedom, just enough not to receive an error message by the computer program. In addition, the biological activity values are defined in an improper manner, the range of the log %ITW values is only 0.24 log units (!), the signs of the log P and (log P)^2 terms are wrong (inverse parabola), the coefficients of the σ terms are unreasonably large, no confidence intervals of the regression coefficients are included, and too many decimal places are given.



(11)

$$\log \% \text{ITW} = 1.4111(\log P)^2 - 0.5971 \log P - 0.1714\pi^2_{\text{Ali}} - \\ - 3.2293\sigma_{\text{Ali}} + 0.9595E_{s\,\text{Ali}} - 6.6199\sigma_I + 1.3249 \qquad (69)$$

$(n = 8; r = 0.9912; s = 0.0310)$

The factors being important for the significance and validity of QSAR relationships (compare chapter 5.2) have been reviewed in several publications [403–409].

4.2. The Additivity Model (Free Wilson Analysis)

The Free Wilson approach [16, 20, 390, 391] is a true structure-activity relationship model. An indicator variable is generated for each structural feature that deviates from an arbitrarily chosen reference compound. Values 1, indicating the presence of a certain substituent or structural feature, and 0, indicating its absence, are correlated with the biological activity values by linear multiple regression analysis. The resulting regression coefficients of the indicator variables are the biological activity contributions of the corresponding structural elements. "Mathematical model", "additivity model", or "*de novo* approach" are synonyms for the Free Wilson method.

The Free Wilson model was in its original formulation [16] not as simple. No reference compound was selected and so-called symmetry equations were generated to avoid the problem of linear dependences between the variables.

The version described by Fujita and Ban (eq. 8, chapter 1.1) [20, 390, 391] is a straightforward application of the additivity concept of group contributions to biological activity values. As nowadays only this modification is used, no details of the original formulation of the Free Wilson model and its complicated symmetry equations are discussed here.

In comparison with the classical version of the Free Wilson analysis, the Fujita Ban variant offers a number of important advantages:

The table for regression analysis can easily be generated.

The addition and elimination of compounds is simple and does not change the values of the other regression coefficients significantly.

Any compound may be chosen as the reference compound.

Substituents which always occur together in two different positions of the molecule can be combined to a pseudosubstituent.

Singularity problems are easily avoided.

The values of the group contributions are directly related to Hansch analysis-derived group contributions (see below, chapter 4.3).

Fujita Ban-type indicator variables can be combined with Hansch analysis to a mixed approach (see chapters 3.8 and 4.3) [22, 390, 391].

If the Free Wilson method is applied to the α-bromophenethylamines (Table 12; chapter 4.1) and if the unsubstituted parent compound (**12**, X = Y = H) is selected as the reference compound, Table 13 results; the regression analysis of these data leads to eq. 70 [390, 391, 393].

$$\begin{aligned}\log 1/C = {} & -0.301(\pm 0.50)\,[m\text{-F}] + 0.207(\pm 0.29)\,[m\text{-Cl}] + \\ & + 0.434(\pm 0.27)\,[m\text{-Br}] + 0.579(\pm 0.50)\,[m\text{-I}] + \\ & + 0.454(\pm 0.27)\,[m\text{-Me}] + 0.340(\pm 0.30)\,[p\text{-F}] + \\ & + 0.768(\pm 0.30)\,[p\text{-Cl}] + 1.020(\pm 0.30)\,[p\text{-Br}] + \\ & + 1.429(\pm 0.50)\,[p\text{-I}] + 1.256(\pm 0.33)\,[p\text{-Me}] + 7.821(\pm 0.27) \qquad (70)\end{aligned}$$

$(n = 22;\ r = 0.969;\ s = 0.194;\ F = 16.99)$

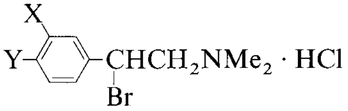
Different regression coefficients are obtained if any other compound is chosen as the reference compound or if the classical Free Wilson model is applied. However, these values are only linearly shifted to the values of eq. 70; all statistical parameters are identical, with the only exception of the 95% confidence intervals [390, 391, 410].

The Free Wilson model is easy to apply. Especially in the early phases of structure-activity analyses it is a simple method to derive substituent contributions and to have a first look on their possible dependence on different physicochemical properties.

However the Free Wilson model also has some shortcomings:

First of all and most important, structural variation is necessary in at least two different positions of substitution; otherwise, meaningless group contributions would result, one for each compound.

Table 13. Antiadrenergic activities of *meta-* and *para*-substituted N,N-dimethyl-α-bromo-phenethylamines (**12**); table for Free Wilson analysis [391, 393]


(12)

meta	*para*	*meta-*					*para-*					**log 1/C**	**log 1/C**
(X)	**(Y)**	**F**	**Cl**	**Br**	**I**	**Me**	**F**	**Cl**	**Br**	**I**	**Me**	**obsd.**	**calc.[a]**
H	H											7.46	7.82
H	F						1					8.16	8.16
H	Cl							1				8.68	8.59
H	Br								1			8.89	8.84
H	I									1		9.25	9.25
H	Me										1	9.30	9.08
F	H	1										7.52	7.52
Cl	H		1									8.16	8.03
Br	H			1								8.30	8.26
I	H				1							8.40	8.40
Me	H					1						8.46	8.28
Cl	F		1				1					8.19	8.37
Br	F			1			1					8.57	8.60
Me	F					1	1					8.82	8.62
Cl	Cl		1					1				8.89	8.80
Br	Cl			1				1				8.92	9.02
Me	Cl					1		1				8.96	9.04
Cl	Br		1						1			9.00	9.05
Br	Br			1					1			9.35	9.28
Me	Br					1			1			9.22	9.30
Me	Me					1					1	9.30	9.53
Br	Me			1							1	9.52	9.51

[a] eq. 70

Every substituent which only once occurs in the data set, leads to a single-point determination; the corresponding group contribution contains the whole experimental error of this one biological activity value.

Only a common activity contribution can be derived for substituents which always occur together in two different positions of the molecule.

In most cases a large number of parameters is needed to describe relatively few compounds, sometimes leading to equations which are statistically not significant (10 parameters are needed in the case of eq. 70 to describe 22 data points, leaving only 11 degrees of freedom).

Only a small number of new analogs can be predicted from a Free Wilson analysis (14 other compounds in the case of eq. 70).

Predictions for substituents which are not included in the analysis are generally impossible; estimated values may be derived from working hypotheses regarding a physicochemical interpretation of the obtained group contributions.

By its definition the Free Wilson analysis is limited to linear additive structure-activity relationships (its application to nonlinear relationships and the combination with Hansch analysis to a mixed approach are described in chapter 4.3).

A detailed discussion of the scope and limitations of the Free Wilson model is given in refs. [390, 391]; some applications are discussed in chapter 8.

The values of the confidence intervals have different relevance in Hansch analysis and in the Free Wilson model [390, 391, 410]. While in Hansch analysis they are a measure of the significance of the corresponding variable, in the Free Wilson analysis they depend on the choice of the reference substituents; they only indicate whether two group contributions (one being *e.g.* the group contribution of the reference substituent, which is arbitrarily set to zero) are significantly different or not. Correspondingly, "nonsignificant" terms are usually not eliminated in presenting the results of a Free Wilson analysis (compare eq. 70).

The use of stepwise principal component regression analysis in the Free Wilson approach has been proposed [411]. Other modifications which consider and include only significant terms [*e.g.* 412, 413], like the "reduced Free Wilson model" [414–416] and the BEL-FREE method [417], have the advantage of a larger number of degrees of freedom and therefore most often a higher statistical significance of the results is obtained. However, these modifications have not generally been accepted. In addition, the elimination of parameters depends on the selection of the reference substituents.

The DARC-PELCO approach [418–423] is a simple application of a hyperstructure concept to the Free Wilson method; while the approach may be appropriate for extremely large data sets, *e.g.* for the derivation of lipophilicity contributions from partition coefficients, it is useless for most structure-activity analyses, due to the much too large number of variables (compare eqs. 199 and 200, chapter 8) [390, 391]. The results from Hansch, Free Wilson, and DARC-PELCO analyses have been compared with each other [421, 422, 424]; no advantages of the latter approach could be seen.

Other hyperstructure approaches, which to some extent are related to the Free Wilson model, are discussed in chapter 4.6.

4.3. The Relationships between Hansch and Free Wilson Analysis (The Mixed Approach)

Hansch analysis and the Free Wilson method differ in their application, but they are nevertheless closely related [390, 391, 394]. From the general formulation of a linear Hansch equation (eq. 71; Φ_i is any physicochemical property) group contributions a_i can be derived for each substituent under consideration (eq. 72; Φ_{ij} is the physicochemical property j of the substituent X_i).

$$\log 1/C = k_1\Phi_1 + k_2\Phi_2 + \ldots + k_n\Phi_n + c = \Sigma k_j\Phi_j + c \tag{71}$$

$$a_i = \Sigma k_j\Phi_{ij} \tag{72}$$

If all physicochemical properties are normalized to $\Phi_j = 0$ for hydrogen and if the Free Wilson group contributions also refer to hydrogen as reference substituent, the values of corresponding activity contributions are (within experimental error) identical [393]. To give an example, eqs. 73 and 74 (the E_s term in eq. 74 only applies to *meta*-substituents) can be derived from the Hansch equations 65 and 68 (chapter 4.1) [390, 391, 393]; the resulting a_i values are numerically equivalent to the corresponding Free-Wilson group contributions (eq. 70).

$$a_i = 1.151\pi_i - 1.464\sigma_i^+ \tag{73}$$

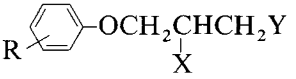
$$a_i = 1.259\pi_i - 1.460\sigma_i^+ + 0.208(E_s^{meta} - E_{s,H}^{meta}) \tag{74}$$

This theoretical relationship between Hansch analysis and the Free Wilson model was first recognized by Singer and Purcell [392]. Although it was questioned by Cammarata [399, 425], later investigations confirmed it theoretically and by practical examples [390, 391, 393, 394].

According to eq. 72, Free Wilson group contributions contain all possible physicochemical contributions of a substituent; correspondingly, a Free Wilson analysis always gives the upper limit of correlation which can be achieved by a linear Hansch analysis [390, 391, 393]. No linear Hansch equation (not including squared terms or nonadditive terms like the dipole moment) can be obtained for the α-bromophenethylamines (chapters 4.1 and 4.2) with a correlation coefficient larger than $r = 0.969$ (the correlation coefficient of the Free Wilson analysis, eq. 70). On the other hand, the standard deviation s may become smaller due to the usually larger number of degrees of freedom in a Hansch analysis (for the definition of the correlation coefficient r and the standard deviation s see chapter 5.1).

A comparison of the results from Hansch and Free Wilson analyses offers some information, whether a certain Hansch model can be considered to be acceptable or not. In most cases the Free Wilson analysis of a data set shows whether a linear additive model is suited for the analysis; only in certain cases is a good fit obtained for nonlinear relationships, especially if there are only few degrees of freedom [22, 390, 391].

An example, how a Hansch equation can be improved by comparing the group contributions with those obtained from Free Wilson analysis, is given below. Eq. 75 was derived for the antifungal activities of phenyl ethers (**13**, X, Y = H, OH) [426]. Eq. 76 results if all members of the series are included, giving an even worse fit [393]; in both equations the σ term is not significant.



(13)

$$\log 1/C = 0.691(\pm 0.14)\log P + 0.428(\pm 0.51)\,\sigma + 1.213 \tag{75}$$

$$(n = 26;\ r = 0.911;\ s = 0.216)$$

$$\log 1/C = 0.665(\pm 0.15)\log P + 0.500(\pm 0.57)\,\sigma + 1.235 \quad (76)$$

$$(n = 28;\ r = 0.879;\ s = 0.241)$$

The Free Wilson analysis of the same data set [393] indicates that the smaller group contributions of the *ortho*-substituents might be explained by a steric effect. Correspondingly, eq. 77 was derived [393], which gives a much better fit than eqs. 75 and 76; eq. 77 includes all compounds and, in addition, only contains significant variables.

$$\log 1/C = 0.741(\pm 0.11)\log P + 0.214(\pm 0.08)\,E_s^{ortho} + 0.846 \quad (77)$$

$$(n = 28;\ r = 0.942;\ s = 0.170)$$

Due to the relationships between Hansch analysis and the Free Wilson model, indicator variables (chapter 3.8) have relatively early been included in Hansch analyses (*e.g.* [21, 427, 428]). Both models can be combined to a mixed approach, in a linear (eq. 78) and a nonlinear form (eq. 79), which offers the advantages of both, Hansch analysis and Free Wilson analysis, and widens their applicability in quantitative structure-activity relationships [22].

$$\log 1/C = k_1\Phi_1 + k_2\Phi_2 + \ldots + k_n\Phi_n + \Sigma a_i + c = \Sigma k_j\Phi_j + \Sigma a_i + c \quad (78)$$

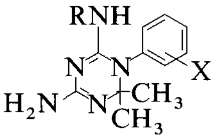
$$\log 1/C = b_1\Phi_1^2 + b_2\Phi_2^2 + \ldots + k_1\Phi_1 + k_2\Phi_2 + \ldots + k_n\Phi_n + \Sigma a_i + c = $$
$$= \Sigma b_j\Phi_j^2 + \Sigma k_j\Phi_j + \Sigma a_i + c \quad (79)$$

It should be mentioned that Bocek and Kopecky, independently and at the same time as Free and Wilson, proposed an additive model with additional interaction terms (eq. 80, reformulated; e_Xe_Y = interaction term) [429, 430].

$$\log 1/C = \Sigma a_i + \Sigma e_Xe_Y + c \quad (80)$$

The Bocek Kopecky model is theoretically and practically related to the nonlinear Hansch model [22, 392, 393] and thus to eq. 79. However, due to the large number of variables this modification has never been used.

The mixed approach allows the description of data sets, when the structural variation is sufficient to derive a Hansch-type relationship for one or several sites of substitution (parameters with index j in eqs. 78 and 79), while for others indicator variables are appropriate because the structural variation is too narrow (group contributions with index i), *e.g.* eq. 60 (chapter 3.8) or eq. 81 for the inhibition of dihydrofolate reductase by dihydrotriazines (**14**) [28, 431, 432].



(14)

$$\begin{aligned}\log 1/C = {} & 0.680(\pm 0.12)\,\pi_3 - 0.118(\pm 0.03)\,\pi_3^2 + 0.230(\pm 0.07)\,\mathrm{MR}_4 - \\ & - 0.0243(\pm 0.009)\,\mathrm{MR}_4^2 + 0.238(\pm 0.12)\,\mathrm{I}_1 - 2.530(\pm 0.27)\,\mathrm{I}_2 - \\ & - 1.991(\pm 0.29)\,\mathrm{I}_3 + 0.877(\pm 0.23)\,\mathrm{I}_4 + 0.686(\pm 0.14)\,\mathrm{I}_5 + \\ & + 0.704(\pm 0.16)\,\mathrm{I}_6 + 6.489(\pm 0.16) \qquad (81)\end{aligned}$$

optimum $\pi_3 = 2.9$ optimum $\mathrm{MR}_4 = 4.7$

$(n = 244;\ r = 0.923;\ s = 0.377)$

For a successful application of the mixed approach it is highly recommended to derive Hansch equations for each subset and to compare whether they correspond to each other or not, before combining them into one equation with the help of indicator variables (*e.g.* eqs. 58–60, chapter 3.8).

Today the mixed approach is the most powerful tool for the quantitative description of large and structurally diverse data sets. Numerous Hansch analyses including Free Wilson-type variables have been published (some more examples are discussed in chapter 7).

4.4. Nonlinear Relationships

Nonlinear relationships between biological activities and lipophilicity are very common. While biological activity values most often linearly increase with increasing lipophilicity [182], such an increase is no longer obtained if a certain range of lipophilicity is surpassed; biological activities remain constant or decrease more or less rapidly with further increase of lipophilicity [7, 19]. Many reviews deal with nonlinear lipophilicity-activity relationships [19, 175, 178, 345, 433].

Different explanations have been given for this effect,

the kinetics of drug transport in biological systems and

the distribution of drugs in different compartments of a biological system

being the most common. In addition,

limited space for the interaction of hydrophobic groups at the binding site,

allosteric effects,

increased metabolism of higher, lipophilic analogs,

end product inhibition by lipophilic products of an enzymatic reaction,

micelle formation or limited solubility of higher analogs, and,

last but not least, the principle of minimum receptor occupation

have been discussed as being responsible for such nonlinear relationships [19].

Hansch formulated a parabolic model (eq. 7, chapter 1.1) [15, 17–19] for the mathematical description of nonlinear relationships. He was aware that the sides of a parabola are always more or less curved, while in most cases at least the "left side" of the structure-activity relationship (*i.e.* the lipophilicity dependence of the more hydrophilic analogs) is strictly linear; equations including a third-order lipophilicity term did not produce much improvement [19]. A computer simulation of the transport of drugs in a biological system, using hypothetical rate constants,

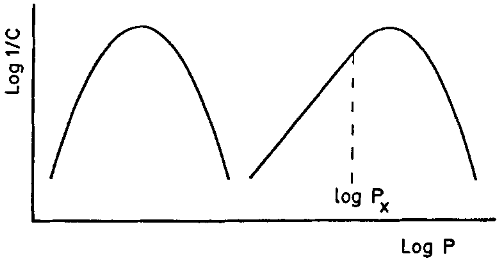


Figure 12: Comparison of the parabolic Hansch model (left curve) and Franke's protein binding model (right curve). Log P_x is the lipophilicity limit, where steric hindrance or other unfavorable interactions cause a change of the linear lipophilicity-activity relationship to a parabola (reproduced from Figure 9 of ref. [175] with permission from Birkhäuser Verlag AG, Basel, Switzerland).

was performed [434] to support the parabolic model. Even this computer simulation showed some systematic deviations at the sides of the parabola.

Franke developed another empirical model to bridge the gap between so many linear relationships and a nonlinear model (Figure 12). He considered binding of ligands at a hydrophobic protein surface of limited size as being responsible for nonlinear lipophilicity-activity relationships and formulated two equations, one for the linear left side (eq. 82) and the other one for the right side, the nonlinear part (eq. 83; log P_x = critical log P value, where the linear relationship changes to a nonlinear one) [435].

$$\log /C = a \log P + c \qquad (\text{if } \log P < \log P_x) \tag{82}$$

$$\log 1/C = \alpha(\log P)^2 + \beta \log P + \gamma \qquad (\text{if } \log P > \log P_x) \tag{83}$$

For practical reasons and to avoid too many parameters, it is better to combine both equations to eq. 84. The term log $[P > P_x]$ in eq. 84 is zero for all log P values being smaller than log P_x, while it is the difference (log P − log P_x) for all log P values being larger than log P_x [175, 345].

$$\log 1/C = a(\log [P > P_x])^2 + b \log P + c \tag{84}$$

The Franke model is a definite improvement as compared to the parabolic model. In many practical cases it gives a much better fit, *e.g.* eqs. 85 and 86 for the spasmolytic activities of mandelic acid esters (Table 14; RA = relative biological activity values, based on cyclandelate (**15**), RA = 100%) [345].

$$\log RA = -0.189(\pm 0.09)(\log P)^2 + 1.566(\pm 0.56)\log P - 1.438 \tag{85}$$

$$\text{optimum } \log P = 4.15$$

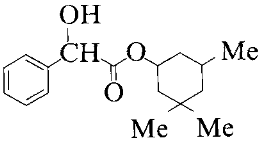
$(n = 11; r = 0.958; s = 0.298)$

$$\log RA = -0.585(\pm 0.15)(\log [P > P_x])^2 + 0.802(\pm 0.11)\log P - 0.901 \tag{86}$$

$\log P_x = 3.43 \qquad \text{optimum } \log P = 4.12$

$(n = 11; r = 0.989; s = 0.164)$

Table 14. Spasmolytic activities of mandelic acid esters in the guinea pig ileum *vs.* histamine [345]; RA = relative biological activity values, based on cyclandelate (**15**) (RA = 100%)



(15)

Compound	log P	log RA
Methyl ester	0.41	-0.52
Ethyl ester	0.91[a)]	-0.22
Propyl ester	1.41	0.20
Butyl ester	1.91	0.59
Pentyl ester	2.41	1.08
Hexyl ester	2.91	1.52
Heptyl ester	3.41	1.70
Octyl ester	3.91	2.18
Nonyl ester	4.41	2.26
Decyl ester	4.91	1.45
Undecyl ester	5.41	1.28

[a)] experimental value, all other values extrapolated

Theoretical approaches were followed by others, the first one laying the foundation for all other theoretical models. McFarland considered the rate constants k (transport from the aqueous phase into an organic phase) and l (transport in the reverse direction) to be related to the probabilities of a molecule to enter either the lipid phase (eq. 87) or the aqueous phase (eq. 88) from an aqueous/lipid interface ($p_{i,j}$ = probabilities) [436].

$$k = p_{0,1} = \frac{P}{P+1} \tag{87}$$

$$l = p_{1,0} = \frac{1}{P+1} \tag{88}$$

From these two equations he derived the probability of a molecule of crossing many different aqueous and lipid barriers and to arrive at the receptor by multiplying the different probabilities (eq. 89; $p_{1,2} = p_{1,0}$, $p_{2,3} = p_{0,1}$, *etc.*). As c_r, the concentration in the receptor phase, is proportional to $p_{0,n}$, eq. 90 results from appropriate transformation of eq. 89 [175, 345, 436].

$$p_{0,n} = p_{0,1} \cdot p_{1,2} \cdot p_{2,3} \cdot \ldots \cdot p_{n-1,n} = \frac{P^{n/2}}{(P+1)^n} \tag{89}$$

$$\log c_r = a \log P - 2a \log (P+1) + c \tag{90}$$

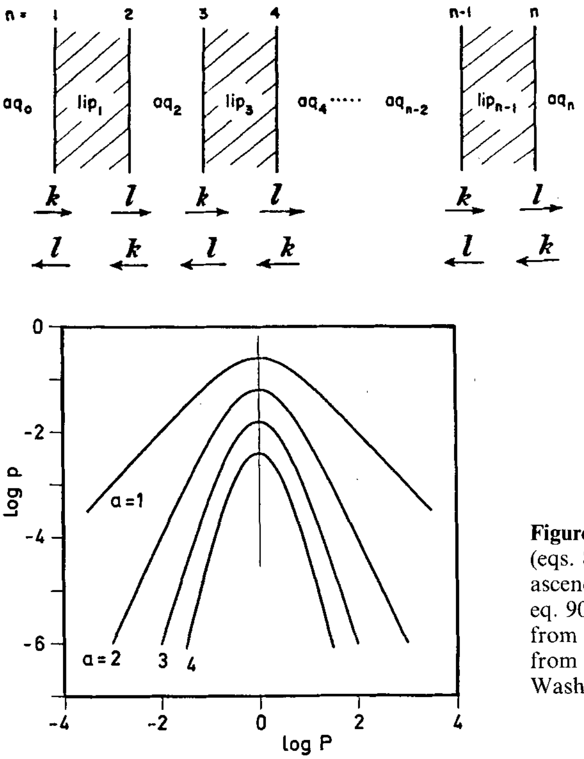


Figure 13: McFarland probability model (eqs. 87–90). Symmetrical curves with linear ascending and descending sides result from eq. 90 (reproduced from Figure 1 and redrawn from Figure 3 of ref. [436] with permission from the American Chemical Society, Washington, DC, USA).

Symmetrical curves with linear ascending and descending sides, having their optimum at log P = 0, result from eq. 90 (Figure 13). No practical application of eq. 90 was possible because most often nonlinear lipophilicity-activity relationships are unsymmetrical and their optimum log P values are different from zero.

Higuchi and Davis considered a biological system under equilibrium conditions [437]. From different simulations they obtained unsymmetrical curves with linear ascending and descending sides; however, no general equation for quantitative structure-activity analyses was derived. This was done by Hyde by considering much simpler model systems [438, 439]. Eq. 91 (slope = 1) and eq. 92 (different slopes at the ascending left side) describe nonlinear relationships with linear left sides, leveling off to a plateau [438, 439]; only such examples of lipophilicity-activity relationships can be described by this model [175, 345].

$$\log 1/C = \log P - \log(aP + 1) + c \qquad (91)$$

$$\log 1/C = b \log P - \log(aP^b + 1) + c \qquad (92)$$

The bilinear model (eq. 93) (Figure 14) was derived from a reconsideration of the McFarland model, taking into account the different volumes of aqueous and organic phases of a biological system [23, 175, 345, 440–442].

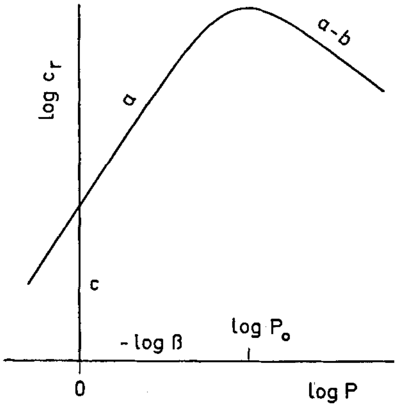


Figure 14: Bilinear model.
$\log c_r = a \log P - b \log(\beta P + 1) + c$
(c_r = concentration in the receptor phase of a hypothetical biological system); for symmetrical curves the lipophilicity optimum is $\log P_0 \approx -\log \beta$.

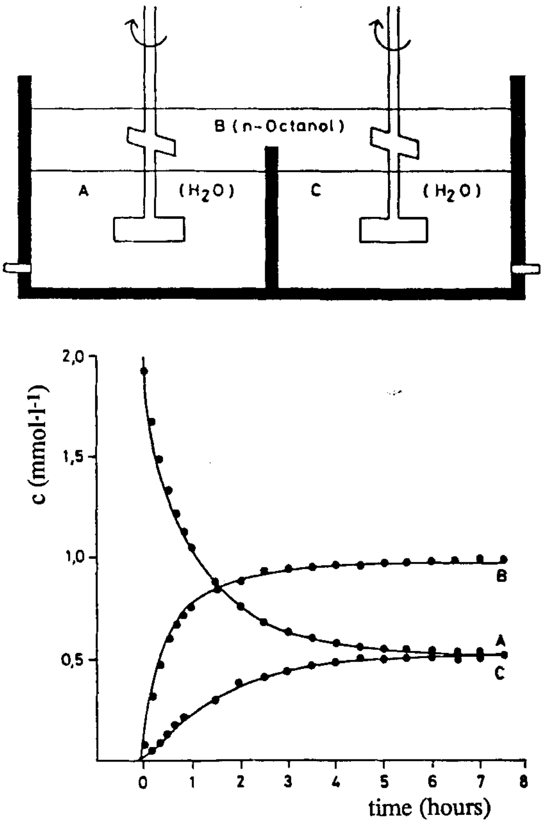


Figure 15: Determination of the rate constants k_1 and k_2 of the transport of homologous quaternary alkylammonium bromides in a three-compartment model system. Water/*n*-octanol/water system (upper diagram) and time dependence of substance concentrations in the different compartments after different times t (lower diagram) (reproduced from Figures 1 and 4 of ref. [443] with permission from Editio Cantor Verlag GmbH, Aulendorf, Germany).

$$\log 1/C = a \log P - b \log (\beta P + 1) + c \tag{93}$$

$$\text{optimum } P = \frac{a}{\beta \cdot (b - a)} \quad \text{(only valid for } b > a)$$

A different formulation of the bilinear model (*e.g.* eq. 94) is used if the physicochemical parameters are already in the logarithmic scale; on the other hand, terms like $10^{\log P}$ (= P), sometimes seen in the QSAR literature, should be avoided.

$$\log 1/C = a\pi - b \log (\beta \cdot 10^{\pi} + 1) + c \tag{94}$$

Eqs. 93 and 94 may be considered as extensions of eqs. 90–92. In contrast to these equations, the bilinear model is generally applicable to the quantitative description of a wide variety of nonlinear lipophilicity-activity relationships. In addition to the parameters that are calculated by linear regression analysis, it contains a nonlinear parameter β, which must be estimated by a stepwise iteration procedure [440, 441]. It should be noted that, due to this nonlinear term, the confidence intervals of a, b, and c refer to the linear regression using the best estimate of the nonlinear term. The additional parameter β is considered in the calculation of the standard deviation s and the F value *via* the number of degrees of freedom (compare chapter 5.1). The term a in eq. 93 is the slope of the left linear part of the lipophilicity-activity relationship, the value (a − b) corresponds to the negative slope on the right side.

The bilinear model is confirmed by simulations, using experimental rate constants of drug transport, which were determined from the time dependence of substance concentrations in the different phases of a three-compartment system water/*n*-octanol/water (Figure 15) [443].

Eqs. 95 and 96 could be derived for the dependence of the transport rate constants k_1 (from the aqueous phase into the organic phase) and the reverse rate constants k_2 (from the organic phase into the aqueous phase) on lipophilicity [444]. According to eq. 95, the rate constants k_1 are thermodynamically controlled, they linearly increase with lipophilicity. With further increase in lipophilicity the diffusion of the solutes becomes rate-limiting; a plateau is reached because now thermodynamic control is replaced by kinetic control. The reverse holds true for the rate constants k_2 (eq. 96) (Figure 16).

$$\log k_1 = \log P - \log (\beta P + 1) + c \tag{95}$$

$$\log k_2 = - \log (\beta P + 1) + c \tag{96}$$

Both equations were confirmed by independent investigations of other series of structurally largely different substances, including neutral compounds, ionized compounds, and quaternary ion pairs, with molecular weights in the range < 100 to > 500 dalton, and in different solvent systems (*e.g.* eqs. 97 and 98) [445–447].

$$\log k_1 = \log P - \log (0.386P + 1) - 3.999 \tag{97}$$

$$(n = 45;\ r = 0.997;\ s = 0.047)$$

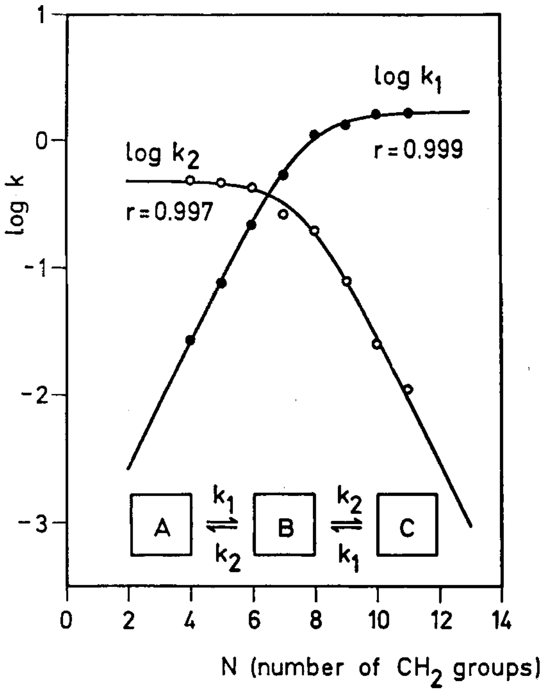


Figure 16: Rate constants k_1 and k_2 of the transport of homologous quaternary alkylammonium bromides in a three-compartment model system; comparison of experimental [443] and calculated values (eqs. 95 and 96) (redrawn from Figure 2 of ref. [444] with permission from the American Pharmaceutical Association, Washington, DC, USA).

$$\log k_2 = -\log(0.385P + 1) - 4.002 \qquad (98)$$

$(n = 45; r = 0.998; s = 0.047)$

Prior assumptions [*e.g.* 433, 434] that the product $k_1 \cdot k_2$ is equal to unity or at least a constant value, are disproved by eqs. 95 and 96; in reality, the product $k_1 \cdot k_2$ is related to lipophilicity by the function $k_1 \cdot k_2 = \text{const. } P/(\beta P + 1)^2$ [442].

Eq. 93 can be derived from kinetically controlled model systems (Figure 17) as well as from equilibrium models (Figure 18), indicating that the bilinear model is valid under diffusion control as well as under equilibrium or pseudoequilibrium conditions [175, 345, 440, 448, 449].

Many data sets can be explained much better with the help of this theoretically derived model than with the empirical parabolic model [23, 175, 345]. Only two examples shall be given here, one (eq. 99) describing the spasmolytic activities of mandelic acid esters (Table 14; eqs. 85 and 86), the other one describing the antifungal activities of aliphatic amines *vs. Rhinocladium beurmanni* (eq. 100) [345]; in the latter case the parabolic model gives r = 0.967, while a combination of the parabolic model with an additional, highly interrelated MW term yielded r = 0.994 [344] (chapter 3.7).

$$\log 1/C = 0.851(\pm 0.12)\log P - 2.259(\pm 0.55)\log(\beta P + 1) - 0.963(\pm 0.29) \qquad (99)$$

$\log \beta = -4.359 \qquad \text{optimum } \log P = 4.14$

$(n = 11; r = 0.990; s = 0.160; F = 109.94)$

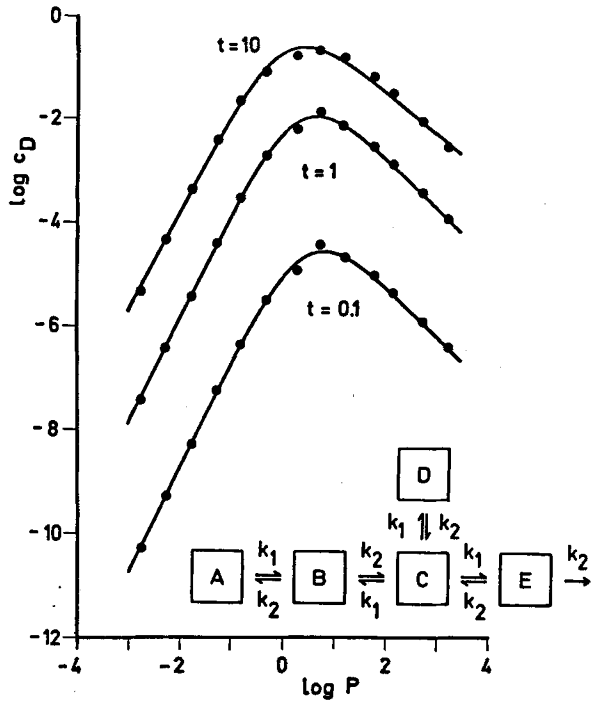


Figure 17: Simulation of the lipophilicity dependence of drug concentrations in compartment D of a five-compartment non-equilibrium model system after different times t; calculated values, using experimental rate constants k_1 and k_2 (Figures 15 and 16), are fitted with the bilinear model (Figure 14; eq. 93); A and C model the outer and inner aqueous phases of a simple biological system, B and E are lipid barriers, and D stands for a receptor compartment. Unsymmetrical curves describe the lipophilicity dependence of the concentration in compartment D (redrawn from Figures 6 and 9 of ref. [442] with permission from Editio Cantor Verlag GmbH, Aulendorf, Germany).

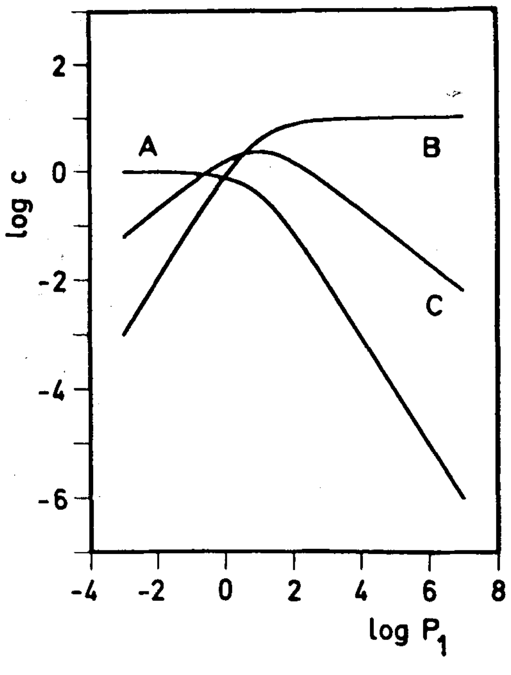


Figure 18: Lipophilicity dependence of drug concentrations in a hypothetical three-compartment equilibrium model system; curve A represents the concentration dependence in the aqueous phase, curve B in a lipid phase, and curve C in a phase of intermediate lipophilicity; log P_1 is the phase C/phase A partition coefficient (reproduced from Figure 3 of ref. [442] with permission from Editio Cantor Verlag GmbH, Aulendorf, Germany).

$$\log 1/C = 0.944(\pm 0.01) \log P - 2.347(\pm 0.05) \log(\beta P + 1) - 0.053(\pm 0.05) \quad (100)$$

$$\log \beta = -5.787 \qquad \text{optimum } \log P = 5.62$$

$$(n = 15;\ r = 1.000;\ s = 0.031;\ F = 7945)$$

However, the parabolic model is still valuable for structure-activity analyses. It is the simpler model, easier to calculate, and most often a sufficient approximation of the true structure-activity relationship. The calculation of bilinear equations is relatively time-consuming, as compared to the parabolic model; strange results may be obtained in ill-conditioned data sets. On the other hand, in many cases the bilinear model gives a better description of the data, especially if additional physicochemical parameters are included in the regression equation. The lipophilicity optimum of symmetrical curves is precisely described by both, the parabolic model (optimum $\log P = -b/2a$) and the bilinear model (optimum $\log P = -\log \beta$). In the case of unsymmetrical curves the site of the lipophilicity optimum is described much better by the bilinear model (optimum $\log P = \log a - \log \beta - \log(b - a)$; eq. 93) than by the parabolic model.

Other models, which are closely related to the bilinear model, have been derived for the quantitative description of nonlinear structure-activity relationships [41, 156, 450–455], *e.g.* eq. 101 [41, 156], eq. 102 [453], eq. 103 [454, 455], and eq. 104 (Figure 19) [455].

$$\log 1/C = a \log P - \log(\beta P^b + 1) + c \quad (101)$$

$$\log 1/C = a \log P + \Sigma b_i \log(\beta_i P + 1) + c \quad (102)$$

$$\log 1/C = a \log P - bP^c/(dP^c + 1) - \log(dP^c + 1) + \text{const.} \quad (103)$$

$$\log 1/C = a \log P - bP^c + \text{const.} \quad (104)$$

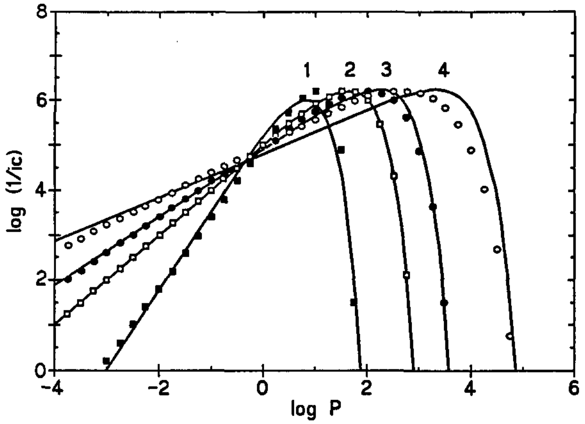


Figure 19: Different functions generated by eq. 104 (dots; c = 1.6, curve 1; c = 1.0, curve 2; c = 0.8, curve 3; c = 0.6, curve 4; ic = C = isoeffective concentration) are fitted by the function log 1/C = a log P − bP + const. (lines; eq. 104, c = 1) (reproduced from Figure 5 of ref. [455a] with permission from the copyright owner).

In certain cases the functions described by eqs. 90–92 and 101–104 may indeed give a better fit of the data. However, none of these models combines general applicability and ease of calculation to the same extent as the parabolic model and the bilinear model do; while some of them contain too many nonlinear parameters (eqs. 101–103), others are only appropriate for certain nonlinear cases, having either a symmetric (eq. 90) or an asymmetric shape (eqs. 91, 92, 104).

The use of lipophilicity similarity matrices for the quantitative description of nonlinear lipophilicity-activity relationships is discussed in chapter 9.4.

Other nonlinear relationships are known in addition to nonlinear lipophilicity-activity relationships. Most common are nonlinear dependences on molar refractivity (*e.g.* resulting from a limited binding site at the receptor; for examples see chapter 7.1), but also other types of nonlinear relationships, *e.g.* with steric parameters, are frequently obtained. Even electronic parameters (eq. 48; chapter 3.5) or molecular weight terms (eq. 56; chapter 3.7) have been used in nonlinear equations.

Cross product terms, like *e.g.* $MR_1 \cdot MR_2 \cdot MR_3$ [456], should be avoided because they are highly interrelated with squared terms (*e.g.* $MR_1 \cdot MR_2 \cdot MR_3$ *vs.* a combination of ΣMR^2, MR_1, MR_2, and MR_3: $n = 71$; $r = 0.993$) [175]. Curious nonlinear models, like hyperbolic regressions [422, 457], sinus [457], or tangens terms [458] shall be mentioned here without giving advice to use them; these approaches can only be characterized by a phrase, created by Hansch, as *"statistical unicorns, beasts that exist on paper but not in reality"* [307].

4.5. Dissociation and Ionization of Acids and Bases

Ions are much more polar than neutral compounds, due to their positive or negative charges. Correspondingly, the degree of dissociation and protonation has a significant influence on the partition coefficients of acids and bases; the dependences of the apparent partition coefficients P_{app} on pH values, pK_a values, P_u values (partition coefficient of the neutral, unionized form) and P_i values (partition coefficient of the ionized form) are described by eq. 105 (acids) (Figure 20) and eq. 106 (bases) (for reviews see refs. [41, 156, 173, 175, 459]).

$$\log P_{app} = \log(P_u \cdot 10^{pK_a} + P_i \cdot 10^{pH}) - \log(10^{pK_a} + 10^{pH}) \qquad (105)$$

$$\log P_{app} = \log(P_u \cdot 10^{pH} + P_i \cdot 10^{pK_a}) - \log(10^{pK_a} + 10^{pH}) \qquad (106)$$

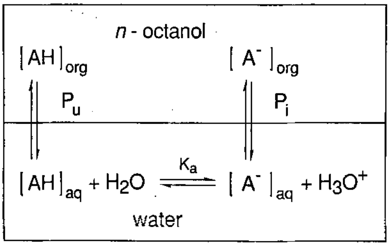


Figure 20: Dissociation and partitioning equilibria of an acid AH in the two-compartment system *n*-octanol/aqueous buffer (redrawn from Figure 2 of ref. [175] with permission from Birkhäuser Verlag AG, Basel, Switzerland).

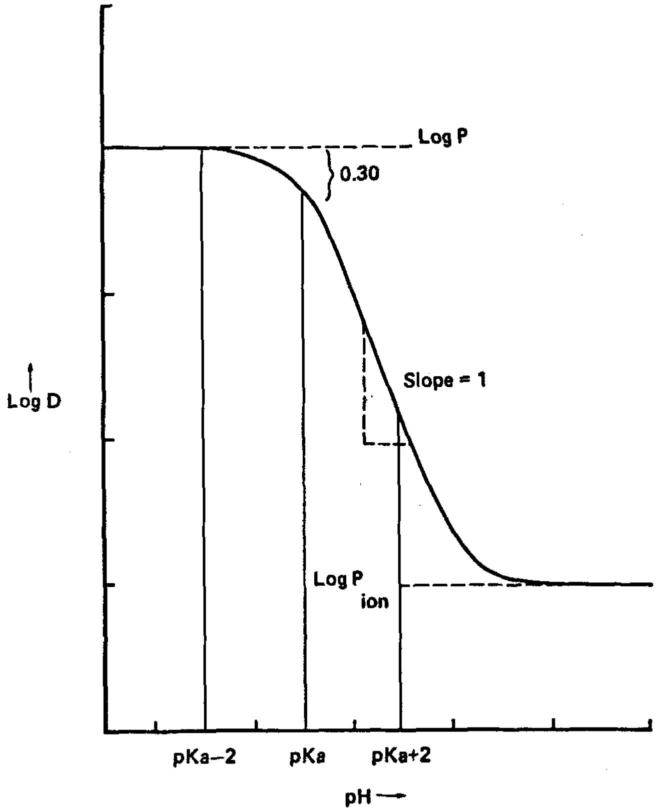


Figure 21: Sigmoidal pH partition profile of an acid AH. Log P_{app} (= log D) is constant at low pH values (log P_{app} = log P_u), decreases with increasing pH values ($P_{app} = 0.5P_u$ at pH = pK_a) with a slope = 1 and reaches again a constant value at log P_{app} = log P_{ion} (reproduced from Figure 2 of ref. [220] with permission from the copyright owner).

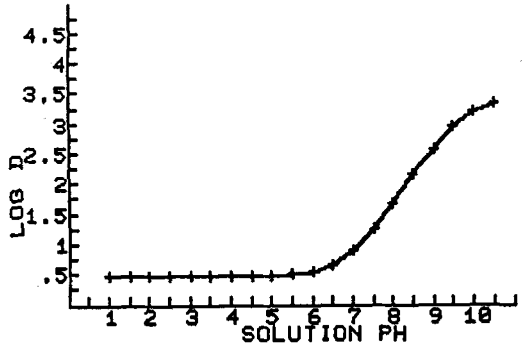
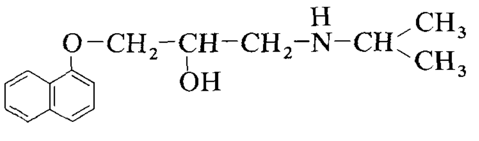


Figure 22: pH Partition profile of propranolol (**16**) calculated from measured values $pK_a = 9.72$, log $P_u = 3.41$, and log $P_i = 0.48$ (reproduced from Figure 3 of ref. [463] with permission from the American Pharmaceutical Association, Washington, DC, USA).



(16)

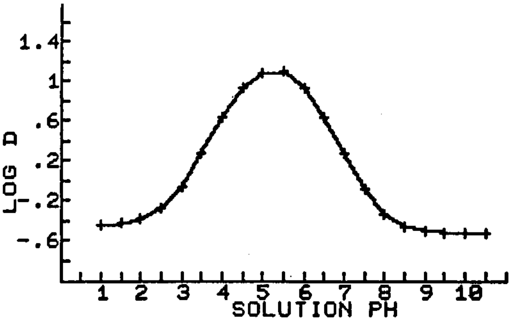
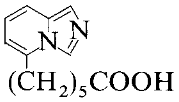


Figure 23: pH Partition profile of the amino acid CGS 13080 (**17**) calculated from measured values $pK_1 = 4.51$, $pK_2 = 6.01$, $\log P_u = 1.46$ and 1.50, $\log P_{cation} = 0.78$, and $\log P_{anion} = 0.65$ (reproduced from Figure 3 of ref. [463] with permission from the American Pharmaceutical Association, Washington, DC, USA).



(17)

Sigmoidal curves are obtained for the pH dependence of the log P_{app} values of most acids and bases (Figures 21 and 22) [184, 460–463]. At pH values, where the neutral form predominates (pH < pK_a for acids; pH > pK_a for bases), the P_{app}values are identical with the $P_u$values. With increasing ionization, *i.e.* with increasing (acids) or decreasing (bases) pH values, the log P_{app} values decrease linearly, until again a constant P_{app} value is obtained, because now only the ionic form contributes to partitioning ($P_{app} = P_i$).

It is not surprising that many drugs are either weak acids or weak bases. They are able to cross membranes in their neutral forms and they are soluble in aqueous phases in their ionized forms. Due to this dual character they are readily absorbed and distributed in the different compartments of a biological system.

Complex pH dependences are obtained for amino acids (*e.g.* [463]) and other compounds containing more than one ionizable group (*e.g.* Figures 23 and 24) [173, 464].

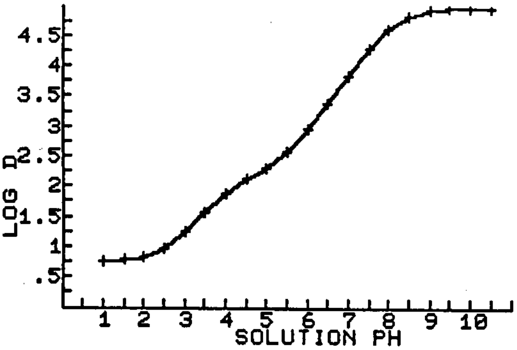
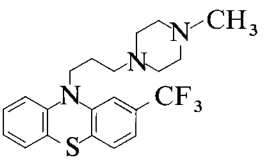


Figure 24: pH Partition profile of trifluoperazine (**18**) calculated from measured values $pK_1 = 4.04$, $pK_2 = 8.08$, $\log P_u = 4.95$, and $\log P_i = 2.04$ and 0.74 (reproduced from Figure 3 of ref. [463] with permission from the American Pharmaceutical Association, Washington, DC, USA).



(18)

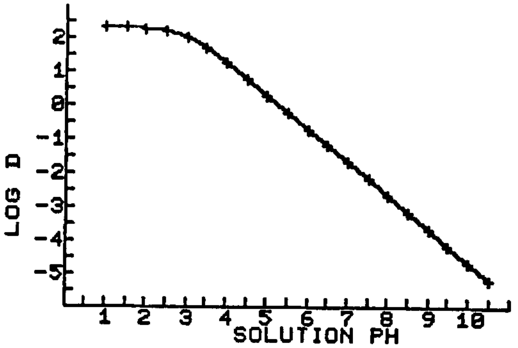
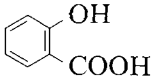


Figure 25: pH Partition profile of salicylic acid (**19**) calculated from measured values $pK_a = 2.95$, $\log P_u = 2.34$, and $\log P_i < -5.0$ (reproduced from Figure 3 of ref. [463] with permission from the American Pharmaceutical Association, Washington, DC, USA).



(19)

Most often P_i values are 3 to 5 decades smaller than the corresponding P_u values; the differences can be larger, *e.g.* for salicylic acid (**19**) which contains two acidic groups (Figure 25) [463], but also much smaller, *e.g.* for chenodesoxycholic acid, which forms micelles [462], or in the presence of lipophilic counter ions, *e.g.* for salicylic acid in the presence of tetrabutyl-ammonium bromide (**20**) (Figure 26) [463]. Lipophilicity and polarizability of the counterion have a significant influence on the partitioning of ion pairs (*e.g.* [465–468]).

The concentrations of A^- and BH^+ in the organic phase may be neglected for compounds which are not too lipophilic; much simpler equations than eqs. 105 and 106 can be used to estimate the log P_{app} values of most acids (eq. 107) and bases (eq. 108) at pH values which are not too far from the pK_a values.

$$\log P_{app} = \log P_u - \log(1 + 10^{pH - pK_a}) \qquad (107)$$

$$\log P_{app} = \log P_u - \log(1 + 10^{pK_a - pH}) \qquad (108)$$

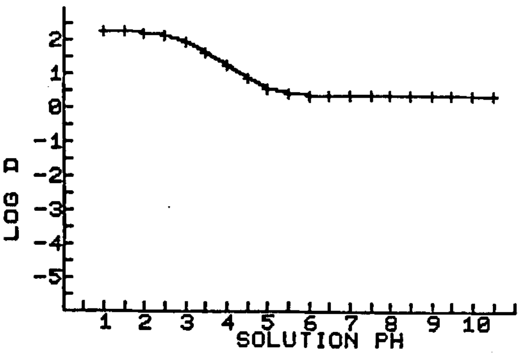
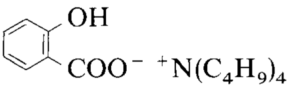


Figure 26: pH Partition profile of the ion pair salicylic acid/tetrabutyl-ammonium bromide (**20**) calculated from measured values $pK_a = 2.95$, $\log P_u = 2.27$, and $\log P_{ion\,pair} = 0.39$ (reproduced from Figure 3 of ref. [463] with permission from the American Pharmaceutical Association, Washington, DC, USA).



(20)

Table 15. Colonic absorption of acidic compounds in a rat *in situ* model at pH = 6.8 [442, 469, 472]

Compound	log P	pK_a - pH	log P_{app}	log %ABS	log k_{abs}
5-Nitrosalicylic acid	1.98	-4.5	-2.52	0.30	-1.69
m-Nitrobenzoic acid	1.83	-3.4	-1.57	1.00	-0.98
Salicylic Acid	2.26	-3.8	-1.54	1.08	-0.89
Benzoic Acid	1.85	-2.6	-0.75	1.28	-0.68
Phenylbutazone	3.22	-2.4	0.82	1.58	-0.32
o-Nitrophenol	1.79	0.2	1.58	1.74	-0.10
Thiopental	2.50	0.8	2.44	1.70	-0.16
p-Hydroxypropiophenone	1.85	1.0	1.81	1.66	-0.21
m-Nitrophenol	2.00	1.4	2.00	1.64	-0.24
Phenol	1.46	3.1	1.46	1.55	-0.35

Scherrer [469—471] calculated log D (= log P_{app}) values by eqs. 107 and 108, respectively, and used them instead of the most often inappropriate (pK_a − pH) terms. The (pK_a − pH) approximation [472] only holds true for pH ranges where the compounds predominantly exist in their ionized forms. Eq. 109 (% ABS = % colonic absorption at pH = 6.8) [472] is incorrect in two respects: first, percentage values are used instead of rate constants, and second, the (pK_a − pH) term produces wrong values (corresponding to more than 100% neutral form of the molecules) for compounds with pK_a values larger than 6.8 (Table 15) [175, 469].

$$\log \% \text{ ABS} = 0.156(\pm 0.08)(pK_a - pH) + 0.366(\pm 0.44)\log P + 0.755 \quad (109)$$

$$(n = 10;\ r = 0.866;\ s = 0.258)$$

A better and more consistent result is obtained if log k_{abs} values are calculated from % ABS values by appropriate transformation and if log P_{app} values and the parabolic [469] or the bilinear model are used, *e.g.* eq. 110 [442].

$$\log k_{abs} = 1.024(\pm 0.31)\log P_{app} - 0.881(\pm 0.36)\log(\beta P_{app} + 1) + 0.935 \quad (110)$$

$$\log \beta = 1.600$$

$$(n = 10;\ r = 0.991;\ s = 0.081)$$

Eq. 111 [345] correlates the buccal absorption rate constants of an acid (*p*-hexylphenylacetic acid, pK_a = 4.36) and a base (propranolol, pK_a = 9.45) at different pH values with log P_{app} values (Table 16) [345, 470].

$$\log k_{abs} = 0.448(\pm 0.05)\log P_{app} - 0.448(\pm 0.05)\log(\beta P_{app} + 1) - 1.689 \quad (111)$$

$$\log \beta = -2.792$$

$$(n = 12;\ r = 0.988;\ s = 0.102)$$

Table 16. Buccal absorption of propranolol and *p*-hexylphenylacetic acid at different pH values [345, 470]

Compound	pH	log P_{app}	log k_{abs}
Propranolol	5.08	-1.04	-2.19
log P = 3.33	6.02	-0.10	-1.71
pK_a = 9.45	7.00	0.88	-1.22
	7.93	1.81	-0.79
	8.94	2.70	-0.53
	9.93	3.21	-0.35
***p*-Hexylphenylacetic acid**	4.0	4.20	-0.46
log P = 4.25	5.0	3.63	-0.54
pK_a = 4.36	6.0	2.72	-0.72
	7.0	1.72	-1.07
	8.0	0.72	-1.44
	9.0	-0.28	-1.78

The use of log P_{app} values is appropriate in all cases where rate constants are involved, but not for binding or other equilibrium systems [173]. The correct approach to equilibrium systems is demonstrated by eq. 112 (K_i = inhibition of monoamine oxidase by amines and alcohols; I = 0 for amines, I = 1 for alcohols), where the biological data have been corrected for the concentration of the unionized form (Table 17) [175].

$$\log 1/K_i^{corr} = \log 1/K_i + \log(1 + 10^{pK_a - pH}) = 3.130(\pm 0.17)\log P - \\ - 3.797(\pm 0.32)\log(\beta P + 1) - 3.507(\pm 0.12)\,I + 3.379(\pm 0.15) \quad (112)$$

$\log \beta = -1.781$ optimum $\log P = 2.45$

$(n = 21;\ r = 0.999;\ s = 0.118;\ F = 1737)$

Deviations from the simple pH partition hypothesis (*i.e.* the pH absorption profiles should be parallel to pH partition profiles) (*e.g.* [473, 474]), called pH shift, are obtained for highly lipophilic compounds; their absorption profiles are shifted to higher (acids) and lower (bases) pH values (Figure 27). The higher the lipophilicity of the neutral species is, the higher is the observed pH shift.

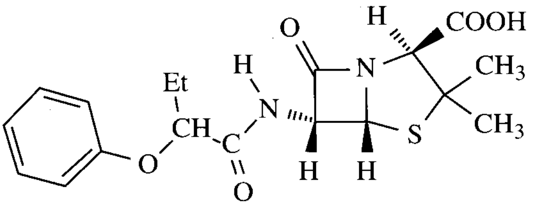
Different explanations were discussed for this effect [175, 475], which even is obtained in simple *n*-octanol/water *in vitro* systems (Figure 28) [476]. Without questioning the relevance of different reasons in certain cases, the pH shift can easily be explained by the assumption of an aqueous diffusion layer at the aqueous/organic interface. Neutral species rapidly enter the organic phase from the aqueous/organic interface. They are steadily regenerated by the dissociation equilibrium from ionized species within the aqueous diffusion layer, much faster than the neutral molecules can diffuse from the bulk solution into this layer.

Table 17. Inhibition of human liver mitochondrial monoamine oxidase by various aliphatic amines and alcohols at different pH values [175]

Compound	log P	pH	K_i, mmol	log $1/K_i$	log $1/K_i^{corr}$ [a]
Ethylamine	- 0.03	8.72	36	1.44	3.39
n-Propylamine	0.47	7.62	25	1.60	4.64
		8.72	2.0	2.70	4.64
n-Butylamine	0.97[b]	7.51	1.2	2.92	6.07
		8.11	0.31	3.51	6.06
		8.72	0.073	4.14	6.08
n-Pentylamine	1.47	7.62	0.044	4.36	7.40
		8.72	0.0035	5.46	7.40
n-Hexylamine	1.97	7.57	0.0092	5.04	8.13
		8.72	0.00068	6.17	8.11
n-Heptylamine	2.47	7.62	0.0075	5.12	8.17
n-Octylamine	2.97	7.48	0.015	4.82	8.00
		7.62	0.010	5.00	8.04
n-Nonylamine	3.47	8.72	0.00096	6.02	7.96
n-Propanol	0.38	8.72	72	1.14	= log $1/K_i$
n-Butanol	0.88[b]	7.51	3.6	2.44	
		8.72	3.6	2.44	
n-Pentanol	1.38	8.72	0.17	3.77	
n-Heptanol	2.38	8.72	0.025	4.60	
n-Octanol	2.88	7.51	0.034	4.47	
		8.72	0.032	4.49	

[a] log $1/K_i^{corr}$ = log $1/K_i$ + log $(1 + 10^{pK_a - pH})$; pK_a (amines) = 10.66

[b] experimental values, all other values extrapolated



(21)

Much more complex mathematical models have been derived for the quantitative description of pharmacokinetic data of acids and bases at different pH values. Applications to practical examples are discussed in chapter 7.3 and in refs. [41, 156, 175, 459, 477–479].

Several methods were described for the simultaneous determination of *n*-octanol/water log P values and pK_a values [463, 464, 480–486].

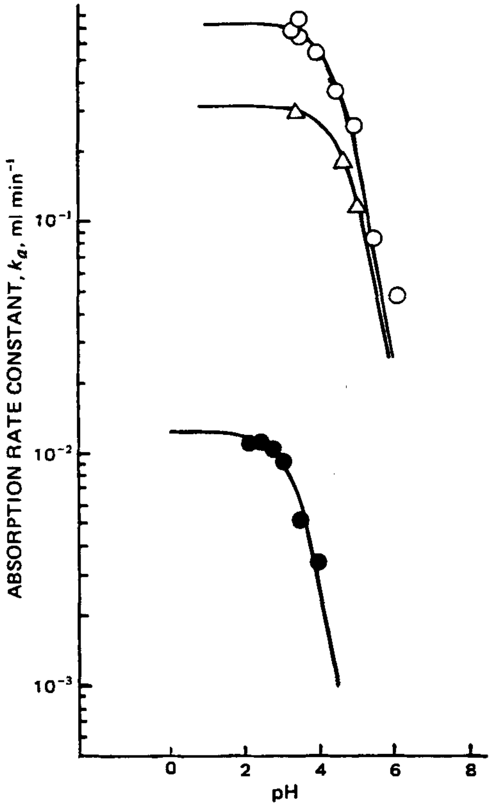


Figure 27: pH Dependence of the absorption rate constants of propicillin (**21**) ($pK_a = 2.76$) in rat stomach and intestine *in situ* models under various conditions. ●, gastric absorption rate constants at a flow rate of 10 ml · min^{-1} ($pK_a^{app} = 3.49$, corresponding to a pH shift of 0.73); ○, intestinal absorption rate constants at a flow rate of 10 ml · min^{-1} ($pK_a^{app} = 4.61$, pH shift = 1.85) and Δ, same experiment with perfusion stopped after 2.5 min ($pK_a^{app} = 4.87$, pH shift = 2.11). The solid lines represent the best fit of the experimental data (reproduced from Figure 4 of ref. [476] with permission from the American Pharmaceutical Association, Washington, DC, USA).

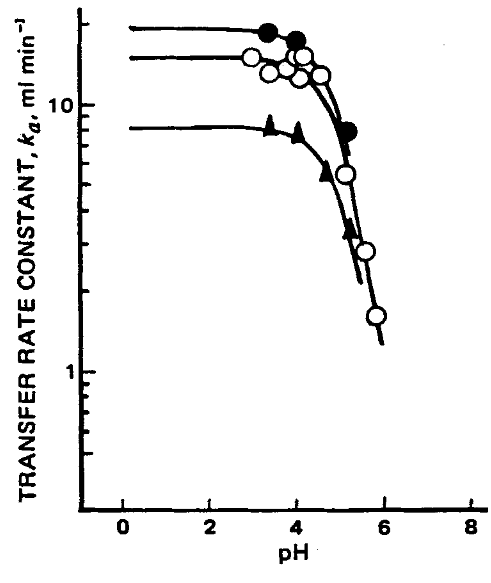


Figure 28: pH Dependence of the transfer rate constants of propicillin (**21**) ($pK_a = 2.76$) in a two-compartment *n*-octanol/water *in vitro* model under various conditions. ▲, at a rotation rate of 10 rpm ($pK_a^{app} = 5.08$, pH shift = 2.32); ○, 20 rpm ($pK_a^{app} = 5.02$, pH shift = 2.26); ●, 38 rpm ($pK_a^{app} = 4.89$, pH shift = 2.13). The solid lines represent the best fit of the experimental data (reproduced from Figure 9 of ref. [476] with permission from the American Pharmaceutical Association, Washington, DC, USA).

4.6. Other QSAR Approaches

Pattern recognition techniques have attracted much attention in the past two decades [24, 25, 58, 487 – 502]. In principle, there is no difference between the classical QSAR methods and pattern recognition, only the number of variables is much larger in pattern recognition than in Hansch analysis. In most cases, no preselection of variables is done on the basis of a biochemical model; a large variety of variables (physicochemical parameters, quantum-chemical indices, topological parameters, indicator variables, *etc.*) is included, but only a part of the data (usually 50 – 70%, the training set) is selected to derive a quantitative model for the prediction of the rest of the data (the test set). Many problems are associated with the selection of a proper training set (*e.g.* [503]) and the limitations arising from stepwise regression analysis (few degrees of freedom, interrelated variables); more consistent results are obtained by using other multivariate methods, like principal component analysis or soft modeling techniques, *e.g.* SIMCA or PLS analysis (see chapter 5.3), instead of regression analysis. An overview of pattern recognition techniques is given in two recent reviews [487, 502].

While the application of pattern recognition methods, at least in combination with a reasonable preselection of variables and use of the PLS method, seems to be justified for groups of congeneric drugs which have the same mechanism of action, its abuse to correlate and predict global toxic, mutagenic, teratogenic, carcinogenic, and other biological properties [504 – 511] must be criticized. By no means it makes sense to correlate the toxicities of drugs as different as *e.g.* HCN, dibenzodioxines, strychnine, tetrodotoxin, and toxic peptides (all having different mechanisms of toxic action) with any combination of physicochemical and/or other parameters. All obtained results must be considered as chance correlations. The common argument that such studies are useful or necessary for risk assessment purposes can easily be refuted: while even the proponents of these approaches admit that there are at least 5 to 10% false positive and false negative predictions (in a qualitative manner, *i.e.* yes or no), 20 to 50% wrong predictions seem to be a more realistic estimate for compounds which are neither included in the training set nor in the test set.

That this estimate is not too pessimistic is confirmed by a recent joint effort of several groups to predict the carcinogenicity of some chemicals considered to be relevant as environmental health hazards [512, 513]. In 1990 Tennant *et al.* [512] published prospective predictions of the potential rodent carcinogenicity for 44 chemicals that were being tested by the U.S. National Toxicology Program (NTP). The journal editor encouraged other groups to make their predictions and as a result in 1993 several published and unpublished investigations, using different approaches, were reviewed at a workshop on rodent carcinogenicity prediction [513]. While human expertise, combining *in vitro* results (*Salmonella* mutagenesis test) with other short-term test data, chemical, and biological experience, led to correct predictions (carcinogenic activity yes/no) for nearly 80% of the compounds, all automated expert systems and other prediction methods, without using such short-term biological test results, more or less failed, being predictive to only

45 – 65%. In evaluating these figures, it must be considered that even in unbalanced groups, *i.e.* independent of the number of active and inactive analogs being included in the test set, there is a 50% probability of success (!) of an unbiased blind guess.

From a merely scientific point of view such investigations may be justified, but for practical purposes they are useless. If the prediction is negative, there is by no means a guarantee that the compound indeed is not cancerogenic or toxic; if the "prediction" gives a false positive answer, then it will be much harder to disprove the negative image of such a compound than without the wrong prediction.

A somewhat more reliable approach seem to be rule-based expert systems which compare and categorize structures of new compounds with respect to the information that is extracted from a large database, *e.g.* the CASE (computer-automated structure evaluation) program [514 – 523]. CASE and MULTICASE (a modification of CASE including a hierarchical selection of descriptors) (Figure 29) [524] are artificial intelligence programs which automatically identify molecular features that contribute to (biophores) or reduce (biophobes) biological activity. In congeneric series these fragments are correlated (corresponding to a stepwise Free Wilson analysis) with biological activities; in noncongeneric series structural features are identified which discriminate the active compounds from the inactive ones.

It is claimed that CASE differs from other techniques in being completely automatic and by learning directly from the crude data, selecting its own descriptors from the practically infinite number of possible structural assemblies and creating an *ad hoc* dictionary without human interference [524]. While this statement is against all prior experience with automated approaches, even from a critical point of view it cannot be ruled out that the CASE approach may for the first time be an approximation of artificial intelligence to the medicinal chemist's intuition and skill.

A machine learning program GOLEM from the field of inductive logic programming was developed and applied to model the structure-activity relationships of dihydrofolate reductase (DHFR) inhibitors [525]. GOLEM uses activity, structure, and stereochemistry information of active and inactive analogs to derive inductive hypotheses, *i.e.* rules relating the structures of arbitrarily chosen pairs of analogs with their activities. The rules are refined with all possible pairs of compounds and rules having a high degree of confidence are added to a knowledge base. The process of rule generation is repeated until no improvement in prediction is produced. 44 DHFR inhibitors were used as a training set to automatically derive rules which where applied to predict the activities of 11 test compounds. It was concluded that rules obtained by GOLEM yield more accurate data than a Hansch regression model, on the training set as well as on the test set [525].

Many different but inherently related QSAR approaches start from a hyperstructure, which is a hypothetical molecule including all structural features of the molecules under investigation. The presence and absence of certain hyperstructure atoms or groups in the individual molecules are correlated with their biological activities in a Free Wilson-like procedure. The DARC-PELCO method [418 – 423] was used to analyze and optimize a large group of anticholinergics [526] by deriving a common superstructure.

In the topological pharmacophore methods [527 – 531], *e.g.* LOCON [527, 528, 530], LOGANA [528 – 530], and EVAL [531], Free Wilson-type indicator variables

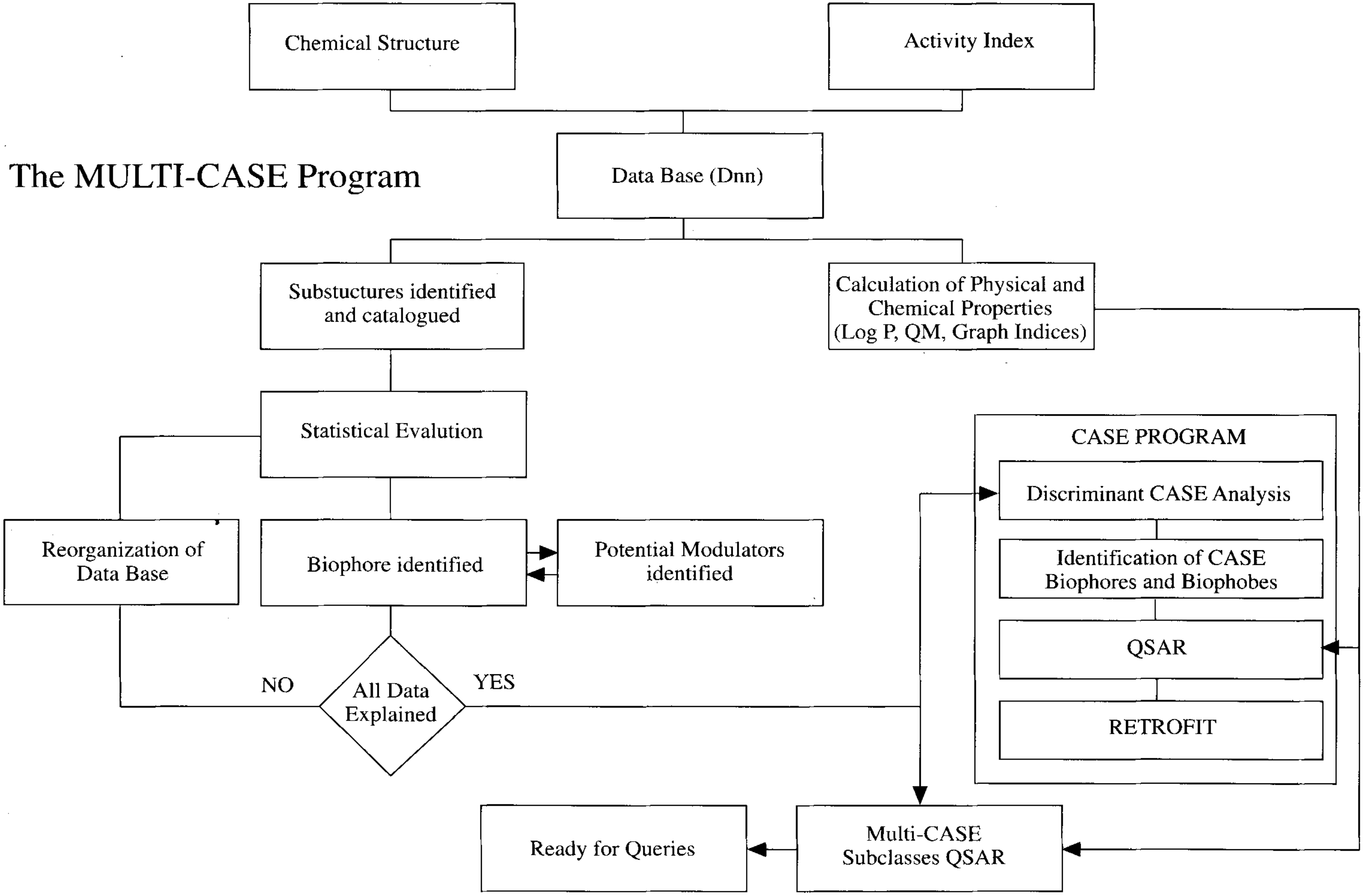

Chemical Structure
Activity Index
The MULTI-CASE Program
Data Base (Dnn)
Substuctures identified and catalogued
Calculation of Physical and Chemical Properties (Log P, QM, Graph Indices)
Statistical Evalution
CASE PROGRAM
Discriminant CASE Analysis
Identification of CASE Biophores and Biophobes
QSAR
RETROFIT
Reorganization of Data Base
Biophore identified
Potential Modulators identified
NO
All Data Explained
YES
Ready for Queries
Multi-CASE Subclasses QSAR

Figure 29: The MULTICASE program (redrawn from Figure 1 of ref. [524] with permission from the copyright owner).

are in a stepwise procedure connected by logical operators, *e.g.* negation (not), conjunction (and), adjunction (inclusive or), disjunction (exclusive or), implication (if then), replication (so if), equivalence (if and only if), Sheffer's function (incompatible with), and Nicod's function (neither nor) [531] to correlate structures and activities. The LOGANA and LOCON methods were applied to phenethylamine agonists and antagonists of the β-adrenergic receptor [527, 531] and to a group of 382 antimalarials [529]. Succinate dehydrogenase inhibitors $R_x - X - Z - Y - R_y$ ($n = 89$; X = *n*-propyl, phenyl, heterocycles; R_x = substituents at the X fragment; $Z = -CONH-$, $-COCH_2-$, etc.; Y = H, phenyl, cyclohexyl; R_y = substituents at the Y fragment) were separated by different logical operators into classes of active and inactive analogs. The conjunction "$-C=C-C=O$ in X and no *ortho*-unsubstituted ring X" is present in 63 of 64 active compounds, but also in 12 inactive compounds. Another conjunction "$-C=C-C=O$ in X and no *ortho*-unsubstituted ring X and no ring in Y and $Z <> -CONH-$ and $X <> -CH=CH-CH_3$ and $R_x <>$ *ortho*-phenyl" is present in 61 of the active compounds and in no inactive compound [528].

Also the minimal steric difference (MSD) [52, 532–534], MCD (Monte Carlo version of MSD) [52, 535–537], the minimal topological difference (MTD) [52, 534, 538–542], and SIBIS [543] methods (for reviews see [38, 52, 287]) define hypothetical hypermolecules. These hypermolecules are generated from an approximate atom-by-atom superposition of all molecules; hydrogen atoms and minor differences in atomic positions, bond lengths, and bond angles are neglected. The definition of the hypermolecule is a relatively easy and straightforward procedure if the series includes a rigid and highly active analog. In many other cases an intuitive approach is applied and arbitrary assumptions have to be made on the connectivity of the hypermolecule; problems associated with the superposition of different ring systems are discussed in [542].

With respect to the binding site, some of the atomic positions of the hypermolecule will be in the receptor cavity ($\varepsilon = -1$), some at the binding site surface (the "cavity wall") ($\varepsilon = +1$), and the rest in aqueous solution ($\varepsilon = 0$). The minimal steric difference (MSD) is defined as the difference in the number of atomic positions that are occupied by each analog, with respect to the analog of highest biological activity value within the series, which is assumed to have an optimal fit to the binding site. A slightly different definition is used for the minimal topological difference (MTD). MTD is the sum of all ε values, multiplied by their occurrence in a certain molecule. As the ε values in the different positions are *a priori* unknown, arbitrary values of $+1$, 0, and -1 (the start map) are assigned to the different atoms (vertices) of the hypermolecule and correlated with biological activity values. In the next step all ε values are systematically changed to achieve a maximum correlation between the optimized maps and the biological activities. This is an ambiguous procedure, leading to a huge number of different permutations and to several optimized receptor maps; a validation test, *e.g.* cross-validation (chapter 5.3), must be performed to prove the significance and predictive ability of the different maps.

Magee [544, 545] combined the hyperstructure concept with the strategies of Hansch analysis and the mixed approach (chapter 4.3). As only several atoms or groups of a molecule modulate biological activity, each position of the hypermolecule can be characterized as being favorable, unfavorable, or indifferent for biological activity. Positional effects for lipophilic interactions, polarizability, electronegativity, steric interactions, and hydrogen bonds may be assigned to certain positions.

While all these methods regard hypermolecules as rigid frames, molecular shape analysis (MSA) [38, 287, 546–556] considers conformational flexibility. Minimum energy conformations are calculated for all molecules within the series; the differences in the overlap volumes to a reference structure are calculated and correlated with biological activities. Also in this approach it is advantageous to start with a rigid analog of high biological activity as the template.

The application of neural networks in structure-activity relationship studies is a recent development in QSAR [557–570]. Neural networks are approaches to model and simulate the information processing in the brain. A multilayer network, as typically applied in QSAR studies, consists of an input layer, a small number of hidden connection layers, bias neurons, and an output layer. Each neuron is connected to every neuron of the adjacent layers and to the bias neurons. In a training phase the network is fed with input values (the independent variables) and the expected output values (the dependent variable). Connection weights and biases of the network are iteratively modified to reduce the differences between observed and predicted values of the dependent variable. By performing a sufficient number of iterations the network "learns" to recognize patterns in the data. After successful training the network may be used to predict the biological activity values of new analogs from new input values.

It seems too early to judge on the real suitability of neural nets for QSAR studies; further investigations which compare classical structure-activity analyses and results from neural networks (*e.g.* [570]) are required to evaluate the scope and limitations of neural nets. Some problems of neural networks, *e.g.* the design of the network, lack of convergence, chance correlations, and overtraining of the network, have been discussed and critically commented [562, 567–570].

QSAR: Hansch Analysis and Related Approaches
by Hugo Kubinyi
© VCH Verlagsgesellschaft mbH, 1993

5. Statistical Methods

5.1. Regression Analysis

Regression analysis [388, 389, 571] correlates independent X variables (*e.g.* physicochemical parameters, indicator variables) with dependent Y variables (*e.g.* biological data) (Figure 30). The dependent variables contain error terms ε, while the independent variables are supposed to contain no such error. In reality, this is only an approximation, because the physicochemical parameters of a QSAR equation indeed contain experimental error; however, in most cases this error is much smaller than the error in the biological data. Only Free Wilson (indicator) variables are error-free terms.

Regression analysis is an exact mathematical procedure, despite the fact that it derives correlations from data containing experimental error (Table 18).

Eqs. 113 and 114 describe a regression model containing two X variables, which is the simplest case of a linear multiple regression analysis.

$$y_{obs} = ax_1 + bx_2 + c + \varepsilon \tag{113}$$

$$y_{calc} = ax_1 + bx_2 + c \tag{114}$$

Since $\Sigma\varepsilon^2 = \Sigma\,\Delta^2 = \Sigma(y_{obs} - y_{calc})^2$ shall be a minimum, the function $f = \Sigma(y_{obs} - ax_1 - bx_2 - c)^2$ is differentiated according to $df/da = df/db = df/dc = 0$ (eqs. 115–117). The so-called normal equations 118–120 result from eqs. 115–117.

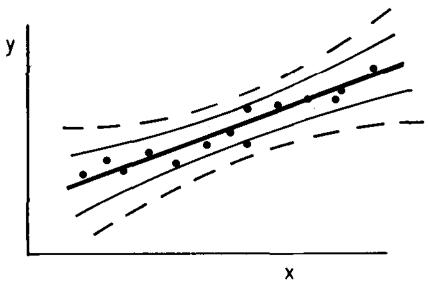


Figure 30: Example of a regression model, $y = ax + b$; the dots are experimental values, the bold line is the regression line (line of best fit to these data), the thin lines are the 95% confidence intervals of the regression line (indicating a 95% probability that by repetition of the experiment the resulting regression line will be within these borders), and the dotted lines are the 95% confidence intervals for new observations.

Table 18. Example of a linear regression. Hemolytic activity of alkanoic acids

Compound	**Log P**	**Log 1/C**
Octanoic acid	-1.20	1.46
Nonanoic acid	-0.70	1.72
Decanoic acid	-0.20	2.37
Undecanoic acid	0.30	2.67
Dodecanoic acid	0.80	3.23
Tetradecanoic acid	1.80	3.93

$x = \log P$ $y = \log 1/C$ n = number of data sets

k = number of x variables (in this case = 1)

$\Sigma x = 0.80$ $\Sigma x^2 = 5.94$

$\Sigma y = 15.38$ $\Sigma y^2 = 43.7136$ $\Sigma xy = 7.0290$

$y = ax + b$

$a = (n \cdot \Sigma xy - \Sigma x \cdot \Sigma y)/(n \cdot \Sigma x^2 - (\Sigma x)^2) = 0.85$

$b = (\Sigma y - a \cdot \Sigma x)/n = 2.45$

$r^2 = (\Sigma xy - \Sigma x \cdot \Sigma y/n)^2/((\Sigma x^2 - (\Sigma x)^2/n) \cdot (\Sigma y^2 - (\Sigma y)^2/n)) = 0.990469$

$r = 0.995$

$s^2 = (1 - r^2) \cdot (\Sigma y^2 - (\Sigma y)^2/n)/(n - k - 1) = 0.0102$

$s = 0.101$

$F = r^2 \cdot (n - k - 1)/(k \cdot (1 - r^2)) = 415.68$

log 1/C = 0.85 log P + 2.45

(n = 6; r = 0.995; s = 0.101; F = 415.68)

$$df/da = 2 \cdot \Sigma(y - ax_1 - bx_2 - c) \cdot x_1 = 0 \tag{115}$$

$$df/db = 2 \cdot \Sigma(y - ax_1 - bx_2 - c) \cdot x_2 = 0 \tag{116}$$

$$df/dc = 2 \cdot \Sigma(y - ax_1 - bx_2 - c) = 0 \tag{117}$$

$$a\Sigma x_1^2 + b\Sigma x_1x_2 + c\Sigma x_1 = \Sigma x_1y \tag{118}$$

$$a\Sigma x_1x_2 + b\Sigma x_2^2 + c\Sigma x_2 = \Sigma x_2y \tag{119}$$

$$a\Sigma x_1 + b\Sigma x_2 + c \cdot n = \Sigma y \tag{120}$$

Most often these equations are written in the form of a matrix:

$$\begin{array}{ccc} a & b & c \\ \left| \begin{array}{ccc} \Sigma x_1^2 & \Sigma x_1x_2 & \Sigma x_1 \\ \Sigma x_1x_2 & \Sigma x_2^2 & \Sigma x_2 \\ \Sigma x_1 & \Sigma x_2 & n \end{array} \right. & & \end{array} \left| = \left| \begin{array}{c} \Sigma x_1y \\ \Sigma x_2y \\ \Sigma y \end{array} \right| \right.$$

Inversion of the symmetrical matrix as shown below or *via* triangularisation [572] gives the values c_{ij}, from which the coefficients a, b, and c of the regression model (eq. 114) are calculated by eqs. 121 – 123.

$$\begin{array}{ccl} \text{matrix} & \text{inverted matrix} & \\ \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} & \begin{vmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \\ c_{31} & c_{32} & c_{33} \end{vmatrix} & \begin{array}{l} \text{for } a_{12} = a_{21}, a_{13} = a_{31} \\ \text{and } a_{23} = a_{32} \\ \text{(symmetrical matrix)} \end{array} \end{array}$$

$$D_{11} = a_{22}a_{33} - a_{23}^2 \qquad D_{22} = a_{11}a_{33} - a_{13}^2$$

$$D_{12} = D_{21} = -(a_{21}a_{33} - a_{23}a_{31}) \qquad D_{23} = D_{32} = -(a_{11}a_{32} - a_{12}a_{31})$$

$$D_{13} = D_{31} = a_{21}a_{32} - a_{22}a_{31} \qquad D_{33} = a_{11}a_{22} - a_{12}^2$$

$$D \text{ (determinant)} = a_{11}D_{11} + a_{12}D_{12} + a_{13}D_{13}$$

$$c_{11} = D_{11}/D \quad c_{12} = D_{12}/D \quad c_{13} = D_{13}/D \quad \ldots \quad c_{ij} = c_{ji} = D_{ij}/D$$

$$a = c_{11}\Sigma x_1 y + c_{12}\Sigma x_2 y + c_{13}\Sigma y \tag{121}$$

$$b = c_{21}\Sigma x_1 y + c_{22}\Sigma x_2 y + c_{23}\Sigma y \tag{122}$$

$$c = c_{31}\Sigma x_1 y + c_{32}\Sigma x_2 y + c_{33}\Sigma y \tag{123}$$

The correlation coefficient r (eq. 124) is a relative measure of the quality of fit of the model because its value depends on the overall variance of the dependent variable (this is illustrated by eqs. 58 – 60, chapter 3.8; while the correlation coefficients r of the two subsets are relatively small, the correlation coefficient derived from the combined set is much larger, due to the increase in the overall variance). The squared correlation coefficient r^2 is a measure of the explained variance, most often presented as a percentage value. The overall (total) variance S_{yy} is defined by eq. 125, the unexplained variance (SSQ = sum of squared error; residual variance; variance not explained by the model) by eq. 126.

$$r^2 = 1 - \Sigma\,\Delta^2/S_{yy} \tag{124}$$

$$S_{yy} = \Sigma(y_{obs} - y_{mean})^2 = \Sigma y^2 - (\Sigma y)^2/n \tag{125}$$

$$\Sigma\,\Delta^2 = SSQ = \Sigma(y_{obs} - y_{calc})^2 \tag{126}$$

Another algorithm (eqs. 127 – 129) is more convenient for the calculation of the correlation coefficient r in computer programs.

$$r^2 = (aS_{1y} + bS_{2y} + \ldots)/S_{yy} \tag{127}$$

$$S_{1y} = \Sigma x_1 y - \Sigma x_1 \cdot \Sigma y/n \tag{128}$$

$$S_{iy} = \Sigma x_i y - \Sigma x_i \cdot \Sigma y/n \tag{129}$$

The standard deviation s (eq. 130) is an absolute measure of the quality of fit, as can be seen from its definition and, *e.g.* from a comparison of the standard deviations of eqs. 58 – 60 (chapter 3.8). Its value considers the number of objects n and the number of variables k. Therefore, s depends not only on the quality of fit

Table 19. Selected F values (DF = n − k − 1; two-sided, 95% significance level)

	k = number of variables in the equation							
DF	**1**	**2**	**3**	**4**	**5**	**....**	**10**	**20**
1	161.4	199.5	215.7	224.6	230.2	....	241.9	248.0
2	18.51	19.00	19.16	19.25	19.30	....	19.40	19.45
3	10.13	9.55	9.28	9.12	9.01	....	8.79	8.66
4	7.71	6.94	6.59	6.39	6.26	....	5.96	5.80
5	6.61	5.79	5.41	5.19	5.05	....	4.74	4.56
....	....	....	....	....	....	....	....	....
10	4.96	4.10	3.71	3.48	3.33	....	2.98	2.77
....	....	....	....	....	....	....	....	....
20	4.35	3.49	3.10	2.87	2.71	....	2.35	2.12
30	4.17	3.32	2.92	2.69	2.53	....	2.16	1.93
60	4.00	3.15	2.76	2.53	2.37	....	1.99	1.75
∞	3.84	3.00	2.60	2.37	2.21	....	1.83	1.57

but also on the number of degrees of freedom, DF = n − k − 1. The larger the number of objects and the smaller the number of variables is, the smaller the standard deviation s will be for a certain value of $\Sigma\,\Delta^2$.

$$s^2 = \frac{\Sigma\,\Delta^2}{n - k - 1} = \frac{(1 - r^2) \cdot S_{yy}}{n - k - 1} \tag{130}$$

The F value (eq. 131) is a measure of the level of statistical significance of the regression model. The number of variables being included to derive the model has an even stronger influence than in the case of the standard deviation s; only F values being larger than the 95% significance limits (Table 19) prove the overall significance of a regression equation.

$$F = \frac{r^2 \cdot (n - k - 1)}{k \cdot (1 - r^2)} \tag{131}$$

The confidence intervals of the regression terms a, b, and c (eq. 114) can be calculated from the standard deviation s, the Student t value (Table 20), and the diagonal terms of the inverted matrix, *e.g.* for a: $\pm s \cdot t \cdot \sqrt{c_{11}}$, for b: $\pm s \cdot t \cdot \sqrt{c_{22}}$, and for c: $\pm s \cdot t \cdot \sqrt{c_{33}}$.

Two different regression models, containing different numbers of variables k_1 (smaller number) and k_2 (larger number), can be compared by a sequential (partial) F test (eq. 132). The use of the model containing the larger number of variables is justified if the resulting partial F value indicates a 95% significance (Table 19) for the introduction of the new variable/s.

$$F = \frac{(r_2^2 - r_1^2) \cdot (n - k_2 - 1)}{(k_2 - k_1) \cdot (1 - r_2^2)} \tag{132}$$

Table 20. Selected t values (DF = n − k − 1; 95% significance level)

DF	$t_{95\%}$
1	12.706
2	4.303
3	3.182
4	2.776
5	2.571
6	2.447
7	2.365
8	2.306
9	2.262
10	2.228
11	2.201
12	2.179
13	2.160
14	2.145
15	2.131
16	2.120
17	2.110
18	2.101
19	2.093
20	2.086
21	2.080
22	2.074
23	2.069
24	2.064
25	2.060
26	2.056
27	2.052
28	2.048
29	2.045
30	2.042
....	....
40	2.021
50	2.009
60	2.000
70	1.994
80	1.990
90	1.987
100	1.984
....	....
∞	1.960

5.2. The Significance and Validity of QSAR Regression Equations

In general, a regression equation can be accepted in QSAR studies,

if the correlation coefficient r is around or better than 0.9 for *in vitro* data and 0.8 for whole animal data (as already discussed, its value depends not only on the quality of fit but also on the overall variance of the biological data; compare eqs. 124–126, chapter 5.1),

if the standard deviation s is not much larger than the standard deviation of the biological data (normally around 0.3, which closely corresponds to the mean error of most biological data; this value may be considerably smaller for *in vitro* data. If natural logarithms of biological activity values are used instead of decadic logarithms, as done in some CoMFA studies, all biological activity values and the standard deviation s are 2.3 times larger),

if its F value indicates that the overall significance level is better than 95%, and

if the confidence intervals of all individual regression coefficients prove that they are justified at the 95% significance level (*i.e.* if their confidence intervals are smaller than the absolute values of the regression coefficients).

In addition, the biological data should cover a range of at least one, better two or even more logarithmic units; they should be well distributed over the whole distance (*i.e.*, no clustering of activity values should occur, as discussed in chapter 2). Also the physicochemical parameters should be spread over a certain range and should be more or less evenly distributed; if a certain parameter has identical values for all but one or two objects, then this parameter must be considered as a hidden indicator variable and should be replaced by such a term. In parabolic and especially in bilinear equations the nonlinear parameter should cover a range of at least two logarithmic units, in order to justify the presence of a nonlinear term.

The equation has to be rejected

if the number of variables included in the regression equation (or used for the selection of variables to be included in the equation) is unreasonably large (see chapter 4.1) or

if the standard deviation s is smaller than the error in the biological data (overprediction by the model).

Eq. 132 (chapter 5.1) is used in automated algorithms to derive a regression model in a stepwise manner (*e.g.* [389]), *e.g.* by backward elimination (starting with all possible variables and in each further step eliminating the variable which is the least significant one), forward selection (starting with the best single variable and adding further significant variables, according to their contribution to the model) and stepwise procedures (using a forward selection procedure with intermediate proof whether already introduced variables are no longer significant at a later stage) (Table 21). While such procedures avoid the testing of hundreds or even thousands of different variable combinations, their results often are ambiguous. In the case of large numbers of (partially interrelated) variables a local optimum may be obtained instead of the global optimum.

Some of the problems in automated stepwise regression analysis can be illustrated by an artificial data set (Table 22). The dependent variable is highly correlated with X-4 ($r = 0.918$), but not with X-1, X-2, or X-3 (all r values <0.25). Stepwise selection of variables first picks X-4 and then stops because a sequential F test indicates no significance of additional variables; multiparameter equations can only be derived if this significance barrier is forced to zero. However, the "best" three-parameter equation contains X-1, X-2, and X-3 ($r = 1.000$), but no longer X-4! Although this example is constructed to demonstrate such an effect, it reflects the situation that all points are located in a plane which is skewed against the X-1, X-2, and X-3 planes; simple projections of the data to the different planes do not indicate any correlations; only the combination of all three variables gives a perfect fit.

A more suitable procedure in QSAR studies is first to establish a physicochemical model for a small subset of compounds, which is in accordance with the biological mechanism of action. Then more and more compounds can be added by introducing new variables for additional structural variation (*e.g.* eqs. 133 – 135) or by combining different subsets with the help of indicator variables (*e.g.* eqs. 58 – 60, chapter 3.8). Eqs. 133 – 135 describe the inhibition of chymotrypsin by various thiophosponates (**22**) [573], starting from a small group of compounds and ending up with the whole series of inhibitors. Eq. 133 includes 19 analogs with structural variation in the ester

Table 21. Stepwise regression analysis of a data set with 4 independent variables (adapted from ref. [389])

Y	**X-1**	**X-2**	**X-3**	**X-4**
78.5	7	26	6	60
74.3	1	29	15	52
104.3	11	56	8	20
87.6	11	31	8	47
95.9	7	52	6	33
109.2	11	55	9	22
102.7	3	71	17	6
72.5	1	31	22	44
93.1	2	54	18	22
115.9	21	47	4	26
83.8	1	40	23	34
113.3	11	66	9	12
109.4	10	68	8	12

Y *vs.* X-1 r = 0.731; s = 10.73; F = 12.60

Y *vs.* X-2 r = 0.816; s = 9.08; F = 21.96

Y *vs.* X-3 r = 0.535; s = 13.28; F = 4.40

Y *vs.* X-4 r = 0.821; s = 8.96; F = 22.80

Correlation matrix of the X variables (r^2 values)

	X-1	X-2	X-3	X-4
X-1	1.000	0.052	0.679	0.060
X-2		1.000	0.019	0.947
X-3			1.000	0.001
X-4				1.000

Stepwise selection of variables

Y *vs.* X-4 r = 0.821; s = 8.96; F = 22.80
Y *vs.* X-1 and X-4 r = 0.986; s = 2.73; F = 176.63
sequential F test: F = 108.22
Y *vs.* X-1, X-2 and X-4 r = 0.991; s = 2.31; F = 166.83
sequential F test: F = 5.03
Y *vs.* X-1 and X-2 r = 0.989; s = 2.41; F = 229.50
sequential F test: F = 1.86

End of the stepwise selection

"best" equation:

Y = 1.468 (±0.27) X-1 + 0.662 (±0.10) X-2 + 52.58 (±5.09)
(n = 13; r = 0.989; s = 2.41; F = 229.50)

Table 22. Regression analysis of a data set with 4 independent variables

Y	X-1	X-2	X-3	X-4
1.0	0	0	1	1
1.0	1	0	0	1
2.0	0	1	1	2
2.0	2	-2	2	3
1.0	0	0	1	1
2.0	1	0	1	2
1.0	1	0	0	1
2.1	0	3	-1	3

Y *vs.* X-1	r = 0.153;	s = 0.586;	F = 0.14
Y *vs.* X-2	r = 0.239;	s = 0.576;	F = 0.36
Y *vs.* X-3	r = 0.096;	s = 0.590;	F = 0.06
Y *vs.* X-4	r = 0.918;	s = 0.236;	F = 31.97

Correlation matrix of the X variables (r^2 values)

	X-1	X-2	X-3	X-4
X-1	1.000	0.527	0.154	0.073
X-2		1.000	0.663	0.030
X-3			1.000	0.002
X-4				1.000

Stepwise selection of variables

Y *vs.* X-4 r = 0.918; s = 0.236; F = 31.97

sequential F test: F = 0.33

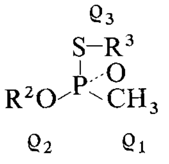
End of the stepwise selection

"best" three-variable equation

Y = 1.027 (±0.05) X-1 + 1.034 (±0.05) X-2 + 1.005 (±0.05) X-3 -
- 0.016 (±0.07)

(n = 8; r = 1.000; s = 0.023; F = 1385.76)

substituent R^2. Then 21 analogs with additional variation in the thioester group R^3 are added (eq. 134). The last equation includes all compounds, the new ones containing an additional charged group (I = 1) in residue R^3 (eq. 135).



$R^2O-P(=O)(CH_3)-S-R^3$ (Q_3, Q_2, Q_1)

(22)

$$\log k_i = 1.60(\pm 0.22)\,MR_2 - 3.85(\pm 1.17)\log(\beta \cdot 10^{MR_2} + 1) - 4.76(\pm 0.51) \quad (133)$$

$\log \beta = -3.86 \qquad \text{optimum } MR_2 = 3.72$

$(n = 19;\ r = 0.978;\ s = 0.258)$

$$\begin{aligned} \log k_i = {} & 1.42(\pm 0.12)\,MR_2 - 3.30(\pm 0.88)\log(\beta \cdot 10^{MR_2} + 1) + \\ & + 0.35(\pm 0.10)\,MR_3 + 1.21(\pm 0.42)\,\sigma_3^* - 5.18(\pm 0.45) \end{aligned} \quad (134)$$

$\log \beta = -3.87 \qquad \text{optimum } MR_2 = 3.75$

$(n = 40;\ r = 0.981;\ s = 0.253)$

$$\begin{aligned} \log k_i = {} & 1.47(\pm 0.10)\,MR_2 - 3.43(\pm 0.74)\log(\beta \cdot 10^{MR_2} + 1) + \\ & + 0.34(\pm 0.09)\,MR_3 + 1.25(\pm 0.19)\,\sigma_3^* - 1.06(\pm 0.31)\,I - \\ & - 5.26(\pm 0.38) \end{aligned} \quad (135)$$

$\log \beta = -3.85 \qquad \text{optimum } MR_2 = 3.71$

$(n = 53;\ r = 0.985;\ s = 0.243)$

Sometimes a certain parameter (*e.g.* a nonlinear term) is only justified by a single activity value. Due to the most often small number of data points being included in a QSAR equation, the best way to deal with this problem is to present both regression equations, one including all variables, the other one eliminating this term and the corresponding activity value which led to its consideration.

Cross-validation, in which objects are eliminated and only the excluded objects are predicted from the resulting model to check its stability and validity (see chapter 5.3 for a detailed description), seems to be a too crude instrument to (automatically) decide on the validity of a QSAR regression equation. Cross-validation may be applied to relatively large data sets. But if only few compounds are included in the QSAR equation, if a certain parameter is mainly based on a single data point, or if the compounds have been selected according to a rational design procedure, *e.g.* a D-optimal design (chapter 6), cross-validation may incorrectly indicate a lack of validity of the QSAR model.

Outliers, *i.e.* data that cannot be explained by the model, constitute a serious problem in QSAR studies. Most often they are omitted from the data set without further comments, which is not a good practice. A lot of information might be derived from the careful inspection and consideration of the residuals of a multiple regression analysis (*e.g.* [574]) and of so-called outliers (*e.g.* [575, 576]).

Some problems related to the proper application of regression analysis and of other multivariate statistical methods in QSAR studies and concerning the validity of the obtained results have recently been reviewed [403, 408, 409] (compare chapter 4.1).

5.3. Partial Least Squares (PLS) Analysis and Other Multivariate Statistical Methods

Discriminant analysis (Figure 31) [41, 487, 577 – 581] separates objects with different properties, *e.g.* active and inactive compounds, by deriving a linear combination of some other features (*e.g.* of different physicochemical properties), which leads to the best separation of the individual classes. Discriminant analysis is also appropriate for semiquantitative data and for data sets, where activities are only characterized in qualitative terms. As in pattern recognition, training sets are used to derive a model and its stability and predictive ability is checked with the help of different test sets.

COMPACT (computer-optimized molecular parametric analysis of chemical toxicity) [582, 583], a discriminant analysis approach, is described to predict carcinogenicity and other forms of toxicity involving the formation of reactive intermediates by determining the structural criteria for substrate specificity towards cytochrome P-450 enzymes; it is claimed that the method is about 75% predictive for rodent carcinogenicity [583]. Recently, a discriminant-regression model was described [584]. It applies stepwise discriminant analysis to form clusters of compounds for which quantitative relationships are derived by multiple regression analysis.

The adaptive least squares (ALS) method [396, 585 – 588] is a modification of discriminant analysis which separates several activity classes (*e.g.* data ordered by a rating score) by a single discriminant function. The method has been compared with ordinary regression analysis, linear discriminant analysis, and other multivariate statistical approaches; in most cases the ALS approach was found to be superior to categorize any numbers of classes of ordered data. ORMUCS (ordered multicategorial classification using simplex technique) [589] is an ALS-related approach which

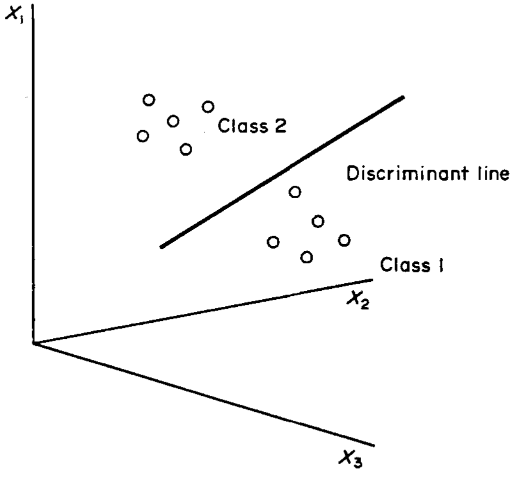


Figure 31: Two classes of objects are separated by a discriminant line (reproduced from Figure 3 of ref. [487] with permission from Pergamon Press Ltd., Headington Hill Hall, Oxford OX3 0BW, UK).

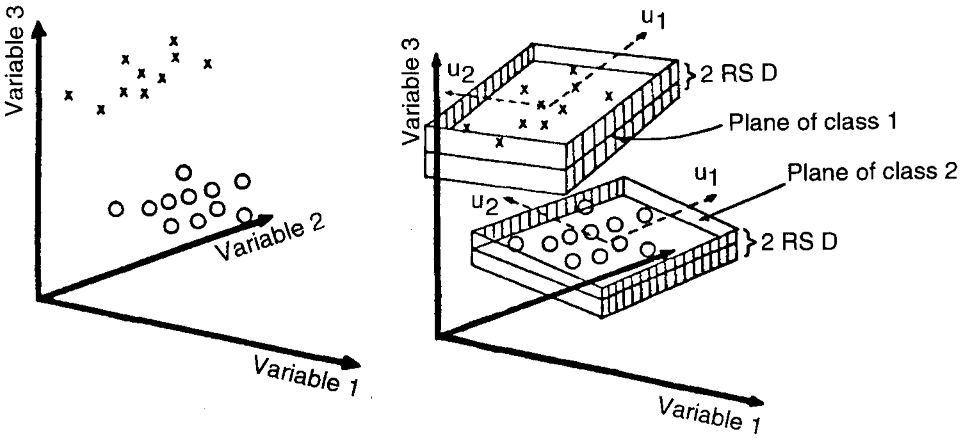


Figure 32: SIMCA description of two different classes. The classification problem (left diagram) and the resulting hyperboxes for each class (right diagram) (reproduced from Figures 1 and 2 of ref. [495] with permission from the American Chemical Society, Washington, DC, USA).

applies a simplex technique for the derivation of the discriminant function. Recently, a fuzzy ALS version was developed and used in QSAR studies [590, 591].

Regression analysis and discriminant analysis are extremely sensitive to interrelated X variables; multivariate statistical methods which reduce the dimensionality of the X block are robust in this respect. Factor analysis (FA) and principal component analysis (PCA) [592—595] derive vectors which are orthogonal and, in decreasing order, contain the maximum amount of information that can stepwise be extracted from the X block. These principal components are then correlated with biological activity values. Applications of FA and PCA in QSAR are illustrated by several studies (*e.g.* [596—600]). Single-class discrimination, using PCA, was used to identify "embedded" classes of active analogs surrounded by diffuse classes of inactive compounds [601]. The application of PCA to extract principal properties, *e.g.* the BC(DEF) parameters, out of a large group of partially interrelated physicochemical properties has already been discussed in chapter 3.

The SIMCA method (SIMCA stands for "similarity, chemistry, and analogy") [407, 487, 495, 499, 602—605] is a class modeling technique which places objects from p-dimensional space into lower dimension boxes. The "box" can have any dimension (*e.g.* a line, plane, or hyperplane); its size is determined by the scatter of the data (Figure 32). Discrimination of objects of different classes is possible by deriving separate principal component models for each class. The NIPALS (nonlinear iterative partial least squares) algorithm [606], an iterative procedure avoiding time-consuming matrix diagonalization, is applied to derive a user-defined, limited number of principal components for each class.

The most promising new approach in multivariate statistical methods is the PLS (partial least squares in latent variables) method [26, 27, 38, 607—610]. Many, even hundreds or thousands of independent variables (the X block) can be correlated with one or several dependent variables (the Y block). PLS analysis is a principal component-like method, with the main difference that the vectors are not indepen-

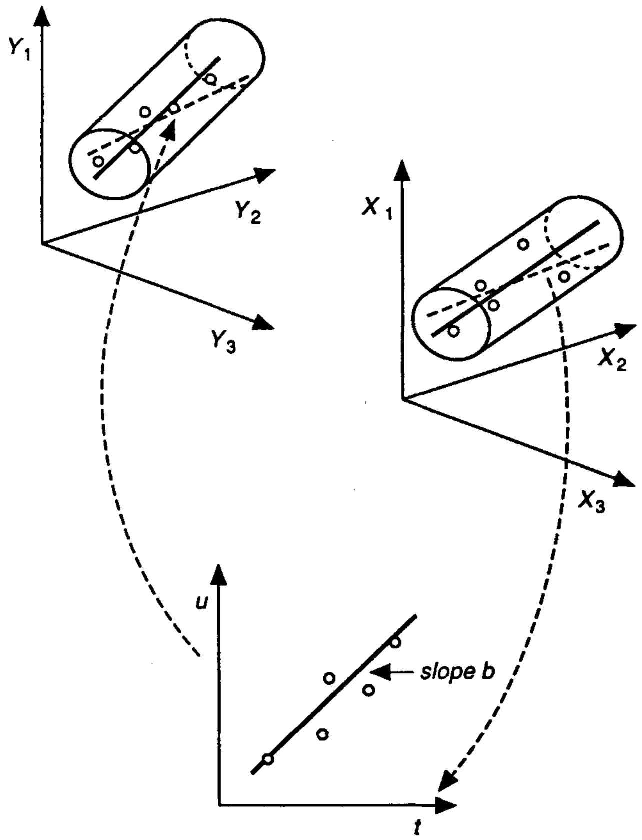


Figure 33: Representation of a PLS regression through the inner relation u = b.t. The solid lines in X- and Y-space are the principal components and the dashed lines are the PLS vectors. These are slightly skewed to account for the correlation between the two data blocks (redrawn from Figure 9 of ref. [487] with permission from Pergamon Press Ltd., Headington Hill Hall, Oxford OX3 0BW, UK).

dently extracted for the X and the Y block. As in the SIMCA method, the NIPALS algorithm is used to derive a certain number of vectors for each block. In contrast to the principal components derived by PCA, the resulting vectors are slightly shifted to their exact positions, in such a manner that the correlation of corresponding X block-derived and Y block-derived vectors is optimized while these vectors are still located in their corresponding boxes (Figure 33). Detailed descriptions of the PLS algorithm have been published [27, 607, 608].

The results of a PLS analysis can be transformed to regression coefficients of the X block variables, most often leading to the curious result that more regression coefficients than objects are obtained. It should be mentioned that in the case of one dependent variable and a number of X variables that equals the number of PLS components, the results from regression analysis and, after appropriate transformation, from PLS analysis are numerically identical.

Depending on the number of components, often perfect correlations are obtained in PLS analyses, due to the usually large number of included X variables. Correspondingly, the quality of fit is no criterion for the validity of a PLS model. A cross-validation procedure (Figure 34) [26, 409, 611] must be used to select the model having the highest predictive ability. In cross-validation many PLS runs are

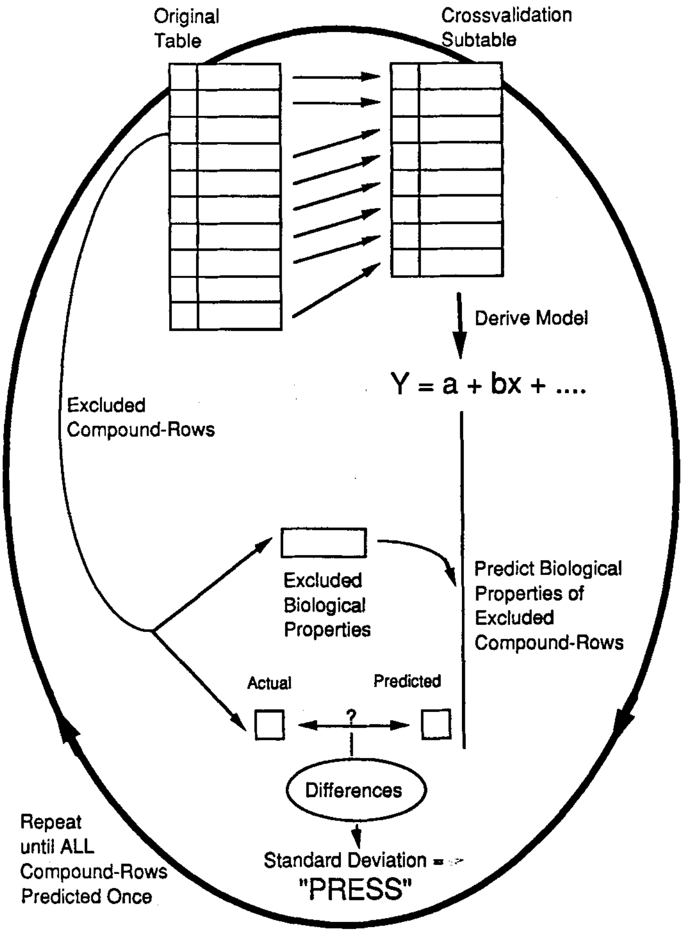

Original Table
Crossvalidation Subtable
Derive Model
Y = a + bx + ...
Excluded Compound-Rows
Excluded Biological Properties
Predict Biological Properties of Excluded Compound-Rows
Actual
Predicted
?
Differences
Standard Deviation = "PRESS"
Repeat until ALL Compound-Rows Predicted Once

Figure 34: Cross-validation procedure (reproduced from Figure 3 of ref. [611] with permission from the copyright owner).

performed in which one (leave-one-out technique) or several objects are eliminated from the data set either randomly or in a systematic manner. Only the excluded objects are predicted by the corresponding model.

The standard deviation s_{PRESS} (calculated from PRESS, the sum of the squared errors of these predictions, divided by the number of degrees of freedom; compare eq. 130, chapter 5.1) [611] is taken as the criterion for the optimum number of components. As long as only significant components are derived in the PLS analysis, PRESS and s_{PRESS} will decrease; if too many components are extracted, overprediction results and PRESS and s_{PRESS} increase. The smallest s_{PRESS} value indicates the optimum number of components. SDEP (standard deviation of the error of predictions) [612, 613] corresponds to s_{PRESS}, the only difference being that the number of degrees of freedom is not considered in the calculation of the SDEP value.

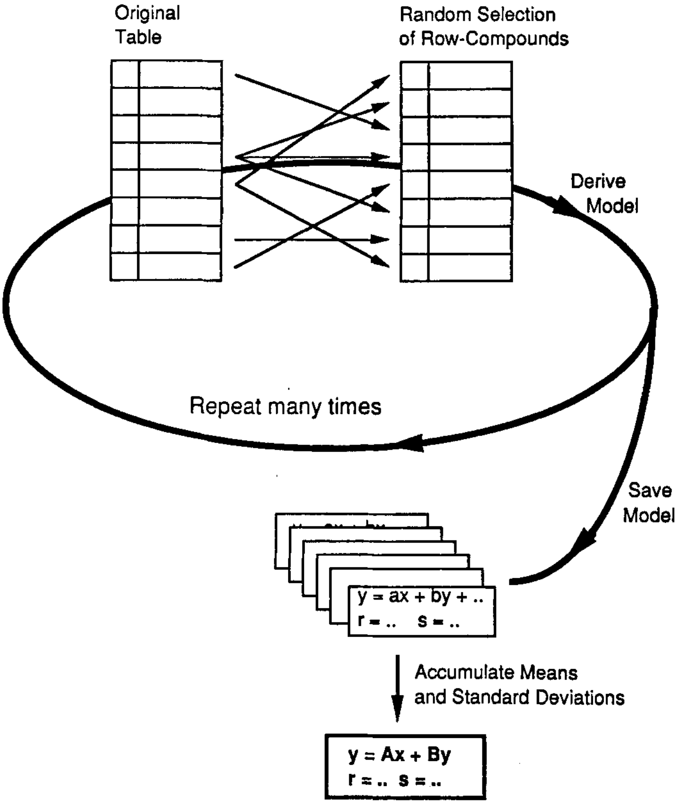


Figure 35: The bootstrapping process (reproduced from Figure 2 of ref. [611] with permission from the copyright owner).

The squared correlation coefficient r^2 is defined in PLS analysis as in regression analysis. Correspondingly, a r^2_{PRESS} value can be formulated for cross-validation runs using PRESS instead of the unexplained variance [611]. Cross-validated r^2 values are always more or less smaller than the r^2 values including all objects (r^2_{FIT}), depending on the quality of the model. In severe cases of overprediction PRESS may become larger than the overall variance of the Y values; then negative r^2_{PRESS} values are obtained, indicating that the predictions from the model are even worse than the y_{mean} values. The significance of cross-validation results has to be commented: in well designed data sets, where a small number of objects is selected to explore the parameter space with a minimum number of objects, cross-validation fails because the eliminated objects cannot be predicted by the model derived from the other objects. On the other hand, the leave-one-out cross-validation procedure gives a much too optimistic result in redundant data sets, where all or at least most objects have close neighbors in multidimensional parameter space. A much more rigorous cross-validation procedure was proposed by Simon [542]. All compounds are ordered according to their biological activity values; odd- and even-numbered analogs are being selected as training and test sets, and *vice versa.*

Bootstrapping (Figure 35) [409, 611] is a procedure in which several times N random selections out of the original set of N objects are performed to simulate different samplings from a larger set of objects. In each run some objects are not included in the PLS analysis, some others are included more than once. Confidence intervals for each term can be estimated from such a procedure, giving an independent measure of the stability of the PLS model.

Although the PLS method is claimed to be a robust modeling technique, experience shows that too many noise variables, *i.e.* variables that do not contribute to prediction, obscure the result. For prediction such additional variables are most often useless or even detrimental. This effect can clearly be demonstrated by different PLS analyses of the data set given in Table 22, using different combinations of the independent variables X-1, X-2, X-3, and X-4 (Table 23); three vectors were chosen as the maximum number of PLS components. If all X variables are included, the results are only slightly better than those obtained from regression analysis using X-4 as the only independent variable (Tables 22 and 23).

Cross-validation does not confirm any of these models; the r^2_{PRESS} values have a maximum at two components but all values are smaller than 0.4. If a fourth component is added, r_{FIT} jumps to 1.000 ($s = 0.007$; $F = 9646.4$; same result as obtained from the regression analysis) and the cross-validated r^2_{PRESS} to 0.992 ($s_{PRESS} = 0.083$). But if X-4 is eliminated from the original data set, PLS analysis gives a much better result. The three-component model indicates high significance ($r = 1.000$; $s = 0.023$; $F = 1385.75$; same result as obtained from the regression analysis) and also high predictive ability, as shown by cross-validation ($r^2_{PRESS} = 0.993$; $s_{PRESS} = 0.070$). PLS analysis of the same data, using X-4 as the only independent variable, gives an identical result as compared to regression analysis; now the cross-validation result is much better ($r^2_{PRESS} = 0.798$; $s_{PRESS} = 0.304$) than the one-component cross-validation result of the original data set, including all X variables ($r^2_{PRESS} = 0.287$; $s_{PRESS} = 0.572$).

A procedure for variable elimination, called GOLPE (generating optimal linear PLS estimations), was developed [38, 614, 615] to solve such problems in cases of extremely large X variable blocks. In GOLPE first a D-optimal design is used to preselect non-redundant variables; a smaller number of variables, having a high degree of orthogonality in multidimensional parameter space, is selected by this procedure. In the next step a fractional factorial design is used to run PLS analyses with different combinations of these variables. The predictive ability of each model is checked by a cross-validation procedure and the effect of every variable can be estimated from a comparison of the PRESS values of the models including this variable and those not containing it. Variables significantly contributing to prediction (*i.e.* lowering PRESS) are kept in the X block, all others are excluded.

Recently, a nonlinear version of PLS analysis was described [616]. The CARSO (computer-aided response surface optimization) procedure [617, 618] aims at obtaining response surfaces for non-designed data sets. Quadratic terms and interaction terms are generated for each independent variable and PLS analysis is used to model the data, due to the fact that regression analysis will fail for data sets with many highly interrelated variables in the X block.

Table 23. PLS analyses of a data set (Table 22), using different combinations of the independent variables

PLS analysis, all X-variables included			
Fit	1 Vector	2 Vectors	3 Vectors
r values	0.926	0.943	0.946
s values	0.224	0.215	0.236
F values	35.95	20.22	11.25
sequential F values		1.50	0.16
Cross-validation	1 Vector	2 Vectors	3 Vectors
r^2_{PRESS}	0.287	0.360	-0.177
s_{PRESS}	0.572	0.594	0.900
PLS analysis, X-4 not included in the analysis			
Fit	1 Vector	2 Vectors	3 Vectors
r values	0.861	0.945	1.000
s values	0.301	0.213	0.023
F values	17.22	20.74	1385.75
sequential F values		7.01	443.63
Cross-validation	1 Vector	2 Vectors	3 Vectors
r^2_{PRESS}	-1.696	-0.270	0.993
s_{PRESS}	1.113	0.836	0.070
PLS analysis, only X-4 included in the analysis			
Fit	1 Vector	2 Vectors	3 Vectors
r values	0.918	-	-
s values	0.236	-	-
F values	31.97	-	-
Cross-validation	1 Vector	2 Vectors	3 Vectors
r^2_{PRESS}	0.798	-	-
s_{PRESS}	0.304	-	-

PLS analysis has a number of important advantages as compared to regression analysis [611]. Nevertheless, only in a few cases has it been used in classical QSAR studies (*e.g.* [27, 619–622]). An explanation might be the complexity of the PLS algorithm and the better availability of computer programs for regression analysis. In addition, no physical meaning is attributed to the components, although the loadings of the X variables and the regression coefficients calculated from these loadings tell about the contributions of the individual X variables; the relative importance of different parameters can be seen at a glance, without the problems that usually are associated with high interrelation of variables and local minima in stepwise regression analysis.

Concerning the application of 3D QSAR methods, *e.g.* comparative molecular field analysis (CoMFA, chapter 9.3), PLS analysis is the method of choice. No other method is suited to correlate one or several dependent variables with thousands of independent variables that are incorporated in this approach. However, PLS analysis of such large data sets also offers some problems. There is clear evidence that, due to the large number of included variables, PLS analysis will not reveal the explanatory effect of an unweighted single variable, even if this variable alone could perfectly correlate the data (compare Table 23). It is hoped that the GOLPE variable selection procedure makes PLS analysis an even more powerful tool in 3D QSAR studies [38].

Cluster analysis separates and arranges different objects into groups, according to their distances in multidimensional space. Different algorithms are used to agglomerate related objects, most often in a hierarchical manner. Its application to the rational selection of substituents in drug design [50, 154] and in QSAR studies (*e.g.* [623 – 626]) has been discussed. The results from cluster analyses and from nonlinear mapping, a method which projects data from multidimensional space to fewer, *e.g.* two dimensions (the nonlinear map), by a principal component-like iterative procedure, have been compared [624].

Cluster significance analysis (CSA) [577, 627 – 631], originally derived from the so-called parameter focusing technique (chapter 6), is a graphical method to look at the clustering of active compounds in a space that is made up of various physicochemical parameters. The advantage of the method is that qualitative or rank-ordered biological data can be used. A significance probability value is calculated to judge whether any cluster must be considered as a chance occurrence or not. After a selection of the descriptors being relevant for class separation by cluster significance analysis, SIMCA can be used to predict the activity classes for new members of a series [631]. Cluster significance analysis is discussed in detail and FORTRAN programs for CSA are given in ref. [577].

QSAR: Hansch Analysis and Related Approaches
by Hugo Kubinyi
© VCH Verlagsgesellschaft mbH, 1993

6. Design of Test Series in QSAR

The statistical analysis of data requires a proper design of experiments to prove or disprove a certain hypothesis which has been formulated in advance. From the viewpoint of a puritanical statistician most QSAR analyses are "forbidden", because they are retrospective studies and, in addition, many different hypotheses (*i.e.* combinations of independent variables) are tested sequentially. Indeed, many problems arise from the application of regression analysis in ill-conditioned data sets. Only in later stages of lead structure optimization are certain hypotheses, *e.g.* on the influence of more lipophilic, electronegative, polar, or bulky substituents in a certain position, systematically tested, now fulfilling the requirements for the proper application of statistical methods.

The most important condition in a QSAR study is the design of a congeneric series, including only compounds having the same mechanism of action. Sometimes both principles are violated. While congenericity is not easy to define, from common experience it is clear that all compounds of a set should have the same molecular frame, *i.e.* an identical parent compound, with structural variation in only one or several positions. Exceptions are compounds having nonspecific biological activities (*e.g.* narcotic activity) which are only caused by their membrane-partitioning properties. Normally, the mechanism of action is not proven for all members within a series. While it is evident that a group of enzyme inhibitors, being active in the nanomolar range, displays the same mechanism of action (even here multiple binding modes may occur and complicate the analyses), in the case of millimolar activities or just toxic, carcinogenic, mutagenic, or teratogenic effects it cannot be presumed that all analogs indeed exhibit the same mechanism of action.

The relationships between structures and specific biological activities are caused by many different physicochemical interactions, they are multivariate. Parameter interrelation in less well designed data sets was recognized as a problem in QSAR analyses from the very beginning (see eqs. 38, 39; chapter 3.5). The first contribution to solve this problem in a proper manner was made by Craig [632]; he proposed to study 2D plots of physicochemical properties (Figure 36) and to select substituents from all different quadrants of the resulting diagram.

This qualitative graphical method is easy to apply and different physicochemical properties can be plotted against each other. Substituents may be chosen by the chemist according to their relative ease of synthesis. If *e.g.* a π *vs.* σ plot is considered, optimum substituents may be H (in the origin of the diagram), Cl or Br (lipophilic electron acceptors), NO_2 (electron acceptor), NH_2 (polar electron donor), $N(Et)_2$ (lipophilic electron donor), and NH(COMe) (polar electron acceptor), all analogs being easily accessible in most series. The Craig method has also been extended to 3D plots [633]. However, the main disadvantage of Craig plots is that they are

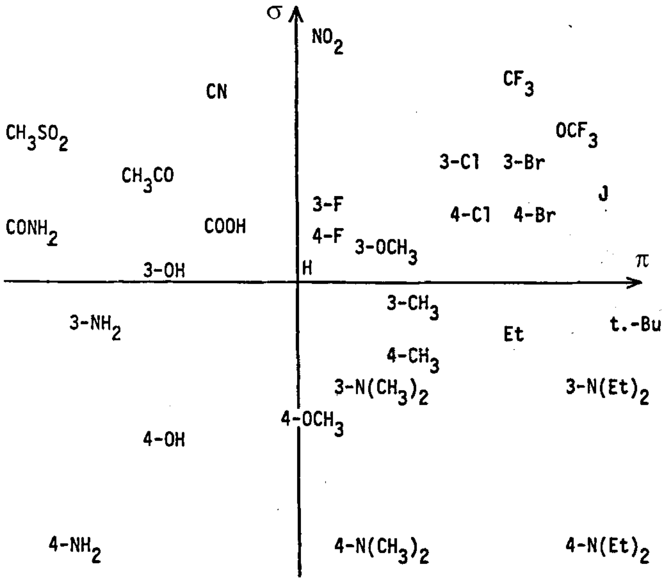


Figure 36: Craig plot of aromatic substituents in a σ vs. π diagram (adapted from Figure 1 of ref. [632] with permission from the American Chemical Society, Washington, DC, USA).

restricted to at most three dimensions while it is *a priori* not known which physicochemical properties will be responsible for the biological activities.

Parameter focusing is a related technique, developed by Magee [634]. Different 2D plots of physicochemical properties are drawn to find out which parameter combination separates active and inactive compounds to the largest extent. Therefore, the method can be applied, in contrast to QSAR analyses, also to qualitative data. Cluster significance analysis (chapter 5.3) was developed from this approach.

A sequential simplex technique was proposed by Darvas [635] to find the most active analog within a series in a stepwise procedure. Any triangle of different substituents may be chosen by starting from a 3D graph where biological activities are plotted against two different physicochemical parameters (Figure 37). A line is drawn from the least active analog in the direction to and beyond the mean value of the other two compounds and a substituent is selected in this area. In the next step the biological activity of this new analog is determined. The least active analog is omitted and it is expected that a new triangle of more active compounds results. The procedure is continued until there is no further increase in biological activities (Figure 38). Although the method has been applied to several examples [635, 636], it has some limitations: first of all, only two physicochemical parameters are considered simultaneously, and secondly, it is only suited for retrospective analyses or for cases in which synthesis is much easier and faster than biological testing.

A real advance in design strategy resulted from the Topliss operational schemes [637, 638]. The Topliss scheme for aromatic substituents (Figure 39) starts from two analogs, *e.g.* a compound bearing an unsubstituted phenyl ring and the corre-

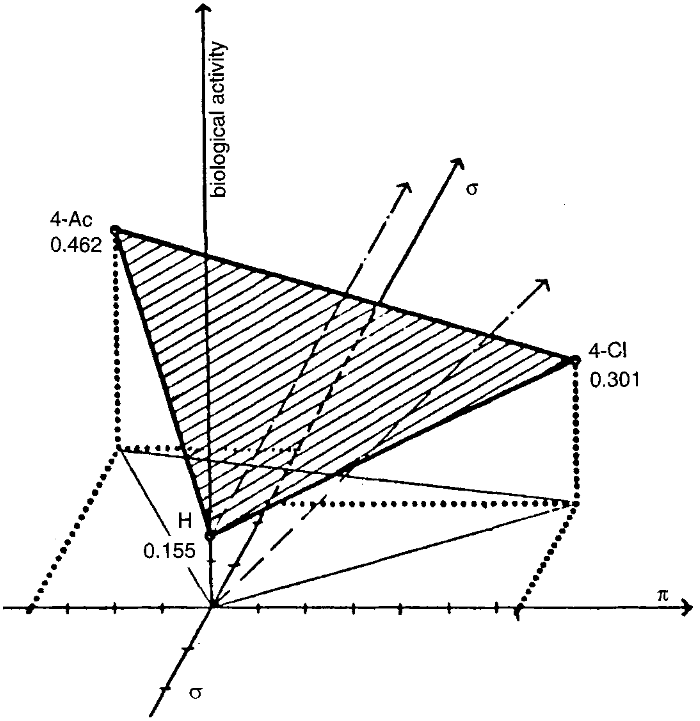


Figure 37: Sequential simplex procedure. A triangle is spanned on the biological activity values (z axis) of the H, 4-Cl, and 4-Ac analogs. The H analog is the least active one and a new analog is searched in the direction between and beyond the mean value of the other two analogs (left dashed line; the right dashed line is the projection onto the π *vs.* σ plane) (reproduced from Figure 4 of ref. [635] with permission from the American Chemical Society, Washington, DC, USA).

sponding *p*-Cl analog. The biological activity of the *p*-Cl analog can be higher, equal, or lower than that of the parent compound. Correspondingly, in the next step more lipophilic analogs (in the case of higher activity of the *p*-Cl analog) or more hydrophilic analogs, *e.g.* *p*-OMe (lower activity of *p*-Cl), are proposed by the operational scheme. If the activity of both analogs is about the same, either the lipophilicity or the electron acceptor properties of the chlorine substituent are unfavorable, while the other property might increase biological activity. Thus, a lipophilic donor substituent, *e.g.* Me, and a not too lipophilic acceptor substituent, *e.g.* the NO_2 group, are the next proposals. If activity is increased in any direction, this guides further syntheses; if not, it has to be considered that large substituents in the *para*-position may be unfavorable and the same procedure as above is applied to the *meta*-position. A corresponding scheme was developed for aliphatic substitution.

A manual method was proposed by Topliss [639] as a modification of his operational schemes: a larger number of substituents is selected in the first step to derive the dependence of biological activities on π (linear and nonlinear) and σ with a minimum number of analogs. This latter approach has been criticized because of collinearity and unbalanced spanning of the parameter space [640].

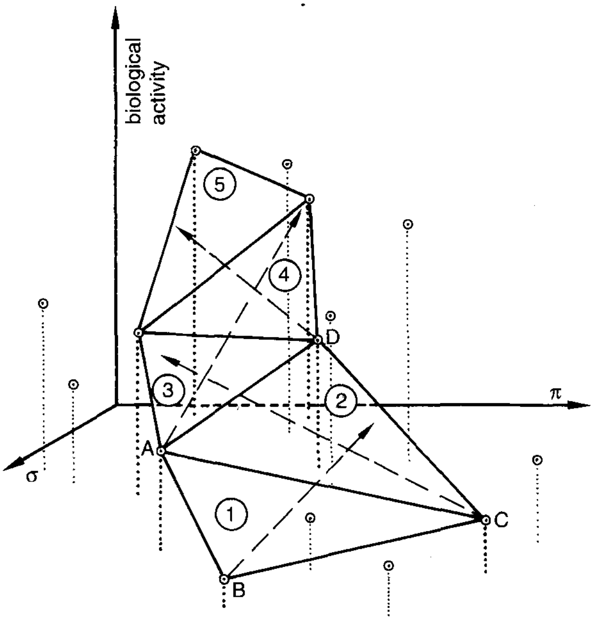


Figure 38: Sequential simplex procedure. In a hypothetical set of compounds a search is made for analogs having the highest biological activity by starting from the compounds A, B, and C (reproduced from Figure 2 of ref. [635] with permission from the American Chemical Society, Washington, DC, USA).

Start with substituents H and 4-Cl

- I. if 4-Cl > H then 3,4-Cl_2
  - I.1. if 3,4-Cl_2 > 4-Cl then 3,4-Br_2, I, CF_3
  - I.2. if 3,4-Cl_2 < 4-Cl then Br, I, 4-NO_2
- II. if 4-Cl < H then 4-OCH_3
  - II.1. If 4-OCH_3 > H then 4-$N(CH_3)_2$
    - a) if 4-$N(CH_3)_2$ > 4-OCH_3 then 4-$N(C_2H_5)_2$
    - b) if 4-$N(CH_3)_2$ < 4-OCH_3 then NH_2, OH
  - II.2. if 4-OCH_3 < H then 3-Cl, *etc.*
- III. if 4-Cl ≈ H then 4-CH_3
  - III.1. if 4-CH_3 > H, 4-Cl then Et, Pr, *i*-Pr, Bu, *t*-Bu
  - III.2. if 4-CH_3 < H, 4-Cl then 3-Cl
    - a) if 3-Cl > 4-CH_3 then 3,5-Cl_2, 3-Br, 3-I
    - b) if 3-Cl < 4-CH_3 then 3-CH_3, 3-$N(CH_3)_2$, 4-F, 4-NO_2, 4-CN, 4-$CONH_2$

Figure 39: Topliss scheme for aromatic substitution (adapted from Scheme 1 of ref. [637] with permission from the American Chemical Society, Washington, DC, USA).

The different approaches proposed by Topliss should not be understood as rigid schemes; they are strategies which have to be adjusted to each problem. A recent review [403] lists more than 50 references where the Topliss methods have been applied, mostly in medicinal chemistry. It was shown that optimum activity would have rapidly been reached in many series of compounds in accordance with the Topliss scheme [641]; on the other hand, there are at least some examples where the Topliss method failed [403, 633].

All stepwise approaches have the disadvantage that iterative synthesis and testing are extremely time-consuming strategies. In most cases a brief study of the parameter table and the correlation matrix provides a QSAR practitioner enough details on the parameter spanning and on the interrelation of the parameters (with the exception of multiple correlations).

2^n-Factorial design techniques are quite common in experimental design [642–644]. They have successfully been applied to series. design in lead structure optimization, using + and −, or +, 0, and − as descriptors for each physicochemical property of the different substituents [645–650]. As compared to graphical methods and the Topliss operational schemes, the advantage of factorial design procedures is that several physicochemical properties can be considered simultaneously. In addition, synthetic accessibility guides the choice of the substituents to be included in the fractional design. The program package EDISFAR [651] extracts physicochemical information out of a database and uses different factorial design techniques to select the substituents.

More objective procedures in series design are clustering methods in multidimensional parameter space; substituents from different clusters are selected for synthesis (chapter 3) [50, 154, 403]. As this approach cannot automatically avoid collinearity or multicollinearity, several different standard sets of aromatic substituents have been proposed (*e.g.* [652, 653]). A distance mapping technique may be used to select further substituents on the basis of a maximum distance to the substituents which already are included [652]. A modification [654] of this approach uses the determinant of the parameter correlation matrix as the criterion for substituent selection.

D-optimal design methods [389, 403, 655, 656] calculate the determinant of the variance-covariance matrix; its value is largest for substituent sets with maximum variance and minimum covariance (linear and multiple correlation) in their physicochemical properties. An information theory approach, which leads to comparable results, has been proposed by Herrmann [657]. In some other approaches synthetic accessibility has been included as an additional selection feature [656, 658, 659].

Other design methods like principal component methods that are combined with multidimensional mapping [163] and a two-dimensional mapping of intraclass correlation matrices [660] are reviewed in [403]. Principal components of properties were also used in fractional design methods (Figure 40) [661, 662] and in D-optimal design [662]. Principal properties of amino acids (so-called z scales) [170, 171] are suited to select minimum analog peptide sets for QSAR studies, according to statistical design methods [663, 664].

Other design strategies have to be applied in Free Wilson analysis [390, 649]. While in Hansch analysis structural modification is sometimes restricted to only one

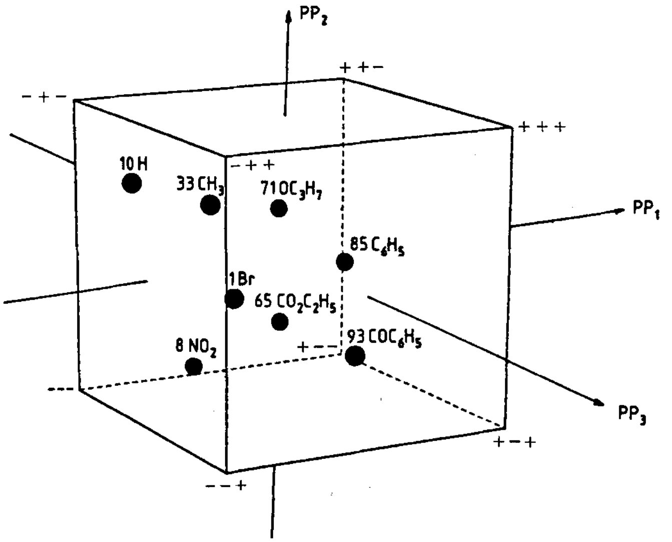


Figure 40: Representation of substituents of Table 2 according to their principal properties. The block represents a two-level factorial design in the three principal components PP_1, PP_2, and PP_3 (reproduced from Figure 1 of ref. [662] with permission from the copyright owner).

position of substitution, including a relatively large number of different substituents, in a Free Wilson analysis at least two different positions of substitution must be included. Each substituent in every position should occur several times (at least twice); otherwise, the group contribution of such a single-point determination contains the whole experimental error of this one biological activity value. Substituents which always occur together must be combined to a pseudosubstituent; only the group contribution of this hypothetical substituent can be derived from the analysis (compare chapter 4.2). Also for Free Wilson analyses a factorial design technique [649] and a quantitative procedure to extract an optimal set out of all possible analogs, based on the maximization of the determinant of the substituent correlation matrix, have been proposed [665].

Experimental design strategies for QSAR studies have been reviewed [38, 403, 648]. They have been compared on the basis of more than 20 different criteria [403] and some of their practical limitations [403, 640, 656] have been discussed.

QSAR: Hansch Analysis and Related Approaches
by Hugo Kubinyi
© VCH Verlagsgesellschaft mbH, 1993

7. Applications of Hansch Analysis

In the original definition of the extrathermodynamic approach [15,17], a linear combination of lipophilic and electronic terms accounted for the drug-receptor interaction, while a nonlinear lipophilicity parameter was included to model the transport and distribution of the drugs. Therefore, Hansch analysis can be used to describe complex biological data, where several different transport processes and equilibria contribute to the overall structure-activity relationship. Correspondingly, all different kinds of biological activities have been correlated with linear free energy-related parameters, using Hansch analysis.

Only in special cases, *e.g.* in enzyme inhibition, is an isolated process measured and described. In the case of membrane-bound receptors, the drug-membrane interaction may influence the ligand binding to the active site (compare chapter 1.2). In other biological systems, like in cells, isolated organs, or whole animal data, much more complex relationships are to be expected, an exception being pharmacokinetic data, where only rate constants or concentrations of different drugs in certain compartments are described in a quantitative manner.

Activity-activity relationships, *i.e.* the comparison of biological activities of a group of compounds in different biological test models, are originally not the domain of QSAR analyses. However, in industrial practice such relationships and their quantitative description are of utmost importance. Instead of wasting thousands of animals, nowadays enzyme inhibition, receptor binding, and cell culture data are used to derive activity profiles of large classes of compounds and to predict the pharmacodynamic effects of new drugs from simple and efficient *in vitro* test models.

Thousands of Hansch equations have been published in the past 30 years. An early review [42] contains several hundred examples of the successful application of Hansch analysis; a recent starting point to collect all relevant information resulted in a database that already contains 3,000 QSAR equations of biological data as well as 3,000 linear free energy relationships in organic chemistry [666]. Such an effort will for the first time allow systematic comparisons of different biological effects, *e.g.* of nonspecific toxicities or of the inhibition of different classes of enzymes, and of the effects of certain groups of compounds on different biological systems; it is evident that comparative structure-activity relationships [667] will provide further insight on the intrinsic nature of drug-receptor interactions.

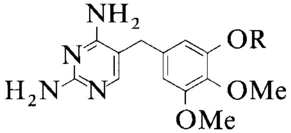
No attempt has been made to present a comprehensive overview of the use of Hansch analysis in medicinal chemistry. Only a subjective selection of typical applications is given in this chapter to demonstrate its proper use and its value for rational drug design. For more examples the reader is referred to refs. [39–44] and to the abstracts services listed in chapter 1.1 [85–88].

7.1. Enzyme Inhibition

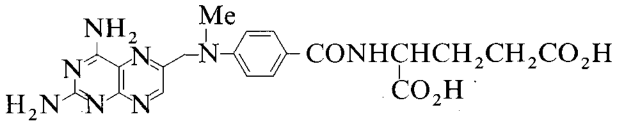
Significant progress in QSAR resulted from Hansch analyses of enzyme inhibitors [432, 456, 668 – 670], especially from the systematic work of Hansch and his group on dihydrofolate reductase and on cysteine and serine proteases. Most of our current knowledge of the quantitative aspects of ligand-protein interactions has been derived from QSAR equations, aided by the interpretation of the 3D structures of enzymes and their inhibitor complexes with molecular graphics [38, 288, 671 – 676].

Dihydrofolate reductase (DHFR) is by far the most extensively investigated enzyme. 3D structures of binary and ternary DHFR complexes from different bacteria and vertebrates have been published and an extremely large number of QSAR equations have been derived, both for the isolated enzyme and for growth inhibition of whole cells [288, 396, 431, 432, 671, 677 – 691]. Due to the central role of DHFR in purine biosynthesis, DHFR inhibitors are therapeutically important as highly selective antibacterial (trimethoprim), antimalarial, and antitumor agents (methotrexate).

DHFR was also the target of one of the very first attempts of a structure-based drug design using crystallographic information. From the chemical similarity of trimethoprim (**23**, R = Me) and methotrexate (**24**) Kuyper and Goodford concluded [692] that replacement of one methoxy group of trimethoprim by an acidic side chain should lead to an additional contact of this newly introduced carboxylate group with the charged guanidinium group of a buried arginine. Indeed, a 50-fold



(23)



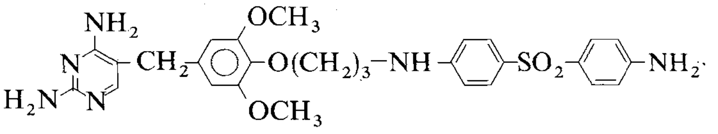
(24)

Table 24. Trimethoprim analogs (**23**) as dihydrofolate reductase inhibitors [692]

R	K_i, nmol
$-CH_3$ (trimethoprim)	1.3
$-CH_2COOH$	2.6
$-(CH_2)_2COOH$	0.37
$-(CH_2)_3COOH$	0.035
$-(CH_2)_4COOH$	0.066
$-(CH_2)_5COOH$	0.024
$-(CH_2)_6COOH$	0.050

Table 25. Cell-free and whole cell inhibitory activities ($\mu mol \cdot l^{-1}$) of trimethoprim (**23**, R = CH_3) and its diphenyl sulfone analog K130 (**25**) [214]

Test system	Trimethoprim	K130
Escherichia coli		
I_{50}, cell-free	0.0045	0.00039
MIC, whole cells	1.50	56
Mycobacterium lufu		
I_{50}, cell-free	0.325	0.034
MIC, whole cells	> 110	1.0

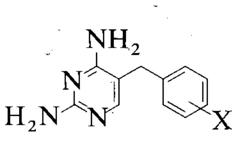


(25)

increase in inhibitory activities was obtained for analogs having the optimum chain length (Table 24); unfortunately, selectivity and membrane permeability significantly decreased.

Seydel [214, 693] followed this approach and designed a hybrid structure (**25**) of trimethoprim and the antileprotic drug 4,4′-diaminodiphenyl sulfone. The new analog is active against *Mycobacterium lufu*, in cell-free systems and against whole cells, most probably due to the replacement of the charged carboxylate group of the inhibitors of Table 24 by the neutral sulfonyl group. On the other hand, although the compound is a potent *in vitro* inhibitor of *Escherichia coli* DHFR, the inhibition of *E. coli* cell growth is low, indicating that both bacteria largely differ in the permeability of their cell walls (Table 25) [214, 693].

Some QSAR equations for 5-(X-benzyl)-2,4-diaminopyrimidines (**26**) as inhibitors of DHFR from different species are given below (eqs. 136–140) [671].



(26)

Escherichia coli DHFR

$$\log 1/K_{i\,app} = 0.75(\pm 0.26)\,\pi_{3,4,5} - 1.07(\pm 0.34)\log(\beta \cdot 10^{\pi_{3,4,5}} + 1) + \\ + 1.36(\pm 0.24)\,MR'_{3,5} + 0.88(\pm 0.29)\,MR'_4 + 6.20 \qquad (136)$$

$\log \beta = 0.12$ optimum $\pi = 0.25$

($n = 43$; $r = 0.903$; $s = 0.290$)

Lactobacillus casei DHFR

$$\log 1/K_{iapp} = 0.31(\pm 0.11)\,\pi_{3,4} - 0.88(\pm 0.24)\log(\beta \cdot 10^{\pi_{3,4}} + 1) + 0.95(\pm 0.21)\,MR'_{3,4} + 5.32 \qquad (137)$$

$\log \beta = -1.33$ optimum $\pi = 1.05$

$(n = 42;\ r = 0.876;\ s = 0.222)$

Chicken liver DHFR

$$\log 1/K_{iapp} = 0.55(\pm 0.19)\,\pi_{3,4,5} - 0.43(\pm 0.35)\log(\beta \cdot 10^{\pi_{3,4,5}} + 1) + 0.20(\pm 0.10)\,MR_3 + 0.32(\pm 0.26)\,\Sigma\sigma + 4.46 \qquad (138)$$

$\log \beta = -0.222$ no optimum $(b < a)$

$(n = 39;\ r = 0.900;\ s = 0.241)$

Bovine liver DHFR

$$\log 1/K_{iapp} = 0.48(\pm 0.11)\,\pi_{3,5} - 1.25(\pm 0.40)\log(\beta \cdot 10^{\pi_{3,5}} + 1) + 0.13(\pm 0.10)\,MR_3 + 0.24(\pm 0.24)\,\Sigma\sigma + 5.43 \qquad (139)$$

$\log \beta = -1.98$ optimum $\pi = 1.52$

$(n = 42;\ r = 0.875;\ s = 0.227)$

Human lymphoblastoid DHFR

$$\log 1/K_{iapp} = 0.59(\pm 0.20)\,\pi_{3,5} - 0.63(\pm 0.59)\log(\beta \cdot 10^{\pi_{3,5}} + 1) + 0.19(\pm 0.14)\,\pi_4 + 0.19(\pm 0.15)\,MR_3 + 0.30(\pm 0.28)\,\Sigma\sigma + 4.03 \qquad (140)$$

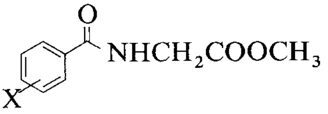
$\log \beta = -0.82$ optimum $\pi = 1.94$

$(n = 38;\ r = 0.879;\ s = 0.266)$

While the equations look the same at a first glance, some striking differences can be seen on a closer inspection. First, the vertebrate, but not the bacterial DHFR equations contain an electronic parameter in addition to lipophilicity and molar refractivity terms. Second, in the case of *L. casei* (eq. 137) the 5-position of the benzyl group does not at all contribute to biological activities. An explanation could be derived by a comparison of the 3D structure of *L. casei* DHFR with the *E. coli* DHFR structure. The active sites of both enzymes are more or less identical in the geometries of the protein backbone and the amino acid side chains. However, there is one significant difference: *E. coli* DHFR contains a methionine side chain in the area where the 5-substituents bind, while there is a relatively rigid leucine side chain in the *L. casei* DHFR which obviously interferes with the 5-substituents. Therefore, the active site of *L. casei* DHFR is sterically more constrained and the positive lipophilicity and polarizability contributions of the 5-substituents are counterbalanced by their steric hindrance [432, 682].

Another well investigated enzyme is the cysteine protease papain (eqs. 58–60, chapter 3.8; eqs. 141–143) [288, 395, 673, 694–700].

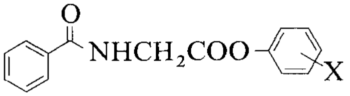
N-(X-Benzoyl)glycine methyl esters (**27**) [694]



(27)

$$\log 1/K_m = 1.01(\pm 0.11)\,\pi + 1.46 \qquad (141)$$
$$(n = 16;\ r = 0.981;\ s = 0.165)$$

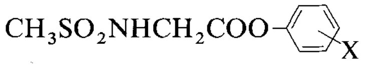
N-Benzoylglycine X-phenyl esters (**28**) [695]



(28)

$$\log 1/K_m = 1.03(\pm 0.25)\,\pi'_3 + 0.57(\pm 0.20)\,\sigma + {} + 0.61(\pm 0.29)\,MR_4 + 3.80(\pm 0.17) \qquad (142)$$
$$(n = 25;\ r = 0.907;\ s = 0.208)$$

N-Mesylglycine X-phenyl esters (**29**) [673]



(29)

$$\log 1/K_m = 0.61(\pm 0.09)\,\pi'_3 + 0.55(\pm 0.20)\,\sigma + {} + 0.46(\pm 0.11)\,MR_4 + 2.00(\pm 0.12) \qquad (143)$$
$$(n = 32;\ r = 0.945;\ s = 0.178)$$

A comparison of eqs. 141, 142, and 143 and an inspection of the 3D structure of papain show that the phenyl ester group binds in a polar environment, while the substituents of the N-benzoyl group (eq. 141) are located in hydrophobic space. This hydrophobic interaction explains the different intercepts in the case of the benzamides (eq. 142) as compared to the much more polar mesylamides (eq. 143; eqs. 58–60, chapter 3.8).

A QSAR study of a larger series of N-(X-benzoyl)glycine pyridyl esters revealed significant differences for $\log 1/K_m$ and $\log k_{cat}$ values ($\log 1/K_m$ *vs.* π_4: $n = 22$; $r = 0.946$; $s = 0.176$. $\log k_{cat}$ *vs.* σ: $n = 23$; $r = 0.933$; $s = 0.094$. $\log k_{cat}/K_m$ *vs.* π_4 and σ: $n = 22$; $r = 0.926$; $s = 0.193$) [700]. Possible explanations for the differences to equations having been derived earlier were discussed.

Hydrophilic *meta*-substituents at the phenyl ester group are not parametrized in eqs. 142 and 143; only the more hydrophobic substituent is considered in the case of two different *meta*-substituents. It seems that the phenyl ring turns around by

180° if hydrophilic *meta*-substituents are repelled by a hydrophobic area at the binding site. The smaller coefficient of the π term in eq. 143, as compared to eq. 142, has been explained correspondingly [288].

Similar equations were derived for the closely related cysteine hydrolases actinidin [288, 696–698], bromelain [694, 697, 698], and ficin [697, 698].

N-Benzoylglycine X-phenyl esters (**28**, eq. 142) were also investigated as inhibitors of the serine proteases chymotrypsin (eq. 144) [288, 675] and trypsin (eq. 145) [288, 701].

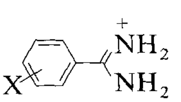
$$\log 1/K_m = 0.28(\pm 0.06)\,\pi'_3 + 0.42(\pm 0.08)\,\sigma^- + 3.87(\pm 0.05) \qquad (144)$$

$$(n = 28;\ r = 0.945;\ s = 0.081)$$

$$\log 1/K_m = 0.71(\pm 0.17)\,\sigma + 3.31(\pm 0.09) \qquad (145)$$

$$(n = 10;\ r = 0.961;\ s = 0.100)$$

The inhibition of trypsin by benzamidines (**30**) and 2-naphthylamidines (**31**) is described by eqs. 146 and 147, respectively [288, 702].

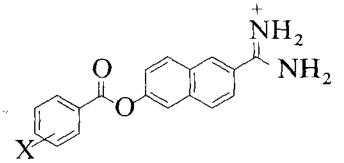


(30)

$$\begin{aligned}\log 1/K_i = {} & -0.59(\pm 0.49)\,MR_4 + 0.88(\pm 0.52)\log(\beta \cdot 10^{MR_4} + 1) + \\ & + 0.23(\pm 0.07)\,\pi'_3 - 0.74(\pm 0.20)\,\sigma + 0.20(\pm 0.30)\,\text{I-M} + \\ & + 0.65(\pm 0.22)\,\text{I-1} + 0.43(\pm 0.19)\,\text{I-2} + 0.51(\pm 0.15)\,\text{I-3} + \\ & + 1.38(\pm 0.28) \qquad (146)\end{aligned}$$

$$\log \beta = -0.73 \qquad \text{optimum } MR_4 = 1.03$$

$$(n = 104;\ r = 0.924;\ s = 0.222)$$



(31)

$$\log 1/K_i = 0.47(\pm 0.19)\,MR_4 - 1.40(\pm 0.40)\,\sigma + 2.59(\pm 0.24) \qquad (147)$$

$$(n = 21;\ r = 0.915;\ s = 0.322)$$

Out of many other Hansch analyses that were derived for different enzymes, only investigations on

ADP-ribosyltransferase [703],
alcohol dehydrogenase [288, 704–707],
butyrylcholinesterase [708],
carbonic anhydrase [288, 709, 710],
chymotrypsin (eqs. 133–135, chapter 5.2) [288, 456, 573, 669, 711],
cytosine nucleoside deaminase [669],

DNA polymerase [712],
β-glucosidase [713],
glutamate dehydrogenase [669],
glyceraldehyde 3-phosphate dehydrogenase [669],
glycolic acid oxidase [714],
guanine deaminase [668, 715],
lactate dehydrogenase [669],
lipoxygenase [716],
malate dehydrogenase (eq. 36, chapter 3.4) [293, 669],
monoamine oxidase (eq. 112, chapter 4.5) [175, 717],
phenol sulfotransferase [718],
thiopurine methyltransferase [719],
thromboxane synthase [720],
thymidilate synthase [669],
thymidine phosphorylase [669],
HSV-thymidine kinase [721],
uridine phosphorylase [669], and
xanthine oxidase [668, 722]

shall be mentioned here. A recent compilation of QSAR studies of enzyme inhibitors [670] reviews more than 400 Hansch equations.

7.2. Other *in vitro* Data

Specific and nonspecific binding of drugs to proteins (other than enzymes) show significant differences. While the serum albumin binding of miscellaneous neutral compounds can be explained by their lipophilicity (eq. 148) but not by their polarizability ($\log 1/C$ *vs.* MR: $r = 0.307$) [18, 723], the highly specific, polar binding of phenyl β-D-glucosides to concanavalin A is described much better by a polarizability term MR (eq. 149) than by their lipophilicity ($\log M_{50}$ *vs.* π: $r = 0.664$) [724].

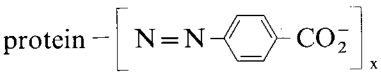
$$\log 1/C = 0.751(\pm 0.07) \log P + 2.300 \qquad (148)$$

$$(n = 42;\ r = 0.960;\ s = 0.159)$$

$$\log M_{50} = 0.019(\pm 0.003)\,\mathrm{MR} + 2.23 \qquad (149)$$

$$(n = 19;\ r = 0.954;\ s = 0.038)$$

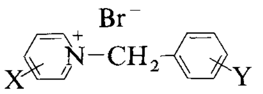
Structure-activity relationships in immunochemistry [21, 335, 396, 725–728] reveal the importance of steric interactions in the QSAR of hapten-antibody (**32**) interactions (eq. 150) [396, 725]; eq. 151 could be derived for the 50% inhibition of complement by benzylpyridinium ions (**33**) [727].



(32)

$$\log K_{rel} = 0.864(\pm 0.07)\,E_s^{ortho} + 0.084(\pm 0.07)\,E_s^{meta} - \\ - 0.436(\pm 0.08)\,E_s^{para} + 0.166(\pm 0.07)\,\Sigma\pi - 0.715(\pm 0.12) \qquad (150)$$

$(n = 22;\ r = 0.989;\ s = 0.120)$



(33)

$$\log 1/C = 0.16(\pm 0.03)\,\pi\text{-}1 + 0.38(\pm 0.10)\,\pi\text{-}2 + 0.86(\pm 0.23)\,\sigma^{+}\text{-}1 + \\ + 0.69(\pm 0.09)\,D\text{-}1 + 0.33(\pm 0.14)\,D\text{-}2 + 2.59(\pm 0.09) \qquad (151)$$

$(n = 132;\ r = 0.953;\ s = 0.197)$

Systematic investigations have been performed for more than one decade on muscarinic receptor ligands [729–732]. Examples of QSAR equations for a combined set of rigid and flexible agonists (eq. 152) [729] and for a group of structurally different antagonists (eq. 153) [730] are given below.

$$pD_2 = 1.13(\pm 0.19)\,\pi_{X_5} - 0.42(\pm 0.08)\,B_4 + 0.81(\pm 0.06)\,I\text{-}X_{3,5,5'} + \\ + 5.70(\pm 0.14) \qquad (152)$$

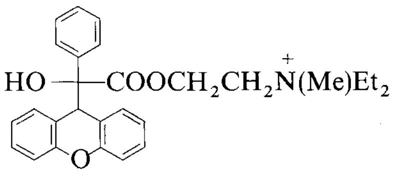
$(n = 62;\ r = 0.968;\ s = 0.196)$

$$\log 1/C = 1.69(\pm 0.32)\,\Sigma\pi_{1,2} - 2.42(\pm 0.50)\log(\beta \cdot 10^{\Sigma\pi_{1,2}} + 1) + \\ + 0.66(\pm 0.20)\,E_s + 1.37(\pm 0.23)\,I\text{-}2 - 0.93(\pm 0.54)\,\Sigma\sigma^{*}_{3,4,5} + \\ + 3.07(\pm 1.92)\,\Sigma MR_{3,4,5} - 0.51(\pm 0.35)\,(\Sigma MR_{3,4,5})^2 - \\ - 4.12(\pm 2.84) \qquad (153)$$

$\log \beta = -3.92$ optimum $\Sigma\pi_{1,2} = 4.3$
optimum $\Sigma MR_{3,4,5} = 3.0$

$(n = 64;\ r = 0.921;\ s = 0.365)$

A QSAR equation for the acetylcholine receptor affinity of quaternary ammonium compounds was used to predict the potency of a new, structurally different analog (**34**) [733]. The compound was synthesized and tested; however, the prediction turned out to be completely wrong [734]; observed and predicted affinities differed by 6 log units (!), once again showing the risk of predictions for structurally dissimilar compounds which are too far outside the included parameter range. A reanalysis of the data led to an equation which gave a much better prediction of the affinity of this compound (error: 1.4 log units) [403].

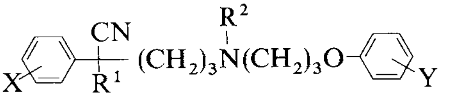


(34)

QSAR studies of
H_1 receptor antagonists [735],
serotonin antagonists and uptake inhibitors [736] as well as of
receptor and thyroxine binding protein (TBP) affinity of thyroid hormone analogs [737] have been reviewed.

In addition to these reviews, Hansch analyses for binding affinities to the
β-adrenergic receptor [738, 739],
benzodiazepine receptor [740],
tetrachlorodibenzodioxin (TCDD) receptor [741, 742],
dopamine receptor [743], and
estrogen receptor [744] as well as for
dopamine-, norepinephrine-, and serotonin-uptake inhibition [745]
shall be mentioned here. Some more QSAR studies of receptor agonists and antagonists are discussed and reviewed in ref. [98].

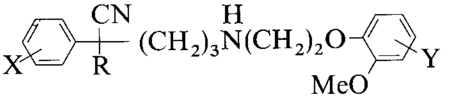
Many QSAR studies have been published for calcium antagonists [349, 396, 746–753]. Eq. 154 was derived for verapamil-type compounds (**35**) [753]; a chemically closely related series of compounds (**36**) turned out to be potent α-adrenergic antagonists (eq. 155) [753]. While it is difficult to decide on the biological significance of these multiparameter equations, one must at least consider the complexity of such drug-receptor interactions.



(35)

$$\begin{aligned} pA_2 = & -0.31(\pm 0.08)\,\Sigma\pi^2 + 0.95(\pm 0.30)\,\Sigma\pi - 0.93(\pm 0.41)\,\Delta MR^2_{meta}(A) + \\ & + 1.45(\pm 0.53)\,\Delta MR_{meta}(A) - 0.30(\pm 0.16)\,\Delta L^2_{para}(A) + \\ & + 0.76(\pm 0.37)\,\Delta L_{para}(A) - 1.74(\pm 0.59)\,F_{ortho}(B) - \\ & - 0.59(\pm 0.17)\,\Delta B_5^{para}(B) + 0.36(\pm 0.26)\,\pi_{para}(B) + \\ & + 1.45(\pm 0.61)\,\Delta B_1(Q) + 5.45(\pm 0.54) \end{aligned} \qquad (154)$$

$(n = 75;\ r = 0.89;\ s = 0.33;\ F = 24.97)$



(36)

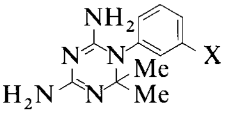
$$\begin{aligned} pA_2 = & -0.22(\pm 0.12)\,\Sigma\pi^2 - 0.17(\pm 0.13)\,\Sigma\pi + 0.35(\pm 0.31)\,I(A) + \\ & + 0.99(\pm 0.29)\,I_{ortho}(B) - 0.53(\pm 0.18)\,\Delta B_5^{meta}(B) - \\ & - 0.65(\pm 0.16)\,\Delta L_{para}(B) + 6.70(\pm 0.34) \end{aligned} \qquad (155)$$

$(n = 59;\ r = 0.92;\ s = 0.37;\ F = 48.25)$

Nonspecific hemolytic, antibacterial, and antifungal *in vitro* activities generally follow linear [18, 182], parabolic [18, 19], and bilinear [23, 345] lipophilicity-activity relationships. Many of the bilinear equations (*e.g.* eq. 100, chapter 4.4) have correlation coefficients r close to unity [345]. It seems that at least in some cases biological activities only depend on the critical micelle concentration (CMC) of the compounds [754]. QSAR studies of antibacterial sulfa drugs (eq. 48, chapter 3.5) [396, 755] and some other examples of specific antibacterial activities [41, 42, 396] have been reviewed.

Much work has been done to elucidate the molecular mechanisms of drug resistance [756]. Comparative QSAR equations were derived for the inhibition of bacterial DHFR, bacterial cell cultures [213, 214, 288, 432, 677, 685], and different tumor cell lines [288, 432, 671, 678, 679, 681, 684, 687], all being sensitive and resistant to methotrexate (MTX), *e.g.* eqs. 156 – 161.

Inhibition of MTX-sensitive *Lactobacillus casei* DHFR by 3′-X-1-phenyl-*s*-triazines (**37**) [677].



(37)

$$\log 1/C = 0.53(\pm 0.10)\,\pi_3 - 0.67(\pm 0.35)\log(\beta \cdot 10^{\pi_3} + 1) + \\ + 0.79(\pm 0.25)\,MR' + 3.13(\pm 0.15) \qquad (156)$$

$\log \beta = -3.46$ optimum $\pi_3 = 4.03$

$(n = 28;\ r = 0.949;\ s = 0.302)$

Inhibition of MTX-resistant *Lactobacillus casei* DHFR by 3′-X-1-phenyl-*s*-triazines (**37**) [677].

$$\log 1/K_{i\,app} = 0.52(\pm 0.10)\,\pi' - 0.64(\pm 0.24)\log(\beta \cdot 10^{\pi'} + 1) + \\ + 1.80(\pm 0.40)\,I + 0.68(\pm 0.63)\,\sigma - 0.27(\pm 0.28)\,MR_Y + \\ + 2.94(\pm 0.25) \qquad (157)$$

$\log \beta = -3.68$ optimum $\pi' = 4.33$

$(n = 44;\ r = 0.958;\ s = 0.308)$

Inhibition of MTX-sensitive *Lactobacillus casei* cell culture by 3′-X-1-phenyl-*s*-triazines (**37**) [288, 677].

$$\log 1/C = 0.80(\pm 0.15)\,\pi'_3 - 1.06(\pm 0.27)\log(\beta \cdot 10^{\pi'_3} + 1) - \\ - 0.94(\pm 0.39)\,MR_Y + 0.80(\pm 0.56)\,I + 4.37(\pm 0.19) \qquad (158)$$

$\log \beta = -2.45$ optimum $\pi'_3 = 2.94$

$(n = 34;\ r = 0.929;\ s = 0.371)$

Inhibition of MTX-resistant *Lactobacillus casei* cell culture by 3′-X-1-phenyl-*s*-triazines (**37**) [288].

$$\log 1/C = 0.42(\pm 0.05)\,\pi + 1.09(\pm 0.33)\,I - 0.48(\pm 0.24)\,MR_Y + \\ + 3.39(\pm 0.14) \qquad (159)$$

estimated optimum $\pi \approx 6$

$(n = 38;\ r = 0.960;\ s = 0.274)$

Inhibition of MTX-sensitive murine leukemia (L5178Y) tumor cells by 3′-X-phenyl-1-*s*-triazines (**37**) [684]

$$\log 1/C = 1.40(\pm 0.23)\,\pi - 1.65(\pm 0.26)\log(\beta \cdot 10^{\pi} + 1) + 0.88(\pm 0.57)\,\sigma + 0.52(\pm 0.20)\,I - 0.25(\pm 0.24)\,OR + 0.63(\pm 0.33)\,DO + 7.94(\pm 0.21) \quad (160)$$

$\log \beta = -0.054$ optimum $\pi = 0.89$

$(n = 64;\ r = 0.904;\ s = 0.298)$

Inhibition of MTX-resistant murine leukemia (L5178Y) tumor cells by 3′-X-phenyl-1-*s*-triazines (**37**) [684].

$$\log 1/C = 0.63(\pm 0.20)\,\pi - 0.26(\pm 0.25)\log(\beta \cdot 10^{\pi} + 1) - 0.17(\pm 0.07)\,MR - 0.33(\pm 0.24)\,OR + 5.11(+0.19) \quad (161)$$

$\log \beta = -0.718$ estimated optimum $\pi \approx 5 - 6$

$(n = 61;\ r = 0.878;\ s = 0.335)$

While the lipophilicity optima of eqs. 156 and 157 are identical, the most striking differences between eqs. 158 and 160 on the one hand and eqs. 159 and 161 on the other hand are the much higher lipophilicity optima of the MTX-resistant cells; even larger differences were obtained in other cases [679, 681, 684]. From this evidence Hansch concluded that a change in the membrane properties should be responsible for the MTX resistance. Seydel [213, 214] found differences in the dose response curves of *Escherichia coli* strains, that are sensitive and resistant to trimethoprim (TMP), to be responsible for TMP resistance. Lipophilic analogs still bind to the enzyme, but they are no longer antagonists; the higher, amphiphilic analogs are nonspecific membrane-perturbing agents. Chloroquine resistance and multidrug resistance were explained in the same manner [213].

Multidrug resistance (MDR) is the acquired resistance of tumor cells to a wide variety of structurally diverse, polar and lipophilic, small and large antitumor drugs, caused by a single antitumor agent. MDR is associated with an over-expression of a membrane-embedded active transport glycoprotein GP-170, which steadily eliminates the antitumor drugs from the resistant cells. The QSAR of multidrug resistance shows linear dependences as well as nonlinear ones on lipophilicity and on the size of the molecules (*e.g.* eq. 56, chapter 3.7) [327, 346, 348]; the nonlinear dependence of MDR on the size of the molecules was explained by the inability of medium-sized molecules to reenter the cells either by passive diffusion (pathway of the small molecules) or by endocytosis (pathway of the large molecules) [346].

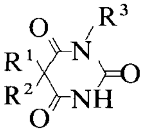
Another extensively investigated field in QSAR are mutagenic agents [325–328, 757–760]. The QSAR equations of a series of 1-(X-phenyl)-3,3-dialkyltriazenes show that mutagenic activity (and presumably carcinogenicity) can be minimized with relatively little loss in antitumor potency. While such hints are useful, they should not be overemphasized: lipophilicity optima can be significantly different in isolated cells and in whole animals.

Recent studies [325–328] resulted in a revival of quantum-chemical indices; their important advantage is that much more heterogeneous sets, *e.g.* aromatic and heteroaromatic compounds, can be combined in a single QSAR equation (*e.g.* eq. 50, chapter 3.5).

7.3. Pharmacokinetic Data

Pharmacokinetics describes the time dependence of transport and distribution of a drug in the different compartments of a biological system, *e.g.* by rate constants of absorption, blood and tissue levels, and metabolism and elimination rate constants. Quantitative structure-pharmacokinetics relationships [433, 442, 451, 452, 472, 761–766] investigate the structural dependence of such parameters within groups of chemically related compounds.

Model simulations (see chapter 4.4) substantiate that the lipophilicity dependence of the rate constants of drug transport should follow bilinear relationships [41, 156, 175, 345, 440, 442]. Indeed, bilinear equations have been derived for the rate constants of drug transport in *n*-octanol/water (eqs. 95–98, chapter 4.4) [444–447] and for the rate constants of the transfer of various barbiturates (**38**) in a Sartorius absorption simulator® from an aqueous phase (pH = 3) through an organic membrane to another aqueous phase (pH = 7.5), modeling the gastric absorption of these compounds (Figure 41) (eq. 162; recalculated optimum log P value) [442].



(38)

$$\log k_{abs} = 0.949(\pm 0.06)\log P - 1.238(\pm 0.11)\log(\beta P + 1) - 3.131 \qquad (162)$$
$$\log \beta = -1.271 \qquad \text{optimum } \log P = 1.79$$
$$(n = 23;\ r = 0.992;\ s = 0.081;\ F = 389.66)$$

The buccal absorption of homologous alkanoic acids can be described by eqs. 163 (at pH = 3.1–3.6) and 164 (at pH = 6) [442].

$$\log k_{abs} = 0.339(\pm 0.08)\log P - 0.318(\pm 0.13)\log(\beta P + 1) - 1.246 \qquad (163)$$
$$\log \beta = -2.450$$
$$(n = 8;\ r = 0.995;\ s = 0.030;\ F = 124.68)$$

$$\log k_{abs} = 0.862(\pm 0.16)\log P - 0.661(\pm 0.20)\log(\beta P + 1) - 3.015 \qquad (164)$$
$$\log \beta = -2.282$$
$$(n = 8;\ r = 0.998;\ s = 0.034;\ F = 431.21)$$

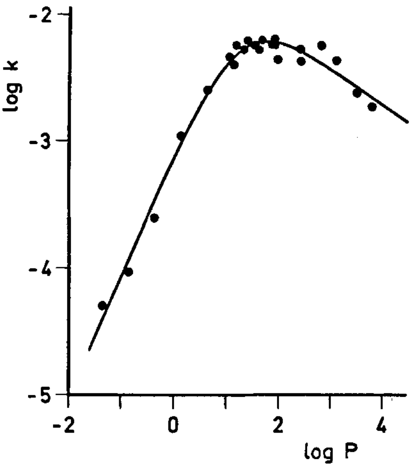


Figure 41: Diffusion rate constants log k of barbiturates (**38**) in the Sartorius absorption simulator® (eq. 162) (reproduced from Figure 11 of ref. [442] with permission from Editio Cantor Verlag GmbH, Aulendorf, Germany).

Much more complex models correlate the buccal absorption rate constants of these acids at different pH values, *e.g.* eq. 165 [479]. Even the rate constants for acidic and basic drugs can be combined to obtain one equation by using log P_{app} instead of log P (eq. 111, chapter 4.5).

$$\log k_{abs} = \log \frac{P^b(af_u + ef_i)}{1 + cP^b(af_u + ef_i)} \tag{165}$$

$a = 0.054(\pm 0.01)$, $b = 0.523(\pm 0.03)$, $c = 2.435(\pm 0.15)$, $e = 0.000212(\pm 0.00007)$

$(n = 71;\ r = 0.97;\ s = 0.12)$

Gastric and intestinal absorption rate constants of neutral carbamates are described by eqs. 166 and 167 [442], respectively. The slopes of the resulting curves are relatively flat because the gastric and intestinal walls are no true lipid membranes but relatively polar, pore-containing tissues (Figure 42). The lipophilicity optimum of the intestinal absorption is shifted to the left because the polar glycocalyx inside the intestinal wall prevents the absorption of lipophilic compounds. On the other hand, the intestinal absorption rate constants are much larger than the gastric absorption rate constants (as can be seen from the intercepts of eqs. 166 and 167 and from Figure 42), due to the larger and morphologically more differentiated surface structure of the intestinal wall.

$$\log k_{ABS} = 0.138(\pm 0.06) \log P - 0.228(\pm 0.16) \log(\beta P + 1) - 2.244 \tag{166}$$

$\log \beta = -1.678$ optimum $\log P = 1.87$

$(n = 8;\ r = 0.971;\ s = 0.030;\ F = 22.14)$

$$\log k_{ABS} = 0.234(\pm 0.10) \log P - 0.502(\pm 0.15) \log(\beta P + 1) - 0.786 \tag{167}$$

$\log \beta = -0.621$ optimum $\log P = 0.56$

$(n = 8;\ r = 0.989;\ s = 0.031;\ F = 61.10)$

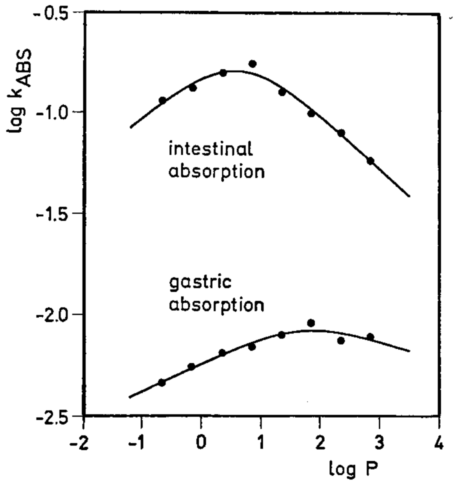


Figure 42: Gastric and intestinal absorption rate constants log k_{ABS} of homologous *n*-alkyl carbamates $R-OCONH_2$ (R = methyl to *n*-octyl; eqs. 166 and 167) (reproduced from Figure 14 of ref. [442] with permission from Editio Cantor Verlag GmbH, Aulendorf, Germany).

The colonic absorption of different acidic compounds is described by eq. 110 (chapter 4.5) [442].

The blood-brain barrier is a tight layer of endothelial cells which inhibits the passage of hydrophilic compounds from the blood into the central nervous system (CNS). For various groups of CNS-active drugs lipophilicity optima for blood-brain barrier penetration at log P values around 2.1 [767], 1.8 – 2.0 [768, 769], and 1.4 – 2.7 [472, 770] have been reported [433, 771]. A dedicated investigation of the extraction of neutral ^{11}C-labeled compounds into adult baboon brain showed that compounds having log P values between 0.9 and 2.5 were completely extracted [772]. The rat-brain capillary permeability coefficients P_c of a wide variety of compounds, ranging from water (MW = 18) to bleomycin (MW = 1400), can be described by eq. 168 [771], giving evidence for the importance of the size of the molecules (approximated by log MW, compare eq. 56, chapter 3.7) for diffusion and pore transport.

$$\log P_c = 0.50(\pm 0.10) \log P - 1.43(\pm 0.58) \log MW - 1.84 \qquad (168)$$

$$(n = 23;\ r = 0.927;\ s = 0.461)$$

The blood-brain barrier penetration of H_2 receptor antihistaminics has been correlated with Δ log P, the difference between *n*-octanol/water and cyclohexane/water partition coefficients (eqs. 23 and 24, chapter 3.1) [199, 200]. An even better correlation could be obtained by using Λ_{alk}, a hydrogen-bonding capability parameter, and V_M, the van der Waals volume (eq. 169) [773].

$$\log (C_{brain}/C_{blood}) = -0.338(\pm 0.03)\, \Lambda_{alk} + 0.007(\pm 0.001)\, V_M + $$
$$+ 1.730(\pm 0.30) \qquad (169)$$

$$(n = 20;\ r = 0.934;\ s = 0.290;\ F = 58)$$

Hansch emphasized the importance of lipophilicity for central nervous system-mediated and other nonspecific (*e.g.* toxic) side effects of drugs [771]; hydrophobic drugs do not only readily pass the blood-brain barrier (*e.g.* sedative side effects of most antihistaminics), they are also slowly eliminated from the biological system, are more inhibitory to biochemical systems than hydrophilic compounds, induce cytochrome P-450 [774], and reactive species may be formed in their metabolism. Thus, without convincing evidence to the contrary, drugs should be made as hydrophilic as possible [771].

Also the placenta has a barrier for hydrophilic and very lipophilic compounds, which is comparable to the blood-brain barrier. Correspondingly, quantitative relationships could be derived for the placental transfer ratios TR of various drugs (eq. 170, recalculated) [775]. The diffusion of drugs into milk and prostatic fluid has been reviewed [472, 761].

$$\log TR = 0.354(\pm 0.06) \log P - 0.469(\pm 0.13) \log(\beta P + 1) - \\ - 0.116(\pm 0.07) \qquad (170)$$

$$\log \beta = -0.658 \qquad \text{optimum } \log P = 1.15$$

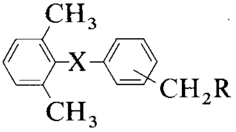
$$(n = 21;\ r = 0.949;\ s = 0.106;\ F = 51.17)$$

The most comprehensive review of quantitative structure-pharmacokinetics relationships [452] tabulates about 100 equations, including absorption, distribution, protein binding, elimination, and metabolism of drugs. Since many of these equations and those included in other reviews (*e.g.* [472, 761]) have been derived before appropriate mathematical models for nonlinear lipophilicity-activity relationships (chapter 4.4) and for the correct consideration of the dissociation and ionization of acids and bases (chapter 4.5, especially eqs. 107–110) were available, some of the older results should be recalculated by using the theoretical models (chapters 4.4 and 4.5) instead of the empirical ones.

7.4. Other Biological Data

Numerous kinds of biological activities have been correlated with physicochemical properties; only a few selected examples will be discussed here to show the wide variety of quantitative relationships that were derived in different therapeutic fields.

The local anesthetic activities of lidocaine analogs (**39**, eq. 171) have been compared with their acute toxicities (eq. 172) [776]. Possible reasons for the differences between both equations were discussed, but no structural proposals were derived for local anesthetics with lower toxicity.



(39)

$$\begin{aligned}\log 1/ED_{50} = {} & 1.24(\pm 0.90)\log P_{app} - 1.80(\pm 1.13)\log(\beta P_{app} + 1) - \\ & - 0.92(\pm 0.27)\,\text{I-NHCO} + 0.36(\pm 0.20)\,\text{I-DIET} + \\ & + 3.22(\pm 1.40)\end{aligned} \tag{171}$$

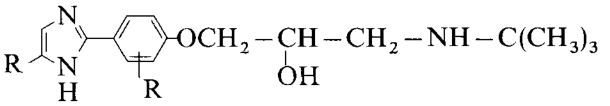
$\log \beta = -1.90 \qquad \text{optimum } \log P_{app} = 2.24$

$(n = 59;\ r = 0.808;\ s = 0.332)$

$$\begin{aligned}\log 1/LD_{50} = {} & 0.18(\pm 0.07)\log P_{app} - 0.30(\pm 0.16)\,\text{I-OCOOC} - \\ & - 0.22(\pm 0.12)\,\text{I-MOR} + 3.86(\pm 0.18)\end{aligned} \tag{172}$$

$(n = 60;\ r = 0.762;\ s = 0.192)$

Antiadrenergic activities of α-bromophenethylamines are described by eqs. 65–68 (chapter 4.1). Eqs. 173 and 174 correlate the β_1- and β_2-antagonistic activities (measured in dogs) of a series of 4-imidazol-2'-yl-phenoxypropanolamines (**40**) [777]; MR must be about 2 for maximum β_1 antagonism (eq. 173), while electron-acceptor substituents increase the β_2-antagonistic potencies (eq. 174).



(40)

$$\begin{aligned}pD_2(\beta_1) = {} & 0.99(\pm 0.43)\,MR - 0.25(\pm 0.12)\,MR^2 + 0.58(\pm 0.26)\,I_s + \\ & + 6.22(\pm 0.34)\end{aligned} \tag{173}$$

$\text{optimum } MR = 1.98$

$(n = 15;\ r = 0.84;\ s = 0.25)$

$$pD_2(\beta_2) = 0.63(\pm 0.26)\,\pi + 1.11(\pm 0.99)\,\sigma_p + 0.50(\pm 0.46)\,I_s + 4.88(\pm 0.44) \tag{174}$$

$(n = 15;\ r = 0.82;\ s = 0.46)$

The structure-activity relationships of antimalarial drugs attracted much attention in the early years of QSAR [778, 779] because of an extensive program of the Walter Reed Army Institute for Medical Research during the Vietnam war. Eq. 175 correlates the antimalarial activities of a large group of structurally diverse phenanthreneaminoalkylcarbinols and related analogs in mice [779]; this equation is one of the two examples which include more than 500 analogs in one equation (the other one being eq. 180, see below).

$$\begin{aligned}\log 1/C = {} & 0.576(\pm 0.09)\,\Sigma\sigma + 0.168(\pm 0.05)\,\Sigma\pi + 0.105(\pm 0.05)\log P - \\ & - 0.167(\pm 0.07)\log(\beta P + 1) - 0.169(\pm 0.10)\,\text{c-side} + \\ & + 0.319(\pm 0.136)\,CNR_2 - 0.139(\pm 0.06)\,AB - \\ & - 0.795(\pm 0.06) < \text{3-cures} + 0.278(\pm 0.11)\,\text{MR-4'-Q} + \\ & + 0.252(\pm 0.18)\,\text{Me-6,8-Q} + 0.084(\pm 0.10)\,\text{2-Pip} + \\ & + 0.151(\pm 0.19)\,\text{NBrPy} - 0.683(\pm 0.22)\,\text{Q2P378} + \\ & + 0.267(\pm 0.11)\,\text{Py} + 2.726(\pm 0.15)\end{aligned} \tag{175}$$

$\log \beta = -3.959 \qquad \text{optimum } \log P = 4.19$

$(n = 646;\ r = 0.898;\ s = 0.309)$

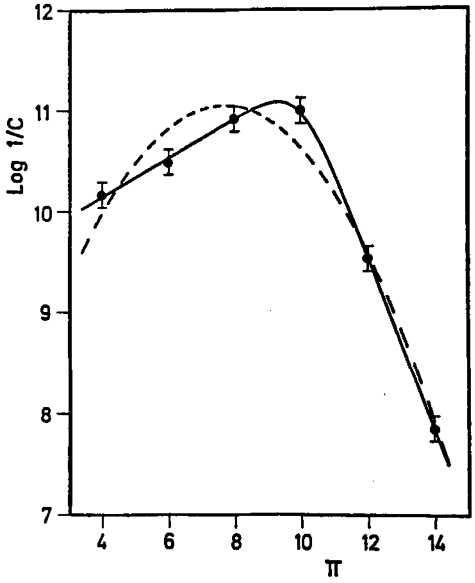
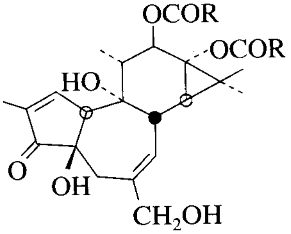


Figure 43: Inflammatory activities of phorbol-12,13-diesters (**41**, R = *n*-propyl to *n*-tridecyl); comparison of the parabolic model (r = 0.978; dashed line) with the bilinear model (r = 1.000; solid line) (redrawn from Figure 3 of ref. [440] with permission from Editio Cantor Verlag GmbH, Aulendorf, Germany).



(41)

Inflammatory activities (eq. 176; topical application at the mouse ear) of phorbol esters (**41**, R = *n*-propyl to *n*-tridecyl) [23, 440, 442] are correlated with their lipophilicity values, giving evidence that in addition to some specific mechanism of action the skin permeation of the compounds is responsible for their biological activity. This example may be criticized due to the small number of compounds. However, these phorbol esters are amongst the compounds having highest *in vivo* activities (up to 10^{-11} mol · kg^{-1}) and also standard deviations of the biological activity values have been determined. In addition, the compounds cover an extremely wide range of lipophilicity values (10 logarithmic units!), clearly showing a typical bilinear lipophilicity-activity dependence (Figure 43).

$$\log 1/C = 0.193(\pm 0.04)\pi - 1.054(\pm 0.09)\log(\beta \cdot 10^{\pi} + 1) + 9.373(\pm 0.30) \qquad (176)$$

$$\log \beta = -9.983 \qquad \text{optimum } \pi = 9.33$$

$$(n = 6;\ r = 1.000;\ s = 0.041;\ F = 1390)$$

Also the relative tumor-promoting activities [nrtpa] of phorbol esters (**41**) have been correlated with their lipophilicity (eq. 177) [780].

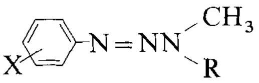
$$\log [\mathrm{nrtpa}] = 0.441(\pm 0.09) \log P - 0.738(\pm 0.16) \log(\beta P + 1) - \\ - 2.571(\pm 0.32) \qquad (177)$$

$$\log \beta = -5.026 \qquad \text{optimum } \log P = 5.20$$

$$(n = 42;\ r = 0.853;\ s = 0.326;\ F = 33.80)$$

Quantitative structure-activity relationships of antitumor drugs have been reviewed [781–783]. Out of many studies performed by Hansch *et al.* (*e.g.* [784–789]) some are discussed here to demonstrate how QSAR results can be used to decide on the probability of success of further research.

The antitumor activities of aniline mustards against Walker 256 solid tumor and L1210 and P388 leukemia are correlated with the hydrolysis rate constants of these compounds; also acute toxicity parallels their antitumor efficacy [784]. The antileukemic activities (L 1210 in mice) of 1-(X-aryl)-3,3-dialkyltriazenes (**42**) can be described by eq. 178 [785], indicating that *ortho*-substituents reduce antileukemic activity, while electron donor substituents X increase it. However, from the hydrolysis rate constants of these compounds (eq. 179) follows that the introduction of electron donor substituents produces unstable drugs; the 4′-OCH_3 analog has a half-life of only about 12 min [785]. In addition, the acute toxicities of these compounds are correlated with their antitumor activities; thus, no more syntheses and testing of new analogs were recommended on the basis of this comparative QSAR study [786].



(42)

$$\log 1/C = 0.100(\pm 0.08) \log P - 0.042(\pm 0.02)(\log P)^2 - \\ - 0.312(\pm 0.11)\, \Sigma\sigma^+ - 0.178(\pm 0.08)\, \mathrm{MR}\text{-}2,6 + \\ + 0.391(\pm 0.18)\, E_s\text{-}R + 4.124(\pm 0.27) \qquad (178)$$

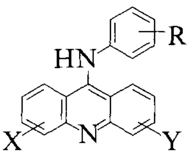
$$\text{optimum } \log P = 1.18$$

$$(n = 61;\ r = 0.836;\ s = 0.191)$$

$$\log k_X/k_H = -4.42(\pm 0.29)\, \sigma - 0.016(\pm 0.13) \qquad (179)$$

$$(n = 14;\ r = 0.995;\ s = 0.171)$$

Eq. 180 was derived for the antitumor activities of 9-anilinoacridines (**43**) [789]. The coefficients of the $\Sigma\pi$ terms indicate that activities fall off more rapidly for the hydrophilic analogs; as the parent compound has a log P value of about 4.8, the lipophilicity optimum can be estimated to be close to log P = 0.



(43)

$$
\begin{aligned}
\log 1/D_{50} = {} & 0.63(\pm 0.27)\,\Sigma\pi - 0.75(\pm 0.23)\log(\beta_1 \cdot 10^{\Sigma\pi} + 1) - \\
& - 1.01(\pm 0.09)\,\Sigma\sigma - 1.21(\pm 0.36)\,R_{BS} - 0.26(\pm 0.16)\,MR_2 + \\
& + 4.95(\pm 0.75)\,MR_3 - 5.13(\pm 0.86)\log(\beta_2 \cdot 10^{MR_3} + 1) - \\
& - 0.67(\pm 0.12)\,I_{3,6} - 1.67(\pm 0.20)\,E_s\text{-}3' - 1.57(\pm 0.21)\,(E_s\text{-}3')^2 + \\
& + 0.58(\pm 0.13)\,I\text{-}NO_2 + 0.87(\pm 0.31)\,I_{DAT} + 0.52(\pm 0.17)\,I_{BS} + \\
& + 9.24(\pm 1.33)
\end{aligned}
\tag{180}
$$

$$
\begin{array}{ll}
\log \beta_1 = 5.64 & \text{optimum } \Sigma\pi = -4.93 \\
\log \beta_2 = 0.01 & \text{optimum } MR_3 = 1.44 \\
 & \text{optimum } E_s\text{-}3' = -0.53
\end{array}
$$

$(n = 509;\ s = 0.893;\ s = 0.305)$

QSAR studies on hallucinogens [790] and on drugs acting at the central nervous system [791] have been reviewed; together both reviews contain about 260 QSAR equations. Also the QSAR of steroids, displaying a wide variety of different biological activities, has been reviewed [792].

Some more examples of quantitative structure-activity relationships are contained in textbooks [40–44] and in dedicated reviews (*e.g.* [18, 396]). QSAR publications are regularly reviewed in the abstracts section of the journal *Quantitative Structure-Activity Relationships* [85] and by some other abstracts services [86–88].

7.5. Activity-Activity Relationships

Activity-activity relationships, *i.e.* the comparison of activities of a group of compounds in different biological test systems, are the orphans in QSAR research. No systematic investigations of the dependence of activity-activity relationships on the physicochemical properties of drugs have been performed, although such studies would be of utmost importance for drug research in pharmaceutical industry. In industrial lead structure optimization hundreds or even thousands of compounds are tested in simple *in vitro* models, *e.g.* in enzyme inhibition, receptor binding, cell culture test models, and isolated organs. *In vivo* activities of the compounds are then estimated from the activity-activity relationships which were obtained from a few standard compounds in more complex test models, *e.g.* in whole animals. In this way the lives of legions of test animals are saved.

Most often linear relationships are obtained between different types of biological activities within a group of related compounds, provided that both activities are caused by the same mechanism of action and that drug transport and distribution do not predominate (*e.g.* Figure 44) [132, 396, 793–795].

Most often lipophilicity plays an important role if different compartments of a biological system or if *in vitro* and *in vivo* data are compared. From a statistical point of view a problem of such activity-activity relationships is that both sides of the equation contain variables including experimental error.

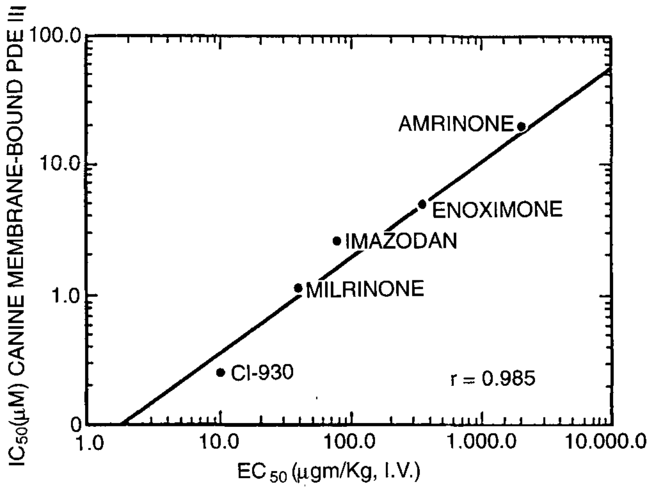
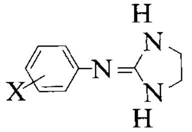


Figure 44: Activity-activity relationship of a series of phosphodiesterase III inhibitors. The *in vitro* inhibitory effects on membrane-bound cAMP-specific phosphodiesterase are plotted against the *in vivo* positive inotropic responses to these drugs in the anesthetized dog (reproduced from Figure 3 of ref. [795] with permission from the American Heart Association, Dallas, TX, USA).

Eqs. 181 and 182 correlate hypotensive and bradycardic activities of α_2-adrenergic clonidine analogs (**44**) [796, 797]. While no log P term was necessary for the ED_{20}(bradycardia) values (eq. 181), it had to be included to describe ED_{20}(hypotension) values by ED_{10}(bradycardia) values (eq. 182) [396].



(44)

$$\log 1/ED_{20}(\text{hypotension}) = 1.118(\pm 0.15) \log 1/ED_{20}(\text{bradycardia}) - 0.824 \quad (181)$$

$$(n = 14;\ r = 0.979;\ s = 0.177;\ F = 283.01)$$

$$\log 1/ED_{20}(\text{hypotension}) = 1.039(\pm 0.26) \log 1/ED_{10}(\text{bradycardia}) + 0.566(\pm 0.19) \log P - 1.511 \quad (182)$$

$$(n = 14;\ r = 0.976;\ s = 0.199;\ F = 111.2)$$

For the same class of compounds Timmermans determined log P_{app} values (n-octanol/buffer, pH = 7.4), binding affinities to α_1 adrenoceptors, $IC_{50}\alpha_1$ (displacement of the α_1 antagonist prazosin), binding affinities to α_2 adrenoceptors, $IC_{50}\alpha_2$ (displacement of the α_2 agonist clonidine), antihypertensive activities (mediated by a central mechanism) in anesthetized normotensive rats ($ED_{25\%}$, i.v. application), and hypertensive activities (mediated by peripheral α stimulation which, in the absence of central nervous system regulation, causes blood vessel contraction) in pithed rats ($ED_{60\,mm}$, i.v. application) [798, 799]. The pC_{60} values are directly related to $IC_{50}\alpha_2$ values (eq. 183), because the site of application of the drugs and the site of action are identical. On the other hand, a parabolic lipophilicity relationship had to be included to describe the relationship between the antihypertensive activities (pC_{25} values) and the $IC_{50}\alpha_2$ values, because the drugs must cross the blood-brain barrier to achieve their central effect. A slightly better description of the data is obtained if the bilinear model is used instead (eq. 184) [800, 801]. As a consequence, the antihypertensive effect and the undesired hypertensive side effect can be correlated by eq. 185.

$$pC_{60} = 1.163(\pm 0.21)\log 1/IC_{50}\alpha_2 - 0.962(\pm 0.39) \tag{183}$$

$$(n = 21;\ r = 0.936;\ s = 0.317;\ F = 135.15)$$

$$\begin{aligned} pC_{25} = {} & 0.805(\pm 0.22)\log P - 3.373(\pm 1.02)\log(\beta P + 1) + \\ & + 1.071(\pm 0.20)\log 1/IC_{50}\alpha_2 - 1.164(\pm 0.39) \end{aligned} \tag{184}$$

$$\log\beta = -1.986 \qquad \text{optimum } \log P = 1.48$$

$$(n = 21;\ r = 0.971;\ s = 0.284;\ F = 65.22)$$

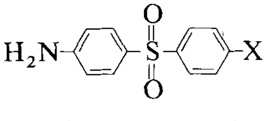
$$\begin{aligned} pC_{25} = {} & 0.784(\pm 0.26)\log P - 3.685(\pm 1.39)\log(\beta P + 1) + \\ & + 0.830(\pm 0.20)\,pC_{60} - 0.189(\pm 0.30) \end{aligned} \tag{185}$$

$$\log\beta = -2.078 \qquad \text{optimum } \log P = 1.51$$

$$(n = 21;\ r = 0.954;\ s = 0.354;\ F = 40.52)$$

Compounds with log P values around 1.5 display the highest antihypertensive selectivity, because they easily penetrate the blood-brain barrier. Much more important is the result that both biological effects can be predicted with a high degree of accuracy from the simple -binding assay (eq. 183), combined with lipophilicity measurements (eq. 184). While the determination of pC_{25} and pC_{60} values of a single compound needs about twenty to fifty (and even more) rats, the brain homogenate of only one rat is sufficient to measure the *in vitro* binding affinities of a large number of analogs.

Similar equations have been derived for bacterial systems, comparing the folate biosynthesis inhibition I_{50} of a series of 4'-substituted 4-aminodiphenyl sulfones (**45**) in a cell-free system with *Escherichia coli* and *Mycobacterium smegmatis* (eq. 186) cell culture inhibition, MIC (log k' is a HPLC capacity factor, closely related to lipophilicity, expressed by π: n = 15; r = 0.994) [596].



(45)

$$\begin{aligned} \log 1/MIC = {} & 0.729(\pm 0.38)\,pI_{50} + 0.933(\pm 0.73)\log k' - \\ & - 1.438(\pm 0.68)\log(\beta k' + 1) + 1.532(\pm 2.33) \end{aligned} \tag{186}$$

$$\log\beta = -0.35 \qquad \text{optimum } \log k' = 0.621$$

$$(n = 17;\ r = 0.899;\ s = 0.329;\ F = 9.22)$$

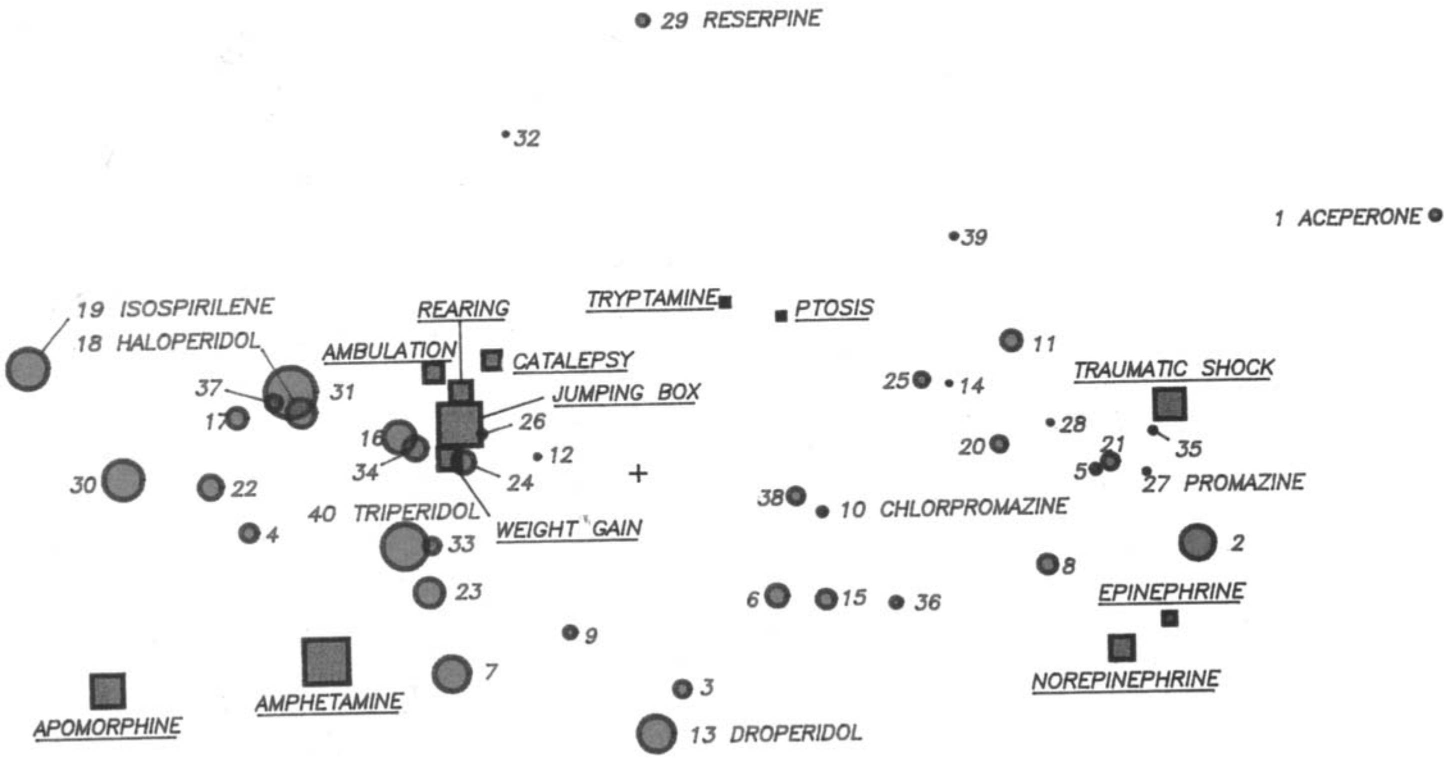


Figure 45: Spectral mapping diagram of neuroleptic drugs. Circles represent the compounds and squares the different test models. The sizes of the symbols indicate the relative activities of the drugs and the sensitivities of the test models; numbers refer to Figure 3 of ref. [802] (diagram provided by courtesy of Dr. Paul Lewi, Janssen Pharmaceutica, Beerse, Belgium).

A spectral mapping technique, based on principal component analysis, has been developed for the two-dimensional interpretation of multidimensional activity data. It was successfully applied to characterize the activity profiles of drugs according to their effects in different pharmacological test models (*e.g.* Figure 45) [802 – 807].

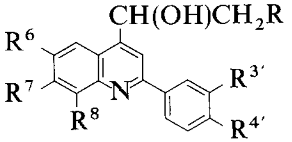
Principal component analysis and QSAR were used to analyze various biological test systems for the quantification of ecotoxic compounds [808]. The QSAR of *in vitro* approaches for developing non-animal methods to substitute the *in vivo* LD_{50} test has been reviewed [809].

QSAR: Hansch Analysis and Related Approaches
by Hugo Kubinyi
© VCH Verlagsgesellschaft mbH, 1993

8. Applications of Free Wilson Analysis and Related Models

The Free Wilson model never became as popular as the Hansch model, due to the limitations inherent in its applicability and predictive ability. Only a few hundred studies [390, 391] have been published since 1964.

From a Free Wilson analysis of antimalarial 2-phenylquinolinylmethanols (**46**), including 40 substituents in 6 different positions ($n = 69$; $r = 0.905$; $s = 0.359$; 34 variables, corresponding to $69 - 34 - 1 = 34$ degrees of freedom), Hansch equations could be derived for the group contributions in different positions of the molecule, *e.g.* in the 4′-position (eq. 187), 7-position (eq. 188), and the 8-position (eq. 189) [810].



(46)

$$a_{4'} = 0.220(\pm 0.35)\,\pi + 0.626(\pm 0.94)\,\sigma_{meta} - 0.232(\pm 0.23) \quad (187)$$

$$(n = 7;\ r = 0.895;\ s = 0.133;\ F = 8.06)$$

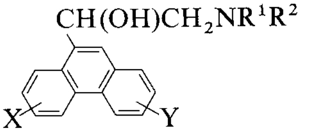
$$a_7 = 1.811(\pm 0.65)\,\sigma_{para} - 0.010(\pm 0.19) \quad (188)$$

$$(n = 5;\ r = 0.981;\ s = 0.123;\ F = 77.7)$$

$$a_8 = 0.959(\pm 0.78)\,\pi - 0.395(\pm 0.57) \quad (189)$$

$$(n = 4;\ r = 0.966;\ s = 0.151;\ F = 27.9)$$

These analyses aided in the stepwise derivation of Hansch equations for the antimalarial activities of structurally related phenanthreneaminoalkylcarbinols (**47**) (eq. 190) [778] and finally led to the derivation of eq. 175 (chapter 7.4) for a much larger group of analogs [779].



(47)

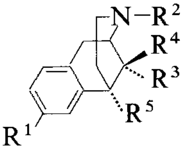
$$\log 1/C = 0.396(\pm 0.134)\,\pi_y + 0.270(\pm 0.105)\,\pi_x + 0.654(\pm 0.280)\,\sigma_x + 0.878(\pm 0.269)\,\sigma_y + 0.137(\pm 0.087)\,\pi_{sum} - 0.015(\pm 0.009)\,(\pi_{sum})^2 + 2.335(\pm 0.194) \quad (190)$$

$$\text{optimum } \pi_{sum} = 4.44$$

$$(n = 102;\ r = 0.913;\ s = 0.258)$$

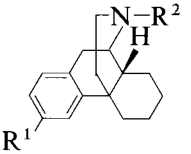
Some other examples of the stepwise derivation of Hansch equations and of the improvement of Hansch equations from the interpretation of Free Wilson analyses have been published (compare eqs. 75–77, chapter 4.3) [22, 390, 393].

Different sets of compounds were used in a Free Wilson analysis of analgesic benzomorphans (**48**); the first one included all compounds (38 variables; $n = 99$; $r = 0.893$; $s = 0.466$), a second one only contained racemic compounds (36 variables; $n = 86$; $r = 0.909$; $s = 0.457$) and a last one excluded all single point determinations (20 variables, $n = 70$; $r = 0.879$; $s = 0.457$) [811].



(48)

Two extra variables in the first analysis accounted for (+)-enantiomers ($a_i = -0.97$) and (−)-enantiomers ($a_i = 0.17$). While the value for the more active (−)-enantiomers is not too far from the theoretical value of 0.3, which results if the (+)-enantiomers are absolutely inactive, general experience shows that the less active enantiomers cannot be expected to differ from the active ones (or the racemates) by a constant value (problems associated with QSAR analyses of optically active compounds are discussed below and in chapter 9.1). The group contributions of the benzomorphans could be used to predict the biological activity values of structurally related morphinans (**49**), which are more active than the benzomorphans by some orders of magnitude (eq. 191) [811].

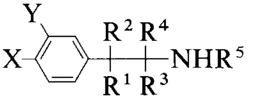


(49)

$$\log 1/C_{obs} = 0.769(\pm 0.35) \log 1/C_{calc} + 4.052(\pm 1.02) \tag{191}$$

$(n = 6;\ r = 0.950;\ s = 0.254;\ F = 37.12)$

In a Free Wilson analysis of norepinephrine-uptake inhibiting phenethylamines (**50**), including achiral analogs, racemates, and pure enantiomers, but also diastereomeric mixtures, different group contributions were attributed to the R and S substituents (9 variables; $n = 30$; $r = 0.963$; $s = 0.276$) [812].

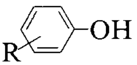


(50)

While the assignment of values of 0.5 to R and S positions (1 to either position in the case of pure enantiomers) is correct for the racemates, ratios of 0.5 : 0.5 : 0.5 : 0.5 were arbitrarily assigned to the R and S positions of the two independent chiral centers of diastereomeric mixtures; however, this is an arbitrary and most often wrong assumption (see chapter 9.1).

A comparative study of the inhibitory activities of benzamidines (37 variables) against the serine proteases thrombin ($n = 83$; $r = 0.90$; $s = 0.39$), plasmin ($n = 82$; $r = 0.96$; $s = 0.24$), and trypsin ($n = 84$; $r = 0.91$; $s = 0.35$) [813] shows differences between the group contributions of different substituents, giving some hints for analogs displaying higher selectivities to any of the three different proteases. The use of the partial least squares (PLS) method (chapter 5.3) has been proposed for comparative analyses [390], but seemingly Free Wilson analyses have not yet been performed by using PLS analysis. Recently, a stepwise principal component regression analysis was used in Free Wilson analysis [411].

The simplest form of a Free Wilson analysis is presented in eq. 192 [22], which describes the antibacterial activities of phenol and isomeric chlorophenols (**51**, R = H, Cl; one to five chlorine atoms) *vs. Staphylococcus aureus*; at least the linearity of the structure-activity relationship can be derived from eq. 192; on the other hand, although most probably lipophilicity is responsible for the variance in the biological activities, no Hansch equation can be derived, because each other physicochemical property of the chlorine atom will give identical results.



(51)

$$\log 1/C = 0.503(\pm 0.13)\,[\mathrm{Cl}] + 2.578 \qquad (192)$$

$$(n = 9;\ r = 0.960;\ s = 0.256;\ F = 83.06)$$

Free Wilson analyses which include too many single-point determinations (substituents which only once occur in the data matrix) suggest a much better fit of the biological data than is actually obtained without these values. For the hallucinogenic properties of phenylalkylamines excellent statistical parameters result if all members are included in the analysis (15 variables; $n = 23$; $r = 0.985$; $s = 0.182$; $F = 15.28$; recalculated F value) [390, 814]. Elimination of the single-point determinations tells the truth (7 variables; $n = 15$; $r = 0.896$; $s = 0.182$; $F = 4.08$) [390]: the correlation coefficient and the overall significance considerably decrease; only the number of degrees of freedom and therefore the standard deviation s remain constant (as for the sets of 86 and 70 benzomorphanes, see above), indicating that s is the only reliable statistical parameter if single-point determinations are included.

Sometimes Free Wilson analyses are presented in graphical form (*e.g.* [413, 815]), which allows an easier interpretation of the results if many variables in different positions are involved.

The use of Free Wilson-type indicator variables in Hansch analysis has been discussed in chapters 3.8, 4.3, and 7 [21, 390, 391, 393]. Nonadditivities in Free Wilson analyses due to nonlinear lipophilicity-activity relationships have been discussed in chapter 4.3 [22, 390–392, 394].

The effect of an additional nonlinear term in a Free Wilson equation is demonstrated by eqs. 193 and 194 [390, 393], which have been derived for the

Table 26. Hypnotic activities of N,N′-diacylureas $R^1NHCONHR^2$; structures and table for the Free Wilson analysis [393]

Substituents[a]		Table for Free Wilson Analysis					Interaction	
R^1	R^2	Prop	But	Pent	Hex	Hept	term, $N_1.N_2$	Log 1/C
Acet	Prop	1					6	1.84
Prop	Prop	2					9	2.06
Acet	But		1				8	2.16
Prop	But	1	1				12	2.23
Acet	Pent			1			10	2.27
But	But		2				16	2.40
Prop	Pent	1		1			15	2.35
Acet	Hex				1		12	2.46
But	Pent		1	1			20	2.38
Prop	Hex	1			1		18	2.25
Acet	Hept					1	14	2.55
Pent	Pent			2			25	2.32
But	Hex		1		1		24	2.28

[a] Acet, Prop, But, Pent, Hex, Hept = acetyl, propionyl, butanoyl, pentanoyl, hexanoyl, and heptanoyl

hypnotic activities of N,N′-diacylureas (Table 26). Although eq. 193 contains 5 parameters for only 13 compounds, it is not significant at the 95% level.

$$\begin{aligned}\log 1/C = {} & -0.003(\pm 0.22)\,[\text{Prop}] + 0.149(\pm 0.22)\,[\text{But}] + \\ & + 0.186(\pm 0.23)\,[\text{Pent}] + 0.242(\pm 0.29)\,[\text{Hex}] + \\ & + 0.511(\pm 0.47)\,[\text{Hept}] + 2.039(\pm 0.34)\end{aligned} \tag{193}$$

$(n = 13;\ r = 0.822;\ s = 0.135;\ F = 2.93)$

A much better and highly significant result is obtained if an interaction term $N_1 \cdot N_2$ (N being the number of carbon atoms of each of the acyl residues) is included to account for a nonlinear dependence of the biological activity on the chain length; all regression coefficients in eq. 194 are significant at the 95% level. In addition, the group contributions of the acyl residues are highly correlated with their chain lengths (a_i *vs.* N, including the acetyl group, $a_i = 0$: $n = 6$; $r = 0.993$; $s = 0.090$). The nonlinear dependence on N results from a nonlinear lipophilicity-activity relationship (eq. 195) [767]; absolutely identical statistical results are obtained if the log P term in eq. 195 is replaced by N [393].

$$\begin{aligned}\log 1/C = {} & 0.430(\pm 0.17)\,[\text{Prop}] + 0.910(\pm 0.28)\,[\text{But}] + \\ & + 1.249(\pm 0.39)\,[\text{Pent}] + 1.487(\pm 0.45)\,[\text{Hex}] + \\ & + 1.813(\pm 0.49)\,[\text{Hept}] - 0.085(\pm 0.03)\,N_1 \cdot N_2 + 1.930(\pm 0.13)\end{aligned} \tag{194}$$

$(n = 13;\ r = 0.982;\ s = 0.049;\ F = 26.59)$

Table 27. Result of a Free Wilson analysis (R^1, R^2, R^3 = alkyl, cycloalkyl; $\mu = 6.499$; n = 128; r = 0.991; s = 0.231) [390, 801] of the affinities of quaternary ammonium compounds to the postganglionic acetylcholine receptor [816]

X in $X\text{-}CH_2CH_2\text{-}N^+(R^1R^2R^3)$	a_i
CH_3CH_2O-	-2.479
$CH_3CH_2CH_2$-	-2.175
$C_6H_5CH_2COO$-	-1.228
$C_6H_5CH_2CH_2O$-	-1.177
$C_6H_5CH_2CH_2CH_2$-	-0.909
$C_6H_{11}CH_2COO$-	-1.035
$C_6H_{11}CH_2CH_2O$-	-0.819
$C_6H_{11}CH_2CH_2CH_2$-	-0.683
$(C_6H_5)_2CHCOO$-	0.872
$(C_6H_5)_2CHCH_2O$-	-0.070
$(C6H5)_2CHCH_2CH_2$-	0.374
$C_6H_5(C_6H_{11})CHCOO$-	2.035
$(C_6H_5)_2C(OH)COO$-	2.047
$(C_6H_{11})_2CHCOO$-	1.467
$(C_6H_{11})_2CHCH_2O$-	0.806
$C_6H_5(C_6H_{11})C(OH)COO$-	2.975

$$\log 1/C = -0.177(\pm 0.09)(\log P)^2 + 0.599(\pm 0.22)\log P + 1.893 \qquad (195)$$

$$(n = 13;\ r = 0.918;\ s = 0.079)$$

Although this example may be criticized because of its too narrow range of activity values, it clearly shows that the Free Wilson model is generally restricted to linear structure-activity relationships.

Another type of hitherto unexplained nonadditivity is observed in a Free Wilson analysis of the affinities of quaternary ammonium compounds to the postganglionic acetylcholine receptor [816]. While the whole set of compounds, including a large number of antagonists, but also some agonists and partial agonists, can excellently be described by only 15 variables (n = 128; r = 0.991; s = 0.231) (Table 27) [390, 801], a closer inspection of the individual group contributions shows that they themselves do not behave in an additive manner.

In all analogs that contain no ring, the introduction of a phenyl group, C_6H_5, or a cyclohexyl group, C_6H_{11}, increases affinity by about 1.2 to 1.7 logarithmic units. But it makes a large difference, whether $R = C_6H_5$ (instead of H) is introduced into $R-CH_2CH_2CH_2-$, $R-CH_2CH_2O-$, $C_6H_5CH(R)COO-$, or $C_6H_{11}CH(R)COO-$; a corresponding effect is observed for the introduction of C_6H_{11} into $R-CH_2CH_2CH_2-$, $R-CH_2CH_2O-$, $C_6H_5CH(R)COO-$, or $C_6H_{11}CH(R)COO-$ (Table 28). If one phenyl group in $(C_6H_5)_2CHCOO-$ is

Table 28. Nonadditivies in the Free Wilson group contributions of quaternary ammonium compounds $X-CH_2CH_2-N^+(R^1R^2R^3)$ [390, 801]

Differences in receptor affinity (logarithmic scale), by changing			
R in group X from:	**H to C_6H_5**	**H to C_6H_{11}**	**C_6H_5 to C_6H_{11}**
R-$CH_2CH_2CH_2$-	1.266	1.492	0.226
R-CH_2CH_2O-	1.302	1.660	0.358
$C_6H_5CH(R)COO$-	2.100	3.263	1.163
$C_6H_{11}CH(R)COO$-	3.070	2.502	-0.568

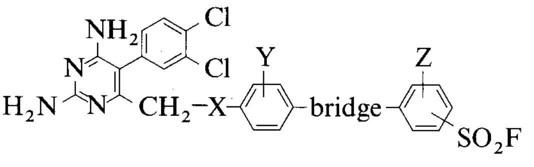
exchanged for C_6H_{11}, the affinity increases by +1.163 log units; on the other hand, the same exchange in $C_6H_5(C_6H_{11})CHCOO-$ reduces the affinity by −0.568 log units [390, 801].

Eq. 196 (K = affinity constant) results if the information from the Free Wilson group contributions is translated into indicator variables [PHE] (number of phenyl groups in the molecule), [*c*-HEX] (number of cyclohexyl groups), I_{OH} (presence of a hydroxyl group), I_{COO} (presence of a sterically hindered carboxylate group, *i.e.* an acetyl group bearing two large substituents), and [INT] for the simultaneous presence of C_6H_5 and C_6H_{11} in the ester group. The coefficient of [INT] indicates that whenever C_6H_5 and C_6H_{11} appear together, affinity is much higher than expected; although some other QSAR analyses were derived for this data set [366, 403, 545, 733, 734], no reasonable explanation for [INT] can be given, even if the entropy of ligand binding or a cooperative effect are taken into account.

$$\begin{aligned}\log K = {} & 1.258(\pm 0.09)\,[PHE] + 1.545(\pm 0.11)\,[c\text{-}HEX] + \\ & + 1.069(\pm 0.19)\,I_{OH} + 0.755(\pm 0.17)\,I_{COO} + \\ & + 0.769(\pm 0.19)\,[INT] + 4.142(\pm 0.12)\end{aligned} \tag{196}$$

$(n = 128;\ r = 0.983;\ s = 0.290;\ F = 711.14)$

Free Wilson analyses may include far fewer variables than substituents, if group contributions being not significant are eliminated. Indicator variables for 28 different structural features and different test models and 15 interaction terms were investigated to describe the inhibition of dihydrofolate reductase by 2,4-diaminopyrimidines (**52**); 9 indicator variables and 2 interaction terms were selected and eq. 197 was derived out of the 2047 theoretically possible linear combinations of any numbers of these variables [412].



(52)

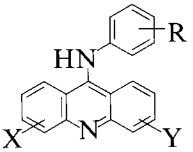
$$\begin{aligned}\log 1/C = {} & 0.365(\pm 0.12)\,\text{I-1} + 1.013(\pm 0.12)\,\text{I-8} - 0.784(\pm 0.19)\,\text{I-9} + \\ & + 0.419(\pm 0.20)\,\text{I-13} - 0.220(\pm 0.09)\,\text{I-15} + \\ & + 0.513(\pm 0.18)\,\text{I-20} + 0.674(\pm 0.23)\,\text{I-4} \cdot \text{I-8} + 7.174(\pm 0.07) \qquad (197)\end{aligned}$$

$(n = 105;\ r = 0.903;\ s = 0.229)$

In a series of ACTH-derived peptides the same quality of fit was obtained if only 11 group contributions ($n = 52$; $r = 0.984$; $s = 0.406$; $F = 112$) were used instead of the original set of 24 variables ($n = 52$; $r = 0.986$; $s = 0.464$; $F = 40$) [413]; the standard deviation s is even smaller, because a slight increase in the sum of squared errors is more than counterbalanced by the much larger number of degrees of freedom.

Mager [414–416] introduced the term "reduced Free Wilson model" for this modification and proposed the use of stepwise regression analysis to derive the equation; some more examples of this approach have been published [417, 544, 545]. However, one should bear in mind that the significance of a certain group contribution not only depends on its confidence interval but also on the selection of the reference substituents [390, 391, 410].

The computer-automated structure evaluation (CASE, chapter 4.6) of 9-anilinoacridines (**53**) is illustrated below. The CASE program automatically generated fragments, containing about 3–10 heavy atoms, to differentiate between active and inactive analogs. Out of nearly 200 descriptors produced in this manner, several biophores and biophobes (*i.e.* fragments increasing or reducing biological activities) were selected by the program and correlated with the antitumor activities of the compounds (eq. 198) [519].

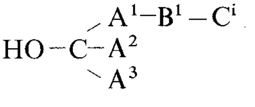


(53)

$$\begin{aligned}\log 1/ED_{50} = {} & 0.63n_1F_1 + 0.31n_2F_2 + 0.20n_3F_3 - 1.85n_4F_4 + 0.16n_5F_5 + \\ & + 0.51n_6F_6 - 0.34n_7F_7 - 0.21n_8F_8 + 0.41n_9F_9 + \\ & + 0.40n_{10}F_{10} + 0.37n_{11}F_{11} - 0.40n_{12}F_{12} - 0.35n_{13}F_{13} - \\ & - 0.26n_{14}F_{14} - 0.08 \log P + 4.11 \qquad (198)\end{aligned}$$

$(n = 461;\ r = 0.805;\ s = 0.46;\ F = 54.52)$

An example of the application of the DARC-PELCO approach is given in eq. 199 ($PC' =$ phenol coefficient) [421], which describes the antibacterial activities of linear and branched alcohols (**54**); however, the result is trivial, because the compounds



$$\mathrm{HO{-}C}\begin{array}{l}\diagup A^1{-}B^1{-}C^i\\ {-}A^2\\ \diagdown A^3\end{array}$$

(54)

are described with about the same quality of fit just by a linear lipophilicity-activity relationship (eq. 200) [817].

$$\log PC' = 0.392(\pm 0.21)\,A1 + 0.268(\pm 0.21)\,A2 + 0.158(\pm 0.21)\,A3 + 0.446(\pm 0.15)\,B1 + 0.533(\pm 0.04)\,Ci - 2.05(\pm 0.16) \quad (199)$$

$(n = 15;\ r = 1.000;\ s = 0.050)$

$$\log PC' = 1.024(\pm 0.06) \log P - 1.536(\pm 0.07) \quad (200)$$

$(n = 15;\ r = 0.996;\ s = 0.090)$

Hypermolecule approaches, like the MSD and MTD method (chapter 4.6), are to some extent related to the Free Wilson model. For a series of progesterone analogs, a hypermolecule was constructed (Figure 46); in a stepwise procedure the different atomic positions were characterized as being beneficial ($\varepsilon = -1$), irrelevant ($\varepsilon = 0$), or detrimental ($\varepsilon = +1$) to biological activities. Afterwards the minimal topological difference (MTD) values *vs.* this optimized map were calculated for each member of the series; together with a side chain-corrected lipophilicity parameter f they were correlated with the relative binding affinities (RBA) to the progesterone receptor (eq. 201) [542].

$\varepsilon = -1$: positions 1, 14, 20, 39

$\varepsilon = 0$: positions 3, 6, 8, 13, 15, 16, 22–27, 30, 33–36, 40

$\varepsilon = +1$: positions 2, 4, 5, 9, 11, 28, 29, 31, 32, 37, 38

$$\log RBA = 0.696(\pm 0.09)\,f - 0.744(\pm 0.13)\,MTD + 3.917(\pm 0.66) \quad (201)$$

$(n = 55;\ r = 0.935;\ s = 0.331)$

Magee's approach to active-site binding analysis [544, 545] may be considered as a combination of the hyperstructure concept (chapter 4.6) with the mixed approach (chapter 4.3), as can *e.g.* be seen from eq. 202 [545]; first a hyperstructure (**55**) was formulated for a series of *para*-substituted phenyl-N-methylcarbamates and then

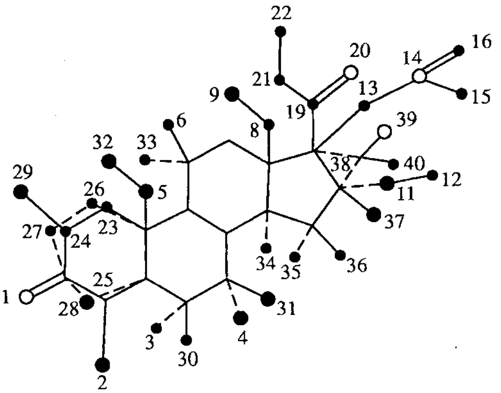
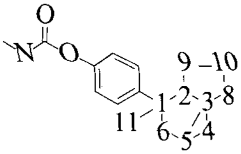


Figure 46: Hypermolecule for progesterone derivatives (eq. 201). Open circles ○ stand for beneficial ($\varepsilon = -1$), filled circles ● for detrimental ($\varepsilon = +1$), and dots • for irrelevant ($\varepsilon = 0$) positions (reproduced from Figure 2 of ref. [542] with permission from the copyright owner).

their inhibitory activities against acetylcholinesterase were correlated with lipophilicity and polarizability in the 1-position and with Free Wilson group contributions in the 2-, 5-, and 6-positions.



(55)

$$\begin{aligned} pI_{50} = {} & 0.221(\pm 0.18)\, f_1 + 0.064(\pm 0.04)\, MR_1 + 0.859(\pm 0.36)\, P_2 + \\ & + 0.864(\pm 0.30)\, P_5 - 0.390(\pm 0.31)\, P_6 + 3.40 \end{aligned} \tag{202}$$

$(n = 40;\ r = 0.865;\ s = 0.341)$

The Free Wilson analysis was also applied to describe
α-adrenoceptor binding [815],
analgesic activities [16, 415, 492, 811, 818, 819],
antibacterial activities [16, 22, 40, 41, 42, 421, 424, 803, 820–825],
anticholinergic activities [826],
antilipemic activities [827],
antiviral activities [828–830],
ATPase inhibition [831],
benzodiazepine receptor binding [832–834],
bradykinin potentiation [835],
butyrylcholinesterase inhibition [393, 836–840],
carboxylesterase inhibition [836],
dopamine β-hydroxylase substrate properties [20, 841],
dopamine receptor binding [815, 842],
glucocorticoid activities [843],
inhibition of oxytocin contraction [844–847],
learning and behavior [413, 848],
muscarinic receptor binding [815],
neuroleptic activities [815],
norepinephrine-uptake inhibition [812, 841, 849],
phenylethanolamine-N-methyltransferase inhibition [841],
phenylethanolamine-N-methyltransferase substrate properties [20, 841, 849],
progestational activities [850],
renin inhibition [851],
serotonin receptor binding [815],
thyromimetic activities [8, 393],
toxicities [16, 414, 415, 425, 429, 430, 827, 836, 852–858], and
tranquillizing activities [853, 859].

More applications of the Free Wilson model are reviewed in textbooks (*e.g.* [40–43]) and in dedicated articles [390, 391]. Some statistical problems arising in the application of the Free Wilson model are discussed in refs. [390, 411].

QSAR: Hansch Analysis and Related Approaches
by Hugo Kubinyi
© VCH Verlagsgesellschaft mbH, 1993

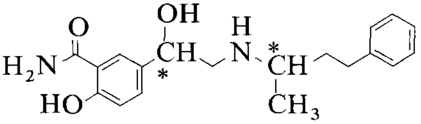
9. 3D QSAR Approaches

9.1. Stereochemistry and Drug Action

The different enantiomers of optically active compounds have identical chemical and physicochemical properties, except their different influence on the rotation of polarized light. However, a binding site is a chiral environment that discriminates between the different enantiomers as if they were completely different molecules; optical enantiomers also differ in their metabolism. Thus, stereochemistry plays an important role in the biological activity of drugs.

Several contributions by Ariëns and Lehmann [860–867] stress the important influence of chirality on biological activity; in a strong effort to push pure enantiomers instead of racemates for human medication, Ariëns polemically categorized racemates as "drugs containing 50% impurity" [861]. The situation is even worse in the case of diastereomeric mixtures, for two reasons: first, 2^n species (n being the number of asymmetric centers) are involved and second, despite the fact that each individual racemate is a 50:50 mixture of the corresponding enantiomers, the amount of the different racemates in the mixture varies largely. The relative ratios of the racemates depend on the reaction conditions used in their synthesis, on the stability and relative energies of the different transition states as well as on the isolation and purification conditions.

The pure enantiomers of labetalol (**56**), a β-antiadrenergic drug containing two different centers of optical asymmetry, show different pharmacological characteristics: while the RS and the SS isomers are only weakly antagonistic against α_1, β_1, and β_2 receptors, the SR isomer carries most of the α_1-antagonistic activity and the RR isomer contributes mainly to the β_1- and β_2-antagonistic activities (Table 29) [861].



(56)

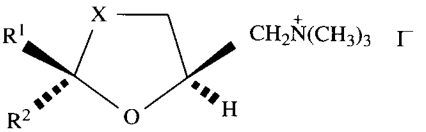
Whenever possible, pure enantiomers are nowadays developed and introduced into the pharmaceutical market; the only exception are compounds, where identical pharmacological profiles are found for both enantiomers or compounds, where the much higher price of one enantiomer, as compared to the racemate, precludes such a selection.

Different QSAR equations have been derived for different enantiomers of phenoxypropionic acids [868], giving evidence for the validity of Pfeiffer's rule

Table 29. Antiadrenergic activities (pA_2 values) of different stereoisomers of labetalol (**56**) [861]

Isomer	α_1 Receptor (rabbit aortic strip)	β_1 Receptor (guinea pig left atrium)	β_2 Receptor (guinea pig tracheal strip)
R,R	5.87	8.26	8.52
R,S	5.5	6.97	6.33
S,R	7.18	6.37	< 6.0
S,S	5.98	6.43	< 6.0

[860, 869] that the activity ratio (the eudismic index) of the active (eutomer) to the less active enantiomer (distomer) increases with increasing activity of the more active one. Because of exceptions to this rule [97] there has been some dispute, whether Pfeiffer's rule indeed is generally valid or not [870–874]. Heterogeneous groups of compounds have to be separated into appropriate subgroups: while for a mixed group of acetylcholine receptor antagonists no correlation could be found ($n = 18$; $r = 0.57$), excellent results were obtained after separation of the 1,3-oxathiolanes (**57**, X = S) ($n = 6$; $r = 0.98$) and the 1,3-oxathiolane 3-oxides (**57**, X = >SO) ($n = 12$; $r = 0.98$), giving evidence that Pfeiffer's rule at least holds true for closely related series (Figure 47) [871–873]. Comparing the eudismic ratios of more heterogeneous series, one should also consider the possibility of multiple binding modes of different analogs.



(57)

Schaper derived equations to describe the nonlinear dependence of the biological activities of racemates at different concentrations on the activities of the pure enantiomers [875]. Not only quantitative but also qualitative differences were observed for the QSARs of different enantiomers of chiral phosphonic acids; a linear dependence of butyrylcholinesterase inhibition on chain length resulted for the (+) enantiomers, while a clear bilinear dependence was observed for the (−) isomers [876].

On the other hand, different enantiomers of chymotrypsin ligands (2 achiral compounds, 19 D-isomers, and 24 L-isomers) could be combined in one equation on the assumption that the group with the largest MR value binds in a hydrophobic cleft (the so-called ϱ_2 area), while the smaller groups bind in ϱ_1 space (eq. 203; MR-L = MR of the larger group, MR-S = MR of the smaller group) [877].

$$\begin{aligned}\log 1/K = {} & 0.72(\pm 0.13)\,\text{MR-L} + 0.230(\pm 0.07)\,\text{MR-S} + \\ & + 0.323(\pm 0.20)\,\text{I-1} + 0.311(\pm 0.15)\,\text{I-2} - 1.062(\pm 0.45) \qquad (203)\end{aligned}$$

($n = 45$; $r = 0.928$; $s = 0.235$)

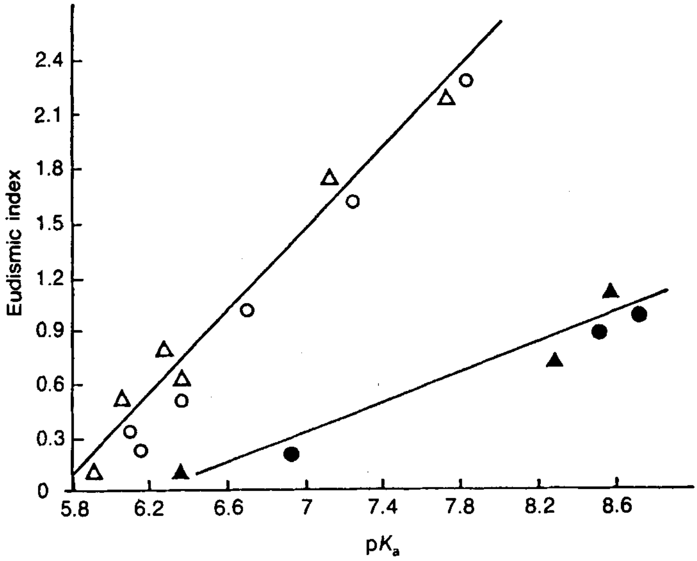


Figure 47: Pfeiffer's rule [860]. Eudismic indices (= logarithm of the ratio of the affinities of the more active analog to the less active analog) are plotted against the affinities pK_a of the eutomers (open symbols, 1,3-oxathiolane 3-oxides **57**, X = >SO; filled symbols, 1,3-oxathiolanes **57**, X = S; circles indicate the affinities to the ileum muscarinic receptor, triangles indicate the affinities to the bladder receptor) (reproduced from Figure 2 of ref. [871] with permission from Elsevier Trends Journals, Cambridge, UK).

The proper parametrization of optically active compounds in Free Wilson analysis has been discussed in chapter 8.

With the exception of the ambiguities in the alignment procedure, stereochemistry is adequately considered in 3D QSAR analyses.

9.2. Active Site Interaction Models

Early attempts to map the properties of an unknown receptor (or any other ligand binding site) started from qualitative structure-activity relationships [878], from MO calculations of preferred conformations of ligands [879] and from the interpretation of multiparameter Hansch equations (*e.g.* Figure 48) [28].

Pharmacophoric pattern searching and receptor mapping [41, 132, 288, 880 – 882] use information from the QSARs in the different positions of the ligands and also from ligands with restricted internal rotations (rigid analogs) to derive the structural elements being necessary for receptor affinity (the pharmacophore) and to draw a conclusion which properties are relevant at the different sites of the receptor surface (the receptor map) (*e.g.* [28, 737, 883 – 889]). Receptor maps can be used to predict the affinities or activities of new analogs in a qualitative manner or, if a QSAR equation has been derived, in quantitative terms. Also hyperstructure models,

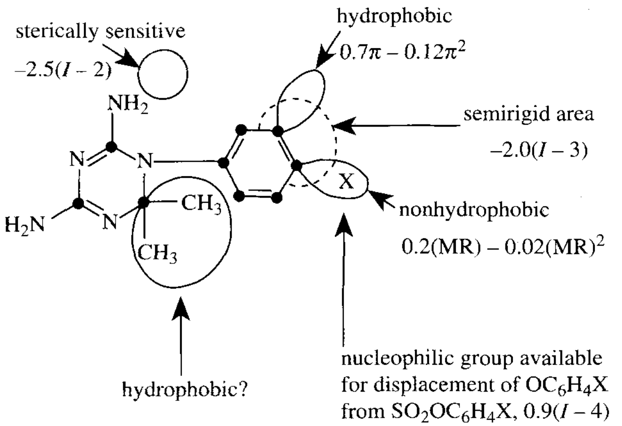

sterically sensitive
$-2.5(I-2)$
NH_2
N
N
H_2N
N
CH_3
CH_3
hydrophobic
$0.7\pi - 0.12\pi^2$
semirigid area
$-2.0(I-3)$
X
nonhydrophobic
$0.2(MR) - 0.02(MR)^2$
hydrophobic?
nucleophilic group available
for displacement of OC_6H_4X
from $SO_2OC_6H_4X$, $0.9(I-4)$

Figure 48: Receptor map of dihydrofolate reductase, derived from eq. 81 (chapter 4.3) (reproduced from Chart I of ref. [28] with permission from the American Chemical Society, Washington, DC, USA).

like MSD, MTD, and related methods (chapter 4.6), may be considered as quantitative approaches to receptor mapping.

Systematic investigations of the interaction energies between ligands and hypothetical binding sites have been performed by Höltje *et al.* for
chloramphenicol binding to ribosomes [29, 30, 890, 891],
pyrimidinone H_2-antihistaminics [891, 892],
acetylcholinesterase substrates [890, 893],
cyclopropylamine inhibitors of monoamine oxidase [894, 895],
antihypertensive benzothiadiazine 1,1-dioxides [896],
norepinephrine-uptake inhibition by phenethylamines [897],
sulfonamide binding to serum albumin [898],
calcium antagonism of verapamil analogs [899, 900],
binding, calcium agonism, and calcium antagonism of 1,4-dihydropyridines [901 – 904], and
5-HT_2 agonism of 2,5-dimethoxyphenethylamines [905].

Simple organic molecules are taken as models of the different amino acid side chains, *e.g.*
propane for aliphatic amino acids,
acetamide for amide side chains,
toluene, *p*-cresol, and 3-methylindole for aromatic amino acids,
n-propylguanidine (positively charged) for basic amino acids,
acetate (negatively charged) for acidic amino acids, and
methanol for serine.

Next, the interaction forces of each molecule are calculated by using several of these probes; all analogs of a series are placed in standard geometries and at certain

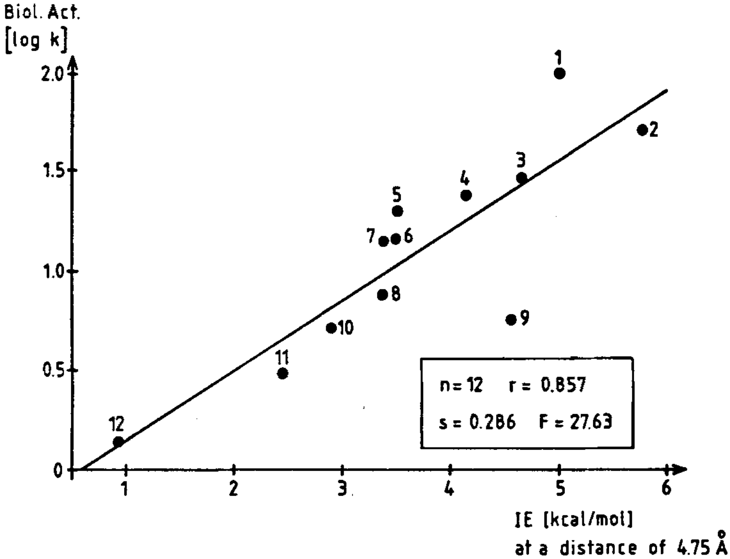
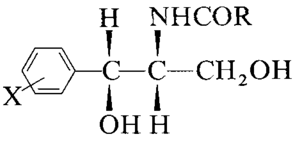


Figure 49: Regression line for the arginine binding site model of the acyl side chain of chloramphenicol (**58**) and analogs. Biological activity values are plotted against the calculated interaction energies IE (kcal instead of kJ and Å instead of nm were given in the original diagram; 1 kcal = 4.18 kJ, 1 Å = 0.1 nm) (reproduced from Figure 2 of ref. [30] with permission from the copyright owner).

distances to the different amino acid side chains (the hypothetical binding sites). The resulting energies are then correlated with receptor affinities or biological activities. 3D coordinates from protein crystallography were used to calculate the interaction energies of sulfonamide inhibitors of erythrocytic carboanhydrase [906] and of methotrexate analogs as inhibitors of dihydrofolate reductase [907].

The binding of the acyl side chain of chloramphenicol (**58**) analogs to ribosomes could be explained by the interaction energies with an arginine side chain (n = 12; r = 0.857; s = 0.286) (Figure 49), while a histidine side chain modeled the interaction with the aromatic ring system of some other analogs (n = 12; r = 0.887; s = 0.205) (Figure 50) [30, 891].



(58)

A hypothetical three-dimensional receptor complex was constructed from the known primary structure of the ribosomal L16 protein by using a short histidine-containing peptide of the L16 sequence in a low-energy conformation (Figure 51) [30].

The interaction energies between an arginine side chain and the aromatic ring close to the chiral center of verapamil analogs could be correlated with their calcium antagonistic activities (n = 10; r = 0.973; s = 0.095) [899].

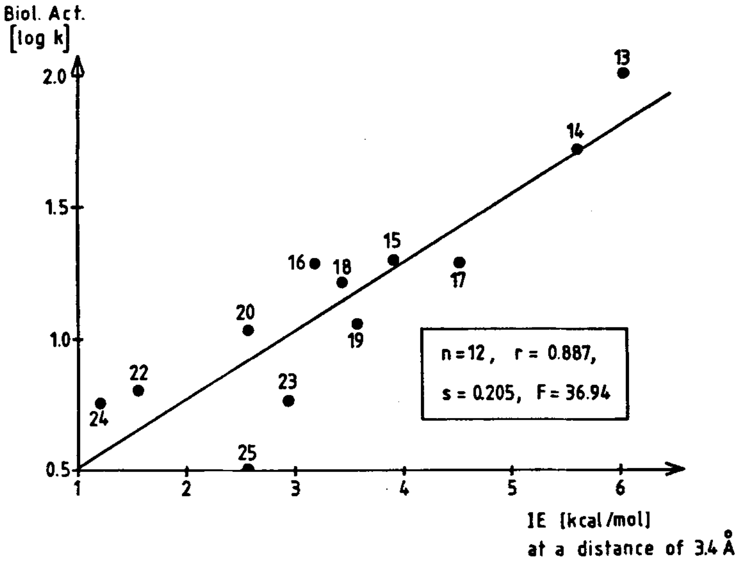


Figure 50: Regression line for the histidine binding site model of the aromatic ring substituents of chloramphenicol (**58**) and analogs. Biological activity values are plotted against the calculated interaction energies IE (kcal instead of kJ and Å instead of nm were given in the original diagram; 1 kcal = 4.18 kJ, 1 Å = 0.1 nm) (reproduced from Figure 3 of ref. [30] with permission from the copyright owner).

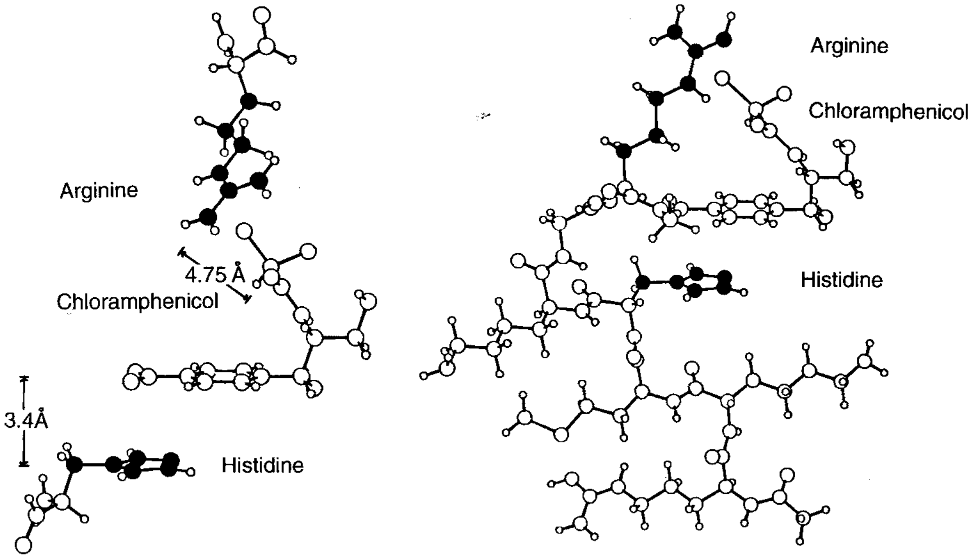

Arginine
4.75 Å
Chloramphenicol
3.4Å
Histidine
Arginine
Chloramphenicol
Histidine

Figure 51: Hypothetical receptor model for chloramphenicol (**58**) binding. Arginine and histidine in their postulated distances (left diagram) and the potential ribosomal protein L16 binding site (right diagram) for chloramphenicol (reproduced from Figures 4 and 5 of ref. [30] with permission from the copyright owner).

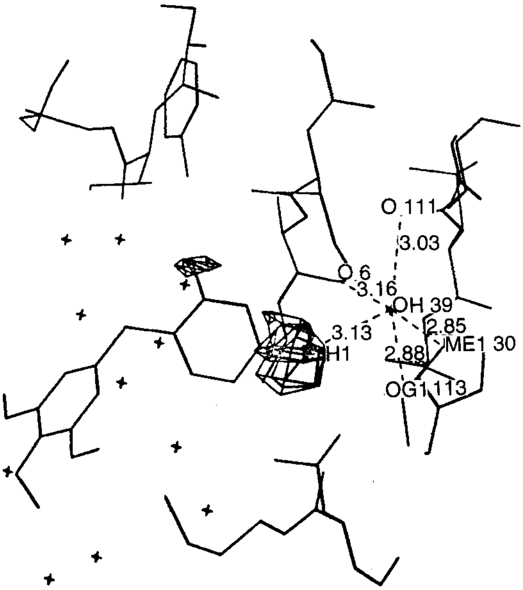


Figure 52: Interaction potential of an amino group with the active site of *Escherichia coli* dihydrofolate reductase, calculated by GRID; the water molecule H39 is treated as a structural part of the enzyme in the calculations. The energy contours are plotted at $-63\ \mathrm{kJ \cdot mol^{-1}}$; they indicate two sites of strong attraction. The position of trimethoprim relative to the active site was adjusted so that the amino nitrogens N2 and N4 are located in these regions (reproduced from Figure 8 of ref. [33] with permission from the American Chemical Society, Washington, DC, USA).

In general, good to excellent correlations are obtained for certain amino acid side chains, while others fail to explain the structure-activity relationship. Quaternary charged and neutral acetylcholine analogs showed a much closer relationship of their acetylcholinesterase affinities to the interaction energies with a benzene ring than with a negatively charged acetate ion [890, 893]; thus prior assumptions that the cationic head of these analogs interacts with an anionic site at the enzyme had to be reconsidered.

The concept of pharmacophore identification based on molecular electrostatic potentials has been reviewed [908]. Some other approaches to correlate biological activities with the interactions at certain positions of the binding site were discussed in chapter 2 (eqs. 17 and 18).

The strategy of interaction energy calculations has been extended by Goodford [33, 130, 909] to the program GRID [910], which calculates the interaction energies of different probe atoms around the surface of a protein of known 3D structure, giving contour maps of energy values (Figure 52).

Contours having negative values can be interpreted as regions of attraction between the probe atom and the protein. Methyl (CH_3), amino groups (NH_2), charged amino groups (NH_3^+), carbonyl oxygens (O), carboxy oxygen (O^-), hydroxyl (OH), and water (H_2O) are used as probe atoms and groups; empirical hydrogen-bond potentials were derived for the determination of energetically favorable binding sites of proteins and of small molecules [38, 909]. Thus, GRID is suited for the design of new ligands, to calculate fields for CoMFA-related 3D QSAR approaches (chapter 9.3), and (in a reverse mode) to model receptor binding sites for series of active analogs [909].

A simple and much faster rule-based algorithm is used in the program LUDI for the *de novo* design of enzyme inhibitors (Figure 53) [38, 911–913]. The program

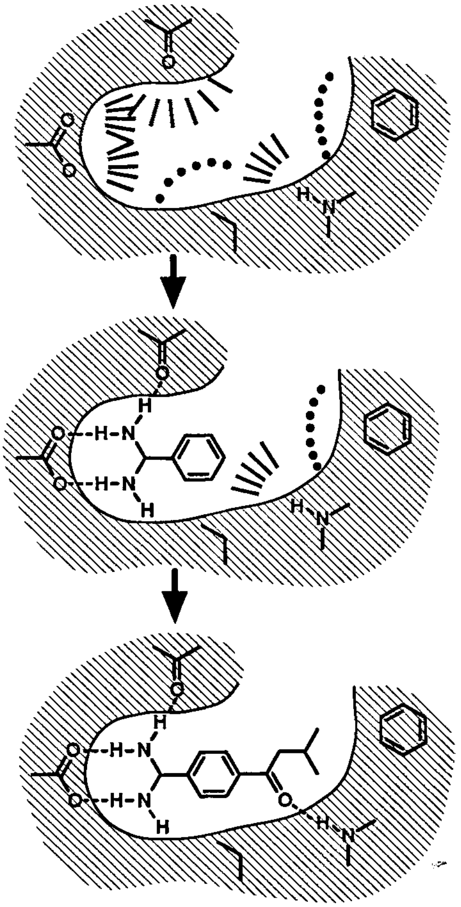


Figure 53: Concept of the program LUDI for the *de novo* design of enzyme inhibitors. The program automatically identifies hydrogen bond donor, acceptor, and lipophilic aliphatic and lipophilic aromatic groups within a predefined area, *e.g.* the active site of an enzyme; in the next step favorable positions of interacting groups are generated (upper diagram) and a database is searched for possible ligands; a hitlist of molecules optimally interacting with these sites is generated (middle diagram; example of a good hit). Another hit, including additional interactions, is shown in the lower diagram; alternatively, this hit can be generated by the program by adding a small fragment to the ligand shown in the middle diagram (reproduced from Figure 3 of ref. [912] with permission from ESCOM Science Publishers BV, Leiden, The Netherlands).

interprets the 3D structure of a binding site in terms of hydrogen bond donor, acceptor, and lipophilic aliphatic and aromatic interaction sites; it assigns complementary sites in standard geometries and distances, derived from the Cambridge database of crystal structures [141] and fits fragments and molecules from its own or external libraries of 3D structures to these sites. Finally, the obtained hits are evaluated according to their quality of fit and their estimated affinity, in this manner proposing new ligands as well as a rational modification of already existing lead structures.

In an active analog mode LUDI may also be used to propose potential new drug candidates. Hydrogen-bonding and hydrophobic interaction sites are derived from a set of active analogs, the internal or external databases of 3D structures are searched for molecules having the same geometry of interaction sites, and ratings are estimated for the affinity values of these new structures.

Atom-based lipophilicity contributions and molar refractivity contributions have been derived for 3D QSAR studies [266–269]. Audry *et al.* defined molecular lipophilicity potentials [914–916] for the determination of lipophilic and hydrophilic regions of a molecule.

Abraham and Leo [917] proposed the conversion of hydrophobic fragmental constants to atomic contributions for the evaluation of hydrophobic interactions between molecules. Hydrophobic atom constants were estimated from published hydrophobic fragmental constants f [50, 917] and applied to calculate the hydrophobic field in a grid around the molecule [918]. The program HINT [38, 918–921] maps such hydrophobic fields of molecules for 3D QSAR studies. HINT may also be used to estimate the log P values of molecules [920]. Additional routines KEY, LOCK, and LOCKSMITH were developed to predict complementary hydrophobicity maps of drug structures from a known receptor surface and *vice versa* receptor maps from drugs [920, 921]. KEY uses the 3D structure of a binding site to model the hydrophobicity profile of the ideal ligand; correspondingly, LOCK uses substrate or drug structures to model the hydrophobic character of the binding site. LOCKSMITH identifies the significant hydrophobic areas of a series of active agents, based on their biological activities, as *e.g.* exemplified for allosteric modifiers of hemoglobin [921].

Distance geometry, originally introduced by Crippen [38, 922], is an approach to calculate 3D coordinates from a set of distances; nowadays, it is routinely used for the calculation of 3D structures of organic compounds, peptides, and small proteins from 2D NMR measurements. The method was extended to quantitative receptor modeling [31, 923–926] and Crippen demonstrated its application for the QSAR of DHFR inhibitors [31, 32, 432, 926–931]. Approximate 3D structures of the ligands are constructed in the distance geometry approach and low-energy conformations are selected. Each ligand is characterized by ligand points, *i.e.* by positions of atoms or groups. Different conformers are considered by defining upper and lower boundaries of the distances between all different points. Also the binding site is defined by points, which can be either empty (a ligand point may be there) or filled (no ligand allowed). Different binding modes may hypothetically be assumed; allowed binding modes result from conformations where all the ligand points occupy empty space. The site points are classified to differentiate the nature or intensity of interaction. Then the smallest set of site points is determined and interaction parameters are calculated by a least squares procedure to minimize the differences between theoretical and observed binding energies. Scope and limitations of the method have been discussed [926]. The method was also applied to β_2-adrenoceptor ligands [738, 925, 932–934].

Another distance approach was developed for the conformational analysis of calcium antagonistic benzothiazepines and related analogs [935, 936]. The most polar groups of the analogs are identified. Then the distances between these groups are calculated in different conformations. After a principal component elimination of redundant values the remaining distances are correlated with biological activity values, leading to a hypothesis on the active conformations at the receptor site.

Voronoi binding site models [937–939] are an approach to correlate binding affinities of ligands with proposed site geometries which are projected onto the surface of polyhedra. The allowed conformational space of each ligand is searched for the conformation and binding mode which is the energetically most favorable one, without further hypotheses concerning how the ligands may bind and which parts of the ligands might interact.

REMOTEDISC (receptor modeling based on the three-dimensional structure and physicochemical properties of the ligand molecules) [940–944] starts from the low-energy conformation of a reference compound. Low-energy conformations of all other analogs are automatically superimposed to achieve a maximum overlap of atom-based physicochemical properties, *e.g.* the molecular lipophilicity potential. The relative importance of the different physicochemical properties of different regions of the active site cavity is then determined by using stepwise regression analysis. The method has been applied to

antiviral ribonucleosides [940],
benzodiazepine receptor ligands [941],
the binding site of the nucleoside transporter protein [942, 943], and
the *in vitro* antitumor activities of purine-6-sulfenamides, -sulfinamides, and -sulfonamides [944].

Like in the CoMFA approach (chapter 9.3), also in this case PLS analysis might be more suitable.

A more sophisticated method, the hypothetical active site lattice (HASL) model, was developed by Doweyko [38, 945]. Minimum energy conformations are calculated for similar or dissimilar ligands and the molecules are placed in a three-dimensional grid. A user-selected physicochemical property, *e.g.* lipophilicity or electron density, is added to it as the fourth dimension. The resulting multidimensional lattices are automatically superimposed by an iterative fitting. In this manner a hypothetical active site lattice is formed, capable of predicting the relative orientations and affinities of the ligands. Wiese and Coats [946] modified the HASL approach by using PLS analysis instead of an iterative fitting and obtained better results, especially in the predictive ability of the models. The HASL approach has been applied to

glyoxalase inhibitors [945],
dihydrofolate reductase inhibitors [945],
glutamine synthetase inhibitors [946],
hexokinase inhibitors [946], and to the
analysis of the sequence specificity of DNA alkylation [947].

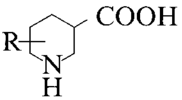
Two-dimensional or even three-dimensional electron-topological matrices of congruity are constructed in the so-called electron-topological approach [948]; they contain both electronic and geometric characteristics of the atoms of a certain drug molecule, *e.g.* atomic charges, polarizabilities, HOMO and LUMO energies, *etc.*; the non-diagonal elements (*i.e.* the elements attributed to two different atoms) can be bond-related indices, if both atoms are covalently linked, or interatomic distances, if they are nonbonded atoms. In the next step the matrices of active and inactive molecules are compared to find the matrix elements which are present in the active analogs and absent in the inactive ones. QSAR applications of this approach have been reviewed [948].

A logico-structural approach to computer-assisted drug design [949, 950] was further developed to a three-dimensional structure-activity expert system Apex-3D [951, 952], which recognizes pharmacophores in biologically active molecules. Various descriptors are used, *e.g.* aromatic ring centers, lipophilic regions, electronic and hydrogen bond donor and acceptor properties, quantum-chemical indices, as well as atomic contributions to hydrophobicity and molar refractivity. The program compares the descriptors and their distances with respect to active and inactive analogs and stores the results as rules in a knowledge base, which can be used to predict the activities of new compounds. Apex-3D is claimed to be more robust than the classical QSAR methods.

9.3. Comparative Molecular Field Analysis (CoMFA)

Comparative molecular field analysis (CoMFA) developed slowly. From the very first formulation of a lattice model to compare molecules by aligning them with a putative pharmacophore and by mapping their surrounding fields to a three-dimensional grid [953], it took nearly 10 years till the CoMFA model, as it is used nowadays, was defined [35–38].

An application of the DYLOMMS (dynamic lattice-oriented molecular modeling system) approach [34, 35, 954, 955], as the method was called till 1987, was only reported once [34, 955]. For GABA-uptake inhibitors potential maps were calculated in a grid, containing 11 × 11 × 11 points; principal component analysis was used to deal with that many variables. The grid points around the 2- and the 5-positions of 17 nipecotic acid derivatives (**59**) were the major contributors to the principal components that were correlated with activity (log $1/IC_{50}$ values; 5 components; $s = 0.39$). It was already recognized that the PLS method might be more efficient to eliminate the redundancy in the X variables.



(59)

A real advance resulted in 1987 [35]; the method was still named DYLOMMS, but now it used grids including several thousands of points, PLS analysis and, most important, a cross-validation procedure (see chapter 5.3) to check the predictive ability of different models. An excellent result was obtained for the binding of steroids to the corticosteroid-binding globulin (5 components; $n = 20$; $r^2_{FIT} = 0.992$; $s_{FIT} = 0.045$; $r^2_{PRESS} = 0.860$; $s_{PRESS} = 0.434$) and other binding, uptake, and enzyme inhibition data.

To illustrate the degree of innovation, it should be mentioned that the book *Quantitative Drug Design* [39], comprising 766 pages and published in 1990, contains less than one page (!) on 3D QSAR methods related to CoMFA [882] while only three years later numerous successful applications have proven the usefulness of the CoMFA approach [38]. Thus, the statement made by Blaney and Hansch that *"there have been many reports in which electrostatic surface potentials have been calculated for ligands and some for enzymes, but it is not yet clear how these can be put into numerical form for use in QSAR"* [288] seems no longer to be true.

Many of the approaches that were reviewed in chapter 9.2 aimed at developing the strategy of this 3D QSAR model. A detailed description of the method was

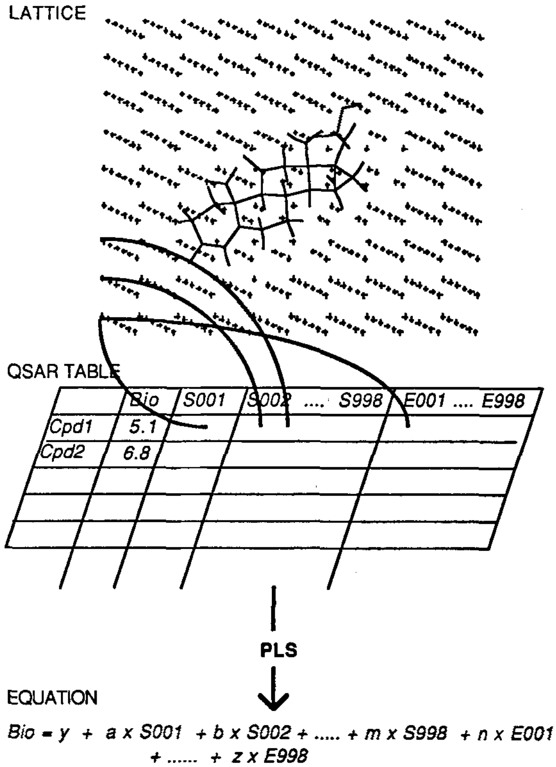


Figure 54: The CoMFA method. Field calculation in a grid and generation of the QSAR table for PLS analysis (reproduced from Figure 1 of ref. [36] with permission from the American Chemical Society, Washington, DC, USA).

given in a later paper, at the time when the method was called comparative molecular field analysis (CoMFA) (Figure 54) [36, 37] and a program became commercially available [956].

There are several important and critical steps in a CoMFA study [36, 38, 957, 958]. First, a group of compounds having a common pharmacophore is selected. Then three-dimensional structures of reasonable conformations must be generated from the 2D structures. Several 2D/3D conversion procedures are in use or have been described in literature, *e.g.* CONCORD [38, 959, 960], AIMB [961], WIZARD and COBRA [962–966], and CORINA [967]. Alternatively, 3D structures derived from crystallographic analyses or 2D NMR studies may be used.

The program MIMUMBA starts from an arbitrary 3D structure and generates other conformations by changing individual torsion angles to values which correspond to different conformational preferences [968]; the distribution profiles of allowed and preferred torsion angles are retrieved from corresponding partial structures in the Cambridge database of crystal structures [141].

Methods for searching the conformational space of small and medium-sized molecules have been reviewed [969]. A general problem in the prediction of 3D structures of flexible drugs is the fact that conformations *in vacuo* (*i.e.* all conformations calculated by force field, semiempirical, or *ab initio* methods) may be significantly different from the conformations in aqueous solution (2D NMR measurements), in the crystal (X-ray structure determination), and at the binding site of a protein. While *in vacuo* intramolecular hydrogen bonds predominate, intermolecular hydrogen bonds (to the surrounding water molecules) determine the geometry of low energy conformations in aqueous solution, in the crystal (to the adjacent molecules), and in the bound state at a receptor surface (to the acceptor and donor groups at the binding site). As a general rule it can be concluded that substrates often are significantly distorted in order to achieve a certain chemical reaction (enzymes are primarily designed to stabilize a transition state, not to display high affinity either to the substrate or to the product of the enzymatic reaction). Thus, enzyme inhibitors should be in their 3D structures as close to the transition state as possible.

The (optionally) energy-minimized structures are stored in a database and fitted to each other according to their chemical similarity by using a pharmacophore hypothesis and postulating orientation rules.

While chemical similarity [970–973] is not easy to define on an objective basis (compare chapter 9.4), every medicinal chemist has some hypothesis or intuition how a pharmacophore can be extracted from the structures of a series of active molecules. Rational methods for the alignment of congeneric molecules are *e.g.* the active analog approach [882, 974–977] and the distance geometry method [31, 926]. In the active analog approach large numbers of different conformations of the molecules are superimposed; often only a few conformations are allowed for all molecules, which lead to a restricted conformational space for their common pharmacophore.

Several other attempts have been made to develop and use computer-assisted or computer-automated procedures (*e.g.* [950–952, 978–988]) for the alignment of molecules. ALADDIN [986] calculates the location of points which may be considered for the superposition of the molecules for all low-energy conformers of a series of compounds (only conformers separated by a certain distance are considered for the same molecule); such points are *e.g.* atoms, ring centers, and projections from the molecule to hydrogen bond donor, acceptor, and charged groups at the binding site. The molecule with the smallest number of possible conformations forms the template; DISCO [987] uses a clique-detection method to find superpositions of the molecules that contain at least one conformation of each compound in the user-defined three-dimensional arrangement of site points. The program CATALYST [988] claims to use a corresponding strategy to derive 3D pharmacophores from sets of active molecules and to search 3D databases for hits which are ordered according to the quality of fit of the molecules to the hypothetical pharmacophore.

In the alignment of molecules it must be considered that multiple binding modes of chemically closely related analogs, once being regarded as rare exceptions, in fact are relatively common. As more and more 3D structures of ligand-protein complexes have become known, more and more examples of multiple binding modes have been identified [38], *e.g.* for

α- and β-N-acetylglucosamine binding to lysozyme [989],
thermolysin inhibitors [990],
dihydrofolate and methotrexate binding to dihydrofolate reductase (Figures 55 and 56) [102, 991, 992],

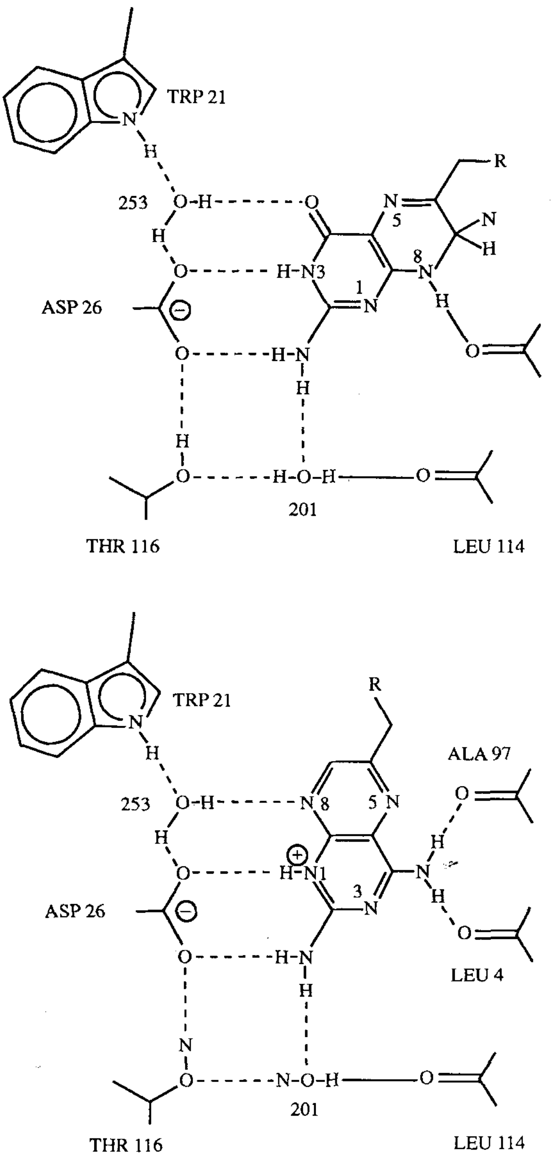

TRP 21
253
ASP 26
THR 116
201
LEU 114
R
N
H
O
H-N3
1
5
8
TRP 21
R
ALA 97
253
N8
5 N
H-N1
3
ASP 26
LEU 4
201
THR 116
LEU 114

Figure 55: Binding mode of the substrate dihydrofolate (upper diagram) and predicted binding mode of the inhibitor methotrexate (lower diagram) to the active site of dihydrofolate reductase; hydrogen bonds are indicated by dashed lines. If the carboxamide oxygen atom of dihydrofolate is replaced by the 4-NH_2 group of methotrexate, an unfavorable donor-donor interaction would result; the rotation of the pteridine ring around the bond between the ring system and the *p*-aminobenzoyl-L-glutamate portion (R) brings the pteridine ring of methotrexate in a much better position, now leading to an additional hydrogen bond to Ala-97 (reproduced from Figure 8 of ref. [991] with permission from the American Society for Biochemistry and Molecular Biology, Baltimore, MD, USA).

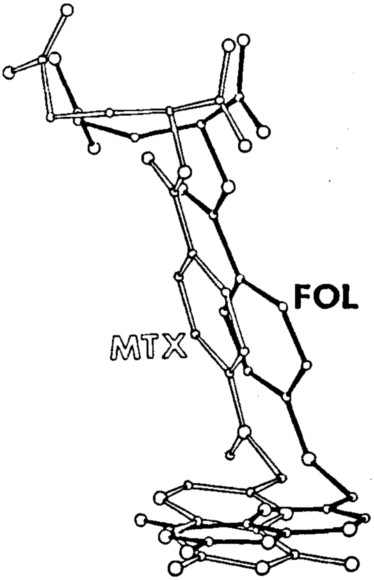


Figure 56: Superimposed binding geometries of the substrate dihydrofolate (FOL) and the inhibitor methotrexate (MTX) to the active site of dihydrofolate reductase, as derived from protein crystallographic analyses (reproduced from Figure 4 of ref. [992] with permission from the American Chemical Society, Washington, DC, USA).

viral coat protein ligands [993 – 995],
binding of isomeric phenylimidazole inhibitors to cytochrome P-450$_{cam}$ (Figure 57) [996],
trimethoprim analogs [997],
carbonic anhydrase inhibitors [998],
purine nucleoside phosphorylase inhibitors (Figure 58) [999],
thrombin inhibitors [1000, 1001], and
thyroid hormone analogs [1002].

In some of these cases only minor changes are observed which, nevertheless, have a significant influence on the resulting structure-activity relationships (*e.g.* Figure 58) [999]. In other cases some analogs bind just in the reverse direction [990, 993 – 995], ring systems turn around by 180° (*e.g.* Figure 56) [991, 992], or totally new, unexpected binding modes are observed (*e.g.* Figure 57) [989, 996].

On the other hand, there is at least one example of an identical binding mode of chemically different analogs. Despite the reverse direction of the amide bonds in thiorphan and *retro*-thiorphan, both molecules bind in the same manner, due to a slight modification in their interaction geometries (Figure 59) [1003]. In addition to the flexibility of the ligands also a certain flexibility of the binding site has to be taken into account, sometimes leading to different positions of amino acid side chains and even to dislocations of the protein backbone.

Once the molecules are aligned, a grid or lattice is established which surrounds the set of analogs in potential receptor space; although 0.2 nm (= 2 Å) is the default value for the distance between the grid points, other values may be chosen; smaller distances seem desirable, but they lead to unreasonably large numbers of grid points.

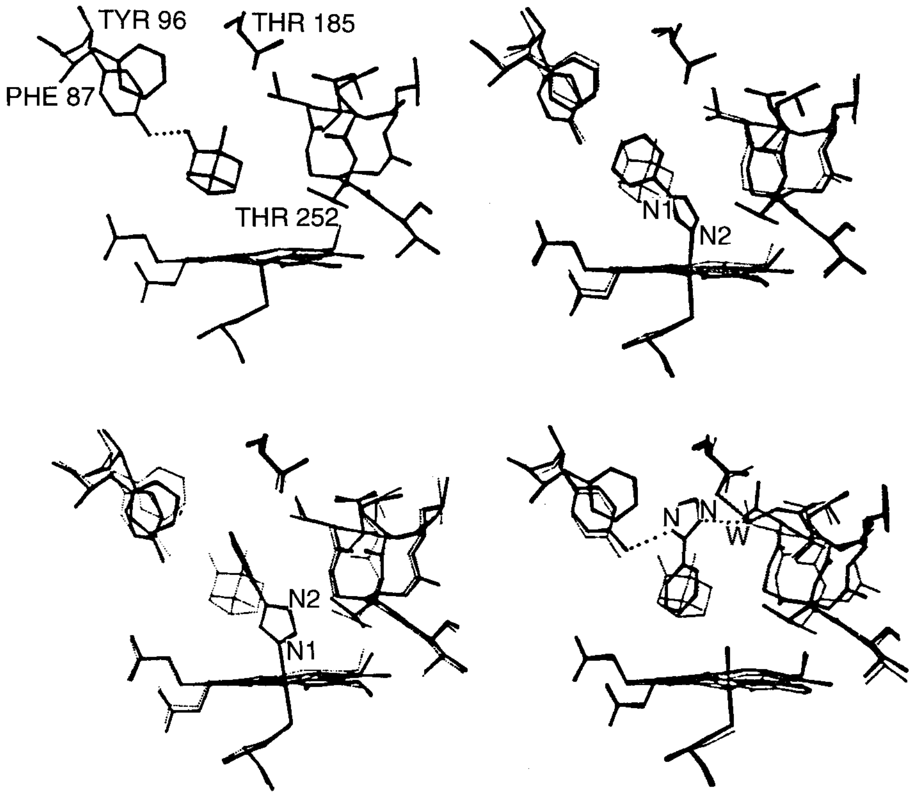


Figure 57: Binding of camphor and three different inhibitors to cytochrome P-450$_{\text{cam}}$, as determined by protein crystallography of the substrate and inhibitor complexes. The camphor molecule is kept in place by hydrophobic interactions and by a hydrogen bond between its carbonyl group and the phenolic hydroxyl group of Tyr-96 (upper left diagram); in all other diagrams the refined structures (solid lines) are superimposed to the native, camphor-bound structure (dotted thin lines). 1-Phenylimidazole (upper right diagram) binds with its N2 to the Fe atom of the heme system and its phenyl ring replaces the camphor molecule; the same holds true for 4-phenylimidazole (N1 binds to the Fe atom; lower left diagram). A completely different binding mode is observed for 2-phenylimidazole (lower right diagram): a N-Fe bond cannot be formed because of unfavorable steric interactions of the phenyl ring with the heme system. Thus, one N atom binds to the hydroxyl group of Tyr-96, the other one binds to an inserted water molecule W, which itself is fixed by hydrogen bonds to Thr-185 and Asp-251; both amino acid side chains move in the direction of this water molecule to form these bonds (reproduced from Figure 4 of ref. [996] with permission from the American Chemical Society, Washington, DC, USA).

The fields which a certain probe atom would experience at every grid point are calculated for each molecule, leading to thousands of columns in the X block. For the steric field a $(a/r^{12} - b/r^6)$ Lennard-Jones potential (Figure 2, chapter 1.2) is calculated, the electrostatic field is a $1/r$ Coulomb potential; the use of $r^{12} - r^6$ van der Waals potentials has been criticized to produce unrealistic results in the

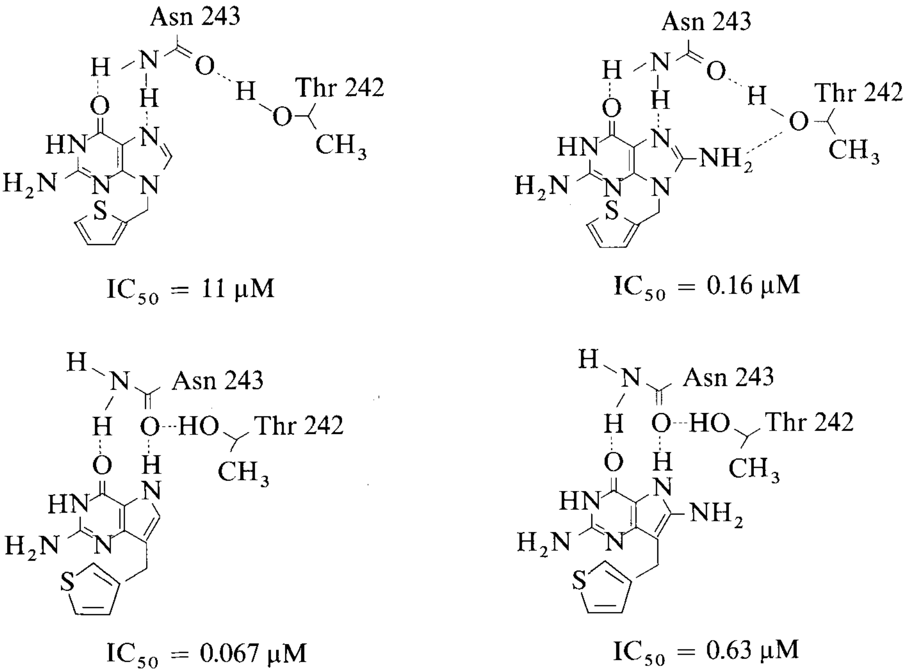


Figure 58: Binding of unsubstituted (left) and 8-amino-substituted (right) purines (upper formulas) and 9-deazapurines (lower formulas) to purine nucleoside phosphorylase. When N7 has an attached hydrogen (9-deazapurines), Asn-243 undergoes a shift to form a highly favorable N(7)-H...O hydrogen bond, increasing the affinity of the unsubstituted analog by a factor of 170 (left formulas). The introduction of an 8-amino group into the purine leads to a new hydrogen bond to Thr-242, thereby increasing the inhibitory potency by a factor of 70 (upper formulas). No such hydrogen bond can be formed to the 8-amino group of the 9-deazapurine. On the contrary, the methyl group of Thr-242 approaches the 8-amino group, generating a hydrophobic environment, which decreases affinity by a factor of 10 (lower formulas). Proposed hydrogen bonds, derived from X-ray structure analyses of the inhibitor complexes, are shown as dotted lines (reproduced from Figure 4 of ref. [999b] with permission from the American Chemical Society, Washington, DC, USA).

case of steric overlap of the ligand and the receptor [1004]. Large positive energy values, *i.e.* grid points inside the molecules, are set constant at certain cut-off values to avoid unreasonably large energy values.

Normally, the steric and electrostatic fields are kept separate for ease of interpretation of the results. Grid points without variance (*e.g.* inside the volume shared by all molecules) or with small variance (*e.g.* in the corners of the box, far away from the van der Waals spheres of the molecules) are eliminated before the PLS analysis is performed.

Other fields than those implemented in the CoMFA program have been proposed for 3D QSAR analyses, *e.g.* different interaction fields calculated by the program GRID [33, 909, 910] or hydrophobic fields derived from HINT [918–921] (chapter 9.2). In addition, any other parameters, *e.g.* physicochemical properties like log P or quantum-chemical indices, may be added to the X block, if they are properly

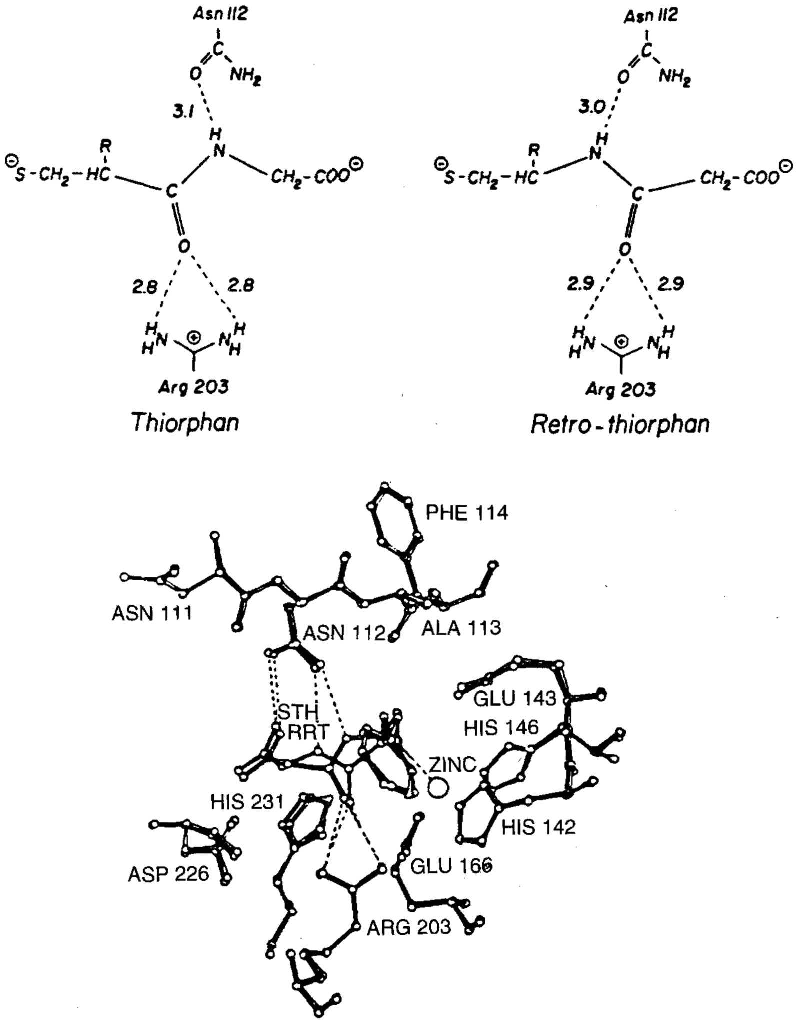

Asn 112
C
O
NH2
3.1
H
N
R
S-CH2-HC
C
CH2-COO
O
2.8
2.8
H
N
H
N
H
H
Arg 203
Thiorphan
Asn 112
3.0
2.9
2.9
Arg 203
Retro-thiorphan
PHE 114
ASN 111
ASN 112
ALA 113
GLU 143
STH
RRT
HIS 146
ZINC
HIS 231
HIS 142
ASP 226
GLU 166
ARG 203

Figure 59: Binding of (S)-thiorphan and (R)-*retro*-thiorphan to the active site of thermolysin; schematic illustration (upper part) and three-dimensional structures (lower part) (reproduced from Figures 3 and 4 of ref. [1003] with permission from the American Chemical Society, Washington, DC, USA).

weighted; as discussed in chapter 5.3, PLS analysis will not reveal the explanatory effect of an unweighted single variable, even if this variable significantly contributes to explanation and prediction of the data.

The molecular alignment, *i.e.* the selection and relative orientation of a certain 3D structure out of several conformers of each molecule, is the most important determinant in a CoMFA study. Recently, a field fit procedure [1005, 1006] has been proposed to improve the alignment; the objective of the field fit is to minimize the residual mean square differences between a fixed template field (consisting of a steric and an electrostatic field) and the fields of the molecules to be aligned. It is claimed that most often better results are obtained by using this field fit after a first, preliminary alignment of the molecules. No better results could be obtained in comparative studies if 3D coordinates from the X-ray structures of ligand-protein complexes were used instead of a user-defined alignment of the ligands [1007], even in the case of analogs binding in different modes (!).

Steric and electrostatic fields are only suitable for the description of the enthalpic contribution of a ligand-protein interaction; as the free energy of binding also involves entropic effects, one cannot *a priori* expect that this default combination of fields is sufficient for the quantitative description of binding affinities or biological activities [1007]. The addition of hydrophobic fields and, optionally, of internal conformational entropy parameters may improve this situation. The combination of global lipophilicity parameters with CoMFA fields (corresponding to a mixed Hansch analysis/CoMFA approach) might be appropriate in certain cases. The CoMFA method should not be applied if transport and distribution predominate, *e.g.* in the case of nonlinear lipophilicity-activity relationships (see below).

The last step in a CoMFA study is a partial least squares (PLS) analysis (chapter 5.3) to determine the minimal set of grid points which is necessary to explain the biological activities of the compounds. Most often good to excellent results are obtained. However, the predictive value of the model must be checked by cross-validation; if necessary, the model is refined and the analysis is repeated until a model of high predictive ability is obtained.

The PLS variant GOLPE [38, 614, 615] seems to be better suited than ordinary PLS analysis because it eliminates variables not contributing to prediction in a stepwise procedure; some recent applications in CoMFA studies confirm [615, 1008] that the predictive power of the CoMFA model increases after reduction of the number of variables according to the GOLPE procedure. An alternative to GOLPE could be the use of genetic algorithms; it has to be proven whether this procedure gives better results than GOLPE without increasing the risk of chance correlations.

The risk of chance correlations seems to be low in CoMFA studies if arbitrary orientations of the molecules are chosen instead of a reasonable alignment [1005, 1006] or if series of random numbers are correlated with biological activities [1009]. On the other hand, CoMFA offers many different options, resulting from slightly different alignments of the molecules or certain side chains, different positions or sizes of the grid box, different distances between the grid points, the use of different fields and additional variables, different values for the cut-off of large positive energy values, *etc.*; while this flexibility of the CoMFA method makes it a powerful tool to perform a QSAR study, the risk of user-generated chance correlations increases considerably. Even cross-validation offers no guarantee to avoid such chance correlations. In large groups of compounds or in groups including several clusters of compounds which are closely related in their respective structures as well as in their biological activities, a leave-one-out cross-validation will not spot such fortunate results.

Biological activities of new compounds can be predicted by transforming the PLS result into a multiple regression equation (*e.g.* [608, 1010]). For a comparison

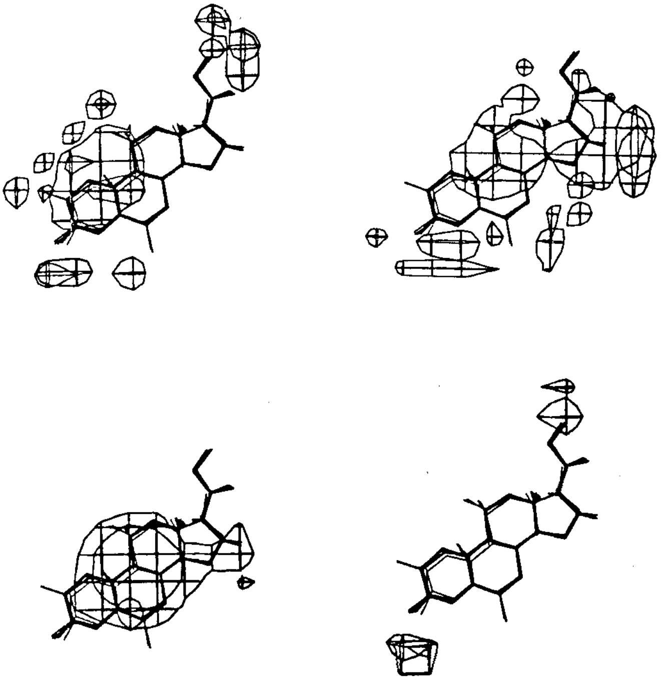


Figure 60: Graphical presentation of the results of CoMFA analyses. Positive (left diagrams) and negative (right diagrams) steric (upper diagrams) and electrostatic (lower diagrams) interactions of steroids with the human corticosteroid-binding globulin are shown as contour maps (reproduced from Figures 4–7 of ref. [1010] with permission from ESCOM Science Publishers BV, Leiden, The Netherlands).

of classical QSAR analyses with 3D QSAR studies the PLS results have been presented in the form of multiple regression equations, including the PLS components as the independent variables (see below).

As in all other QSAR methods, predictive ability depends on the distances of the compounds to be predicted from the other congeners of the series. One cannot expect reliable predictions of biological activity values for analogs having additional side chains or groups with significantly different electronic properties.

Most often the results of a CoMFA study are presented in graphical form, with contours for favorable and unfavorable regions of the different fields (*e.g.* Figure 60). Difference maps were proposed as tools to analyze and identify areas of interest with respect to activity and selectivity, if two different types of biological activities are compared [1011].

CoMFA results are difficult to compare with each other because of the different fields, box sizes, and other options. In addition to this, prior CoMFA versions (up to 5.4) contained an error in the calculation of the electrostatic fields [1012]. Autoscaling of variables should be avoided; PLS analysis may produce wrong results if individual grid points largely reduce their variance in the cross-validation, which seemingly occurs quite often [1013]; in the CoMFA cross-validation wrong

r^2 values (too large, due to a wrong correction factor) resulted in version 5.3 [1012]; the revised r^2 values are still wrong (slightly too small) because now identical overall y_{mean} values are used for all cross-validation groups instead of the correct y_{mean} values of each individual group [1013].

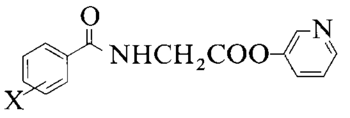
In applying 3D QSAR methods related to CoMFA one should bear in mind that *"a computer based method of designing a molecule which will bind to a larger molecule ... comprising the following steps ... synthesizing a molecule with atoms arranged to occupy or not occupy ... the three-dimensional spaces/volumes indicated in the display as being critical to binding of the molecule to the larger molecule"* is protected by a US patent [37]. Whether it is really possible to protect not only the software but also the design and synthesis of possible drug candidates, without claiming at least general structures and biological actions in the patent, remains open to discussion.

The use of additional (or other) fields than the default steric and electronic fields of the original CoMFA method, together with PLS analysis or GOLPE, is quite common as a valuable extension of the CoMFA program, but it also constitutes an alternative to the relatively expensive commercial software. A combination of shape, lipophilic, steric, and electrostatic potentials in comparative analyses was termed comparative molecular potential analysis (CoMPA) [1014].

CoMFA and related 3D QSAR approaches have been applied to correlate various physicochemical properties. Equilibrium constants of the hydration of carbonyl groups could be explained by a combination of C=O bond order, steric, and electrostatic fields [1005]. 3D QSAR studies that correlate σ, inductive, and resonance parameters of benzoic acids [1015, 1016] as well as pK_a values of clonidine analogs [1017] show that a H^+ field precisely describes such electronic parameters, *e.g.* $\sigma_{m,p}$ of benzoic acids ($n = 49$; $r_{FIT} = 0.976$; $s_{FIT} = 0.082$; $s_{PRESS} = 0.093$). Steric parameters of benzoic acids, like surface area and van der Waals volume can be described by a steric field alone, while E_s values of acetic acid methyl esters need a combination of both steric and electrostatic fields ($n = 21$; $r_{FIT} = 0.984$; $s_{FIT} = 0.133$; $s_{PRESS} = 0.209$) [1016].

CoMFA fields were also proposed and used to derive new steric and electronic parameters for classical QSAR studies [38, 1018].

Results from CoMFA studies have been compared with those from Hansch analyses [38, 1019–1023] and the minimal topological difference (MTD) method [1024]. Examples for the comparison of Hansch equations with CoMFA studies are *e.g.* the papain hydrolysis of N-(X-benzoyl)glycine pyridyl esters (**60**) (eqs. 204, 205; *Zn* = PLS component *n* of the corresponding field; compare chapter 7.1) [1019, 1020], the emulsin-catalyzed hydrolysis of phenyl-β-D-glucosides [1020], the mutagenic activities of substituted (*o*-phenylenediamine)platinum dichlorides [1020], dihydrofolate reductase (DHFR) inhibition [1020], and some other biological activities [38, 1021–1023].



(60)

$$\log 1/K_m = 0.40(\pm 0.06)\,\pi_4 + 4.40(\pm 0.09) \quad (204)$$
$$(n = 22;\ r = 0.946;\ s = 0.176)$$

$$\log 1/K_m = 0.056(\pm 0.003)\,Z1_{H_2O} + 0.051(\pm 0.006)\,Z2_{H_2O} + \\ + 0.026(\pm 0.005)\,Z3_{H_2O} + 0.040(\pm 0.008)\,Z4_{H_2O} + \\ + 0.037(\pm 0.012)\,Z5_{H_2O} + 4.616(\pm 0.022) \quad (205)$$
$$(n = 22;\ r = 0.986;\ s = 0.103;\ s_{PRESS} = 0.334)$$

The results show that a H^+ field is suitable for the description of the interaction of ligands with a positively charged receptor group, while a H_2O probe in GRID describes hydrophobic as well as hydrogen-bonding effects [1019, 1023].

The CoMFA methodology was also used to describe nonlinear lipophilicity-activity relationships, *e.g.* the inhibitory activities of quaternary alkylbenzyldimethylammonium compounds *vs. Clostridium welchii* (eqs. 206–208) [1025], other antibacterial and hemolytic activities [1026, 1027], and toxic activities of alkanes in mice (eqs. 209–211) [1026]; the results of classical QSAR studies (eqs. 206, 207, 209, and 210) [23, 440] were compared with the corresponding CoMFA results (eqs. 208 and 211) [1025–1027]; only homologous series of compounds were investigated.

$$\log MIC = -0.17(\pm 0.06)(\log P)^2 + 0.91(\pm 0.21)\log P + 3.87(\pm 0.21) \quad (206)$$
$$\text{optimum } \log P = 2.68$$
$$(n = 12;\ r = 0.966;\ s = 0.230)$$

$$\log MIC = 0.942(\pm 0.26)\log P - 1.274(\pm 0.31)\log(\beta P + 1) + 3.774(\pm 0.18) \quad (207)$$
$$\log \beta = -1.800 \qquad \text{optimum } \log P = 2.25$$
$$(n = 12;\ r = 0.983;\ s = 0.172)$$

$$\log MIC = 0.071(\pm 0.003)\,Z1_{H_2O} + 0.066(\pm 0.005)\,Z2_{H_2O} + 4.410(\pm 0.025) \quad (208)$$
$$\text{optimum } \log P = 2.82$$
$$(n = 12;\ r = 0.995;\ s = 0.088;\ s_{PRESS} = 0.156)$$

$$\log 1/LD_{100} = -0.107(\pm 0.05)(\log P)^2 + 0.939(\pm 0.47)\log P + 0.200 \quad (209)$$
$$\text{optimum } \log P = 4.37$$
$$(n = 11;\ r = 0.930;\ s = 0.148)$$

$$\log 1/LD_{100} = 0.958(\pm 0.11)\log P - 1.309(\pm 0.13)\log(\beta P + 1) - 0.654 \quad (210)$$
$$\log \beta = -3.52 \qquad \text{optimum } \log P = 3.96$$
$$(n = 11;\ r = 0.996;\ s = 0.039)$$

$$\log 1/LD_{100} = 0.038(\pm 0.004)\,Z1_{H_2O} + 0.038(\pm 0.006)\,Z2_{H_2O} + + 1.940(\pm 0.029) \qquad (211)$$

$$\text{optimum } \log P = 4.75$$

$$(n = 11;\ r = 0.992;\ s = 0.054;\ s_{PRESS} = 0.170)$$

However, despite the simplicity of the analyses and the good correlations obtained in these studies, a ligand interaction-based model like the CoMFA method should not be used to model nonlinear effects arising from transport and distribution; no reasonable results can be expected for sets of compounds which are no homologous series. Better and theoretically more consistent alternatives would be the addition of suitably weighted log P values to the CoMFA table, the use of lipophilicity similarity matrices (chapter 9.4), or the correlation with log P values in the classical manner, applying either the parabolic or the bilinear model.

Due to the definition of the method, many CoMFA studies and related 3D QSAR analyses, where GRID and other molecular fields are implemented in a CoMFA-like model, deal with the quantitative description of ligand-protein interactions, like enzyme inhibition, *e.g.*

angiotensin converting enzyme inhibition [35, 958, 1028 – 1030],
prolyl endopeptidase inhibition [1031]
prostaglandin synthase inhibition [958],
renin inhibition [1007],
thermolysin inhibition [1007], and
tyrosine kinase inhibition [1032];

substrate properties of enzyme ligands, *e.g.*

the toxic activation of MPTP analogs by monoamine oxidase [1033, 1034];

receptor binding, *e.g.*

binding of α_1-adrenergic agonists [1008],
benzodiazepine receptor binding [1035 – 1038],
binding of tetrahydrocannabinol analogs to the cannabinoid receptor [1039],
binding of CCK-A antagonists [1040],
affinity of phenyltropane carboxylic acids to the cocaine binding site [1041],
D_2 receptor affinity of clozapine analogs [1042] and salicylamides [1043],
muscarinic receptor binding [1044, 1045],
non-competitive GABA receptor antagonism [1046],
non-NMDA antagonist binding [1047],
σ receptor ligands [1048],
serotonin receptor ligands [1014, 1049 – 1051], and
the affinities of various halogenated dibenzofurans, dibenzodioxins, and biphenyls to the tetrachlorodibenzodioxin receptor [615, 1052];

other ligand-protein interactions, *e.g.*

the binding of steroids to various carrier proteins [35, 36, 919, 958, 1010, 1011, 1024, 1053],
GABA-uptake inhibition [35],
ligand binding to viral coat proteins [1007, 1054], and
antigenic complexes between peptides and a histocompatibility glycoprotein [1055, 1056].

CoMFA was also applied to correlate various other biological activities, *e.g.*

the induction of human leukemia cell differentiation [1057],

response of tumor cells to estrogens [1058],

diuretic activities of sulfonamides in the isolated perfused tubulus of rabbit kidney [1059],

various biological activities of prostaglandin analogs [1060], and

bone resorption regulation by biphosphonate esters [1061, 1062].

However, the more distant a biological model from the pure ligand-receptor interaction is, *e.g.* the toxicities of alkanes in mice [1026] or the anticoccidial activities of triazines in chicken [1063], the more unrealistic becomes the application of 3D QSAR methods.

Not all the above-mentioned studies fulfill the necessary statistical requirements, especially proper PLS component selection by cross-validation. The crucial influence of different CoMFA options on the obtained results has been demonstrated [38, 1059]. More objective alignment procedures, additional fields (*e.g.* a better consideration of polarizability), and more efficient variable selection procedures are needed. In addition, further research is necessary to find out which of the different CoMFA options should be selected to obtain stable and reproducible results. Theoretical and practical aspects, methodology and applications, as well as some limitations of CoMFA and related approaches have recently been reviewed in a book on 3D QSAR methods [38].

9.4. Molecular Similarity QSAR Analyses

Most often structurally similar molecules have a comparable range of physicochemical properties and biological activities, while structurally dissimilar molecules differ in their properties both qualitatively and quantitatively. Classical QSAR methods as well as 3D QSAR approaches are implicitly based on this trivial fact.

Richards *et al.* [1064, 1065] recently developed a new method to correlate the 3D similarities of molecules in a direct manner with their biological activities. In the first step 3D structures of reasonable conformations of all molecules are generated and aligned in space, as in the CoMFA approach. Then similarity indices between all pairs of molecules are calculated and the resulting N × N similarity matrices are correlated with biological activities.

The Carbo index R_{AB} was originally proposed as a method of comparing molecules in terms of their electron density ϱ [1066]. In a more general version (eq. 212) this similarity index can be applied to compare any structural properties P_A and P_B of two molecules A and B.

$$R_{AB} = \frac{\int P_A P_B \, dv}{(\int P_A^2 \, dv)^{1/2} (\int P_B^2 \, dv)^{1/2}} \qquad (212)$$

An alternative to the Carbo index is the Hodgkin index H_{AB} (eq. 213) [1067]. While the Carbo index is sensitive to the shape of the property distribution, the Hodgkin index reflects more its magnitude.

$$H_{AB} = \frac{2 \int P_A P_B \, dv}{\int P_A^2 \, dv + \int P_B^2 \, dv} \tag{213}$$

Both indices have been applied to other molecular properties, *e.g.* electrostatic potentials and fields, and to describe the shape similarity of molecules [1068, 1069].

The program ASP (automatic similarity package) has been developed to compute the electrostatic potential and field at certain points on a grid [1070, 1071]. A modification of the ASP program uses Gaussian function approximations rather than grids to calculate electrostatic potential and shape similarity [1064, 1072, 1073]. Singularities (due to the fact that 1/r approaches infinity values near atom centers) are avoided by considering properties only outside the van der Waals volumes of the molecules.

In addition, linear and exponential similarity indices, L_{AB} and E_{AB} (eqs. 214 and 215), have been defined [972, 1065, 1074] to evaluate electrostatic similarities in grid points; the sum over all individual grid point values is divided by the number of grid points n.

$$L_{AB} = \frac{\sum_{i=1}^{n} \left(1 - \frac{|P_A - P_B|}{\max(|P_A|, |P_B|)}\right)}{n} \tag{214}$$

$$E_{AB} = \frac{\sum_{i=1}^{n} \exp^{-\frac{|P_A - P_B|}{\max(|P_A|, |P_B|)}}}{n} \tag{215}$$

Molecular similarity indices can be optimized by fixing one molecule and translating and rotating the other one until their similarity reaches an optimum. This ASP-implemented procedure constitutes another promising tool for the rational alignment of molecules in 3D space, also considering the flexibility of molecules [1075].

In a preliminary study [1070], Hodgkin electrostatic potential molecular similarity indices of nitromethylene insecticides were calculated for all analogs in comparison with the most active compound. After the elimination of three data points a plot of the similarity indices *vs.* biological activity values showed a nice linear correlation.

The current version of this approach uses N × N matrices (N being the number of molecules included in the data set), in which the 3D similarity indices of each molecule as compared to every other molecule of the whole set are the non-diagonal elements of the similarity matrix [1064, 1065]. In an application to the affinities of 31 steroids to corticosteroid- and testosterone-binding globulins (same data as used in CoMFA studies, *e.g.* [36]), Carbo index similarity matrices were calculated for the shape and the electrostatic potentials of all molecules. Using the similarity matrices as input data, a neural network (see chapter 4.6) was applied to qualitatively analyze the binding affinities. In the resulting plots a clear separation of high affinity, intermediate affinity, and low-affinity compounds was achieved [1064].

PLS analysis was used to quantitatively correlate the similarity matrices with the affinities to the different steroid-binding proteins. As in the corresponding CoMFA study [36], compounds 1–21 were used as the test set and compounds 22–31 were predicted from the analyses. The obtained r^2_{PRESS} and s_{PRESS} values [1064] were somewhat inferior, but nevertheless close to the values resulting from the CoMFA studies. Much better results were obtained after the elimination of noise variables (*i.e.* variables not contributing to prediction) before performing the PLS analysis, by using the program GOLPE [614, 615] (chapter 5.3).

In a more detailed study of technique validation and of the comparison of different similarity evaluation methods [1065],

several Carbo indices (also using Gaussian function approximations),
the Hodgkin index (eq. 213),
linear and exponential indices (eqs. 214 and 215), and
the Spearman rank correlation coefficient

were tested as coefficients in the shape and electrostatics similarity matrices. All different types of N × N similarity matrices and also combined electrostatic and shape 2N × N similarity matrices were applied to correlate several data sets [1065] for which CoMFA analyses have been published, *i.e.*

Hammett σ constants of substituted benzoic acids [1015],
pK_a values of clonidine analogs and substituted imidazolines [1017],
binding of steroids to different carrier proteins [36],
the toxic activation of MPTP analogs by monoamine oxidase [1033],
binding to the benzodiazepine inverse agonist receptor site [1035],
the affinity of phenyltropane carboxylic acids to the cocaine binding site [1041], and
anticoccidial activities of triazines in chicken [1063],

using PLS analysis with and without elimination of variables not contributing to prediction. As good r^2_{PRESS} values as in the CoMFA studies and in some cases even better r^2_{PRESS} values resulted [1065]. Especially after GOLPE variable selection (Table 30) the reduced matrices produced improved r^2_{PRESS} values in fewer components and with far less variables, as compared to the original 2N × N matrices.

Gaussian electrostatic evaluations of the Carbo index provided better results than grid-based calculations, while in the case of shape the grid-based evaluations seemed to be more suitable. Grid distances of 0.05 nm for shape and 0.2 nm for electrostatics were sufficient to give a satisfactory description of the biological data.

3D QSAR analyses based on similarity matrices offer a valuable new tool for the quantitative description of structure-activity relationships. Also hydrophobic fields and interaction fields with different probe atoms may be implemented, like in CoMFA studies. It is hoped that the preliminary results [1064, 1065] stimulate active research in this field to achieve further methodological improvements. Several CoMFA-inherent problems apparently do not arise in the molecular similarity matrices approach, *e.g.* the cut-off selection, a proper grid spacing, and the elimination of variables having low variance.

Table 30. 3D QSAR analyses of several data sets, using electronic and shape similarity matrices [1065]. Cross-validated r^2_{PRESS} values and the number of significant components (except last three rows, see below) of different analyses are given for comparison

Data set	Complete matrix[a]	GOLPE Reduced Matrix[b]	CoMFA Results[c]
Steroids, testosterone-binding globulin	0.73/3	0.77/2	0.44/4
Steroids, corticosteroid-binding globulin	0.79/2	0.82/2	0.69/2
Anticoccidial triazines	0.74/5	0.73/3	0.47/2
Benzodiazepine receptor ligands	0.69/4	0.72/3	0.59/4
Cocaine receptor ligands	0.54/2	0.64/2	0.57/4
Toxic activation of MPTP analogs by MAO	0.50/2	0.56/2	0.57/4
pK_a Values of clonidine analogs	- [d]	0.32	0.27
pK_a Values of imidazolines	- [d]	0.90	0.69
Hammett σ values of substituted benzoic acids	- [d]	0.12	0.05

[a] combined electronic and shape similarity matrices, except last three rows
[b] similarity matrices after variable selection by GOLPE
[c] results from different groups (for references see text)
[d] only electronic similarity matrices have been used in these cases; as cross-validated r^2_{PRESS} values were not provided in the original CoMFA publications, standard errors of prediction are listed instead

In addition to the similarity indices described above, other similarity indices may be defined and used in QSAR studies. A simple lipophilicity similarity index $a_{ij} = -|\log P_i - \log P_j|$ ($\log P_i$, $\log P_j$ = logarithms of the partition coefficients of molecules i and j) can be applied to describe nonlinear lipophilicity-activity relationships of any type by the corresponding lipophilicity similarity matrices [1013]. For different data sets excellent results were obtained (Table 31), not only in homologous series (as in CoMFA studies [1025—1027]) but also in heterogeneous sets of compounds, where 3D QSAR approaches must fail. A selection procedure based on genetic algorithms was developed for fast and efficient variable elimination in the PLS analyses [1013]. Also in these examples the similarity matrices produced improved r^2_{PRESS} values in fewer components after elimination of variables which did not contribute to prediction (Table 31).

While such lipophilicity similarity matrices do not consider the 3D structures of the molecules, they seem to be appropriate for the incorporation of nonlinear lipophilicity-activity relationships in 3D QSAR analyses, *e.g.* in CoMFA studies. At least from a theoretical point of view lipophilicity similarity matrices should be preferred when the nonlinear lipophilicity-activity relationship does not result from binding but from transport and distribution of the drugs in the biological system, which most often is the case.

Even the combination of extrathermodynamic parameters with similarity matrices, leading to a mixed Hansch analysis/molecular similarity approach, seems to be reasonable and should be further investigated.

Table 31. Nonlinear lipophilicity-activity relationships. Comparison of the bilinear model (r^2 values are given instead of r values) with results from PLS analyses, using lipophilicity similarity matrices $a_{ij} = -|\log P_i - \log P_j|$ [1013]

Data set	Lipophilicity similarity matrices		
(bilinear model)	Parameter[a]	Complete Matrix	Reduced Matrix[b]
Mandelic acid esters,	components	3	1
eq. 99 (chapter 4.4)	r^2_{FIT}	0.982	0.984
(n = 11; r^2 = 0.979;	s_{FIT}	0.150	0.124
s = 0.160)	r^2_{PRESS}	0.956	0.981
	s_{PRESS}	0.255	0.148
Antifungal amines,	components	5	3
eq. 100 (chapter 4.4)	r^2_{FIT}	1.000	0.999
(n = 15; r^2 = 1.000;	s_{FIT}	0.027	0.039
s = 0.031)	r^2_{PRESS}	0.998	0.998
	s_{PRESS}	0.085	0.073
Buccal absorption,	components	2	1
acid and base,	r^2_{FIT}	0.965	0.976
eq. 111 (chapter 4.5)	s_{FIT}	0.124	0.098
(n = 12; r^2 = 0.977;	r^2_{PRESS}	0.939	0.971
s = 0.102)	s_{PRESS}	0.179	0.117
Monoamine oxidase	components	7	4
inhibitors[c],	r^2_{FIT}	1.000	0.999
eq. 112 (chapter 4.5)	s_{FIT}	0.059	0.062
(n = 21; r^2 = 0.998;	r^2_{PRESS}	0.991	0.999
s = 0.118)	s_{PRESS}	0.269	0.079
Barbiturates, transfer	components	3	3
rate constants,	r^2_{FIT}	0.990	0.993
eq. 162 (chapter 7.3)	s_{FIT}	0.066	0.053
(n = 23; r^2 = 0.984;	r^2_{PRESS}	0.984	0.991
s = 0.081)	s_{PRESS}	0.084	0.063
Carbamates, gastric	components	2	1
absorption	r^2_{FIT}	0.939	0.952
eq. 166 (chapter 7.3)	s_{FIT}	0.028	0.023
(n = 8; r^2 = 0.943;	r^2_{PRESS}	0.886	0.932
s = 0.030)	s_{PRESS}	0.044	0.031
Carbamates, intestinal	components	2	2
absorption,	r^2_{FIT}	0.981	0.993
eq. 167 (chapter 7.3)	s_{FIT}	0.026	0.015
(n = 8; r^2 = 0.979;	r^2_{PRESS}	0.965	0.990
s = 0.031)	s_{PRESS}	0.040	0.022

Table 31. (continued)

Placental transfer ratios of various drugs, eq. 170 (chapter 7.3) ($n = 21$; $r^2 = 0.900$; $s = 0.106$)	components	1	1
	r^2_{FIT}	0.901	0.910
	s_{FIT}	0.100	0.095
	r^2_{PRESS}	0.878	0.897
	s_{PRESS}	0.116	0.107
Inflammatory activity, phorbol esters, eq. 176 (chapter 7.4) ($n = 6$; $r^2 = 1.000$; $s = 0.041$)	components	3	2
	r^2_{FIT}	0.999	0.996
	s_{FIT}	0.065	0.098
	r^2_{PRESS}	0.908	0.987
	s_{PRESS}	0.677	0.211
Antihypertensive clonidine analogs[d] eq. 184 (chapter 7.5) ($n = 21$; $r^2 = 0.942$; $s = 0.284$)	components	4	5
	r^2_{FIT}	0.955	0.968
	s_{FIT}	0.251	0.217
	r^2_{PRESS}	0.920	0.956
	s_{PRESS}	0.352	0.268

[a] components = number of significant components, as determined by cross-validation s_{press} values.

[b] variable elimination by a genetic algorithm, based on the predictive ability of the tested variable combinations [1013]

[c] including an indicator variable I as additional parameter

[d] including $IC_{50}\alpha_2$ values as additional parameter

© VCH Verlagsgesellschaft mbH, 1993

10. Summary and Conclusions

Classical QSAR methods still play an important role in drug design, despite the progress in protein crystallography, molecular modeling [1076], and structure-based drug design [911–913, 983, 1077–1092]. QSAR methods are cheap and efficient tools to derive and prove hypotheses on structure-activity relationships in a quantitative manner, especially in those cases where the 3D structure of the biological target is not known. In addition, 3D structure-based drug design is only applicable to ligand design; the quantitative description of transport, distribution, metabolism, and elimination of drugs still remains the domain of classical QSAR methods.

Predictions from QSAR studies (*e.g.* [43, 396, 403, 656, 1093–1098]) and success stories of QSAR-guided drug design [396, 1099, 1100] have been reviewed, *e.g.* the design of

antiallergic purinones [396] and pyrenamines [396, 633],
antidepressant 4-anilinopyrimidines [396, 1101],
cerebral vasodilating benzyldiphenylmethylpiperazines [396, 1100, 1102],
the development of norfloxacin, a novel quinolone carboxylic acid with antibacterial activity [396, 1099, 1103],
antihypertensive quinazolines [396, 1100],
antiinflammatory and analgesic furoindolecarboxamides [396], and
antiulcer benzylpiperazineacetic acid esters [1100, 1104].

Success stories in agrochemistry are the development of the fungicide myclobutanil and the herbicides metamitron and bromobutide [396, 1099].

In most cases QSAR does not directly contribute to the development of a new drug. However, with the increasing evidence of the importance of lipophilicity, dissociation, polarizability, electronic and hydrogen-bonding interactions, and steric fit on drug action, our ability and performance in drug design and in the optimization of lead structures improve. QSAR neither brings the solution of all our problems, nor can it only be considered as an academic game; *"the great advantage of the QSAR paradigm lies not in the extrapolations which can be made from known QSAR to fantastically potent new drugs, but in the less spectacular slow development of science in medicinal chemistry"* [1094].

As stated earlier, prediction is not the main goal of a QSAR analysis. Much more often general conclusions on the reduction of toxic properties, on selectivity, on optimum lipophilicity to pass the blood-brain barrier or, on the other hand, to avoid CNS side effects, are more important for the optimization of a lead structure. As it still is industrial praxis (and will remain for patent reasons) to synthesize and test large numbers of closely related analogs, QSAR is also an important tool to decide when to stop a synthetic program (compare *e.g.* eqs. 178, 179; chapter 7.4) [786, 1095].

Another example of a proper QSAR application is given in eqs. 183–185 (chapter 7.5), which describe the central antihypertensive and peripheral hypertensive activities of clonidine analogs. While no predictions for more active analogs can be derived, three most important conclusions can be drawn from these equations: first, the analogs can be tested in simple *in vitro* systems instead of whole animal models, second, log P values around 1.5 are optimal for the central nervous system-mediated antihypertensive activity, and third, one cannot expect to separate the antihypertensive activity from the hypertensive side effect.

Speculating about the future of QSAR, Hansch [1105] stated that the situation in QSAR may be compared with the development of the Hammett σ constant till 1968, when 43 variations of σ existed. Today, there is rather broad agreement that only four of them, σ, σ^0, σ^-, and σ^+, are needed; however, still no agreement exists how σ can be factorized into field and resonance effects. In deriving a single QSAR equation, one can never be sure that the relationship is a causal one; neither statistical tools nor any other criteria can help in this respect. Only the comparison of QSAR equations from different but related sources [666, 667, 1105] can assure the relevance of certain parameters, *e.g.* lipophilicity for cytochrome P-450 induction by different types of compounds and in different systems [774, 1105] or the importance of a σ term in the description of the rate constants of the hydrolysis of glycine ester amides by different proteases (Table 32) [667, 1105]. Understanding of true structure-activity relationships will depend on lateral validation of QSAR, *i.e.* relating a new QSAR to a matrix of self-consistent structure-activity relationships [1105].

The problem of validation of QSAR studies has also been approached from a statistical point of view [1106]:

Since the biological activity of a compound is most often a combination of several known and unknown subeffects, it is important to measure a fairly large number of different biological effects in different test systems.

Table 32. Comparison of the ρ coefficients in QSAR equations of the hydrolysis of $X\text{-}C_6H_4OCOCH_2NHCOC_6H_5$ (I) and $X\text{-}C_6H_4OCOCH_2NHSO_2CH_3$ (II) by various enzymes; ρ for the uncatalyzed reaction at pH 6 is 1.91 [667, 1105]

Enzyme	Substrate	ρ	pH	Class of Hydrolase
Papain	I	0.57	6	Cysteine
Papain	II	0.55	6	Cysteine
Ficin	I	0.57	6	Cysteine
Ficin	II	0.62	6	Cysteine
Actinidin	I	0.74	6	Cysteine
Bromelain B	I	0.70	6	Cysteine
Bromelain B	II	0.68	6	Cysteine
Bromelain D	I	0.63	6	Cysteine
Subtilisin	I	0.49	7	Serine
Chymotrypsin	I	0.42	6.9	Serine
Trypsin	I	0.71	7	Serine

The aggregation of these effects to one magic number, "the biological activity", reduces the information content.

With certain exceptions, *e.g.* nonspecific, lipophilicity-related biological activities, separate QSARs must be derived for each structural class of chemicals to distinguish between compounds with different mechanisms of biological action.

As most often the structural factors determining biological activities are not directly known, multiparameter tables should be used together with a multivariate characterization approach.

The training set for the development of the model should be selected by a statistical design.

An appropriate data modeling and data analysis system should be used, *e.g.* partial least squares analysis or, in the case of a single activity variable, multiple regression.

A validation procedure must be performed (the use of a fixed number of components in cross-validation runs is criticized because it gives too optimistic results).

The real criterion for the validity of a model is always synthesis and testing of new analogs.

Biological hypotheses can only be tested on a quantitative basis.

We have to live with the fact that only few QSAR studies fulfill these (by no means exaggerated) demands.

The state of the art and some other aspects of the future development of QSAR [39, 1105, 1107] as well as computer-aided drug design [39, 1021, 1107 – 1109] have recently been reviewed.

With the largely increasing number of 3D structures of enzymes and enzyme inhibitor complexes our understanding of the corresponding QSAR equations will further increase [288, 674, 676]. CoMFA and CoMFA-related 3D QSAR approaches [38] are powerful new tools but they will not displace classical QSAR studies. New and hopefully better QSAR parameters will be derived from comparative molecular field analyses [38, 1018]. On the other hand, CoMFA cannot describe biological activities other than ligand-protein interactions, if no global physicochemical parameters are included in the data tables; if they are, then CoMFA imitates classical Hansch analysis. From the current limitations of both methodologies it seems that they may approach each other and may even grow together in the future. Molecular similarity matrices [1064] constitute another promising tool in this respect.

It is the combination of so many different effects which contribute to biological activity that makes the formulation of a sound QSAR model so difficult. To repeat only the most important factors:

Lipophilicity and ionization are responsible for the transport and the distribution of the drug in the biological system.

The drug-receptor interaction is a highly specialized hydrophobic, polar, electronic, and steric interaction; the lipophilicity pattern, the electron density distribution, and the polarizability pattern at the surface of both the drug and its binding site contribute to the interaction energy.

Neither the drug nor its binding site are completely rigid systems; a flexible fit occurs during the binding of the drug.

The topology of the drug molecule must be considered in three dimensions, with restricted rotations or at least conformational barriers around certain bonds.

It is evident that a drug cannot bind in a conformation which is energetically highly unfavorable; however, this does not imply that only minimum energy conformations of a drug are able to interact with the receptor.

The conversion of a favorable conformation to a less favorable one has to be taken into account in those cases, where a higher net free energy results from this new conformation of the ligand.

Even structurally closely related analogs may bind in a completely different manner (multiple binding modes).

The solvation-desolvation balance may be favorable or unfavorable for binding.

The insertion of water molecules between the ligand and its binding site has to be considered.

Entropy effects, *e.g.* the freezing of conformational degrees of freedom, play an important, but much too often neglected role.

Hansch formulated that *"nobody in the world is condemned to work with as many variables as the medicinal chemist"* (with the exception of the people being responsible for the weather forecast, but in their case the working hypothesis is proven within a few days). The multivariate nature of biological systems leads to enormous complications and to the risk of totally wrong conclusions. *"Not all scientists are brave enough to withstand the temptation of accepting a hypothesis as being true only because it more or less fits the experimental data, or to discard their own old concepts for new ones"* [783]. In addition, the elimination of outliers (are they indeed outliers or do they only disprove an arbitrarily chosen, subjective model?) is an often played game to polish the results of QSAR equations. In any case it is the responsibility of the scientist to list the outliers, to look for other reasonable explanations, and to perform further experiments to prove or disprove different hypotheses [783].

We have the ability to intuitively extract high-level information from facts at different levels, without being programmed like a computer. QSAR cannot and will never substitute the creativity and intuition of an experienced medicinal chemist or biologist. But our logical reasoning is limited to one or two, at most three dimensions. QSAR aims at giving an objective interpretation of multidimensional results in medicinal chemistry and at deriving new hypotheses to an extent which is far beyond the intellectual capacity of the human mind. In drug design theoretical chemistry, physical organic chemistry, QSAR studies, protein crystallography, and molecular modeling work hand in hand. They cannot predict drugs, but they lead to deeper insights, they generate new ideas, and they rationalize and shorten the way from the lead structure to an active analog with the desired biological activity, selectivity, and pharmacokinetic properties – the new remedy for a certain disease.

Addendum

This book is a greatly extended version of a text being published as a chapter *"The Quantitative Analysis of Structure-Activity Relationships"* in *Burger's Medicinal Chemistry and Drug Discovery* [1110]. Chemometric methods in QSAR (chapters 4.6 and 5.3) are discussed in more detail in Volume **2** of this series [1111].

QSAR: Hansch Analysis and Related Approaches
by Hugo Kubinyi
© VCH Verlagsgesellschaft mbH, 1993

References

[1] Crum-Brown, A., and Fraser, T. R., *Trans. Roy. Soc. Edinburgh* **25**, 151 – 203, 693 – 739 (1868 – 69)
[2] Richet, M. C., *Compt. Rend. Soc. Biol. (Paris)* **45**, 775 – 776 (1893)
[3] Meyer, H., *Arch. Exp. Path. Pharm.* **42**, 109 – 118 (1899)
[4] Overton, E., *Studien über die Narkose, zugleich ein Beitrag zur allgemeinen Pharmakologie*, G. Fischer, Jena, 1901; English translation by Lipnick, R. L., ed., *Studies on Narcosis, Charles Ernest Overton*, Chapman and Hall, London, 1991
[5] Lipnick, R. L., *Trends Pharmacol. Sci.* **7**, 161 – 164 (1986); **10**, 265 – 269 (1989)
[6] Fühner, H., and Neubauer, E., *Arch. Exp. Path. Pharm.* **56**, 333 – 345 (1907)
[7] Ferguson, J., *Proc. Roy. Soc., Ser.* **B 127**, 387 – 404 (1939)
[8] Bruice, T. C., Kharasch, N., and Winzler, R. J., *Arch. Biochem. Biophys.* **62**, 305 – 317 (1956)
[9] Zahradnik, R., and Chvapil, M., *Experientia* **16**, 511 – 512 (1960)
[10] Zahradnik, R., *Arch. Int. Pharmacodyn. Ther.* **135**, 311 – 329 (1962)
[11] Zahradnik, R., *Experientia* **18**, 534 – 536 (1962)
[12] Hammett, L. P., *Physical Organic Chemistry. Reaction Rates, Equilibria and Mechanism*, 2nd Ed., McGraw-Hill, New York, 1970
[13] Hansen, O. R., *Acta Chem. Scand.* **16**, 1593 – 1600 (1962)
[14] Hansch, C., Maloney, P. P., Fujita, T., and Muir, R. M., *Nature* **194**, 178 – 180 (1962)
[15] Hansch, C., and Fujita, T., *J. Am. Chem. Soc.* **86**, 1616 – 1626 (1964)
[16] Free Jr., S. M., and Wilson, J. W., *J. Med. Chem.* **7**, 395 – 399 (1964)
[17] Hansch, C., *Acc. Chem. Res.* **2**, 232 – 239 (1969)
[18] Hansch, C., *Drug Design*, Vol. **I**, Ariëns, E. J., ed., Academic Press, New York, 1971, p. 271 – 342
[19] Hansch, C., and Clayton, J. M., *J. Pharm. Sci.* **62**, 1 – 21 (1973)
[20] Fujita, T., and Ban, T., *J. Med. Chem.* **14**, 148 – 152 (1971)
[21] Hansch, C., and Yoshimoto, M., *J. Med. Chem.* **17**, 1160 – 1167 (1974)
[22] Kubinyi, H., *J. Med. Chem.* **19**, 587 – 600 (1976)
[23] Kubinyi, H., *J. Med. Chem.* **20**, 625 – 629 (1977)
[24] Brugger, W. E., Stuper, A. J., and Jurs, P. C., *J. Chem. Inf. Comput. Sci.* **16**, 105 – 110 (1976)
[25] Kirschner, G. L., and Kowalski, B. R., *Drug Design*, Vol. **VIII**, Ariëns, E. J., ed., Academic Press, New York, 1979, p. 73 – 131
[26] Wold, S., Ruhe, A., Wold, H., and Dunn III, W. J., *SIAM J. Sci. Stat. Comput.* **5**, 735 – 743 (1984)
[27] Dunn III, W. J., Wold, S., Edlund, U., Hellberg, S., and Gasteiger, J., *Quant. Struct.-Act. Relat.* **3**, 131 – 137 (1984); erratum **4**, 82 (1985)
[28] Fukunaga, J. Y., Hansch, C., and Steller, E. E., *J. Med. Chem.* **19**, 605 – 611 (1976)
[29] Höltje, H.-D., and Kier, L. B., *J. Med. Chem.* **17**, 814 – 819 (1974)
[30] Höltje, H.-D., and Tintelnot, M., *Quant. Struct.-Act. Relat.* **3**, 6 – 9 (1984)
[31] Crippen, G. M., *J. Med. Chem.* **22**, 988 – 997 (1979)
[32] Ghose, A. K., and Crippen, G. M., *J. Med. Chem.* **25**, 892 – 899 (1982)
[33] Goodford, P. J., *J. Med. Chem.* **28**, 849 – 857 (1985)
[34] Wise, M., *QSAR and Strategies in the Design of Bioactive Compounds* (Proceedings of the 5th European QSAR Symposium, Bad Segeberg, 1984), Seydel, J. K., ed., VCH, Weinheim, 1985, p. 19 – 29
[35] Cramer III, R. D., and Bunce, J. D., *QSAR in Drug Design and Toxicology* (Proceedings of the 6th European QSAR Symposium, Portoroz, 1986), Hadzi, D., and Jerman-Blazic, B., eds., *Pharmacochem. Libr.* **10**, Elsevier, Amsterdam, 1987, p. 3 – 12

[36] Cramer III, R. D., Patterson, D. E., and Bunce, J. D., *J. Am. Chem. Soc.* **110**, 5959–5967 (1988)

[37] Cramer, R. D., and Wold, S. B., U.S. Pat. 5 025 388 (June 18, 1991)

[38] Kubinyi, H., ed., *3D QSAR in Drug Design. Theory, Methods, and Applications*, ESCOM Science Publishers, Leiden, 1993

[39] Ramsden, C. A., ed., *Quantitative Drug Design*, Vol. **4** of *Comprehensive Medicinal Chemistry. The Rational Design, Mechanistic Study & Therapeutic Application of Chemical Compounds*, Hansch, C., Sammes, P. G., and Taylor, J. B., eds., Pergamon Press, Oxford, 1990

[40] Purcell, W. P., Bass, G. E., and Clayton, J. M., *Strategy of Drug Design. A Molecular Guide to Biological Activity*, Wiley, New York, 1973

[41] Martin, Y. C., *Quantitative Drug Design. A Critical Introduction* (Medicinal Research Series, Vol. **8**), Marcel Dekker, New York, 1978

[42] Seydel, J. K., and Schaper, K.-J., *Chemische Struktur und biologische Aktivität von Wirkstoffen. Methoden der Quantitativen Struktur-Wirkung-Analyse*, Verlag Chemie, Weinheim, 1979

[43] Franke, R., *Optimierungsmethoden in der Wirkstoff-Forschung. Quantitative Struktur-Wirkungs-Analyse*, Akademie-Verlag, Berlin, 1980; revised English translation: Franke, R., *Theoretical Drug Design Methods*, *Pharmacochem. Libr.* **7**, Elsevier, Amsterdam, 1984

[44] Topliss, J. G., ed., *Quantitative Structure-Activity Relationships of Drugs*, Academic Press, New York, 1983

[45] Golberg, L., ed., *Structure-Activity Correlation as a Predictive Tool in Toxicology. Fundamentals, Methods, and Applications*, Hemisphere, Washington, 1983

[46] Jolles, G., and Woolridge, K. R. H., eds., *Drug Design: Fact or Fantasy?*, Academic Press, London, 1984

[47] Karcher, W., and Devilers, J., eds., *Practical Application of Quantitative Structure-Activity Relationships (QSAR) in Environmental Chemistry and Toxicology* (Chemical and Environmental Sciences, Vol. **1**), Kluwer Academic Publishers for the European Communities, Dordrecht, 1990

[48] Kier, L. B., and Hall, L. H., *Molecular Connectivity in Chemistry and Drug Research*, Academic Press, New York, 1976

[49] Rekker, R. F., *The Hydrophobic Fragmental Constant. Its Derivation and Application. A Means of Characterizing Membrane Systems*, *Pharmacochem. Libr.* **1**, Elsevier, Amsterdam, 1977

[50] Hansch, C., and Leo, A., *Substituent Constants for Correlation Analysis in Chemistry and Biology*, Wiley, New York, 1979

[51] Yalkowsky, S. H., Sinkula, A. A., and Valvani, S. C., eds., *Physical Chemical Properties of Drugs* (Medicinal Research Series, Vol. **10**), Marcel Dekker, New York, 1980

[52] Balaban, A. T., Chiriac, A., Motoc, I., and Simon, Z., *Steric Fit in Quantitative Structure-Activity Relations* (Lecture Notes in Chemistry **15**), Springer-Verlag, Berlin, 1980

[53] Perrin, D. D., Dempsey, B., and Serjeant, E. P., pK_a Prediction for Organic Acids and Bases, Chapman and Hall, London, 1981

[54] Dunn III, W. J., Block, J. H., and Pearlman, R. S., eds., *Partition Coefficient: Determination and Estimation*, Pergamon, New York, 1986

[55] McGowan, J. C., and Mellors, A., *Molecular Volumes in Chemistry and Biology: Applications Including Partitioning and Toxicity*, Ellis Horwood Ltd., Chichester, 1986

[56] Rekker, R. F., and Mannhold, R., *Calculation of Drug Lipophilicity. The Hydrophobic Fragmental Constant Approach*, VCH, Weinheim, 1992

[57] Chapman, N. B., and Shorter, J., eds., *Correlation Analysis in Chemistry: Recent Advances*, Plenum Press, New York, 1978

[58] Stuper, A. J., Brugger, W. E., and Jurs, P. C., *Computer Assisted Studies of Chemical Structure and Biological Function*, Wiley, New York, 1979

[59] Dean, P. M., *Molecular Foundations of Drug-Receptor Interaction*, Cambridge University Press, Cambridge, 1987

[60] Martin, Y. C., Kutter, E., and Austel, V., *Modern Drug Research. Paths to Better and Safer Drugs* (Medicinal Research Series, Vol. **12**), Marcel Dekker, New York, 1989

[61] Richards, W. G., ed., *Computer-Aided Molecular Design*, IBC Technical Services, London, 1989

[62] Perun, T. J., and Propst, C. L., *Computer-Aided Drug Design. Methods and Applications*, Marcel Dekker, New York, 1989

[63] Johnson, M. A., and Maggiora, G. M., eds., *Concepts and Applications of Molecular Similarity*, Wiley, New York, 1990

[64] Wermuth, C. G., Koga, N., König, H., and Metcalf, B. W., eds., *Medicinal Chemistry for the 21st Century*, Blackwell Scientific Publications, Oxford, 1992

[65] Ariëns, E. J., ed., *Drug Design*, Vol. **I** (1971) – Vol. **X** (1980), Academic Press, New York

[66] Jucker, E., ed., *Fortschr. Arzneimittelforsch. (Prog. Drug Res.)*, Vol. **1** (1959) ff., Birkhäuser-Verlag, Basel

[67] *Pharmacochem. Libr.* **1** (1977) ff.

[68] Gould, R. F., ed., *Biological Correlations – The Hansch Approach, Advan. Chem. Ser.* **114** (1972)

[69] Tichy, M., ed., *Quantitative Structure-Activity Relationships* (Proceedings of the Conference on Chemical Structure-Biological Activity Relationships: Quantitative Approaches, Prague, 1973), Experientia Suppl. **23**, Birkhäuser Verlag, Basel, 1976

[70] Keverling Buisman, J. A., ed., *Biological Activity and Chemical Structure* (Proceedings of the IUPAC-IUPHAR Symposium, Noordwijkerhout, 1977), *Pharmacochem. Libr.* **2**, Elsevier, Amsterdam, 1977

[71] Franke, R., and Oehme, P., eds., *Quantitative Structure-Activity Analysis* (Proceedings of the Second Symposium on Chemical Structure-Biological Activity Relationships: Quantitative Approaches, Suhl, 1976), Akademie-Verlag, Berlin, 1978

[72] Olsen, E. C., and Christofferson, P. E., eds., *Computer-Assisted Drug Design* (ACS Symp. Series **112**), Am. Chem. Soc., Washington, 1979

[73] Knoll, J., and Darvas, F., eds., *Chemical Structure-Biological Activity Relationships. Quantitative Approaches* (Proceedings of the 3rd European QSAR Symposium, Budapest, 1979), Akadémiai Kiadó, Budapest, 1980

[74] Keverling Buisman, J. A., ed., *Strategy in Drug Research* (Proceedings of the Second IUPAC-IUPHAR Symposium, Noordwijkerhout, 1981), *Pharmacochem. Libr.* **4**, Elsevier, Amsterdam, 1982

[75] Dearden, J. C., ed., *Quantitative Approaches to Drug Design* (Proceedings of the 4th European Symposium on Chemical Structure – Biological Activity: Quantitative Approaches, Bath, 1982), *Pharmacochem. Libr.* **6**, Elsevier, Amsterdam, 1983

[76] Kuchar, M., ed., *QSAR in Design of Bioactive Compounds* (Proceedings of the 1st Telesymposium on Medicinal Chemistry, 1984), Prous Science Publishers, Barcelona, 1984

[77] Tichy, M., ed., *QSAR in Toxicology and Xenobiochemistry* (Proceedings of a Symposium, Prague, 1984), *Pharmacochem. Libr.* **8**, Elsevier, Amsterdam, 1985

[78] Seydel, J. K., ed., *QSAR and Strategies in the Design of Bioactive Compounds* (Proceedings of the 5th European Symposium on QSAR, Bad Segeberg, 1984), VCH, Weinheim, 1985

[79] Harms, A. F., ed., *Innovative Approaches in Drug Research* (Proceedings of the Third Noordwijkerhout Symposium on Medicinal Chemistry, Noordwijkerhout, 1985), *Pharmacochem. Libr.* **9**, Elsevier, 1986

[80] Hadzi, D., and Jerman-Blazic, B., eds., *QSAR in Drug Design and Toxicology* (Proceedings of the 6th European Symposium on QSAR, Portoroz, 1986), *Pharmacochem. Libr.* **10**, Elsevier, Amsterdam, 1987

[81] Kaiser, K. L. E., ed., *QSAR in Environmental Toxicology*, Reidel, Dordrecht, 1987

[82] Fauchère, J. L., ed., *QSAR: Quantitative Structure-Activity Relationships in Drug Design* (Proceedings of the 7th European Symposium on QSAR, Interlaken, 1988), *Progr. Clin. Biol. Research* **291**, Alan R. Liss, New York, 1989

[83] Silipo, C., and Vittoria, A., eds., *QSAR: Rational Approaches to the Design of Bioactive Compounds* (Proceedings of the 8th European Symposium on QSAR, Sorrento, 1990), *Pharmacochem. Libr.* **16**, Elsevier, Amsterdam, 1991

[84] Wermuth, C. G., ed., *Trends in QSAR and Molecular Modelling 92* (Proceedings of the 9th European Symposium on Structure-Activity Relationships: QSAR and Molecular Modelling, Strasbourg, 1992), ESCOM Science Publishers, Leiden, 1993

[85] Darvas, F., ed., Abstracts of Publications Related to QSAR, *Quant. Struct.-Act. Relat.* **1** (1982) ff.

[86] CAS Selects "*Structure-Activity Relationships*", Chemical Abstracts Service (service code 04S), Columbus, Ohio, U. S. A.

[87] Ringdoc Drug Information, Profile **38**, "*Structure Activity*", Derwent Publications Ltd., London, England

[88] *Molecular Modelling & Computational Chemistry Results* **1** (1992) ff., MMCC Publishing, Cambridge, MA, U. S. A.

[89] a) *J. Med. Chem.* b) *Eur. J. Med. Chem.* c) *Med. Chem. Res.* d) *J. Comput.-Aided Mol. Des.*

[90] Craig, P. N., *Drug Inf. J.* **18**, 123–30 (1984)

[91] Tute, M. S., ref. 39, p. 1–31

[92] Rekker, R. F., *Quant. Struct.-Act. Relat.* **11**, 195–199 (1992)

[93] Van de Waterbeemd, H., *Quant. Struct.-Act. Relat.* **11**, 200–204 (1992)

[94] Ehrlich, P., *Lancet* **II**, 445–451 (1913)

[95] Langley, J. N., *J. Physiol. (London)* **1**, 339–369 (1878)

[96] Fischer, E., *Ber. Dtsch. Chem. Ges.* **27**, 2985–2993 (1894)

[97] Ariëns, E. J., *Drug Design*, Vol. **I**, Ariëns, E. J., ed., Academic Press, New York, 1971, p. 1–270

[98] Emmett, J. C., ed., *Membranes & Receptors*, Vol. **3** of *Comprehensive Medicinal Chemistry. The Rational Design, Mechanistic Study & Therapeutic Application of Chemical Compounds*, Hansch, C., Sammes, P. G., and Taylor, J. B., eds., Pergamon Press, Oxford, 1990

[99] Dickerson, R. E., and Geis, I., *Struktur und Funktion der Proteine*, Verlag Chemie, Weinheim, 1971

[100] De Stevens, G., *Fortschr. Arzneimittelforsch. (Prog. Drug. Res.)* **30**, 189–203 (1986)

[101] Williams, R. J. P., *Angew. Chem. Int. Ed. Engl.* **16**, 766–777 (1977)

[102] Roberts, G. C. K., ref. 75, p. 91–98

[103] Deisenhofer, J., and Michel, H., *Science* **245**, 1463–1480 (1989)

[104] Henderson, R., Baldwin, J. M., Ceska, T. A., Zemlin, F., Beckmann, E., and Downing, K. H., *J. Mol. Biol.* **213**, 899–929 (1990)

[105] Weiss, M. S., Abele, U., Weckesser, J., Welte, W., Schiltz, E., and Schulz, G. E., *Science* **254**, 1627–1630 (1991)

[106] Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer Jr., E. F., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T., and Tasumi, M., *J. Mol. Biol.* **112**, 535–542 (1977)

[107] Blundell, T. L., and Johnson, L. N., *Protein Crystallography*, Academic Press, London, 1976

[108] Branden, C., and Tooze, J., *Introduction to Protein Structure*, Garland Publishing Company, New York, 1991

[109] Bode, W., and Huber, R., *Eur. J. Biochem.* **204**, 433–451 (1992)

[110] Bode, W., and Huber, R., ref. 64, p. 73–118

[111] Findlay, J., and Eliopoulos, E., *Trends Pharmacol. Sci.* **11**, 492–499 (1990)

[112] a) Trumpp-Kallmeyer, S., Hoflack, J., Bruinvels, A., and Hibert, M., *J. Med. Chem.* **35**, 3448–3462 (1992) ; b) Hibert, M. F., Trumpp-Kallmeyer, S., Hoflack, J., and Bruinvels, A., *Trends Pharmacol. Sci.* **14**, 7–12 (1993)

[113] Humblet, C., and Mirzadegan, T., *Ann. Rep. Med. Chem.* **27**, 291–300 (1992)

[114] Herbette, L. G., Chester, D. W., and Rhodes, D. G., *Biophys. J.* **49**, 91–94 (1986)

[115] Mason, R. P., Rhodes, D. G., and Herbette, L. G., *J. Med. Chem.* **34**, 869–877 (1991)

[116] Andrews, P. R., and Tintelnot, M., ref. 39, p. 321–347

[117] Tintelnot, M., and Andrews, P. R., *J. Comput.-Aided Mol. Des.* **3**, 67–84 (1989)

[118] Williams, D. H., *Aldrichimica Acta* **24**, 71–80 (1991)

[119] Williams, D. H., *Aldrichimica Acta* **25**, 9 (1992)

[120] Murray-Rust, P., and Glusker, J. P., *J. Am. Chem. Soc.* **106**, 1018–1025 (1984)

[121] Fersht, A. R., Shi, J.-P., Knill-Jones, J., Lowe, D. M., Wilkinson, A. J., Blow, D. M., Brick, P., Carter, P., Waye, M. M. Y., and Winter, G., *Nature* **314**, 235–238 (1985)

[122] Fersht, A. R., *Trends Biochem. Sci.* **12**, 301–304 (1987)

[123] Ward, W. H. J., Timms, D., and Fersht, A. R., *Trends Pharmacol. Sci.* **11**, 280–284 (1990)

[124] Wolfenden, R., and Kati, W. M., *Acc. Chem. Res.* **24**, 209–215 (1991)

[125] a) Doig, A. J., and Williams, D. H., *J. Am. Chem. Soc.* **114**, 338–343 (1992) ; b) Searle, M. S., and Williams, D. H., ref. 84, p. 86–92

[126] Sharp, K. A., Nicholls, A., Friedman, R., and Honig, B., *Biochemistry* **30**, 9686–9697 (1991)

[127] Hitzemann, R., *Trends Pharmacol. Sci.* **9**, 408–411 (1988)

[128] Andrews, P. R., Craik, D. J., and Martin, J. L., *J. Med. Chem.* **27**, 1648–1657 (1984)

[129] Andrews, P., *Trends Pharmacol. Sci.* **7**, 148–151 (1986)

[130] Goodford, P. J., ref. 83, p. 49–55

[131] Williams, M., and Sills, M. A., ref. 98, p. 45–80

[132] Martin, Y. C., Bush, E. N., and Kyncl, J. J., ref. 39, p. 349–373

[133] Hansch, C., *Intra-Science Chem. Rept.* **8**, 17–25 (1974)

[134] Thornber, C. W., *Chem. Soc. Rev.* **8**, 563–580 (1979)

[135] Burger, A., *Fortschr. Arzneimittelforsch. (Prog. Drug. Res.)* **37**, 287–371 (1991)

[136] Bartlett, P. A., and Marlowe, C. K., *Science* **235**, 569–571 (1987)

[137] Morgan, B. P., Scholtz, J. M., Ballinger, M. D., Zipkin, I. D., and Bartlett, P. A., *J. Am. Chem. Soc.* **113**, 297–307 (1991)

[138] Tronrud, D. E., Holden, H. M., and Matthews, B. W., *Science* **235**, 571–574 (1987)

[139] Bash, P. A., Singh, U. C., Brown, F. K., Langridge, R., and Kollman, P. A., *Science* **235**, 574–576 (1987)

[140] Klebe, G., *Struct. Chem.* **1**, 597–616 (1990)

[141] Allen, F. H., Davies, J. E., Galloy, J. J., Johnson, O., Kennard, O., Macrae, C. F., Mitchell, E. M., Mitchell, G. F., Smith, J. M., and Watson, D. G., *J. Chem. Inf. Comput. Sci.* **31**, 187–204 (1991)

[142] Page, M. I., *Angew. Chem. Int. Ed. Engl.* **16**, 449–459 (1977)

[143] Page, M. I., ref. 75, p. 109–119

[144] Jencks, W. P., *Proc. Natl. Acad. Sci. USA* **78**, 4046–4050 (1981)

[145] Wells, J. A., *Biochemistry* **29**, 8509–8517 (1990)

[146] Fothergill, M. D., and Fersht, A. R., *Biochemistry* **30**, 5157–5164 (1991)

[147] Fersht, A. R., *Biochemistry* **26**, 8031–8037 (1987)

[148] Fersht, A. R., Knill-Jones, J. W., Bedouelle, H., and Winter, G., *Biochemistry* **27**, 1581–1587 (1988)

[149] Beddell, C. R., Goodford, P. J., Stammers, D. K., and Wootton, R., *Br. J. Pharmac.* **65**, 535–543 (1979)

[150] Bohacek, R. S., and McMartin, C., *J. Med. Chem.* **35**, 1671–1684 (1992)

[151] Hansch, C., Leo, A., Unger, S. H., Kim, K. H., Nikaitani, D., and Lien, E. J., *J. Med. Chem.* **16**, 1207–1216 (1973)

[152] Hansch, C., Rockwell, S. D., Jow, P. Y. C., Leo, A., and Steller, E. E., *J. Med. Chem.* **20**, 304–306 (1977)

[153] Norrington, F. E., Hyde, R. M., Williams, S. G., and Wootton, R., *J. Med. Chem.* **18**, 604–607 (1975)

[154] Hansch, C., Unger, S. H., and Forsythe, A. B., *J. Med. Chem.* **16**, 1217–1222 (1973)

[155] Verloop, A., *Drug Design*, Vol. **III**, Ariëns, E. J., ed., Academic Press, New York, 1972, p. 133–187

[156] Martin, Y. C., *Drug Design*, Vol. **VIII**, Ariëns, E. J., ed., Academic Press, New York, 1979, p. 1–72

[157] Franke, R., ref. 78, p. 59–78

[158] Van de Waterbeemd, H., and Testa, B., *Adv. Drug. Res.* **16**, 85–225 (1987)

[159] Van de Waterbeemd, H., Testa, B., Carrupt, P.-A., and El Tayar, N., ref. 82, p. 123–126

[160] Van de Waterbeemd, H., El Tayar, N., Carrupt, P.-A., and Testa, B., *J. Comput.-Aided Mol. Des.* **3**, 111–132 (1989)

[161] Van de Waterbeemd, H., Carrupt, P.-A., and El Tayar, N., ref. 82, p. 101–103

[162] Boyd, D. B., and Seward, C. M., ref. 83, p. 167–170

[163] Streich, W. J., Dove, S., and Franke, R., *J. Med. Chem.* **23**, 1452–1456 (1980)

[164] Cramer III, R. D., *J. Am. Chem. Soc.* **102**, 1837–1849, 1849–1859 (1980)

[165] Cramer III, R. D., *Quant. Struct.-Act. Relat.* **2**, 7–12, 13–19 (1983)

[166] Cramer III, R. D., ref. 75, p. 3–14

[167] Hyde, R., ref. 82, p. 91–95

[168] Alunni, S., Clementi, S., Edlund, U., Johnels, D., Hellberg, S., Sjöström, M., and Wold, S., *Acta Chem. Scand.* **B 37**, 47–53 (1983)

[169] Skagerberg, B., Bonelli, D., Clementi, S., Cruciani, G., and Ebert, C., *Quant. Struct.-Act. Relat.* **8**, 32–38 (1989)

[170] Sjöström, M., and Wold, S., *J. Mol. Evol.* **22**, 272–277 (1985)

[171] Jonsson, J., Eriksson, L., Hellberg, S., Sjöström, M., and Wold, S., *Quant. Struct.-Act. Relat.* **8**, 204–209 (1989)

[172] Dill, K. A., *Science* **250**, 297 (1990)

[173] Taylor, P. J., ref. 39, p. 241–294

[174] Leo, A., Hansch, C., and Elkins, D., *Chem. Rev.* **71**, 525–616 (1971)

[175] Kubinyi, H., *Fortschr. Arzneimittelforsch. (Prog. Drug Res.)* **23**, 97–198 (1979)

[176] Coats, E. A., ref. 51, p. 111–139

[177] Chou, J. T., and Jurs, P. C., ref. 51, p. 163–199

[178] Dearden, J. C., *Environ. Health Perspect.* **61**, 203–228 (1985)

[179] Leo, A. J., ref. 39, p. 295–319

[180] Smith, R. N., Hansch, C., and Ames, M. M., *J. Pharm. Sci.* **64**, 599–606 (1975)

[181] Hansch, C., and Leo, A., Partition Coefficient Data Bank, MEDCHEM Project, Pomona College, Claremont, CA, USA

[182] Hansch, C., and Dunn III, W. J., *J. Pharm. Sci.* **61**, 1–19 (1972)

[183] Collander, R., *Acta Chem. Scand.* **5**, 774–780 (1951)

[184] Yeh, K. C., and Higuchi, W. I., *J. Pharm. Sci.* **65**, 80–86 (1976)

[185] Leo, A., and Hansch, C., *J. Org. Chem.* **36**, 1539–1544 (1971)

[186] Leo, A., ref. 68, p. 51–60

[187] Seeman, P., Roth, S., and Schneider, H., *Biochim. Biophys. Acta* **225**, 171–184 (1971)

[188] Dunn III, W. J., and Wold, S., *Acta Chem. Scand.* **B 32**, 536–542 (1978)

[189] Franke, R., Kühne, R., and Dove, S., ref. 75, p. 15–32

[190] Koehler, M. G., Grigoras, S., and Dunn III, W. J., *Quant. Struct.-Act. Relat.* **7**, 150–159 (1988)

[191] Seiler, P., *Eur. J. Med. Chem.* **9**, 473–479 (1974)

[192] El Tayar, N., Tsai, R.-S., Testa, B., Carrupt, P.-A., and Leo, A., *J. Pharm. Sci.* **80**, 590–598 (1991)

[193] Fan, W., El Tayar, N., Testa, B., and Kier, L. B., *J. Phys. Chem.* **94**, 4764–4766 (1990)

[194] Fan, W., El Tayar, N., Testa, B., Carrupt, P.-A., and Kier, L. B., ref. 83, p. 67–70

[195] Wolfenden, R., *Science* **222**, 1087–1093 (1983)

[196] Fujita, T., Nishioka, T., and Nakajima, M., *J. Med. Chem.* **20**, 1071–1081 (1977)

[197] Hadzi, D., Kidric, J., Koller, J., and Mavri, J., *J. Mol. Struct. (Theochem)* **237**, 139–150 (1990)

[198] Raevsky, O. A., Grigor'ev, V. U., Solov'ev, V. P., Kireev, D. B., Sapegin, A. M., and Zefirov, N. S., ref. 83, p. 135–138

[199] Young, R. C., Mitchell, R. C., Brown, T. H., Ganellin, C. R., Griffiths, R., Jones, M., Rana, K. K., Saunders, D., Smith, I. R., Sore, N. E., and Wilks, T. J., *J. Med. Chem.* **31**, 656–671 (1988)

[200] Ganellin, C. R., Brown, T. H., Griffiths, R., Jones, M., Mitchell, R. C., Rana, K., Saunders, D., Smith, I. R., Sore, N. E., Wilks, T. J., and Young, R. C., ref. 83, p. 103–110

[201] Leeson, P. D., Carling, R. W., James, K., Smith, J. D., Moore, K. W., Wong, E. H. F., and Baker, R., *J. Med. Chem.* **33**, 1296–1305 (1990)

[202] Abraham, M. H., Whiting, G. S., Alarie, Y., Morris, J. J., Taylor, P. J., Doherty, R. M., Taft, R. W., and Nielsen, G. D., *Quant. Struct.-Act. Relat.* **9**, 6–10 (1990)

[203] Altomare, C., Tsai, R.-S., El Tayar, N., Testa, B., Carrupt, P.-A., Carotti, A., and De Benedetti, P. G., ref. 83, p. 139–142

[204] Leahy, D. E., Taylor, P. J., and Wait, A. R., *Quant. Struct.-Act. Relat.* **8**, 17–31 (1989)

[205] Leahy, D. E., Morris, J. J., Taylor, P. J., and Wait, A. R., *J. Chem. Soc. Perkin Trans.* **2**, 705–722, 723–731 (1992)

[206] Rhodes, D. G., Sarmiento, J. G., and Herbette, L. G., *Mol. Pharmacol.* **27**, 612–623 (1985)

[207] Herbette, L. G., and Gruner, S. M., *Dev. Cardiovasc.* **68**, 353–365 (1987)

[208] Herbette, L. G., Vant Erve, Y. M. H., and Rhodes, D. G., *J. Mol. Cell. Cardiol.* **21**, 187–201 (1989)

[209] Mason, R. P., Gonye, G. E., Chester, D. W., and Herbette, L. G., *Biophys. J.* **55**, 769–778 (1989)

[210] Moring, J., and Herbette, L. G., *Biophys. J.* **57**, 523a (1990)

[211] Herbette, L. G., Rhodes, D. G., and Mason, R. P., *Drug Des. Delivery* **7**, 75–118 (1991)

[212] Seydel, J. K., Cordes, H.-P., Wiese, M., Chi, H., Croes, N., Hanpft, R., Lüllmann, H., Mohr, K., Patten, M., Padberg, Y., Lüllmann-Rauch, R., Vellguth, S., Meindl, W. R., and Schönenberger, H., *Quant. Struct.-Act. Relat.* **8**, 266–278 (1989)

[213] Seydel, J. K., Wiese, M., Cordes, H. P., Chi, H. L., Schaper, K.-J., Coats, E. A., Kunz, B., Engel, J., Kutscher, B., and Emig, H., ref. 83, p. 367–376

[214] Seydel, J. K., Albores Velasco, M., Coats, E. A., Cordes, H. P., Kunz, B., and Wiese, M., *Quant. Struct.-Act. Relat.* **11**, 205–210 (1992)

[215] Schwyzer, R., *J. Receptor Res.* **11**, 45–57 (1991)

[216] Schwyzer, R., *Biopolymers* **31**, 785–792 (1991)

[217] Dearden, J. C., and Bresnen, G. M., *Quant. Struct.-Act. Relat.* **7**, 133–144 (1988)

[218] Leo, A. J., *Methods Enzymol.* **202**, 544–591 (1991)

[219] Tomlinson, E., Davis, S. S., Parr, G. D., James, M., Farraj, N., Kinkel, J. F. M., Gaisser, D., and Wynne, H. J., ref. 54, p. 83–99

[220] Hersey, A., Hill, A. P., Hyde, R. M., and Livingstone, D. J., *Quant. Struct.-Act. Relat.* **8**, 288–296 (1989)

[221] Davis, S. S., Elson, G., Tomlinson, E., Harrison, G., and Dearden, J. C., *Chem. Ind. (London)*, 677–683 (1976)

[222] Menges, R. A., Bertrand, G. L., and Armstrong, D. W., *J. Liquid Chromatogr.* **13**, 3061–3077 (1990)

[223] Vallat, P., El Tayar, N., Testa, B., Slacanin, I., Marston, A., and Hostettmann, K., *J. Chromatogr.* **504**, 411–419 (1990)

[224] Tsai, R.-S., El Tayar, N., Testa, B., and Ito, Y., *J. Chromatogr.* **538**, 119–123 (1991)

[225] El Tayar, N., Tsai, R.-S., Vallat, P., Altomare, C., and Testa, B., *J. Chromatogr.* **556**, 181–194 (1991)

[226] Boyce, C. B. C., and Milborrow, B. V., *Nature* **208**, 537–539 (1965)

[227] Tomlinson, E., *J. Chromatogr.* **113**, 1–45 (1975)

[228] Biagi, G. L., Recanatini, M., Barbaro, A. M., Guerra, M. C., Sapone, A., Borea, P. A., and Pietrogrande, M. C., ref. 83, p. 83–90

[229] Dross, K. P., Mannhold, R., and Rekker, R. F., *Quant. Struct.-Act. Relat.* **11**, 36–44 (1992)

[230] Mirrlees, M. S., Moulton, S. J., Murphy, C. T., and Taylor, P. J., *J. Med. Chem.* **19**, 615–619 (1976)

[231] Unger, S. H., Cook, J. R., and Hollenberg, J. S., *J. Pharm. Sci.* **67**, 1364–1367 (1978)

[232] Unger, S. H., and Chiang, G. H., *J. Med. Chem.* **24**, 262–270 (1981)

[233] Lewis, S. J., Mirrlees, M. S., and Taylor, P. J., *Quant. Struct.-Act. Relat.* **2**, 1–6, 100–111 (1983)

[234] Haky, J. E., and Young, A. M., *J. Liquid Chromatogr.* **7**, 675–689 (1984)

[235] Smith, R. M., and Burr, C. M., *J. Chromatogr.* **475**, 75–83 (1989)

[236] Kaliszan, R., *Quant. Struct.-Act. Relat.* **9**, 83–87 (1990)

[237] Fujita, T., Iwasa, J., and Hansch, C., *J. Am. Chem. Soc.* **86**, 5175–5180 (1964)

[238] Hansch, C., Leo, A., and Nikaitani, D., *J. Org. Chem.* **37**, 3090–3092 (1972)

[239] Nys, G. G., and Rekker, R. F., *Chim. Ther.* **8**, 521–535 (1973)

[240] Nys, G. G., and Rekker, R. F., *Eur. J. Med. Chem.* **9**, 361–375 (1974)

[241] Rekker, R. F., and De Kort, H. M., *Eur. J. Med. Chem.* **14**, 479–488 (1979)

[242] Davis, S. S., Higuchi, T., and Rytting, J. H., *Adv. Pharm. Sci.* **4**, 73–261 (1974)

[243] Taylor, P. J., ref. 83, p. 123–126

[244] Leo, A., Jow, P. Y. C., Silipo, C., and Hansch, C., *J. Med. Chem.* **18**, 865–868 (1975)

[245] Leo, A. J., *Eur. J. Med. Chem.* **15**, 484 (1980)

[246] Pleiss, M. A., and Grunewald, G. L., *J. Med. Chem.* **26**, 1760–1764 (1983)

[247] Van de Waterbeemd, H., and Testa, B., *Int. J. Pharm.* **14**, 29–41 (1983)

[248] Chou, J. T., and Jurs, P. C., *J. Chem. Inf. Comput. Sci.* **19**, 172–178 (1979)

[249] Leo, A. J., ref. 78, p. 294–298

[250] Leo, A. J., *J. Pharm. Sci.* **76**, 166–168 (1987)

[251] Leo, A. J., ref. 82, p. 53–58

[252] Leo, A. J., ref. 83, p. 349–352

[253] MEDCHEM Software, Releases 3.54 (VMS version), 1989, and 4.2 (UNIX version), 1992, Daylight Chemical Information Systems, Irvine, CA, USA

[254] Weininger, D., *J. Chem. Inf. Comput. Sci.* **28**, 31–36 (1988)

[255] Weininger, D., Weininger, A., and Weininger, J. L., *J. Chem. Inf. Comput. Sci.* **29**, 97–101 (1989)

[256] Weininger, D., and Weininger, J. L., ref. 39, p. 59–82

[257] Mayer, J. M., van de Waterbeemd, H., and Testa, B., *Eur. J. Med. Chem.* **17**, 17–25 (1982)

[258] a) Mannhold, R., Dross, K. P., and Rekker, R. F., *Quant. Struct.-Act. Relat.* **9**, 21–28 (1990); b) Rekker, R. F., ter Laak, A. M., and Mannhold, M., *Quant. Struct.-Act. Relat.* **12**, 152–157 (1993)

[259] Van de Waterbeemd, H., *Hydrophobicity of Organic Compounds. How to Calculate it by PC's*, Booksoft, Vol. **1**, Darvas, F., ed., CompuDrug, Vienna, 1986

[260] Takeuchi, K., Kuroda, C., and Ishida, M., *J. Chem. Inf. Comput. Sci.* **30**, 22–26 (1990)

[261] Darvas, F., Erdos, I., and Teglas, G., ref. 80, p. 70–73

[262] Bradshaw, J., and Taylor, P. J., *Quant. Struct.-Act. Relat.* **8**, 279–287 (1989)

[263] Carrupt, P.-A., Testa, B., Bechalany, A., El Tayar, N., Descas, P., and Perrissoud, D., ref. 83, p. 541–544

[264] Hine, J., and Mookerjee, P. D., *J. Org. Chem.* **40**, 292–298 (1975)

[265] Broto, P., Moreau, G., and Vandycke, C., *Eur. J. Med. Chem.* **19**, 71–78 (1984)

[266] Ghose, A. K., and Crippen, G. M., *J. Comput. Chem.* **7**, 565–577 (1986)

[267] Ghose, A. K., and Crippen, G. M., *J. Chem. Inf. Comput. Sci.* **27**, 21–35 (1987)

[268] Ghose, A. K., Pritchett, A., and Crippen, G. M., *J. Comput. Chem.* **9**, 80–90 (1988)

[269] Viswanadhan, V. N., Ghose, A. K., Revankar, G. R., and Robins, R. K., *J. Chem. Inf. Comput. Sci.* **29**, 163–172 (1989)

[270] Suzuki, T., and Kudo, Y., *J. Comput.-Aided Mol. Des.* **4**, 155–198 (1990)

[271] Suzuki, T., *J. Comput.-Aided Mol. Des.* **5**, 149–166 (1991)

[272] Leo, A., Hansch, C., and Jow, P. Y. C., *J. Med. Chem.* **19**, 611–615 (1976)

[273] Bultsma, T., *Eur. J. Med. Chem.* **15**, 371–374 (1980)

[274] Valvani, S. C., and Yalkowsky, S. H., ref. 51, p. 201–229

[275] Pearlman, R. S., ref. 51, p. 321–347

[276] Iwase, K., Komatsu, K., Hirono, S., Nakagawa, S., and Moriguchi, I., *Chem. Pharm. Bull.* **33**, 2114–2121 (1985)

[277] Dunn III, W. J., Koehler, M. G., and Grigoras, S., *J. Med. Chem.* **30**, 1121–1126 (1987)

[278] Kasai, K., Umeyama, H., and Tomonaga, A., *Bull. Chem. Soc. Jpn.* **61**, 2701–2706 (1988)

[279] Dunn III, W. J., ref. 82, p. 47–51

[280] Bodor, N., Gabanyi, Z., and Wong, C.-K., *J. Am. Chem. Soc.* **111**, 3783–3786 (1989)

[281] De Bruijn, J., and Hermens, J., *Quant. Struct.-Act. Relat.* **9**, 11–21 (1990)

[282] Dunn III, W. J., Nagy, P. I., Collantes, E. R., Glen, W. G., Alagona, G., and Ghio, C., ref. 83, p. 59–65

[283] Sasaki, Y., Kubodera, H., Matuszaki, T., and Umeyama, H., *J. Pharmacobio-Dyn.* **14**, 207–214 (1991)

[284] Bodor, N., and Huang, M.-J., *J. Pharm. Sci.* **81**, 272–281 (1992)

[285] Moriguchi, I., *Chem. Pharm. Bull.* **23**, 247–257 (1975)

[286] Dunn III, W. J., *Eur. J. Med. Chem.* **12**, 109–112 (1977)

[287] Silipo, C., and Vittoria, A., ref. 39, p. 153–204

[288] Blaney, J. M., and Hansch, C., ref. 39, p. 459–496

[289] Gryllaki, M., van de Waterbeemd, H., and Testa, B., ref. 78, p. 273–276

[290] Ahmad, P., Fyfe, C. A., and Mellors, A., *Biochem. Pharmacol.* **24**, 1103–1109 (1975)

[291] Dearden, J. C., Bradburne, S. J. A., and Abraham, M. H., ref. 83, p. 143–150

[292] Hansch, C., and Blaney, J. M., ref. 46, p. 185–208

[293] Shah, K. J., and Coats, E. A., *J. Med. Chem.* **20**, 1001–1006 (1977)

[294] Chapman, N. B., and Shorter, J., eds., *Advances in Linear Free Energy Relationships*, Plenum Press, London, 1972

[295] Perrin, D. D., ref. 51, p. 1–48
[296] Bowden, K., ref. 39, p. 205–239
[297] Loew, G. H., and Burt, S. K., ref. 39, p. 105–123
[298] Swain, C. G., and Lupton Jr., E. C., *J. Am. Chem. Soc.* **90**, 4328–4337 (1968)
[299] Swain, C. G., Unger, S. H., Rosenquist, N. R., and Swain, M. S., *J. Am. Chem. Soc.* **105**, 492–502 (1983)
[300] Swain, C. G., *J. Org. Chem.* **49**, 2005–2010 (1984)
[301] Exner, O., ref. 294, p. 1–69
[302] Shorter, J., ref. 57, p. 119–173
[303] Reynolds, W. F., and Topsom, R. D., *J. Org. Chem.* **49**, 1989–1992 (1984)
[304] Hoefnagel, A. J., Oosterbeek, W., and Wepster, B. M., *J. Org. Chem.* **49**, 1993–1997 (1984)
[305] Charton, M., *J. Org. Chem.* **49**, 1997–2001 (1984)
[306] Hansch, C., Leo, A., and Taft, R. W., *Chem. Rev.* **91**, 165–195 (1991)
[307] Unger, S. H., and Hansch, C., *J. Med. Chem.* **16**, 745–749 (1973)
[308] Taft Jr., R. W., *Steric Effects in Organic Chemistry*, Newman, M. S., ed., Wiley, New York, 1956, p. 556–675
[309] Silipo, C., and Vittoria, A., *Farmaco, Ed. Sci.* **34**, 858–868 (1979)
[310] Seydel, J. K., *Drug Design*, Vol. **I**, Ariëns, E. J., ed., Academic Press, New York, 1971, p. 343–379
[311] Fauchère, J.-L., and Lauterwein, J., *Quant. Struct.-Act. Relat.* **4**, 11–13 (1985)
[312] De Benedetti, P. G., *Adv. Drug. Res.* **16**, 227–279 (1987)
[313] Hetnarski, B., and O'Brien, R. D., *J. Med. Chem.* **18**, 29–33 (1975)
[314] Foster, R., Hyde, R. M., and Livingstone, D. J., *J. Pharm. Sci.* **67**, 1310–1313 (1978)
[315] Livingstone, D. J., Hyde, R. M., and Foster, R., *Eur. J. Med. Chem.* **14**, 393–397 (1979)
[316] Lien, E. J., Liao, R. C. H., and Shinouda, H. G., *J. Pharm. Sci.* **68**, 463–465 (1979)
[317] Lien, E. J., Guo, Z.-R., Li, R.-L., and Su, C.-T., *J. Pharm. Sci.* **71**, 641–655 (1982)
[318] Li, W.-Y., Guo, Z.-R., and Lien, E. J., *J. Pharm. Sci.* **73**, 553–558 (1984)
[319] Hansch, C., Vittoria, A., Silipo, C., and Jow, P. Y. C., *J. Med. Chem.* **18**, 546–548 (1975)
[320] Taft, R. W., Gurka, D., Joris, L., von R. Schleyer, P., and Rakshys, J. W., *J. Am. Chem. Soc.* **91**, 4801–4808 (1969)
[321] Mitsky, J., Joris, L., and Taft, R. W., *J. Am. Chem. Soc.* **94**, 3442–3445 (1972)
[322] Gerhards, J., and Mehler, E. L., ref. 78, p. 153–161
[323] Dearden, J. C., and Nicholson, R. M., *Pestic. Sci.* **17**, 305–310 (1986)
[324] Dearden, J. C., and Nicholson, R. M., ref. 80, p. 307–312
[325] Shusterman, A. J., Debnath, A. K., Hansch, C., Horn, G. W., Fronczek, F. R., Greene, A. C., and Watkins, S. F., *Mol. Pharmacol.* **36**, 939–944 (1989)
[326] Lopez de Compadre, R. L., Debnath, A. K., Shustermann, A. J., and Hansch, C., *Environ. Mol. Mutagen.* **15**, 44–55 (1990)
[327] Hansch, C., ref. 83, p. 3–10
[328] Debnath, A. K., Lopez de Compadre, R. L., Debnath, G., Shusterman, A. J., and Hansch, C., *J. Med. Chem.* **34**, 786–797 (1991)
[329] a) Ebert, C., Katritzky, A. R., and Musumarra, G., *Quant. Struct.-Act. Relat.* **10**, 101–106, 107–109 (1991); b) Caruso, L., Musumarra, G., and Katritzky, A. R., *Quant. Struct.-Act. Relat.* **12**, 146–151 (1993)
[330] Gasteiger, J., and Marsili, M., *Tetrahedron* **36**, 3219–3288 (1980)
[331] Gasteiger, J., Hutchings, M. G., Marsili, M., and Saller, H., ref. 78, p. 90–97
[332] Gasteiger, J., Hutchings, M. G., Christoph, B., Gann, L., Hiller, C., Löw, P., Marsili, M., Saller, H., and Yuki, K., *Topics Curr. Chem.* **137**, 19–73 (1987)
[333] Unger, S. H., and Hansch, C., *Prog. Phys. Org. Chem.* **12**, 91–118 (1976)
[334] Fujita, T., *Pure & Appl. Chem.* **50**, 987–994 (1978)
[335] Verloop, A., Hoogenstraaten, W., and Tipker, J., *Drug Design*, Vol. **VII**, Ariëns, E. J., ed., Academic Press, New York, 1976, p. 165–207
[336] Verloop, A., *The STERIMOL Approach to Drug Design*, Marcel Dekker, New York, 1987
[337] Charton, M., *J. Am. Chem. Soc.* **97**, 1552–1556 (1975)
[338] Charton, M., *J. Org. Chem.* **41**, 2217–2220 (1976)
[339] Charton, M., *Topics Curr. Chem.* **114**, 57–91 (1983)

[340] Hancock, C. K., Meyers, E. A., and Yager, B. J., *J. Am. Chem. Soc.* **83**, 4211–4213 (1961)

[341] Fujita, T., and Iwamura, H., *Topics Curr. Chem.* **114**, 119–157 (1983)

[342] Austel, V., Kutter, E., and Kalbfleisch, W., *Arzneim.-Forsch. (Drug Res.)* **29**, 585–587 (1979)

[343] Verloop, A., ref. 78, p. 98–104

[344] Lien, E. J., and Wang, P. H., *J. Pharm. Sci.* **69**, 648–650 (1980)

[345] Kubinyi, H., ref. 76, p. 321–346

[346] Selassie, C. D., Hansch, C., and Khwaja, T. A., *J. Med. Chem.* **33**, 1914–1919 (1990)

[347] Kitahara, S., Heinz, E., and Stahlmann, C., *Nature* **208**, 187–189 (1965)

[348] Selassie, C. D., personal communication

[349] Seidel, W., Meyer, H., Born, L., Kazda, S., and Dompert, W., ref. 78, p. 366–369

[350] Avbelj, F., and Hadzi, D., *Mol. Pharmacol.* **27**, 466–470 (1985)

[351] Lopez de Compadre, R. L., Pearlstein, R. A., Hopfinger, A. J., and Seydel, J. K., *J. Med. Chem.* **30**, 900–906 (1987); erratum **31**, 2315 (1988)

[352] Hopfinger, A. J., Lopez de Compadre, R. L., Koehler, M. G., Emery, S., and Seydel, J. K., *Quant. Struct.-Act. Relat.* **6**, 111–117 (1987)

[353] Liljefors, T., Bengtsson, M., Bogeso, K. P., Hansson, B. S., and Pettersson, I., *Frontiers in Drug Research* (Alfred Benzon Symposium **28**), Jensen, B., Jorgensen, F. S.,, and Kofod, H., eds., Munksgaard, Copenhagen, 1990, p. 139–149

[354] Randic, M., *J. Am. Chem. Soc.* **97**, 6609–6615 (1975)

[355] Randic, M., *Int. J. Quantum Chem., Quantum Biol. Symp.* **5**, 245–255 (1978)

[356] Kier, L. B., Hall, L. H., Murray, W. J., and Randic, M., *J. Pharm. Sci.* **64**, 1971–1974 (1975)

[357] Hall, L. H., Kier, L. B., and Murray, W. J., *J. Pharm. Sci.* **64**, 1974–1977 (1975)

[358] Murray, W. J., Hall, L. H., and Kier, L. B., *J. Pharm. Sci.* **64**, 1978–1981 (1975)

[359] Kier, L. B., Murray, W. J., and Hall, L. H., *J. Med. Chem.* **18**, 1272–1274 (1975)

[360] Murray, W. J., Kier, L. B., and Hall, L. H., *J. Med. Chem.* **19**, 573–578 (1976)

[361] Kier, L. B., Murray, W. J., Randic, M., and Hall, L. H., *J. Pharm. Sci.* **65**, 1226–1230 (1976)

[362] Kier, L. B., and Hall, L. H., *J. Pharm. Sci.* **65**, 1806–1809 (1976)

[363] Kier, L. B., and Hall, L. H., *Eur. J. Med. Chem.* **12**, 307–312 (1977)

[364] Hall, L. H., and Kier, L. B., *J. Pharm. Sci.* **66**, 642–644 (1977)

[365] Murray, W. J., *J. Pharm. Sci.* **66**, 1352–1354 (1977)

[366] Kier, L. B., and Hall, L. H., *J. Pharm. Sci.* **67**, 1408–1412 (1978)

[367] Kier, L. B., and Hall, L. H., *J. Pharm. Sci.* 68, 120–122 (1979)

[368] Kier, L. B., ref. 51, p. 277–319

[369] Kier, L. B., and Hall, L. H., *Molecular Connectivity in Structure-Activity Analysis*, Research Studies Press, Letchworth, 1986

[370] Osmialowski, K., and Kaliszan, R., *Quant. Struct.-Act. Relat.* **10**, 125–134 (1991)

[371] Hall, L. H., and Kier, L. B., *Reviews in Computational Chemistry* **II**, Lipkowitz, K. B., and Boyd, D. B., eds., VCH Publishers, New York, 1991, p. 367–422

[372] Kier, L. B., and Hall, L. H., *J. Pharm. Sci.* **70**, 583–589 (1981)

[373] Kier, L. B., and Hall, L. H., *J. Pharm. Sci.* **72**, 1170–1173 (1983)

[374] Kier, L. B., ref. 80, p. 13–17

[375] Kier, L. B., *Quant. Struct.-Act. Relat.* **6**, 8–12, 117–122 (1987)

[376] Mokrosz, J. L., *Quant. Struct.-Act. Relat.* **8**, 305–309 (1989)

[377] Randic, M., and Jurs, P. C., *Quant. Struct.-Act. Relat.* **8**, 39–48 (1989)

[378] Kier, L. B., ref. 82, p. 105–109

[379] Kier, L. B., *Quant. Struct.-Act. Relat.* **8**, 221–225 (1989)

[380] Hall, L. H., and Aaserud, D., *Quant. Struct.-Act. Relat.* **8**, 296–304 (1989)

[381] Hall, L. H., and Kier, L. B., *Quant. Struct.-Act. Relat.* **9**, 115–131 (1990)

[382] Wiener, H., *J. Am. Chem. Soc.* **69**, 17–20, 2636–2638 (1947)

[383] Lukovits, I., *Quant. Struct.-Act. Relat.* **9**, 227–231 (1990)

[384] Gombar, V. K., and Enslein, K., *Quant. Struct.-Act. Relat.* **9**, 321–325 (1990)

[385] a) Hall, L. H., Mohney, B., and Kier, L. B., *Quant. Struct.-Act. Relat.* **10**, 43–51 (1991); b) Hall, L. H., and Kier, L. B., *Med. Chem. Res.* **2**, 497–502 (1992)

[386] Kier, L. B., and Hall, L. H., *Quant. Struct.-Act. Relat.* **10**, 134–140 (1991)

[387] Lopez de Compadre, R. L., Compadre, C. M., Castillo, R., and Dunn III, W. J., *Eur. J. Med. Chem.* **18**, 569–571 (1983)

[388] Daniel, C., and Wood, F. S., *Fitting Equations to Data*, Wiley, New York, 1980

[389] Draper, N. R., and Smith, H., *Applied Regression Analysis*, 2nd ed., Wiley, New York, 1981

[390] Kubinyi, H., ref. 39, p. 589–643

[391] Kubinyi, H., *Quant. Struct.-Act. Relat.* **7**, 121–133 (1988)

[392] Singer, J. A., and Purcell, W. P., *J. Med. Chem.* **10**, 1000–1002 (1967)

[393] Kubinyi, H., and Kehrhahn, O.-H., *J. Med. Chem.* **19**, 578–586 (1976)

[394] Schaad, L. J., Hess Jr., B. A., Purcell, W. P., Cammarata, A., Franke, R., and Kubinyi, H., *J. Med. Chem.* **24**, 900–901 (1981)

[395] Hansch, C., and Calef, D. F., *J. Org. Chem.* **41**, 1240–1243 (1976)

[396] Fujita, T., ref. 39, p. 497–560

[397] Graham, J. D. P., and Karrar, M. A., *J. Med. Chem.* **6**, 103–107 (1963)

[398] Hansch, C., and Lien, E. J., *Biochem. Pharmacol.* **17**, 709–720 (1968)

[399] Cammarata, A., *J. Med. Chem.* **15**, 573–577 (1972)

[400] Topliss, J. G., and Costello, R. J., *J. Med. Chem.* **15**, 1066–1068 (1972)

[401] Topliss, J. G., and Edwards, R. P., *J. Med. Chem.* **22**, 1238–1244 (1979)

[402] Jha, T., Debnath, A. K., Mazumdar, A., Sengupta, C., and De, A. U., *Ind. J. Chem.* **25 B**, 169–174 (1986)

[403] Pleiss, M. A., and Unger, S. H., ref. 39, p. 561–587

[404] Cammarata, A., Allen, R. C., Seydel, J. K., and Wempe, E., *J. Pharm. Sci.* **59**, 1496–1499 (1970)

[405] Craig, P. N., Hansch, C. H., McFarland, J. W., Martin, Y. C., Purcell, W. P., and Zahradnik, R., *J. Med. Chem.* **14**, 447 (1971)

[406] Purcell, W. P., *Eur. J. Med. Chem.* **10**, 335–339 (1975)

[407] Wold, S., and Dunn III, W. J., *J. Chem. Inf. Comput. Sci.* **23**, 6–13 (1983)

[408] Benigni, R., and Giuliani, A., *Quant. Struct.-Act. Relat.* **10**, 99–100 (1991)

[409] Wold, S., *Quant. Struct.-Act. Relat.* **10**, 191–193 (1991)

[410] Kubinyi, H., and Kehrhahn, O.-H., *J. Med. Chem.* **19**, 1040–1049 (1976)

[411] Duewer, D. L., *J. Chemometrics* **4**, 299–321 (1990)

[412] Hansch, C., Silipo, C., and Steller, E. E., *J. Pharm. Sci.* **64**, 1186–1191 (1975)

[413] Kelder, J., and Greven, H. M., *Rec. Trav. Chim. Pays-Bas* **98**, 168–172 (1979)

[414] Mager, P. P., Mager, H., and Barth, A., *Sci. Pharm.* **47**, 265–297 (1979)

[415] Mager, P. P., *Med. Res. Rev.* **3**, 435–498 (1983)

[416] Mager, P. P., *Pharmazie* **45**, 359–360 (1990)

[417] Darvas, F., Röhricht, J., Budai, Z., and Bordás, B., *Advances in Pharmacological Research and Practice* (Proceedings of the 3rd Congress of the Hungarian Pharmacological Society, Budapest, 1979), Vol. **3**, Knoll, J., ed., Pergamon Press, 1980, p. 25–38

[418] Dubois, J.-E., Laurent, D., and Aranda, A., *J. Chim. Phys.* **70**, 1608–1615, 1616–1624 (1973)

[419] Aranda, A., *Compt. Rend. Hebd. Seances Acad. Sci. Paris, Ser.* **C 276**, 1301–1304 (1973)

[420] Dubois, J.-E., Laurent, D., Bost, P., Chambaud, S., and Mercier, C., *Eur. J. Med. Chem.* **11**, 225–236 (1976)

[421] Duperray, B., Chastrette, M., Makabeh, M. C., and Pacheco, H., *Eur. J. Med. Chem.* **11**, 323–336, 433–437 (1976)

[422] Mercier, C., and Dubois, J. E., *Eur. J. Med. Chem.* **14**, 415–423 (1979)

[423] Dubois, J.-E., Mercier, C., and Panaye, A., *Acta Pharm. Jugosl.* **36**, 135–169 (1986)

[424] Hall, L. H., and Kier, L. B., *Eur. J. Med. Chem.* **13**, 89–92 (1978)

[425] Cammarata, A., and Bustard, T. M., *J. Med. Chem.* **17**, 981–985 (1974)

[426] Hansch, C., and Lien, E. J., *J. Med. Chem.* **14**, 653–670 (1971)

[427] Hansch, C., *J. Org. Chem.* **35**, 620–621 (1970)

[428] Martin, Y. C., and Lynn, K. R., *J. Med. Chem.* **14**, 1162–1166 (1971)

[429] Bocek, K., Kopecky, J., Krivucova, M., and Vlachova, D., *Experientia* **20**, 667–668 (1964)

[430] Kopecky, J., Bocek, K., and Vlachova, D., *Nature* **207**, 981 (1965)

[431] Silipo, C., and Hansch, C., *J. Am. Chem. Soc.* **97**, 6849–6861 (1975)

[432] Blaney, J. M., Hansch, C., Silipo, C., and Vittoria, A., *Chem. Rev.* **84**, 333–407 (1984)

[433] Dearden, J. C., ref. 39, p. 375–411

[434] Penniston, J. T., Beckett, L., Bentley, D. L., and Hansch, C., *Mol. Pharmacol.* **5**, 333–341 (1969)

[435] Franke, R., and Schmidt, W., *Acta. Biol. Med. Germ.* **31**, 273–287 (1973)

[436] McFarland, J. W., *J. Med. Chem.* **13**, 1192–1196 (1970)

[437] Higuchi, T., and Davis, S. S., *J. Pharm. Sci.* **59**, 1376–1383 (1970)

[438] Hyde, R. M., *J. Med. Chem.* **18**, 231–233 (1975)

[439] Hyde, R. M., *Chem. Ind. (London)*, 859–862 (1977)

[440] Kubinyi, H., *Arzneim.-Forsch. (Drug Res.)* **26**, 1991–1997 (1976)

[441] Kubinyi, H., and Kehrhahn, O.-H., *Arzneim.-Forsch. (Drug Res.)* **28**, 598–601 (1978)

[442] Kubinyi, H., *Arzneim.-Forsch. (Drug Res.)* **29**, 1067–1080 (1979)

[443] Lippold, B. C., and Schneider, G. F., *Arzneim.-Forsch. (Drug Res.)* **25**, 843–852, 1683–1686 (1975)

[444] Kubinyi, H., *J. Pharm. Sci.* **67**, 262–263 (1978)

[445] Van de Waterbeemd, H., van Boeckel, S., Jansen, A., and Geeritsma, K., *Eur. J. Med. Chem.* **15**, 279–282 (1980)

[446] Van de Waterbeemd, H., van Bakel, P., and Jansen, A., *J. Pharm. Sci.* **70**, 1081–1082 (1981)

[447] De Haan, F. H. N., de Vringer, T., van de Waterbeemd, J. T. M., and Jansen, A. C. A., *Int. J. Pharm.* **13**, 75–87 (1982)

[448] Cooper, E. R., Berner, B., and Bruce, R. D., *J. Pharm. Sci.* **70**, 57–59 (1981)

[449] Berner, B., and Cooper, E. R., *J. Pharm. Sci.* **73**, 102–106 (1984)

[450] Yalkowski, S. H., and Flynn, G. L., *J. Pharm. Sci.* **62**, 210–217 (1973)

[451] Seydel, J. K., ref. 74, p. 179–201

[452] Seydel, J. K., and Schaper, K.-J., *Pharmac. Ther.* **15**, 131–182 (1982)

[453] Baláz, S., Sturdík, E., Hrmová, M., Breza, M., and Liptaj, T., *Eur. J. Med. Chem.* **19**, 167–171 (1984)

[454] Baláz, S., Wiese, M., Chi, H.-L., and Seydel, J. K., *Anal. Chim. Acta* **235**, 195–207 (1990)

[455] (a) Baláz, S., Wiese, M., and Seydel, J. K., *Quant. Struct.-Act. Relat.* **11**, 45–49 (1992); (b) *J. Pharm. Sci.* **81**, 849–857 (1992)

[456] Hansch, C., Grieco, C., Silipo, C., and Vittoria, A., *J. Med. Chem.* **20**, 1420–1435 (1977)

[457] Di Paolo, T., *J. Pharm. Sci.* **67**, 564–566, 566–568 (1978)

[458] Mager, P. P., *Pharmazie* **35**, 708–709 (1980)

[459] Martin, Y. C., ref. 51, p. 49–110

[460] Le Petit, G., *Pharmazie* **32**, 289–291 (1977)

[461] Le Petit, G., *Pharmazie* **35**, 696–698 (1980)

[462] Klemm, K., and Krüger, U., *Arzneim.-Forsch.* (*Drug. Res.*) **29**, 2–11 (1979)

[463] Clarke, F. H., and Cahoon, N. M., *J. Pharm. Sci.* **76**, 611–620 (1987)

[464] Schaper, K.-J., *J. Chem. Research* (*S*), 357 (1979)

[465] Murthy, K. S., and Zografi, G., *J. Pharm. Sci.* **59**, 1281–1285 (1970)

[466] Lippold, B. C., and Schneider, G. F., *Pharmazie* **31**, 237–239 (1976)

[467] Lippold, B. C., and Lettenbauer, W. A., *Pharmazie* **33**, 221–225 (1978)

[468] Wang, P.-H., and Lien, E. J., *J. Pharm. Sci.* **69**, 662–668 (1980)

[469] Scherrer, R. A., and Howard, S. M., *J. Med. Chem.* **20**, 53–58 (1977)

[470] Scherrer, R. A., and Howard, S. M., ref. 72, p. 507–526

[471] Scherrer, R. A., *Pesticide Synthesis Through Rational Approaches*, Magee, P. S., Kohn, G. K., and Menn, J. J., eds., ACS Symp. Series **255**, 225 -246 (1984)

[472] Lien, E. J., *Drug Design*, Vol. V, Ariëns, E. J., ed., Academic Press, 1975, p. 81–132

[473] Shore, P. A., Brodie, B. B., and Hogben, C. A. M., *J. Pharmac. Exptl. Therap.* **119**, 361–369 (1957)

[474] Wagner, J. G., *Fundamentals of Clinical Pharmacokinetics*, Drug Intelligence Publications, Inc., Hamilton, 1975

[475] Winne, D., *J. Pharmacokin. Biopharm.* **5**, 53–94 (1977)

[476] Tsuji, A., Miyamoto, E., Hashimoto, N., and Yamana, T., *J. Pharm. Sci.* **67**, 1705–1711 (1978)

[477] Martin, Y. C., and Hackbarth, J. J., *J. Med. Chem.* **19**, 1033–1039 (1976)

[478] Martin, Y. C., ref. 71, p. 351–358

[479] Schaper, K.-J., *Quant. Struct.-Act. Relat.* **1**, 13–27 (1982)

[480] Davis, S. S., and Elson, G., *J. Pharm. Pharmacol.* **26** S, 90 P (1974)

[481] Seiler, P., *Eur. J. Med. Chem.* **9**, 663–665 (1974)

[482] Kaufman, J. J., Semo, N. M., and Koski, W. S., *J. Med. Chem.* **18**, 647–655 (1975)

[483] Ezumi, K., and Kubota, T., *Chem. Pharm. Bull.* **28**, 85–91 (1980)

[484] Clarke, F. H., *J. Pharm. Sci.* **73**, 226–230 (1984)

[485] Clarke, F. H., ref. 78, p. 264–267

[486] a) Avdeef, A., ref. 83, 119–122 ; b) Avdeef, A., *Quant. Struct.-Act. Relat.* **11**, 510–517 (1992); d) Avdeef, A., *J. Pharm. Sci.* **82**, 183–190 (1993); d) Avdeef, A., Comer, J. E. A., and Thomson, S. J., *Anal. Chem.* **65**, 42–49 (1993)

[487] Dunn III, W. J., and Wold, S., ref. 39, p. 691–714

[488] Fukunaga, K., *Introduction to Statistical Pattern Recognition*, Academic Press, New York, 1972

[489] Varmuza, K., *Pattern Recognition in Chemistry* (Lecture Notes in Chemistry **21**), Springer-Verlag, Berlin, 1980

[490] Kowalski, B. R., and Bender, C. F., *J. Am. Chem. Soc.* **94**, 5632–5639 (1972)

[491] Stuper, A. J., and Jurs, P. C., *J. Am. Chem. Soc.* **97**, 182–187 (1975)

[492] Cammarata, A., and Menon, G. K., *J. Med. Chem.* **19**, 739–748 (1976)

[493] Menon, G. K., and Cammarata, A., *J. Pharm. Sci.* **66**, 304–314 (1977)

[494] Wijnne, H. J., ref. 70, p. 211–229

[495] Dunn III, W. J., Wold, S., and Martin, Y. C., *J. Med. Chem.* **21**, 922–930 (1978)

[496] Stuper, A. J., and Jurs, P. C., *J. Pharm. Sci.* **67**, 745–751 (1978)

[497] Jurs, P. C., Chou, J. T., and Yuan, M., ref. 72, p. 103–129

[498] Jurs, P. C., Chou, J. T., and Yuan, M., *J. Med. Chem.* **22**, 476–483 (1979)

[499] Dunn III, W. J., and Wold, S., *J. Med. Chem.* **23**, 595–599 (1980)

[500] Rose, S. L., and Jurs, P. C., *J. Med. Chem.* **25**, 769–776 (1982)

[501] Jurs, P. C., Stouch, T. R., Czerwinski, M., and Narvaez, J. N., *J. Chem. Inf. Comput. Sci.* **25**, 296–308 (1985)

[502] a) Livingstone, D. J., *Methods Enzymol.* **203**, 613–638 (1991); b) Miyashita, Y., Li, Z., and Sasaki, S.-I., *Trends Anal. Chem.* **12**, 50–60 (1993)

[503] Mathews, R. J., *J. Am. Chem. Soc.* **97**, 935–936 (1975)

[504] Craig, P. N., ref. 39, p. 645–666

[505] Enslein, K., and Craig, P. N., *J. Environ. Path. Toxicol.* **2**, 115–121 (1978)

[506] Enslein, K., and Craig, P. N., *J. Toxicol. Environ. Health* **10**, 521–530 (1982)

[507] Enslein, K., Lander, T. R., and Strange, J. L., *Teratog. Carcinog. Mutagen.* **3**, 289–309 (1983)

[508] Enslein, K., Blake, B. W., Tomb, M. E., and Borgstedt, H. H., *In Vitro Toxicol.* **1**, 33–44 (1986)

[509] Klopman, G., and Raychaudhury, C., *J. Comput. Chem.* **9**, 232–243 (1988)

[510] Blake, B. W., Enslein, K., Gombar, V. K., and Borgstedt, H. H., *Mutat. Res.* **241**, 261–271 (1990)

[511] Gombar, V. K., Borgstedt, H. H., Enslein, K., Hart, J. B., and Blake, B. W., *Quant. Struct.-Act. Relat.* **10**, 306–332 (1991)

[512] Tennant, R. W., Spalding, J., Stasiewicz, S., and Ashby, J., *Mutagenesis* **5**, 3–14 (1990)

[513] Bristol, D. W., personal comunication; Workshop "*Predicting Chemical Carcinogenesis in Rodents*", National Institute of Environmental Health, Research Triangle Park, NC, U. S. A., May 1993

[514] Klopman, G., *J. Am. Chem. Soc.* **106**, 7315–7321 (1984)

[515] Klopman, G., and Rosenkranz, H. S., *Mutat. Res.* **126**, 227–238 (1984)

[516] Klopman, G., *Environ. Health Perspect.* **61**, 269–274 (1985)

[517] Klopman, G., Frierson, M. R., and Rosenkranz, H. S., *Environ. Mutagen.* **7**, 625–644 (1985)

[518] Frierson, M. R., Klopman, G., and Rosenkranz, H. S., *Environ. Mutagen.* **8**, 283–327 (1986)

[519] Klopman, G., and Macina, O. T., *Mol. Pharmacol.* **31**, 457–476 (1987)

[520] Klopman, G., Kalos, A. N., and Rosenkranz, H. S., *Mol. Toxicol.* **1**, 61–81 (1987)

[521] Rosenkranz, H. S., and Klopman, G., *Mutat. Res.* **199**, 95–101 (1988)

[522] Rosenkranz, H. S., and Klopman, G., *Toxicol. Ind. Health* **4**, 533–540 (1988)

[523] Klopman, G., and Dimayuga, M. L., *J. Comput.-Aided Mol. Des.* **4**, 117–130 (1990)

[524] a) Klopman, G., *Quant. Struct.-Act. Relat.* **11**, 176–184 (1992); b) Klopman, G., and Ptchelintsev, *Med. Chem. Res.* **3**, 67–77 (1993)

[525] King, R. D., Muggleton, S., Lewis, R. A., and Sternberg, M. J. E., *Proc. Natl. Acad. Sci. USA* **89**, 11322–11326 (1992)

[526] Mercier, C., Troullier, G., and Dubois, J.-E., *Quant. Struct.-Act. Relat.* **9**, 88–93 (1990)

[527] Hübel, S., Rösner, T., and Franke, R., *Pharmazie* **35**, 424–433 (1980)

[528] Franke, R., Huebel, S., and Streich, W. J., *Environ. Health Perspect.* **61**, 239–255 (1985)

[529] Streich, W. J., and Franke, R., *Quant. Struct.-Act. Relat.* **4**, 13–18 (1985)

[530] Franke, R., and Streich, W. J., *Quant. Struct.-Act. Relat.* **4**, 51–63, 63–69 (1985)

[531] Hübel, S., and Franke, R., ref. 83, p. 177–180

[532] Simon, Z., and Szabadai, Z., *Stud. Biophys.* **39**, 123–132 (1973)

[533] Simon, Z., *Angew. Chem. Int. Ed. Engl.* **13**, 719–727 (1974)

[534] Simon, Z., Chiriac, A., Holban, S., Ciubotariu, D., and Mihalas, G. I., *Minimum Steric Difference – The MTD Method for QSAR Studies*, Research Studies Press, Letchworth, 1984

[535] Simon, Z., *Stud. Biophys.* **51**, 49–57 (1975)

[536] Motoc, I., Holban, S., Vancea, R., and Simon, Z., *Stud. Biophys.* **66**, 75–78 (1977)

[537] Motoc, I., *Topics Curr. Chem.* **114**, 93–105 (1983)

[538] Simon, Z., Chiriac, A., Motoc, I., Holban, S., Ciubotariu, D., and Szabadai, Z., *Stud. Biophys.* **55**, 217–226 (1976)

[539] Simon, Z., Holban, S., Motoc, I., Mracec, M., Chiriac, C., Kerek, F., Ciubotariu, D., Szabadai, Z., Pop, R. D., and Schwartz, I., *Stud. Biophys.* **59**, 181–197 (1976)

[540] Simon, Z., Badilescu, I. I., and Racovitan, T., *J. Theor. Biol.* **66**, 485–495 (1977)

[541] Popoviciu, V., Holban, S., Badilescu, I. I., and Simon, Z., *Stud. Biophys.* **69**, 75–76 (1978)

[542] Simon, Z., and Bohl, M., *Quant. Struct.-Act. Relat.* **11**, 23–28 (1992)

[543] Motoc, I., *Quant. Struct.-Act. Relat.* **3**, 43–47, 47–51 (1984)

[544] Magee, P. S., ref. 83, p. 549–552

[545] Magee, P. S., *Quant. Struct.-Act. Relat.* **9**, 202–215 (1990)

[546] Hopfinger, A. J., *J. Am. Chem. Soc.* **102**, 7196–7206 (1980)

[547] Battershell, C., Malhotra, D., and Hopfinger, A. J., *J. Med. Chem.* **24**, 812–818 (1981)

[548] Hopfinger, A. J., *J. Med. Chem.* **24**, 818–822 (1981)

[549] Hopfinger, A. J., *Arch. Biochem. Biophys.* **206**, 153–163 (1981)

[550] Hopfinger, A. J., and Potenzone Jr., R., *Mol. Pharmacol.* **21**, 187–195 (1982)

[551] Hopfinger, A. J., *J. Med. Chem.* **26**, 990–996 (1983)

[552] Hopfinger, A. J., *Quant. Struct.-Act. Relat.* **3**, 1–5 (1984)

[553] Mabilia, M., Pearlstein, R. A., and Hopfinger, A. J., *Eur. J. Med. Chem.* **20**, 163–174 (1985)

[554] Walters, D. E., and Hopfinger, A. J., *J. Mol. Struct.* (*Theochem*) **134**, 317–323 (1986)

[555] Hopfinger, A. J., and Burke, B. J., ref. 82, 151–159

[556] Hopfinger, A. J., and Burke, B. J., ref. 63, p. 173–209

[557] Aoyama, T., Suzuki, Y., and Ichikawa, H., *J. Med. Chem.* **33**, 905–908, 2583–2590 (1990)

[558] Andrea, T. A., and Kalayeh, H., ref. 83, p. 209–212

[559] Rose, V. S., MacFie, H. J. H., and Croall, I. F., ref. 83, p. 213–216

[560] Rose, V. S., Croall, I. F., and MacFie, H. J. H., *Quant. Struct.-Act. Relat.* **10**, 6–15 (1991); erratum **10**, 141 (1991)

[561] Wiese, M., *Quant. Struct.-Act. Relat.* **10**, 369–371 (1991)

[562] Andrea, T. A., and Kalayeh, H., *J. Med. Chem.* **34**, 2824–2836 (1991)

[563] Aoyama, T., and Ichikawa, H., *Chem. Pharm. Bull.* **39**, 358–366, 372–378 (1991)

[564] Chastrette, M., and de Saint Laumer, J. Y., *Eur. J. Med. Chem.* **26**, 829–833 (1991)

[565] Livingstone, D. J., Hesketh, G., Clayworth, D., *J. Mol. Graphics* **9**, 115–118 (1991)

[566] Livingstone, D. J., and Salt, D. W., *BioMed. Chem. Lett.* **2**, 213–218 (1992)

[567] Salt, D. W., Yildiz, N., Livingstone, D. J., and Tinsley, C. J., *Pestic. Sci.* **36**, 161–170 (1992)

[568] a) Manallack, D. T., and Livingstone, D. J., *Med. Chem. Res.* **2**, 181–190 (1992); b) Manallack, D. T., and Livingstone, D. J., ref. 84, p. 128–131

[569] Livingstone, D. J., and Manallack, D. T., *J. Med. Chem.* **36**, 1295–1297 (1992)

[570] So, S.-S., and Richards, W. G., *J. Med. Chem.* **35**, 3201–3207 (1992)

[571] Snedecor, G. W., and Cochran, W. G., *Statistical Methods*, The Iowa State University Press, Ames, 1973

[572] Natrella, M. G., *Experimental Statistics*, National Bureau of Standards Handbook **91**, US Government Printing Office, Washington, 1963
[573] Silipo, C., Hansch, C., Grieco, C., and Vittoria, A., *Arch. Biochem. Biophys.* **194**, 552–557 (1979)
[574] Jeng, Y. J., and Martin, A., *J. Pharm. Sci.* **74**, 1053–1057 (1985)
[575] Seydel, J. K., and Schaper, K.-J., ref. 74, p. 337–354
[576] Lipnick, R. L., *Sci. Total Environ.* **109–110**, 131–153 (1991)
[577] McFarland, J. W., and Gans, D. J., ref. 39, p. 667–689
[578] Martin, Y. C., Holland, J. B., Jarboe, C. H., and Plotnikoff, N., *J. Med. Chem.* **17**, 409–413 (1974)
[579] Prakash, G., and Hodnett, E. M., *J. Med. Chem.* **21**, 369–374 (1978)
[580] Henry, D. R., and Block, J. H., *J. Med. Chem.* **22**, 465–472 (1979)
[581] Smith, C. C., Genther, C. S., and Coats, E. A., *Eur. J. Med. Chem.* **14**, 271–276 (1979)
[582] Lewis, D. F. V., Ioannides, C., and Parke, D. V., *Mutagenesis* **5**, 433–435 (1990)
[583] Lewis, D. F. V., Ioannides, C., and Parke, D. V., ref. 83, p. 525–527
[584] Raevsky, O. A., Sapegin, A. M., and Zefirov, N. S., ref. 83, p. 189–192
[585] Moriguchi, I., and Komatsu, K., *Chem. Pharm. Bull.* **25**, 2800–2802 (1977)
[586] Moriguchi, I., Komatsu, K., and Matsushita, Y., *J. Med. Chem.* **23**, 20–26 (1980)
[587] Moriguchi, I., *J. Pharmacobio-Dyn.* **3**, S3 (1980)
[588] Moriguchi, I., Komatsu, K., and Matsushita, Y., *Anal. Chim. Acta* **133**, 625–636 (1981)
[589] Takahashi, Y., Miyashita, Y., Abe, H., and Sasaki, S., *Bunseki Kagaku* **33**, E487–E494 (1984); ref. *C. A.* **102**, 105 720 n (1985)
[590] Moriguchi, I., Hirono, S., Liu, Q., Matsushita, Y., and Nakagawa, T., *Chem. Pharm. Bull.* **38**, 3373–3379 (1990)
[591] Moriguchi, I., Hirono, S., Liu, Q., and Nakagome, I., *Quant. Struct.-Act. Relat.* **11**, 325–331 (1992)
[592] Malinowski, E. R., and Howery, D. G., *Factor Analysis in Chemistry*, Wiley, New York, 1980
[593] Dillon, W. R., and Goldstein, M., *Multivariate Analysis: Methods and Applications*, Wiley, New York, 1984
[594] Joliffe, I. T., *Principal Component Analysis*, Springer-Verlag, New York, 1986
[595] Wold, S., Esbensen, K., and Geladi, P., *Chemom. Intell. Lab. Syst.* **2**, 37–52 (1987)
[596] Coats, E. A., Cordes, H.-P., Kulkarni, V. M., Richter, M., Schaper, K.-J., Wiese, M., and Seydel, J. K., *Quant. Struct.-Act. Relat.* **4**, 99–109 (1985)
[597] Schaper, K.-J., and Seydel, J. K., ref. 78, p. 173–189
[598] Laass, W., ref. 78, p. 285–289
[599] Kubota, T., Hanamura, J., Kano, K., and Uno, B., *Chem. Pharm. Bull.* **33**, 1488–1495 (1985)
[600] Wiese, M., Seydel, J. K., Pieper, H., Krüger, G., Noll, K. R., and Keck, J., *Quant. Struct.-Act. Relat.* **6**, 164–172 (1987)
[601] Rose, V. S., Wood, J., and MacFie, H. J. H., *Quant. Struct.-Act. Relat.* **10**, 359–368 (1991)
[602] Wold, S., *Pattern Recognition* **8**, 127–139 (1976)
[603] Dunn III, W. J., and Wold, S., *J. Med. Chem.* **21**, 1001–1007 (1978)
[604] Dunn III, W. J., and Wold, S., *Bioorg. Chem.* **9**, 505–523 (1980)
[605] Wold, S., Dunn III, W. J., and Hellberg, S., ref. 46, p. 95–117
[606] Wold, H., *Perspectives in Probability and Statistics, Papers in Honour of M. S. Bartlett*, Gani, J., ed., Academic Press, London, 1975
[607] Wold, S., Martens, H., and Wold, H., *Matrix Pencils* (Lecture Notes in Mathematics) Ruhe, A., and Kagström, B., eds., Springer-Verlag Heidelberg, 1983, p. 286–293
[608] Nordén, B., Edlund, U., Johnels, D., and Wold, S., *Quant. Struct.-Act. Relat.* **2**, 73–76 (1983)
[609] Wold, S., Albano, C., Dunn III, W. J., Esbensen, K., Hellberg, S., Johansson, E., and Sjöström, M., *Food Research and Data Analysis*, Martens, H., and Russwurm Jr., H., eds., Applied Science Publishers, London, 1983, p. 147–188
[610] Wold, S., Albano, C., Dunn III, W. J., Edlund, U., Esbenson, K., Geladi, P., Hellberg, S., Johansson, E., Lindberg, W., and Sjöström, M., *Chemometrics, Mathematics and Statistics in Chemistry*, Kowalski, B. R., ed., Reidel Publishing Company, Dordrecht, 1984, p. 17–95
[611] Cramer III, R. D., Bunce, J. D., Patterson, D. E., and Frank, I. E., *Quant. Struct.-Act. Relat.* 7, 18–25 (1988); erratum 7, 91 (1988)

[612] Cruciani, G., Baroni, M., Bonelli, D., Clementi, S., Ebert, C., and Skagerberg, B., *Quant. Struct.-Act. Relat.* **9**, 101–107 (1990)

[613] Cruciani, G., Baroni, M., Clementi, S., Costantino, G., Riganelli, D., and Skagerberg, B., *J. Chemometrics* **6**, 335–346 (1992)

[614] Baroni, M., Clementi, S., Cruciani, G., Costantino, G., Riganelli, D., and Oberrauch, E., *J. Chemometrics* **6**, 347–356 (1992)

[615] Baroni, M., Costantino, G., Cruciani, G., Riganelli, D., Valigi, R., and Clementi, S., *Quant. Struct.-Act. Relat.* **12**, 9–20 (1993)

[616] Wold, S., Kettaneh-Wold, N., and Skagerberg, B., *Chemom. Intell. Lab. Syst.* **7**, 53–65 (1989)

[617] Barbagianni, N., Bonelli, D., Cesareo, D., Clementi, S., Cruciani, G., and Skagerberg, B., ref. 82, p. 135–138

[618] Clementi, S., Cruciani, G., Curti, G., and Skagerberg, B., *J. Chemometrics* **3**, 499–509 (1989)

[619] Hellberg, S., Wold, S., Dunn III, W. J., Gasteiger, J., and Hutchings, M. G., *Quant. Struct.-Act. Relat.* **4**, 1–11 (1985)

[620] Skagerberg, B., Dunn III, W. J., Hellberg, S., and Wold, S., ref. 78, p. 305–310

[621] Hellberg, S., Sjöström, M., and Wold, S., *Acta Chem. Scand.* **B 40**, 135–140 (1986)

[622] Cocchi, M., Menziani, M. C., Rastelli, G., and De Benedetti, P. G., *Quant. Struct.-Act. Relat.* **9**, 340–345 (1990)

[623] Dunn III, W. J., Greenberg, M. J., and Callejas, S. S., *J. Med. Chem.* **19**, 1299–1301 (1976)

[624] Takahashi, Y., Miyashita, Y., Abe, H., and Sasaki, S.-I., *Anal. Chim. Acta* **122**, 241–247 (1980)

[625] Willett, P., *Anal. Chim. Acta* **136**, 29–37 (1982)

[626] Jurs, P. C., and Lawson, R. G., *Chemom. Intell. Lab. Syst.* **10**, 81–83 (1991)

[627] McFarland, J. W., and Gans, D. J., *J. Med. Chem.* **29**, 505–514 (1986)

[628] McFarland, J. W., and Gans, D. J., ref. 80, p. 25–29

[629] McFarland, J. W., and Gans, D. J., *J. Med. Chem.* **30**, 46–49 (1987)

[630] McFarland, J. W., and Gans, D. J., ref. 82, p. 199–202

[631] McFarland, J. W., and Gans, D. J., *Drug. Information J.* **24**, 705–711 (1990)

[632] Craig, P. N., *J. Med. Chem.* **14**, 680–684 (1971)

[633] Cramer III, R. D., Snader, K. M., Willis, C. R., Chakrin, L. W., Thomas, J., and Sutton, B. M., *J. Med. Chem.* **22**, 714–725 (1979)

[634] Magee, P. S., *IUPAC Pesticide Chemistry: Human Welfare and the Environment*, Vol. **1**, Miyamoto, J., and Kearney, P. C., eds., Pergamon Press, Oxford, 1983, p. 251–260

[635] Darvas, F., *J. Med. Chem.* **17**, 799–804 (1974)

[636] Gilliom, R. D., Purcell, W. P., and Bosin, T. R., *Eur. J. Med. Chem.* **12**, 187–192 (1977)

[637] Topliss, J. G., *J. Med. Chem.* **15**, 1006–1011 (1972)

[638] Topliss, J. G., and Martin, Y. C., *Drug Design*, Vol. V, Ariëns, E. J., ed., Academic Press, New York, 1975, p. 1–21

[639] Topliss, J. G., *J. Med. Chem.* **20**, 463–469 (1977)

[640] Martin, Y. C., and Panas, H. N., *J. Med. Chem.* **22**, 784–791 (1979)

[641] Martin, Y. C., and Dunn III, W. J., *J. Med. Chem.* **16**, 578–579 (1973)

[642] Cochran, W. G., and Cox, B. M., *Experimental Designs*, 2nd Ed., Wiley, New York, 1957

[643] Petersen, R. G., *Design and Analysis of Experiments*, Marcel Dekker, New York, 1985

[644] Deming, S. N., and Morgan, S. L., *Experimental Design: A Chemometric Approach*, Elsevier, Amsterdam, 1987

[645] Austel, V., *Eur. J. Med. Chem.* **17**, 9–16, 339–347 (1982)

[646] Austel, V., ref. 75, p. 223–229

[647] Austel, V., *Quant. Struct.-Act. Relat.* **2**, 59–65 (1983)

[648] Austel, V., *Drugs of the Future* **9**, 349–365 (1984)

[649] Austel, V., ref. 78, p. 247–250

[650] Austel, V., *Topics Curr. Chem.* **114**, 7–19 (1983)

[651] Pastor, M., and Alvarez-Builla, J., *Quant. Struct.-Act. Relat.* **10**, 350–358 (1991)

[652] Wootton, R., Cranfield, R., Sheppey, G. C., and Goodford, P. J., *J. Med. Chem.* **18**, 607–613 (1975)

[653] Woolridge, K. R. H., *Eur. J. Med. Chem.* **15**, 63–66 (1980)

[654] Wootton, R., *J. Med. Chem.* **26**, 275–277 (1983)

[655] Mitchell, T. J., *Technometrics* **16**, 203–210 (1974)

[656] Unger, S. H., *Drug Design*, Vol. **IX**, Ariëns, E. J., ed., Academic Press, New York, 1980, p. 47 – 119

[657] Herrmann, E. C., ref. 75, p. 231 – 234

[658] Schaper, K.-J., ref. 75, p. 235 – 236

[659] Borth, D. M., McKay, R. J., and Elliott, J. R., *Technometrics* **27**, 25 – 35 (1985)

[660] Dove, S., Streich, W. J., and Franke, R., *J. Med. Chem.* **23**, 1456 – 1459 (1980)

[661] Skagerberg, B., Clementi, S., Sjöström, M., Tosato, M.-L., and Wold, S., ref. 82, p. 127 – 130

[662] Clementi, S., Cruciani, G., Baroni, M., and Skagerberg, B., ref. 83, p. 217 – 222

[663] Hellberg, S., Eriksson, L., Jonsson, J., Lindgren, F., Rännar, S., Sandberg, M., Sjöström, M., Skagerberg, B., and Wold, S., ref. 83, p. 357 – 360

[664] Hellberg, S., Eriksson, L., Jonsson, J., Lindgren, F., Sjöström, M., Skagerberg, B., Wold, S., and Andrews, P., *Int. J. Peptide Protein Res.* **37**, 414 – 424 (1991)

[665] Cativiela, C., Elguero, J., Mathieu, D., Melendez, E., and Phan Tan Luu, R., *Eur. J. Med. Chem.* **18**, 359 – 363 (1983)

[666] Hansch, C., Leo, A., and Zhang, L., eds., Comprehensive Quantitative Structure-Activity Relationships: C-QSAR, Pomona College, Claremont, CA, USA, 1992

[667] Hansch, C., ref. 82, p. 23 – 30

[668] Silipo, C., and Hansch, C., *J. Med. Chem.* **19**, 62 – 71 (1976)

[669] Yoshimoto, M., and Hansch, C., *J. Med. Chem.* **19**, 71 – 98 (1976)

[670] Gupta, S. P., *Chem. Rev.* **87**, 1183 – 1253 (1987)

[671] Li, R.-L., Hansch, C., Matthews, D., Blaney, J. M., Langridge, R., Delcamp, T. J., Susten, S. S., and Freisheim, J. H., *Quant. Struct.-Act. Relat.* **1**, 1 – 7 (1982)

[672] Hansch, C., *Drug Intell. Clin. Pharm.* **16**, 391 – 396 (1982)

[673] Carotti, A., Smith, R. N., Wong, S., Hansch, C., Blaney, J. M., and Langridge, R., *Arch. Biochem. Biophys.* **229**, 112 – 125 (1984)

[674] Hansch, C., and Klein, T. E., *Acc. Chem. Res.* **19**, 392 – 400 (1986)

[675] Morgenstern, L., Recanatini, M., Klein, T. E., Steinmetz, W., Yang, C.-Z., Langridge, R., and Hansch, C., *J. Biol. Chem.* **262**, 10767 – 10772 (1987)

[676] Hansch, C., and Klein, T. E., *Methods Enzymol.* **202**, 512 – 543 (1991)

[677] Coats, E. A., Genther, C. S., Dietrich, S. W., Guo, Z.-R., and Hansch, C., *J. Med. Chem.* **24**, 1422 – 1429 (1981)

[678] Khwaja, T. A., Pentecost, S., Selassie, C. D., Guo, Z.-R., and Hansch, C., *J. Med. Chem.* **25**, 153 – 156 (1982)

[679] Selassie, C. D., Guo, Z.-R., Hansch, C., Khwaja, T. A., and Pentecost, S., *J. Med. Chem.* **25**, 157 – 161 (1982)

[680] Li, R.-L., Hansch, C., and Kaufman, B. T., *J. Med. Chem.* **25**, 435 – 440 (1982)

[681] Selassie, C. D., Li, R.-L., Hansch, C., Khwaja, T. A., and Dias, C. B., *J. Med. Chem.* **25**, 518 – 522 (1982)

[682] Hansch, C., Li, R.-L., Blaney, J. M., and Langridge, R., *J. Med. Chem.* **25**, 777 – 784 (1982)

[683] Hansch, C., Hathaway, B. A., Guo, Z.-R., Selassie, C. D., Dietrich, S. W., Blaney, J. M., Langridge, R., Volz, K. W., and Kaufman, B. T., *J. Med. Chem.* **27**, 129 – 143 (1984)

[684] Selassie, C. D., Hansch, C., Khwaja, T. A., Dias, C. B., and Pentecost, S., *J. Med. Chem.* **27**, 347 – 357 (1984)

[685] Coats, E. A., Genther, C. S., Selassie, C. D., Strong, C. D., and Hansch, C., *J. Med. Chem.* **28**, 1910 – 1916 (1985)

[686] Selassie, C. D., Fang, Z.-X., Li, R.-L., Hansch, C., Klein, T., Langridge, R., and Kaufman, B. T., *J. Med. Chem.* **29**, 621 – 626 (1986)

[687] Selassie, C. D., Strong, C. D., Hansch, C., Delcamp, T. J., Freisheim, J. H., and Khwaja, T. A., *Cancer Res.* **46**, 744 – 756 (1986)

[688] Booth, R. G., Selassie, C. D., Hansch, C., and Santi, D. V., *J. Med. Chem.* **30**, 1218 – 1224 (1987)

[689] Li, R.-L., and Poe, M., *J. Med. Chem.* **31**, 366 – 370 (1988)

[690] Selassie, C. D., Li, R. L., Hansch, C. H., Klein, T., Langridge, R., Kaufman, B. T., Freisheim, J., and Khwaja, T., ref. 82, p. 341 – 344

[691] Selassie, C. D., Li, R.-L., Poe, M., and Hansch, C., *J. Med. Chem.* **34**, 46 – 54 (1991)

[692] Kuyper, L. F., Roth, B., Baccanari, D. P., Ferone, R., Beddell, C. R., Champness, J. N., Stammers, D. K., Dann, J. G., Norrington, F. E., Baker, D. J., and Goodford, P. J., *J. Med. Chem.* **28**, 303–311 (1985)

[693] Czaplinsky, K.-H., Kansy, M., Seydel, J. K., and Haller, R., *Quant. Struct.-Act. Relat.* **6**, 70–72 (1987)

[694] Hansch, C., Smith, R. N., Rockoff, A., Calef, D. F., Jow, P. Y. C., and Fukunaga, J. Y., *Arch. Biochem. Biophys.* **183**, 383–392 (1977)

[695] Smith, R. N., Hansch, C., Kim, K. H., Omiya, B., Fukumura, G., Selassie, C. D., Jow, P. Y. C., Blaney, J. M., and Langridge, R., *Arch. Biochem. Biophys.* **215**, 319–328 (1982)

[696] Carotti, A., Hansch, C., Mueller, M. M., and Blaney, J. M., *J. Med. Chem.* **27**, 1401–1405 (1984); erratum **28**, 261 (1985)

[697] Carotti, A., Casini, G., and Hansch, C., *J. Med. Chem.* **27**, 1427–1431 (1984)

[698] Carotti, A., Raguseo, C., and Hansch, C., *Chem.-Biol. Interactions* **52**, 279–288 (1985)

[699] Compadre, C. M., Hansch, C., Klein, T. E., and Langridge, R., *Biochim. Biophys. Acta* **1038**, 158–163 (1990)

[700] Compadre, C. M., Hansch, C., Klein, T. E., Petridou-Fischer, J., Selassie, C. D., Smith, R. N., Steinmetz, W., Yang, C.-Z., and Yang, G.-Z., *Biochim. Biophys. Acta* **1079**, 43–52 (1991)

[701] Selassie, C. D., Chow, M., and Hansch, C., *Chem.-Biol. Interactions* **68**, 13–25 (1988)

[702] Recanatini, M., Klein, T., Yang, C.-Z., McClarin, J., Langridge, R., and Hansch, C., *Mol. Pharmacol.* **29**, 436–446 (1986)

[703] Soman, G., Narayanan, J., Martin, B. L., and Graves, D. J., *Biochemistry* **25**, 4113–4119 (1986)

[704] Cornell, N. W., Hansch, C., Kim, K. H., and Henegar, K., *Arch. Biochem. Biophys.* **227**, 81–90 (1983)

[705] Hansch, C., Klein, T., McClarin, J., Langridge, R., and Cornell, N. W., *J. Med. Chem.* **29**, 615–620 (1986)

[706] Hansch, C., Schaeffer, J., and Kerley, R., *J. Biol. Chem.* **247**, 4703–4710 (1972)

[707] Hansch, C., and Bjorkroth, J.-P., *J. Org. Chem.* **51**, 5461–5462 (1986)

[708] Barbato, F., Cappello, B., Silipo, C., and Vittoria, A., *Farmaco, Ed. Sci.* **37**, 519–536 (1982)

[709] Hansch, C., McClarin, J., Klein, T., and Langridge, R., *Mol. Pharmacol.* **27**, 493–498 (1985)

[710] Carotti, A., Raguseo, C., Campagna, F., Langridge, R., and Klein, T. E., *Quant. Struct.-Act. Relat.* **8**, 1–10 (1989)

[711] Grieco, C., Hansch, C., Silipo, C., Smith, R. N., Vittoria, A., and Yamada, K., *Arch. Biochem. Biophys.* **194**, 542–551 (1979)

[712] Wright, G. E., and Gambino, J. J., *J. Med. Chem.* **27**, 181–185 (1984)

[713] Dale, M. P., Ensley, H. E., Kern, K., Sastry, K. A. R., and Byers, L. D., *Biochemistry* **24**, 3530–3539 (1985)

[714] Rooney, C. S., Randall, W. C., Streeter, K. B., Ziegler, C., Cragoe Jr., E. J., Schwam, H., Michelson, S. R., Williams, H. W. R., Eichler, E., Duggan, D. E., Ulm, E. H., and Noll, R. M., *J. Med. Chem.* **26**, 700–714 (1983)

[715] Silipo, C., and Hansch, C., *Mol. Pharmacol.* **10**, 954–962 (1974)

[716] Maitani, Y., Nakagaki, M., and Nagai, T., *Int. J. Pharm.* **74**, 105–116 (1991)

[717] Martin, Y., Martin, W. B., and Taylor, J. D., *J. Med. Chem.* **18**, 883–888 (1975)

[718] Campbell, N. R. C., Van Loon, J. A., Sundaram, R. S., Ames, M. M., Hansch, C., and Weinshilboum, R., *Mol. Pharmacol.* **32**, 813–819 (1987)

[719] Ames, M. M., Selassie, C. D., Woodson, L. C., Van Loon, J. A., Hansch, C., and Weinshilboum, R., *J. Med. Chem.* **29**, 354–358 (1986)

[720] Manley, P. W., Allanson, N. M., Booth, R. F. G., Buckle, P. E., Kuzniar, E. J., Lad, N., Lai, S. M. F., Lunt, D. O., and Tuffin, D. P., *J. Med. Chem.* **30**, 1588–1595 (1987)

[721] Gambino, J., Focher, F., Hildebrand, C., Maga, G., Noonan, T., Spadari, S., and Wright, G., *J. Med. Chem.* **35**, 2979–2983 (1992)

[722] Silipo, C., and Hansch, C., *Farmaco, Ed. Sci.* **30**, 35–46 (1975)

[723] Leo, A., Hansch, C., and Church, C., *J. Med. Chem.* **12**, 766–771 (1969)

[724] Hansch, C., *Adv. Pharmac. Chemother.* **13**, 45–81 (1975)

[725] Kutter, E., and Hansch, C., *Arch. Biochem. Biophys.* **135**, 126–135 (1969)

[726] Yoshimoto, M., Hansch, C., and Jow, P. Y. C., *Chem. Pharm. Bull.* **23**, 437–444 (1975)

[727] Hansch, C., Yoshimoto, M., and Doll, M. H., *J. Med. Chem.* **19**, 1089–1093 (1976)

[728] Hansch, C., and Moser, P., *Immunochemistry* **15**, 535–540 (1978)

[729] Pratesi, P., Villa, L., Ferri, V., De Micheli, C., De Amici, M., Grana, E., Barbone, M. G. S., Silipo, C., Vittoria, A., and Capello, B., *Farmaco, Ed. Sci.* **39**, 171–188 (1984)

[730] Capello, B., Silipo, C., Vittoria, A., and Pratesi, P., *Farmaco, Ed. Sci.* **39**, 991–1007 (1984)

[731] Pratesi, P., Caliendo, G., Silipo, C., and Vittoria, A., *Quant. Struct.-Act. Relat.* **11**, 1–17 (1992)

[732] Pratesi, P., Caliendo, G., Silipo, C., and Vittoria, A., *Quant. Struct.-Act. Relat.* **11**, 151–161 (1992)

[733] Lien, E. J., Ariëns, E. J., and Beld, A. J., *Eur. J. Pharmacol.* **35**, 245–252 (1976)

[734] Lambrecht, G., Moser, U., and Mutschler, E., *Eur. J. Med. Chem.* **15**, 305–310 (1980)

[735] Ter Laak, A. M., van Drooge, M. J., Timmerman, H., and Donné-op den Kelder, G. M., *Quant. Struct.-Act. Relat.* **11**, 348–363 (1992)

[736] Hansch, C., and Caldwell, J., *J. Comput.-Aided Mol. Des.* **5**, 441–453 (1991)

[737] Dietrich, S. W., Bolger, M. B., Kollman, P. A., and Jorgensen, E. C., *J. Med. Chem.* **20**, 863–880 (1977)

[738] IJzerman, A. P., Aué, G. H. J., Bultsma, T., Linschoten, M. R., and Timmerman, H., *J. Med. Chem.* **28**, 1328–1334 (1985)

[739] IJzerman, A. P., Bultsma, T., and Timmerman, H., *J. Med. Chem.* **29**, 549–554 (1986)

[740] Borea, P. A., Pietrogrande, M. C., and Biagi, G. L., *Biochem. Pharmacol.* **37**, 3953–3957 (1988)

[741] Bandiera, S., Sawyer, T. W., Campbell, M. A., Fujita, T., and Safe, S., *Biochem. Pharmacol.* **32**, 3803–3813 (1983)

[742] Safe, S., Bandiera, S., Sawyer, T., Zmudzka, B., Mason, G., Romkes, M., Denomme, M. A., Sparling, J., Okey, A. B., and Fujita, T., *Environ. Health Perspect.* **61**, 21–33 (1985)

[743] El Tayar, N., Kilpatrick, G. J., van de Waterbeemd, H., Testa, B., Jenner, P., and Marsden, C. D., *Eur. J. Med. Chem.* **23**, 173–182 (1988)

[744] Gupta, S. P., and Handa, A., *Res. Commun. Chem. Pathol. Pharmacol.* **55**, 357–366 (1987)

[745] Singh, P., and Goyal, A., *Arzneim.-Forsch. (Drug Res.)* **37**, 51–54 (1987)

[746] Mannhold, R., *Drugs of Today* **20**, 69–90 (1984)

[747] Mahmoudian, M., and Richards, W. G., *J. Pharm. Pharmacol.* **38**, 272–276 (1986)

[748] Goll, A., Glossmann, H., and Mannhold, R., *Arch. Pharmacol.* **334**, 303–312 (1986)

[749] Mannhold, R., Bayer, R., Ronsdorf, M., and Martens, L., *Arzneim.-Forsch. (Drug Res.)* **37**, 419–424 (1987)

[750] Coburn, R. A., Wierzba, M., Suto, M. J., Solo, A. J., Triggle, A. M., and Triggle, D. J., *J. Med. Chem.* **31**, 2103–2107 (1988)

[751] Mager, P. P., Coburn, R. A., Solo, A. J., Triggle, D. J., and Rothe, H., *Drug Des. Discovery* **8**, 273–289 (1992)

[752] Mitani, K., Yoshida, T., Suzuki, T., Koshinaka, E., Kato, H., Ito, Y., and Fujita, T., *Chem. Pharm. Bull.* **36**, 776–783 (1988)

[753] Mitani, K., Sakurai, S., Suzuki, T., Morikawa, K., Koshinaka, E., Kato, H., Ito, Y., and Fujita, T., *Chem. Pharm. Bull.* **36**, 4103–4120, 4121–4135 (1988)

[754] Bass, G. E., Powers, L. J., and Dillingham, E. O., *J. Pharm. Sci.* **65**, 1525–1527 (1976)

[755] De Benedetti, P. G., *Fortschr. Arzneimittelforsch.* (*Prog. Drug Res.*) **36**, 361–417 (1991)

[756] Hayes, J. D., and Wolf, C. R., *Biochem. J.* **272**, 281–285 (1990)

[757] Venger, B. H., Hansch, C., Hatheway, G. J., and Amrein, Y. U., *J. Med. Chem.* **22**, 473–476 (1979)

[758] Hansch, C., Venger, B. H., and Panthananickal, A., *J. Med. Chem.* **23**, 459–461 (1980)

[759] Leo, A., Panthananickal, A., Hansch, C., Theiss, J., Shimkin, M., and Andrews, A. W., *J. Med. Chem.* **24**, 859–864 (1981)

[760] Hansch, C., *Sci. Total Environ.* **109–110**, 17–30 (1991)

[761] Lien, E. J., *Ann. Rev. Pharmacol. Toxicol.* **21**, 31–61 (1981)

[762] Rowland, M., ref. 75, p. 155–161

[763] Seydel, J. K., ref. 75, p. 163–181

[764] Austel, V., and Kutter, E., ref. 44, p. 437–496

[765] Lien, E. J., *Fortschr. Arzneimittelforsch.* (*Prog. Drug Res.*) **29**, 67–95 (1985)

[766] Mayer, J. M., and van de Waterbeemd, H., *Environ. Health Perspect.* **61**, 295–306 (1985)

[767] Hansch, C., Steward, A. R., Anderson, S. M., and Bentley, D., *J. Med. Chem.* **11**, 1–11 (1968)

[768] Hansch, C., *Farmaco, Ed. Sci.* **23**, 293–320 (1968)

[769] Hansch, C., Steward, A. R., and Iwasa, J., *Mol. Pharmacol.* **1**, 87–92 (1965)

[770] Lien, E. J., Tong, G. L., Chou, J. T., and Lien, L. L., *J. Pharm. Sci.* **62**, 246–250 (1973)

[771] Hansch, C., Björkroth, J. P., and Leo, A., *J. Pharm. Sci.* **76**, 663–687 (1987)

[772] Dischino, D. D., Welch, M. J., Kilbourn, M. R., and Raichle, M. E., *J. Nucl. Med.* **24**, 1030–1038 (1983)

[773] Van de Waterbeemd, H., and Kansy, M., *Chimia* **46**, 299–303 (1992)

[774] Hansch, C., Sinclair, J. F., and Sinclair, P. R., *Quant. Struct.-Act. Relat.* **9**, 223–226 (1990)

[775] Akbaraly, J. P., Leng, J. J., Bozler, G., and Seydel, J. K., ref. 78, p. 313–317

[776] Recanatini, M., Valenti, P., and Da Re, P., *Quant. Struct.-Act. Relat.* **7**, 12–18 (1988)

[777] Dove, S., Koch, A., and Franke, R., *Acta Pharm. Jugosl.* **36**, 119–134 (1986)

[778] Craig, P. N., and Hansch, C. H., *J. Med. Chem.* **16**, 661–667 (1973)

[779] Kim, K. H., Hansch, C., Fukunaga, J. Y., Steller, E. E., Jow, P. Y. C., Craig, P. N., and Page, J., *J. Med. Chem.* **22**, 366–391 (1979)

[780] Rippmann, F., *Quant. Struct.-Act. Relat.* **9**, 1–5 (1990)

[781] Hansch, C., *Farmaco, Ed. Sci.* **34**, 89–104 (1979)

[782] Rekker, R. F., *Dev. Pharmacol.* **3**, 23–46 (1983)

[783] Kubinyi, H., *J. Cancer Res. Clin. Oncol.* **116**, 529–537 (1990)

[784] Panthananickal, A., Hansch, C., Leo, A., and Quinn, F. R., *J. Med. Chem.* **21**, 16–26 (1978)

[785] Hatheway, G. J., Hansch, C., Kim, K. H., Milstein, S. R., Schmidt, C. L., Smith, R. N., and Quinn, F. R., *J. Med. Chem.* **21**, 563–574 (1978)

[786] Hansch, C., Hatheway, G. J., Quinn, F. R., and Greenberg, N., *J. Med. Chem.* **21**, 574–577 (1978)

[787] Panthananickal, A., Hansch, C., and Leo, A., *J. Med. Chem.* **22**, 1267–1269 (1979)

[788] Hansch, C., Leo, A., Schmidt, C., Jow, P. Y. C., and Montgomery, J. A., *J. Med. Chem.* **23**, 1095–1101 (1980)

[789] Denny, W. A., Cain, B. F., Atwell, G. J., Hansch, C., Panthananickal, A., and Leo, A., *J. Med. Chem.* **25**, 276–315 (1982)

[790] Gupta, S. P., Singh, P., and Bindal, M. C., *Chem. Rev.* **83**, 633–649 (1983)

[791] Gupta, S. P., *Chem. Rev.* **89**, 1765–1800 (1989)

[792] Zeelen, F. J., *Quant. Struct.-Act. Relat.* **5**, 131–137 (1986)

[793] Seeman, P., Tittler, M., Tedesco, J., and Lee, T., ref. 70, p. 131–145

[794] Squires, R. F., Saederup, E., Crawley, J. N., Skolnick, P., and Paul, S. M., *Life Sci.* **35**, 1439–1444 (1984)

[795] Weishaar, R. E., Kobylarz-Singer, D. C., Steffen, R. P., and Kaplan, H. R., *Circulation Research* **61**, 539–547 (1987)

[796] Timmermans, P. B. M. W. M., and van Zwieten, P. A., *Arch. Int. Pharmacodyn. Ther.* **228**, 237–250 (1977)

[797] Timmermans, P. B. M. W. M., de Jonge, A., and van Zwieten, P. A., *Quant. Struct.-Act. Relat.* **1**, 8–12 (1982)

[798] Timmermans, P. B. M. W. M., de Jonge, A., van Meel, J. C. A., Slothorst-Grisdijk, F. P., Lam, E., and van Zwieten, P. A., *J. Med. Chem.* **24**, 502–507 (1981)

[799] Timmermans, P. B. M. W. M., de Jonge, A., Thoolen, M. J. M. C., Wilffert, B., Batink, H., and van Zwieten, P. A., *J. Med. Chem.* **27**, 495–503 (1984)

[800] Kubinyi, H., *Chemie in unserer Zeit* **20**, 191–202 (1986)

[801] Kubinyi, H., *Physical Property Prediction in Organic Chemistry*, Jochum, C., Hicks, M. G., and Sunkel, J., eds., Springer-Verlag Berlin, 1988, p. 235–247

[802] Lewi, P. J., *Arzneim.-Forsch.* (*Drug Res.*) **26**, 1295–1300 (1976)

[803] Lewi, P. J., *Drug Design*, Vol. **VII**, Ariëns, E. J., ed., Academic Press, New York, 1976, p. 209–278

[804] Lewi, P. J., *Pharmac. Ther.* **B** 3, 481–537 (1978)

[805] Lewi, P. J., *Biostatistics in Pharmacology*, Vol. 3, Delaunois, A. L., ed., Pergamon Press, Oxford, 1979, p. 1149–1227

[806] Lewi, P. J., *Chemometrics. Mathematics and Statistics in Chemistry*, Kowalski, B. R., ed., Reidel Publishing Company, Dordrecht, 1984, p. 351–376

[807] Lewi, P. J., *Chemom. Intell. Lab. Syst.* **5**, 105–116 (1989)

[808] Nendza, M., and Seydel, J. K., *Quant. Struct.-Act. Relat.* **7**, 165–174 (1988); erratum **7**, 250 (1988)

[809] Phillips, J. C., Gibson, W. B., Yam, J., Alden, C. L., and Hard, G. C., *Food Chem. Toxicol.* **28**, 375–394 (1990)

[810] Craig, P. N., *J. Med. Chem.* **15**, 144–149 (1972)

[811] Katz, R., Osborne, S. F., and Ionescu, F., *J. Med. Chem.* **20**, 1413–1419 (1977)

[812] Ban, T., and Fujita, T., *J. Med. Chem.* **12**, 353–356 (1969)

[813] Labes, D., and Hagen, V., *Pharmazie* **34**, 554–556 (1979)

[814] Bindal, M. C., Singh, P., and Gupta, S. P., *Arzneim.-Forsch. (Drug Res.)* **32**, 719–721 (1982)

[815] Kelder, J., de Boer, T., de Graaf, J. S., and Wieringa, J. H., ref. 78, p. 162–169

[816] Abramson, F. B., Barlow, R. B., Mustafa, M. G., and Stephenson, R. P., *Br. J. Pharmac.* **37**, 207–233 (1969)

[817] Lien, E. J., Hansch, C., and Anderson, S. M., *J. Med. Chem.* **11**, 430–441 (1968)

[818] Mager, P. P., and Seese, A., *Pharmazie* **36**, 427–429 (1981)

[819] Mager, P. P., and Seese, A., *Pharmazie in unserer Zeit* **10**, 97–108 (1981)

[820] Grieco, C., Silipo, C., and Vittoria, A., *Farmaco, Ed. Sci.* **31**, 607–626 (1976)

[821] Cammarata, A., and Yau, S. J., *J. Med. Chem.* **13**, 93–97 (1970)

[822] Kubinyi, H., *Arzneim.-Forsch. (Drug Res.)* **27**, 750–758 (1977)

[823] Rösner, T., Franke, R., and Kühne, R., *Pharmazie* **33**, 226–228 (1978)

[824] Martin, Y. C., Jones, P. H., Perun, T. J., Grundy, W. E., Bell, S., Bower, R. R., and Shipkowitz, N. L., *J. Med. Chem.* **15**, 635–638 (1972)

[825] Noel-Artis, A.-M., Berge, G., Fulcrand, P., and Castel, J., *Eur. J. Med. Chem.* **20**, 25–32 (1985)

[826] Tinland, B., *Farmaco, Ed. Sci.* **30**, 935–936 (1975)

[827] Gombar, V., Kapoor, V. K., and Singh, H., *Arzneim.-Forsch. (Drug Res.)* **32**, 7–9 (1982)

[828] Franke, R., Labes, D., Tonew, M., Zschiesche, W., and Heinisch, L., *Acta Biol. Med. Ger.* **34**, 491–499 (1975)

[829] Thomas, J., Berkoff, C. E., Flagg, W. B., Gallo, J. J., Haff, R. F., Pinto, C. A., Pellerano, C., and Savini, L., *J. Med. Chem.* **18**, 245–250 (1975)

[830] Tinland, B., *Res. Commun. Chem. Pathol. Pharmacol.* **8**, 571–574 (1974)

[831] Schönfeld, W., and Repke, K. R. H., *Quant. Struct.-Act. Relat.* **7**, 160–165 (1988); erratum **7**, 250 (1988)

[832] Borea, P. A., *Arzneim.-Forsch. (Drug Res.)* **33**, 1086–1088 (1983)

[833] Borea, P. A., and Ferretti, V., *Biochem. Pharmacol.* **35**, 2836–2839 (1986)

[834] Schauzu, H. G., and Mager, P. P., *Pharmazie* **38**, 490 (1983)

[835] Schaper, K. J., *Eur. J. Med. Chem.* **15**, 449–452 (1980)

[836] Waisser, K., Machacek, M., and Celadnik, M., ref. 76, p. 425–432

[837] Purcell, W. P., *Biochim. Biophys. Acta* **105**, 201–204 (1965)

[838] Beasley, J. G., and Purcell, W. P., *Biochim. Biophys. Acta* **178**, 175–176 (1969)

[839] Clayton, J. M., and Purcell, W. P., *J. Med. Chem.* **12**, 1087–1088 (1969)

[840] Purcell, W. P., and Clayton, J. M., *Molecular Orbital Studies in Chemical Pharmacology*, Kier, L. B., ed., Springer-Verlag, Berlin, 1970, p. 145–155

[841] Lukovits, I., ref. 76, p. 359–364

[842] Schauzu, H. G., and Mager, P. P., *Pharmazie* **38**, 562 (1983)

[843] Justice Jr., J. B., *J. Med. Chem.* **21**, 465–468 (1978)

[844] Pliska, V., *Experientia* **34**, 1190–1192 (1978)

[845] Pliska, V., *Perspectives in Peptide Chemistry*, Eberle, A., Geiger, R., and Wieland, T., eds., Karger, Basel, 1981, p. 221–235

[846] Pliska, V., *J. Steroid Biochem.* **20**, 1512 (1984)

[847] Pliska, V., and Heiniger, J., ref. 80, p. 263–267

[848] Greven, H. M., and de Wied, D., *Frontiers of Hormone Research*, Vol. **4**, van Wimersma Greidanus, T. B., ed., Karger, Basel 1977, p. 140–152

[849] Lukovits, I., *Mol. Pharmacol.* **22**, 725–731 (1982)

[850] Zeelen, F. J., ref. 70, p. 147–160

[851] Nisato, D., Wagnon, J., Callet, G., Mettefeu, D., Assens, J. L., Plouzane, C., Tonnerre, B., and Fauchere, J. L., ref. 80, p. 277–284

[852] Mizuta, E., Suzuki, N., Miyake, Y., Nishikawa, M., and Fujita, T., *Chem. Pharm. Bull.* **23**, 5–12 (1975)

[853] Darvas, F., Lopata, A., Budai, Z., and Petocz, L., ref. 77, p. 199–210

[854] Dillingham, O. E., Mast, R. W., Bass, G. E., and Autian, J., *J. Pharm. Sci.* **62**, 22–30 (1973)

[855] Hall, L. H., Kier, L. B., and Phipps, G., *Environ. Toxicol. Chem.* **3**, 355–365 (1984)

[856] Mager, P. P., Seese, A., Hikino, H., Ohta, T., Ogura, M., Ohizumi, Y., Konno, C., and Takemoto, T., *Pharmazie* **36**, 717 (1981)

[857] Mager, P. P., Seese, A., and Takeya, K., *Pharmazie* **36**, 381–382 (1981)

[858] Meister, A., Tschäpe, M., and Schrötter, E., *Pharmazie* **32**, 174–177 (1977)

[859] Darvas, F., Lopata, A., Budai, Z., and Petöcz, L., ref. 78, p. 324–327

[860] Lehmann, P. A., Rodrigues de Miranda, J. F., and Ariëns, E. J., *Fortschr. Arzneimittelforsch. (Prog. Drug Res.)* **20**, 101–142 (1976)

[861] Ariëns, E. J., *Eur. J. Clin. Pharmacol.* **26**, 663–668 (1984)

[862] Ariëns, E. J., *Trends Pharmacol. Sci.* **7**, 200–205 (1986)

[863] Lehmann F., P. A., *Trends Pharmacol. Sci.* **7**, 281–285 (1986)

[864] Ariëns, E. J., *Med. Res. Rev.* **7**, 367–387 (1987)

[865] Ariëns, E. J., ref. 82, p. 3–6

[866] Ariëns, E. J., *Chirality in Drug Design and Synthesis*, Brown, C., ed., Academic Press, New York, 1990, p. 29–43

[867] Ariëns, E. J., *Quant. Struct.-Act. Relat.* **11**, 190–194 (1992)

[868] Lien, E. J., Rodrigues de Miranda, J. F., and Ariëns, E. J., *Mol. Pharmacol.* **12**, 598–604 (1976)

[869] Pfeiffer, C. C., *Science* **124**, 29–30 (1956)

[870] Barlow, R. B., *Trends Pharmacol. Sci.* **11**, 148–150 (1990)

[871] Gualtieri, F., *Trends Pharmacol. Sci.* **11**, 315–316 (1990)

[872] Gualtieri, F., Romanelli, M. N., and Teodori, E., *Chirality* **2**, 79–84 (1990)

[873] Teodori, E., Romanelli, M. N., and Gualtieri, F., ref. 83, p. 451–456

[874] Lehmann F., P. A., ref. 83, p. 457–460

[875] Schaper, K.-J., ref. 82, p. 41–44

[876] Schaper, K.-J., ref. 83, p. 25–32

[877] Yoshimoto, M., and Hansch, C., *J. Org. Chem.* **41**, 2269–2273 (1976)

[878] Beckett, A. H., and Casy, A. F., *J. Pharm. Pharmacol.* **6**, 986–999 (1954)

[879] Kier, L. B., *Fundamental Concepts in Drug-Receptor Interactions*, Danielli, J. F., Moran, J. F., and Triggle, D. J., eds., Academic Press, New York, 1968, p. 15–45

[880] Gund, P., *Ann. Rep. Med. Chem.* **14**, 299–308 (1979)

[881] Humblet, C., and Marshall, G. R., *Ann. Rep. Med. Chem.* **15**, 267–276 (1980)

[882] Marshall, G. R., and Naylor, C. B., ref. 39, p. 431–458

[883] Gund, P., and Shen, T. Y., *J. Med. Chem.* **20**, 1146–1152 (1977)

[884] Andrea, T. A., Jorgensen, E. C., and Kollman, P. A., *Int. J. Quantum. Chem., Quantum Biol. Symp.* **5**, 191–200 (1978)

[885] Olson, G. L., Cheung, H.-C., Morgan, K. D., Blount, J. F., Todaro, L., Berger, L., Davidson, A. B., and Boff, E., *J. Med. Chem.* **24**, 1026–1034 (1981)

[886] Sufrin, J. R., Dunn, D. A., and Marshall, G. R., *Mol. Pharmacol.* **19**, 307–313 (1981)

[887] Humber, L. G., Bruderlein, F. T., Philipp, A. H., Götz, M., and Voith, K., *J. Med. Chem.* **22**, 761–767 (1986)

[888] Hibert, M. F., Gittos, M. W., Middlemiss, D. N., Mir, A. K., and Fozard, J. R., *J. Med. Chem.* **31**, 1087–1093 (1988)

[889] Ohta, M., and Koga, H., *J. Med. Chem.* **34**, 131–139 (1991)

[890] Höltje, H.-D., *Pharmazie in unserer Zeit* **4**, 109–117 (1975)

[891] Höltje, H.-D., Tintelnot, M., and Baranowski, P., *Proceedings of the 9th International Congress of Pharmacology*, Vol. **2**, Paton, W., Mitchell, J., and Turner, P., eds., The McMillan Press, London, 1984, p. 75–79

[892] Höltje, H.-D., Baranowski, P., Spengler, J.-P., and Schunack, W., *Arch. Pharm.* **318**, 542–548 (1985)

[893] Höltje, H.-D., and Kier, L. B., *J. Pharm. Sci.* **64**, 418–420 (1975); compare Sussman, J. L., Harel, M., Frolow, F., Oefner, C., Goldman, A., Toker, L., and Silman, I., *Science* **253**, 872–879 (1991)

[894] Höltje, H.-D., *Pharm. Acta Helv.* **54**, 125–134 (1979)

[895] Höltje, H.-D., *Arch. Pharm.* **308**, 438–444 (1975)

[896] Höltje, H.-D., *Arch. Pharm.* **309**, 480–485 (1976)

[897] Höltje, H.-D., and Vogelgesang, L., *Arch. Pharm.* **312**, 578–586 (1979)

[898] Höltje, H.-D., and Adler, P., *Arch. Pharm.* **320**, 234–240 (1987)

[899] Höltje, H.-D., *Arch. Pharm.* **315**, 317–323 (1982)

[900] Höltje, H.-D., and Baranowski, P., *Arch. Pharm.* **316**, 154–160 (1983)

[901] Höltje, H.-D., and Marrer, S., *J. Comput.-Aided Mol. Des.* **1**, 23–30 (1987)

[902] Höltje, H.-D., and Marrer, S., *Quant. Struct.-Act. Relat.* **7**, 174–178 (1988)

[903] Höltje, H.-D., ref. 82, p. 237–241

[904] Höltje, H.-D., *Quant. Struct.-Act. Relat.* **11**, 224–227 (1992)

[905] Höltje, H.-D., and Briem, H., ref. 83, p. 245–252

[906] Höltje, H.-D., and Simon, H., *Arch. Pharm.* **317**, 506–516 (1984)

[907] Höltje, H.-D., and Zunker, P., *J. Mol. Struct. (Theochem)* **134**, 429–436 (1986)

[908] Höltje, H.-D., ref. 64, p. 181–189

[909] a) Boobbyer, D. N. A., Goodford, P. J., McWhinnie, P. M., and Wade, R. C., *J. Med. Chem.* **32**, 1083–1094 (1989); b) Wade, R. C., Clark, K. J., and Goodford, P. J., *J. Med. Chem.* **36**, 140–147 (1993); c) Wade, R. C., and Goodford, P. J., *J. Med. Chem.* **36**, 148–156 (1993)

[910] GRID Program, V7, Molecular Discovery Ltd., West Way House, Elms Parade, Oxford, England

[911] Böhm, H.-J., *J. Comput.-Aided Mol. Des.* **6**, 61–78; 593–606 (1992)

[912] Böhm, H.-J., ref. 84, p. 27–34

[913] LUDI, BIOSYM Technologies, Inc., San Diego, CA, U. S. A.

[914] Audry, E., Dubost, J. P., Colleter, J. C., and Dallet, P., *Eur. J. Med. Chem.* **21**, 71–72 (1986)

[915] Croizet, F., Langlois, M. H., Dubost, J. P., Braquet, P., Audry, E., Dallet, P., and Colleter, J. C., *J. Mol. Graphics* **8**, 153–155 (1990)

[916] Croizet, F., Dubost, J. P., Langlois, M. H., and Audry, E., *Quant. Struct.-Act. Relat.* **10**, 211–215 (1991)

[917] Abraham, D. J., and Leo, A. J., *Proteins: Struct. Funct. Genetics* **2**, 130–152 (1987)

[918] Wireko, F. C., Kellog, G. E., and Abraham, D. J., *J. Med. Chem.* **34**, 758–767 (1991)

[919] Kellog, G. E., Semus, S. F., and Abraham, D. J., *J. Comput.-Aided Mol. Des.* **5**, 545–552 (1991)

[920] Kellog, G. E., Joshi, G. S., and Abraham, D. J., *Med. Chem. Res.* **1**, 444–453 (1992)

[921] Kellogg, G. E., and Abraham, D. J., *J. Mol. Graphics* **10**, 212–217 (1992)

[922] Crippen, G. M., *J. Comput. Phys.* **24**, 96–107 (1977)

[923] Ghose, A. K., and Crippen, G. M., ref. 78, p. 116–119

[924] Donné-Op den Kelder, G. M., *J. Comput.-Aided Mol. Des.* **1**, 257–264 (1987)

[925] Donné-Op den Kelder, G. M., *Trends Pharmacol. Sci.* **9**, 391–393 (1988)

[926] Ghose, A. K., and Crippen, G. M., ref. 39, p. 715–733

[927] Crippen, G. M., *J. Med. Chem.* **23**, 599–606 (1980)

[928] Ghose, A. K., and Crippen, G. M., *J. Med. Chem.* **26**, 996–1010 (1983)

[929] Ghose, A. K., and Crippen, G. M., ref. 75, p. 99–108

[930] Ghose, A. K., and Crippen, G. M., *J. Med. Chem.* **27**, 901–914 (1984)

[931] Ghose, A. K., and Crippen, G. M., *J. Med. Chem.* **28**, 333–346 (1985)

[932] Bultsma, T., and Linschoten, M. R., ref. 78, p. 256–259

[933] Linschoten, M. R., Bultsma, T., IJzerman, A. P., and Timmerman, H., *J. Med. Chem.* **29**, 278–286 (1986)

[934] Donné-Op den Kelder, G. M., Bultsma, T., Timmerman, H., and Rademaker, B., *J. Med. Chem.* **31**, 1069–1079 (1988)

[935] Belvisi, L., Brossa, S., Salimbeni, A., Scolastico, C., and Todeschini, R., ref. 83, p. 423–426

[936] Belvisi, L., Brossa, S., Salimbeni, A., Scolastico, C., and Todeschini, R., *J. Comput.-Aided Mol. Des.* **5**, 571–584 (1991)

[937] Crippen, G. M., *J. Comput. Chem.* **8**, 943–955 (1987)

[938] Boulu, L. G., and Crippen, G. M., *J. Comput. Chem.* **10**, 673–682 (1989)

[939] Boulu, L. G., Crippen, G. M., Barton, H. A., Kwon, H., and Marletta, M. A., *J. Med. Chem.* **33**, 771–775 (1990)

[940] Ghose, A. K., Crippen, G. M., Revankar, G. R., McKernan, P. A., Smee, D. F., and Robins, R. K., *J. Med. Chem.* **32**, 746–756 (1989)

[941] Ghose, A. K., and Crippen, G. M., *Mol. Pharmacol.* **37**, 725–734 (1990)

[942] Viswanadhan, V. N., Weinstein, J. N., and Ghose, A. K., *Biophys. J.* **57**, 208a (1990)

[943] Viswanadhan, V. N., Ghose, A. K., and Weinstein, J. N., *Biochem. Biophys. Acta* **1039**, 356–366 (1990)

[944] Viswanadhan, V. N., Ghose, A. K., Hanna, N. B., Matsumoto, S. S., Avery, T. L., Revankar, G. R., and Robins, R. K., *J. Med. Chem.* **34**, 526–532 (1991)

[945] Doweyko, A. M., *J. Med. Chem.* **31**, 1396–1406 (1988)

[946] Wiese, M., and Coats, E. A., ref. 83, p. 343–346

[947] Doweyko, A. M., and Mattes, W. B., *Biochemistry* **31**, 9388–9392 (1992)

[948] Bersuker, I. B., and Dimoglo, A. S., *Reviews in Computational Chemistry* **II**, Lipkowitz, K. B., and Boyd, D. B., eds., VCH Publishers, New York, 1991, p. 423–460

[949] Golender, V. E., and Rozenblit, A. B., *Drug Design*, Vol. **IX**, Ariëns, E. J., ed., Academic Press, New York, 1980, p. 299–337

[950] Golender, V. E., and Rozenblit, A. B., Logical and Combinatorial Algorithms for Drug Design, Research Studies Press, Letchworth, 1983

[951] Vorpagel, E., Herd, C., and Golender, V., personal communication

[952] Apex-3D, BIOSYM Technologies, Inc., San Diego, CA, USA

[953] Cramer III, R. D., and Milne, M., Abstracts of Papers, Am. Chem. Soc., April 1979, Computer Chemistry Section, no. 44

[954] Wise, M., Cramer, R. D., Smith, D., and Exman, I., ref. 75, p. 145–146

[955] Wise, M., *Molecular Graphics and Drug Design* (Top. Mol. Pharmacol. **3**), Burgen, A. S. V., Roberts, G. C. K., and Tute, M. S., eds., Elsevier, New York, 1986, p. 183–194

[956] SYBYL/QSAR, Version 5.5, February 1992; TRIPOS Associates, Inc., St. Louis, MO, USA

[957] Marshall, G. R., and Cramer III, R. D., *Trends Pharmacol. Sci.* **9**, 285–289 (1988)

[958] Cramer III, R. D., Patterson, D. E., and Bunce, J. D., ref. 82, p. 161–165

[959] Pearlman, R. S., *Chem. Design Automation News* **2** (1), 1–7 (1987); **8** (8), 3–15 (1993)

[960] Rusinko III, A., Skell, J. M., Balducci, R., McGarity, C. M., and Pearlman, R. S., CONCORD, Tripos Associates, Inc., St. Louis, MO, USA

[961] Wipke, W. T., and Hahn, M. A., *Tetrahedron Comp. Methodology* **1**, 141–167 (1988)

[962] Dolata, D. P., and Carter, R. E., *J. Chem. Inf. Comput. Sci.* **27**, 36–47 (1987)

[963] Dolata, D. P., Leach, A. R., and Prout, K., *J. Comput.-Aided Mol. Des.* **1**, 73–85 (1987)

[964] Leach, A. R., Prout, K., and Dolata, D. P., *J. Comput.-Aided Mol. Des.* **2**, 107–123 (1988); **4**, 271–282 (1990)

[965] Leach, A. R., Prout, K., and Dolata, D. P., *J. Comput. Chem.* **11**, 680–693 (1990)

[966] Leach, A. R., and Prout, K., *J. Comput. Chem.* **11**, 1193–1205 (1990)

[967] a) Gasteiger, J., Rudolph, C., and Sadowski, J., *Tetrahedron Comp. Methodology* **3**, 537–547 (1990); b) Sadowski, J., Rudolph, C., and Gasteiger, J., *Anal. Chim. Acta* **265**, 233–241 (1992); c) Sadowski, J., and Gasteiger, J., Chem. Rev., in print

[968] Klebe, G., and Mietzner, T., *Organic Crystal Chemistry*, Jones, D. W., ed., Oxford University Press, Oxford, in print

[969] Leach, A. R., *Reviews in Computational Chemistry* **II**, Lipkowitz, K. B., and Boyd, D. B., eds., VCH Publishers, New York, 1991, p. 1–55

[970] Dean, P. M., ref. 63, p. 211–238

[971] Martin, Y. C., Bures, M. G., and Willett, P., *Reviews in Computational Chemistry*, Lipkowitz, K. B., and Boyd, D. B., eds., VCH Publishers, New York, 1990, p. 213–263

[972] Reynolds, C. A., Burt, C., and Richards, W. G., *Quant. Struct.-Act. Relat.* **11**, 34–35 (1992)

[973] Good, A. C., Hodgkin, E. E., and Richards, W. G., *J. Comput.-Aided Mol. Des.* **6**, 513–520 (1992)

[974] Marshall, G. R., Barry, C. D., Bosshard, A. E., Dammkoehler, R. A., and Dunn, D. A., ref. 72, p. 205–226

[975] Humblet, C., and Marshall, G. R., *Drug. Dev. Res.* **1**, 409–434 (1981)

[976] Mayer, D., Naylor, C. B., Motoc, I., and Marshall, G. R., *J. Comput.-Aided Mol. Des.* **1**, 3–16 (1987)

[977] Dammkoehler, R. A., Karasek, S. F., Shands, E. F. B., and Marshall, G. R., *J. Comput.-Aided Mol. Des.* **3**, 3–21 (1989)

[978] Brint, A. T., and Willett, P., *J. Chem. Inf. Comput. Sci.* **27**, 152–158 (1987)

[979] Namasivayam, S., and Dean, P. M., *J. Mol. Graphics* **4**, 46–50 (1986)

[980] Dean, P. M., and Chau, P.-L., *J. Mol. Graphics* **5**, 152–158 (1987)

[981] Dean, P. M., Callow, P., and Chau, P.-L., *J. Mol. Graphics* **6**, 28–34 (1988)

[982] Kato, Y., Itai, A., and Iitaka, Y., *Tetrahedron* **43**, 5229–5236 (1987)

[983] Itai, A., Tomioka, N., Kato, Y., Nishibata, Y., and Saito, S., ref. 64, p. 191–212

[984] a) Manaut, F., Sanz, F., José, J., and Milesi, M., *J. Comput.-Aided Mol. Des.* **5**, 371–380 (1991); b) Sanz, F., Manaut, F., Rodríguez, J., Lozoya, E., and Lopez-de-Brinãs, E., *J. Comput.-Aided Mol. Des.*, **7**, 337–347 (1993)

[985] Hermann, R. B., and Herron, D. K., *J. Comput.-Aided Mol. Des.* **5**, 511–524 (1991)

[986] Van Drie, J. H., Weininger, D., and Martin, Y. C., *J. Comput.-Aided Mol. Des.* **3**, 225–251 (1989)

[987] Martin, Y. C., Bures, M. G., Danaher, E. A., DeLazzer, J., Lico, I., and Pavlik, P. A., *J. Comput.-Aided Mol. Des.* **7**, 83–102 (1993)

[988] CATALYST, BioCAD Corporation, Mountain View, CA, USA

[989] Imoto, T., Johnson, L. N., North, A. C. T., Phillips, D. C., and Rupley, J. A., *The Enzymes*, 3rd Edition, Vol. **VII**, Boyer, P. D., ed., Academic Press, New York, 1972, p. 665–868

[990] Kester, W. R., and Matthews, B. W., *Biochemistry* **16**, 2506–2516 (1977)

[991] Bolin, J. T., Filman, D. J., Matthews, D. A., Hamlin, R. C., and Kraut, J., *J. Biol. Chem.* **257**, 13650–13662 (1982)

[992] Bystroff, C., Oatley, S. J., and Kraut, J., *Biochemistry* **29**, 3263–3277 (1990)

[993] Smith, T. J., Kremer, M. J., Luo, M., Vriend, G., Arnold, E., Kamer, G., Rossmann, M. G., McKinlay, M. A., Diana, G. D., and Otto, M. J., *Science* **233**, 1286–1293 (1986)

[994] Badger, J., Minor, I., Kremer, M. J., Oliveira, M. A., Smith, T. J., Griffith, J. P., Guerin, D. M. A., Krishnaswamy, S., Luo, M., Rossmann, M. G., McKinlay, M. A., Diana, G. D., Dutko, F. J., Fancher, M., Rueckert, R. R., and Heinz, B. A., *Proc. Natl. Acad. Sci. USA* **85**, 3304–3308 (1988)

[995] Diana, G. D., Treasurywala, A. M., Bailey, T. R., Oglesby, R. C., Pevear, D. C., and Dutko, F. J., *J. Med. Chem.* **33**, 1306–1311 (1990)

[996] Poulos, T. L., and Howard, A. J., *Biochemistry* **26**, 8165–8174 (1987)

[997] Cody, V., and Sutton, P. A., ref. 82, p. 247–250

[998] Baldwin, J. J., Ponticello, G. S., Anderson, P. S., Christy, M. E., Murcko, M. A., Randall, W. C., Schwam, H., Sugrue, M. F., Springer, J. P., Gautheron, P., Grove, J., Mallorga, P., Viader, M.-P., McKeever, B. M., and Navia, M. A., *J. Med. Chem.* **32**, 2510–2513 (1989)

[999] a) Ealick, S. E., Babu, Y. S., Bugg, C. E., Erion, M. D., Guida, W. C., Montgomery, J. A., and Secrist III, J. A., *Proc. Natl. Acad. Sci. USA* **88**, 11540–11544 (1991); b) Montgomery, J. A., Niwas, S., Rose, J. D., Secrist III, J. A., Babu, Y. S., Bugg, C. E., Erion, M. D., Guida, W. C., and Ealick, S. E., *J. Med.* Chem. **36**, 55–69 (1993)

[1000] Banner, D. W., and Hadváry, P., *J. Biol. Chem.* **266**, 20085–20093 (1991)

[1001] Brandstetter, H., Turk, D., Hoeffken, H. W., Grosse, D., Stürzebecher, J., Martin, P. D., Edwards, B. F. P., and Bode, W., *J. Mol. Biol.* **226**, 1085–1099 (1992)

[1002] Wojtczak, A., Luft, J., and Cody, V., *J. Biol. Chem.* **267**, 353–357 (1992)

[1003] Roderick, S. L., Fournie-Zaluski, M. C., Roques, B. P., and Matthews, B. W., *Biochemistry* **28**, 1493–1497 (1989)

[1004] Simon, Z., *Rev. Roum. Chim.* **37**, 323–325 (1992)

[1005] Clark, M., Cramer III, R. D., Jones, D. M., Patterson, D. E., and Simeroth, P. E., *Tetrahedron Comp. Methodology* **3**, 47–59 (1990)
[1006] Cramer III, R. D., Clark, M., Simeroth, P., and Patterson, D. E., ref. 83, p. 239–242
[1007] Klebe, G., and Abraham, U., *J. Med. Chem.* **36**, 70–80 (1993)
[1008] Cocchi, M., Cruciani, G., Menziani, M. C., and De Benedetti, P. G., ref. 84, p. 527–529
[1009] Clark, M., and Cramer III, R. D., *Quant. Struct.-Act. Relat.*, **12**, 137–145 (1993)
[1010] Norinder, U., *J. Comput.-Aided Mol. Des.* **4**, 381–389 (1990)
[1011] Norinder, U., *J. Comput.-Aided Mol. Des.* **5**, 419–426 (1991)
[1012] TRIPOS Associates, Inc., St. Louis, MO, USA, personal communication
[1013] Kubinyi, H., unpublished results
[1014] Floersheim, P., Nozulak, J., and Weber, H. P., ref. 84, p. 227–232
[1015] Kim, K. H., and Martin, Y. C., *J. Org. Chem.* **56**, 2723–2729 (1991)
[1016] Kim, K. H., and Martin, Y. C., ref. 83, p. 151–154
[1017] Kim, K. H., and Martin, Y. C., *J. Med. Chem.* **34**, 2056–2060 (1991)
[1018] Van de Waterbeemd, H., Carrupt, P. A., Testa, B., and Kier, L. B., ref. 84, p. 69–75
[1019] Kim, K. H., *Med. Chem. Res.* **1**, 259–264 (1991)
[1020] Kim, K. H., *Quant. Struct.-Act. Relat.* **11**, 127–134 (1992)
[1021] Martin, Y. C., Kim, K.-H., and Bures, M. G., ref. 64, p. 295–317
[1022] Kim, K. H., *Quant. Struct.-Act. Relat.* **11**, 453–460 (1992)
[1023] Kim, K. H., ref. 84, p. 245–251
[1024] Oprea, T. I., Ciubotariu, D., Sulea, T. I., and Simon, Z., *Quant. Struct.-Act. Relat.* **12**, 21–26 (1993)
[1025] Kim, K. H., *Med. Chem. Res.* **2**, 22–27 (1992)
[1026] Kim, K. H., *Quant. Struct.-Act. Relat.* **11**, 309–317 (1992)
[1027] Kim, K. H., *J. Comput.-Aided Mol. Des.* **7**, 71–82 (1993)
[1028] Marshall, G. R., Mayer, D., Naylor, C. B., Hodgkin, E. E., and Cramer III, R. D., ref. 82, p. 287–295
[1029] Marshall, G. R., *Eur. J. Pharmacol.* **183**, 15 (1990)
[1030] DePriest, S. A., Shands, E. F. B., Dammkoehler, R. A., and Marshall, G. R., ref. 83, p. 405–414
[1031] Langer, T., and Wermuth, C. G., *J. Comput.-Aided Mol. Des.* **7**, 253–262 (1993)
[1032] Nicklaus, M. C., Milne, G. W. A., and Burke Jr., T. R., *J. Comput.-Aided Mol. Des.* **6**, 487–504 (1992)
[1033] Maret, G., El Tayar, N., Carrupt, P.-A., Testa, B., Jenner, P., and Baird, M., *Biochem. Pharmacol.* **40**, 783–792 (1990)
[1034] Altomare, C., Carrupt, P. A., Gaillard, P., El Tayar, N., Testa, B., and Carotti, A., *Chem. Res. Toxicol.* **5**, 366–375 (1992)
[1035] Allen, M. S., Tan, Y.-C, Trudell, M. L., Narayanan, K., Schindler, L. R., Martin, M. J., Schultz, C., Hagen, T. J., Koehler, K. F., Codding, P. W., Skolnick, P., and Cook, J. M., *J. Med. Chem.* **33**, 2343–2357 (1990)
[1036] Diaz-Arauzo, H., Koehler, K. F., Hagen, T. J., and Cook, J. M., *Life Sci.* **49**, 207–216 (1991)
[1037] Allen, M. S., LaLoggia, A. J., Dorn, L. J., Martin, M. J., Costantino, G., Hagen, T. J., Koehler, K. F., Skolnick, P., and Cook, J. M., *J. Med. Chem.* **35**, 4001–4010 (1992)
[1038] Greco, G., Novellino, E., Silipo, C., and Vittoria, A., *Quant. Struct.-Act. Relat.* **11**, 461–477 (1992)
[1039] Thomas, B. F., Compton, D. R., Martin, B. R., and Semus, S. F., *Mol. Pharmacol.* **40**, 656–665 (1991)
[1040] Rault, S., Bureau, R., Pilo, J. C., and Robba, M., *J. Comput.-Aided Mol. Des.* **6**, 553–568 (1992)
[1041] Carroll, F. I., Gao, Y., Rahman, M. A., Abraham, P., Parham, K., Lewin, A. H., Boja, J. W., and Kuhar, M. J., *J. Med. Chem.* **34**, 2719–2725 (1991)
[1042] Liegeois, J. F., Dupont, L., and Delarge, J., *J. Pharm. Belg.* **47**, 100–108 (1992)
[1043] Norinder, U., and Högberg, T., *Quant. Struct.-Act. Relat.* **10**, 1–5 (1991); erratum **10**, 141 (1991)
[1044] Greco, G., Novellino, E., Silipo, C., and Vittoria, A., ref. 83, p. 293–296

[1045] Greco, G., Novellino, E., Silipo, C., and Vittoria, A., *Quant. Struct.-Act. Relat.* **10**, 289–299 (1991)

[1046] Calder, J. A., Wyatt, J. A., Frenkel, D. A., and Casida, J. E., *J. Comput.-Aided Mol. Des.* **7**, 45–60 (1993)

[1047] Naerum, L., and Jensen, J. S., ref. 83, p. 489–492

[1048] Ablordeppey, S. Y., El-Ashmawy, M. B., and Glennon, R. A., *Med. Chem. Res.* **1**, 425–438 (1991)

[1049] Westkaemper, R. B., and Glennon, R. A., *Pharmacology, Biochemistry and Behaviour* **40**, 1019–1031 (1991)

[1050] El-Bermawy, M. A., Lotter, H., and Glennon, R. A., *Med. Chem. Res.* **2**, 290–297 (1992)

[1051] Agarwal, A., and Taylor, E. W., *J. Comput. Chem.* **14**, 237–245 (1993)

[1052] Waller, C. L., and McKinney, J. D., *J. Med. Chem.* **35**, 3660–3666 (1992)

[1053] Loughney, D. A., and Schwender, C. F., *J. Comput.-Aided Mol. Des.* **6**, 569–581 (1992)

[1054] Diana, G. D., Kowalczyk, P., Treasurywala, A. M., Oglesby, R. C., Pevear, D. C., and Dutko, F. J., *J. Med. Chem.* **35**, 1002–1008 (1992)

[1055] Rognan, D., Reddehase, M. J., Koszinowski, U. H., and Folkers, G., *Proteins* **13**, 70–85 (1992)

[1056] Folkers, G., Merz, A., and Rognan, D., ref. 84, p. 233–244

[1057] Harpalani, A. D., Egorin, M. J., and Callery, P. S., Abstracts of Papers, Am. Chem. Soc., August 1992, Medicinal Chemistry Section, no. 124

[1058] Wiese, T. E., Palomino, E., Horwitz, J. P., and Brooks, S. C., Abstracts of Papers, Am. Chem. Soc., April 1991, Medicinal Chemistry Section, no. 150

[1059] Thibaut, U., Folkers, G., and Roth, H. J., ref. 83, p. 431–434

[1060] Koehler, K. F., Gasiecki, A. F., Kramer, S. W., Shone, R. L., Collins, P. W., Tsai, H., and Bianchi, R., Abstracts of Papers, Am. Chem. Soc., August 1990, Medicinal Chemistry Section, no. 75

[1061] Björkroth, J.-P., and Pohjala, E., ref. 83, p. 545–548

[1062] Björkroth, J.-P., Pakkanen, T. A., Lindroos, J., Pohjala, E., Hanhijärvi, H., Laurén, L., Hannuniemi, R., Juhakoski, A., Kippo, K., and Kleimola, T., *J. Med. Chem.* **34**, 2338–2343 (1991)

[1063] McFarland, J. W., *J. Med. Chem.* **35**, 2543–2550 (1992)

[1064] Good, A. C., So, S.-S., and Richards, W. G., *J. Med. Chem.* **36**, 433–438 (1993)

[1065] Good, A. C., Peterson, S. J., and Richards, W. G., *J. Med. Chem.*, submitted for publication

[1066] Carbo, R., Leyda, L., and Arnau, M., *Int. J. Quantum Chem.* **17**, 1185–1189 (1980)

[1067] Hodgkin, E. E., and Richards, W. G., *Int. J. Quantum Chem., Quantum Biol. Symp.* **14**, 105–110 (1987)

[1068] Meyer, A. Y., and Richards, W. G., *J. Comput.-Aided Mol. Des.* **5**, 427–439 (1991)

[1069] Richards, W. G., ref. 84, p. 203–206

[1070] Burt, C., Richards, W. G., and Huxley, P., *J. Comput. Chem.* **11**, 1139–1146 (1990)

[1071] Automated Similarity Package, Oxford Molecular Ltd., Oxford, U. K.

[1072] Good, A. C., Hodgkin, E. E., and Richards, W. G., *J. Chem. Inf. Comput. Sci.* **32**, 188–191 (1992)

[1073] Good, A. C., and Richards, W. G., *J. Chem. Inf. Comput. Sci.* **33**, 112–116 (1993)

[1074] Good, A. C., *J. Mol. Graphics* **10**, 144–151 (1992)

[1075] Burt, C., and Richards, W. G., *J. Comput.-Aided Mol. Des.* **4**, 231–238 (1990)

[1076] Cohen, N. C., Blaney, J. M., Humblet, C., Gund, P., and Barry, D. C., *J. Med. Chem.* **33**, 883–894 (1990)

[1077] a) Borman, S., *C & EN* (August 10, 1992), 18–26; b) Dixon, J. S., *Trends Biotechnol.* **10**, 357–363 (1992); c) Richards, W. G., *Proc. Royal Soc. Edinburgh* **99B** (1/2), 105–111 (1992); d) Beddell, C. R., ed., *The Design of Drugs to Macromolecular Targets*, Wiley, Chichester, 1992

[1078] a) Hol, W. G. J., *Angew. Chem. Int. Ed. Engl.* **25**, 767–778 (1986); b) Verlinde, C. L. M. J., Rudenko, G., and Hol, W. G. J., *J. Comput.-Aided Mol. Des.* **6**, 131–147 (1992)

[1079] DesJarlais, R. L., Sheridan, R. P., Seibel, G. L., Dixon, J. S., Kuntz, I. D., and Venkataraghavan, R., *J. Med. Chem.* **31**, 722–729 (1988)

[1080] Bartlett, P. A., Shea, G. T., Telfer, S. J., and Waterman, S., *Molecular Recognition: Chemical and Biological Problems*, Roberts, S. M., ed., Royal Society of London, London, 1989, p. 182–196

[1081] Lauri, G., Shea, G. T., Waterman, S., Telfer, S. J., and Bartlett, P. A., *CAVEAT: A Program to Facilitate the Design of Organic Molecules, Principles of Operation, Version 2.1*, Department of Chemistry, University of California, Berkeley, 1992

[1082] Moon, J. B., and Howe, W. J., *Tetrahedron Comp. Methodology* **3**, 681–696 (1990)

[1083] Appelt, K., Bacquet, R. J., Bartlett, C. A., Booth, C. L. J., Freer, S. T., Fuhry, M. A. M., Gehring, M. R., Herrmann, S. M., Howland, E. F., Janson, C. A., Jones, T. R., Kan, C.-C., Kathardekar, V., Lewis, K. K., Marzoni, G. P., Matthews, D. A., Mohr, C., Moomaw, E. W., Morse, C. A., Oatley, S. J., Ogden, R. C., Reddi, M. R., Reich, S. H., Schoettlin, W. S., Smith, W. W., Varney, M. D., Villafranca, J. E., Ward, R. W., Webber, S., Webber, S. E., Welsh, K. M., and White, J., *J. Med. Chem.* **34**, 1925–1934 (1991)

[1084] Varney, M. D., Marzoni, G. P., Palmer, C. L., Deal, J. G., Webber, S., Welsh, K. M., Bacquet, R. J., Bartlett, C. A., Morse, C. A., Booth, C. L. J., Herrmann, S. M., Howland, E. F., Ward, R. W., and White, J., *J. Med. Chem.* **35**, 663–676 (1992)

[1085] Lawrence, M. C., and Davis, P. C., *Proteins* **12**, 31–41 (1992)

[1086] Lewis, R. A., Roe, D. C., Huang, C., Ferrin, T. E., Langridge, R., and Kuntz, I. D., *J. Mol. Graphics* **10**, 66–78 (1992)

[1087] a) Kuntz, I. D., Science **257**, 1078–1082 (1992); b) Shoichet, B. K., Stroud, R. M., Santi, D. V., Kuntz, I. D., and Perry, K. M., *Science* **259**, 1445–1450 (1993)

[1088] Shoichet, B. K., Stroud, R. M., Santi, D. V., Kuntz, I. D., and Perry, K. M., *Science* **259**, 1445–1450 (1993)

[1089] Navia, M. A., and Murcko, M. A., *Current Opinion in Structural Biology* **2**, 202–210 (1992)

[1090] Rotstein, S. H., and Murcko, M. A., *J. Comput.-Aided Mol. Des.* **7**, 23–43 (1993)

[1091] Nishibata, Y., and Itai, A., *Tetrahedron* **47**, 8985–8990 (1991)

[1092] Yamada M., Tomioka, N., and Itai, A., ref. 84, p. 404–406

[1093] Hansch, C., *J. Med. Chem.* **19**, 1–6 (1976)

[1094] Hansch, C., ref. 70, p. 47–61

[1095] Hansch, C., *Drug Dev. Res.* **1**, 267–309 (1981)

[1096] Martin, Y. C., *J. Med. Chem.* **24**, 229–237 (1981)

[1097] Fujita, T., ref. 46, p. 19–33

[1098] Hopfinger, A. J., *J. Med. Chem.* **28**, 1133–1139 (1985)

[1099] Boyd, D. B., *Reviews in Computational Chemistry*, Lipkowitz, K. B., and Boyd, D. B., eds., VCH Publishers, New York, 1990, p. 355–371

[1100] Fujita, T., *QSAR in Design of Bioactive Compounds*, Kuchar, M., ed., Prous Science Publishers, Barcelona, 1992, p. 3–22

[1101] Watanabe, H., Miyamoto, S., Yoshimoto, M., Kamioka, T., Nakayama, I., Kobayashi, T., and Honda, T., *Chem. Pharm. Bull.* **35**, 1452–1459 (1987)

[1102] Ohtaka, H., Kanazawa, T., Ito, K., and Tsukamoto, G., *Chem. Pharm. Bull.* **35**, 3270–3275 (1987)

[1103] Koga, H., *Structure-Activity Relationships—Quantitative Approaches: Applications to Drug Design and Mode-of-Action Studies*, Fujita, T., ed., Nankodo, Tokyo, 1982, p. 177–202

[1104] Ohtaka, H., Yoshida, K., Suzuki, K., Shimohara, K., Tajima, S., and Ito, K., *Chem. Pharm. Bull.* **36**, 4825–4833 (1988)

[1105] a) Hansch, C., ref. 64, p. 281–293; b) Hansch, C., *Acc. Chem. Res.* **26**, 147–153 (1993)

[1106] Wold, S., Berntsson, P., Eriksson, L., Geladi, G., Hellberg, S., Johansson, E., Jonsson, J., Kettaneh-Wold, N., Lindgren, F., Rännar, S., Sandberg, M., and Sjöström, M., ref. 83, p. 15–24

[1107] Martin, Y. C., *Methods Enzymol.* **203**, 587–613 (1991)

[1108] Marshall, G. R., ref. 64, p. 163–178

[1109] Mannhold, R., *Quant. Struct.-Act. Relat.* **11**, 232–236 (1992)

[1110] Kubinyi, H., *Burger's Medicinal Chemistry and Drug Discovery*, Vol. **I**, 5th Edition, Wolff, M. E., ed., Wiley-Interscience, New York, in print

[1111] Van de Waterbeemd, H., ed., *QSAR: Chemometric Methods in Molecular Design* (Methods and Principles in Medicinal Chemistry, Vol. **2**), VCH, Weinheim, in print

QSAR: Hansch Analysis and Related Approaches
by Hugo Kubinyi
© VCH Verlagsgesellschaft mbH, 1993

Index

2D/3D conversion procedures 160
2D NMR 157, 160
3D database searches 36, 161
3D QSAR methods 6, 106, 149*ff*, 181
- applications 106, 156, 169*ff*
- history 7, 159
- similarity matrices 172*ff*
3D similarity indices 172*ff*
3D structures
- DHFR complexes 116
- enzymes 181
- enzyme inhibitor complexes 8, 116, 161*ff*, 181
- from crystallographic analyses 160
- from 2D NMR studies 160
- generation, from 2D structures 160
- ligand-protein complexes 161*ff*
- papain 119
- prediction 160

α-adrenergic
- antagonists, QSAR 123
- binding, CoMFA 171
- clonidine analogs 134
α adrenoceptors
- binding affinities, QSAR 134, 147
α antagonism, labetalol 149, 150
α-glucosidase 7
ab initio methods 47, 160
absorption
- models 81, 84
- pH dependence 82
- profiles, pH shift 82
- QSAR 129, 176
- rate constants 15, 57, 126, 128
- simulator®, Sartorius 126, 127
ACE, CoMFA 171
acetic acid methyl esters, CoMFA 169
acetylglucosamines, multiple binding modes 161
acetylcholine
- analogs 155
- log P values 31
- receptor
  - affinity, QSAR 122
  - *antagonists, enantiomers* 150
  - binding, QSAR 143*ff*
acetylcholinesterase
- affinities 155
- inhibition, QSAR 147
- substrates 152
acids
- buccal absorption 127
- dissociation and partitioning 77*ff*
ACTH-derived peptides, QSAR 145
actinidin inhibition
- comparative QSAR 180
- QSAR 120
active analog approach 161
active analog mode, LUDI 156
active conformations 157
active site 115
- binding analysis 146
- DHFR 118
- interaction models 151
active site-directed enzyme inhibitors 10
activities, *in vivo* 133
activity-activity relationships 115, 133*ff*
- nonlinear 135
activity classes
- separation 100, 101
- SIMCA 107
- multidimensional 137

activity profiles 115
- of drugs 137
adaptive least squares, see ALS method
additivity
- binding energies 19
- group contributions 2, 4, 15*ff*, 63
- model 62
adjunction (inclusive or) 88
ADP-ribosyltransferase, QSAR 120
affinity 1*ff*
- binding site 2
- constants 17
- data 15
- receptor site models 153
agonists 20
AKUFVE method 32
ALADDIN program, alignment 161
albumin binding 121
alcohol dehydrogenase, QSAR 120
alcohols
- log P values 29
- MAO inhibition, QSAR 82
- QSAR 83, 145
alignment 6, 161, 166
- ALADDIN program 161
- by similarity indices 173
- computer-assisted 161
- in CoMFA 166
- of enantiomers 151
- of molecules 161
- procedures, CoMFA 172
aliphatic
- σ constants (σ^* values) 45
- substituent constants 22
alkane/water f values 34
alkanoic acids
- buccal absorption 126
- hemolytic activities 92
alkyl carbamates, absorption 128
alkylammonium compounds
- transport 72, 74
- CoMFA 170
alkylpyridines, partitioning 28
allosteric
- effects 1, 20, 68
- modifiers, hemoglobin 157
ALS method 100
- fuzzy 100
ambiguity of results, Hansch analyses 61
Ames test 46
amines, QSAR 49, 74, 82, 83, 176
amino acids
- pH partition profile 79
- principal component analysis 26
- side-chain properties 26
- z scales 26, 113
amino probe, GRID potentials 155
aminodiphenyl sulfones, QSAR 135
ammonium compounds 4
- CoMFA 170
- QSAR 122, 143, 144
amphiphilic drugs 31, 125
amphiprotic regions, of membranes 30
analgesics, QSAR 140, 147
anchor principle 19
angiotensin converting enzyme inhibition, CoMFA 171
aniline mustards, antitumor activities, 132
anilinoacridines
- antitumor activities 132, 133
- CASE program 145
anilinopyrimidines, QSAR 179
animal data 95, 115
animal models 133
- and *in vitro* systems 180
antagonists 20
antiadrenergic activities, QSAR 58*ff*, 63*ff*, 130
antiallergics, QSAR 179
antibacterial activities
- CoMFA 170
- QSAR 123, 141, 145, 147
- specific 124
- sulfonamides 45
- sulfa drugs 124
anticholinergics, QSAR 86, 147
anticoccidial activities, triazines
- CoMFA 172
- similarity matrices 174, 175
antidepressants, QSAR 179

antifungal activities, QSAR 49, 66, 74, 123, 176
antigenic complexes, CoMFA 171
antihistaminics 123
– blood-brain barrier 30, 128
– sedative side effects 129
antihypertensive activities
– QSAR 134, 179, 180
– selectivity 135
antileukemic activities, QSAR 132
antilipemic activities, QSAR 147
antimalarial activities, QSAR 88, 130, 139
antitumor drugs 10, 158
– QSAR 62, 132*ff*, 145
– tumor cell resistance 125
antiviral activities
– QSAR 147
– ribonucleosides 158
Apex-3D expert system 158
apparent partition coefficients, P_{app} 77*ff*
applications of
– CoMFA 169*ff*
– Free Wilson analysis 63*ff*, 139*ff*, 147
– Hansch analysis 58*ff*, 115*ff*
aqueous diffusion layer 82
aqueous/organic interface 70, 82
aqueous solvation energy 22
aromatic substituent parameters 21, 23*ff*
aromaticities 47
artificial bilayers 31
artificial intelligence 86
aryldialkyltriazenes, QSAR 132
ascites tumor cell inhibition, QSAR 41, 42
ASP program, similarity 173
association constants 15
asymmetric centers 149
atom-based
– lipophilicity contributions 156
– molar refractivity contributions 156
atom-by-atom superposition 88
atomic charge difference parameter 46
atomic molar refractivity contributions 42
ATPase inhibition, QSAR 147
attractive forces 10
automated regression algorithms 96

β-adrenoceptor ligands
– binding, QSAR 88, 123
– distance geometry 157
β-antagonistic activities
– labetalol 149, 150
– QSAR 130
β-glucosidase 7
– QSAR 121
B (bulk) parameter 22
baboon brain, lipophilicity optimum 128
back folding, and lipophilicity 34
backward elimination, stepwise regression 96
bacterial
– cell culture inhibition, QSAR 124
– DHFR inhibition, QSAR 118, 124
bacteriorhodopsin 8
barbiturates, transfer rate constants 126, 127, 176
bases
– buccal absorption 127
– ionization 77*ff*
BC(DEF) parameters 22, 40, 41, 101
BEL-FREE method 65
benzamidines
– QSAR 141
– trypsin inhibition 120
benzodiazepine receptor binding 158
– CoMFA 171, 175
– QSAR 123, 147
– similarity matrices 174, 175
benzoic acids
– electronic parameters
– – CoMFA 169
– – similarity matrices 174, 175
– toxicities, QSAR 43
benzomorphans, QSAR 140, 141
benzothiadiazine dioxides 152
benzothiazepines 157

benzoylglycine
– methyl esters, papain inhibition 119
– phenyl esters
– – chymotrypsin inhibition 120
– – papain inhibition, QSAR 54, 119
– pyridyl esters
– – CoMFA 169
– – papain inhibition 119
benzylamines, amphiphilic 31
benzyldiaminopyrimidines, DHFR inhibition 117
benzyldiphenylmethylpiperazines, QSAR 179
benzylpiperazineacetic acid esters, QSAR 179
benzylpyridinium ions, complement inhibition 121
best equation, in QSAR analyses 58*ff*, 61, 96
bias neurons, neural networks 89
bicyclooctane-carboxylic acids 44
bilayer membranes 9, 31
bilayers, artificial 31
bilinear activity-activity relationships 135
bilinear model 6, 50, 58, 71*ff*, 81, 96, 123, 126, 131, 176, 177
– and parabolic model 76, 131
binding
– affinities
– – α adrenoceptors 134
– – CoMFA 167
– – progesterone receptor 146
– – receptor 123
– constants 15, 57
– energies
– – additivity 19
– – AVERAGE ΔG 17
– – functional groups 19
– – hydrogen bonds 10, 11
– enthalpy, ΔH 10
– modes 10
– – favorable 157
– – ligands 49
– – multiple 20, 109, 161*ff*, 182
– – thiorphan and retro-thiorphan 166
– pH dependence 82
– pocket 10
– site 1, 9, 88, 146, 149, 151, 152*ff*
– – analysis 146
– – flexibility 163
– – geometry 49
– – GRID 155
– – hydrophobic character 157
– – hypothetical 152*ff*
– – limited 68, 77
– – models 153*ff*, 157
– – of papain 119
– steroids, 3D QSAR 159
bioavailability 15
biochemical model 85
biodegradability, QSAR 46
bioisosterism 17, 18
biological
– activity values 1, 16*ff*
– – additivity 15*ff*
– – and chirality 149
– – correlations 57, 115
– – cut-off 4
– – group contributions 4, 15*ff*, 54, 62
– – lipophilicity correlations 57
– – nonspecific 2, 5, 17, 29, 123
– – QSAR studies 129
– – rank-ordered 107
– – specific 109
– – variance 95
– Hammett equation 4, 17, 57
– hypotheses, in QSAR 181
– mechanism of action 96
– membrane/buffer partition coefficients 31
– oxygen demand (BOD) 46
– parameters, in QSAR studies 16*ff*
– response 5
– test models 115
biomembranes 32
biophobes, CASE program 86, 145
biophores, CASE program 86, 145
biotin 17
biphenyls, CoMFA 171
biphosphonate esters, CoMFA 172
bleomycin 128

blood-brain barrier 3, 15, 128*ff*, 135, 179
- penetration 30, 128*ff*, 135
blood levels 126
Bocek Kopecky model 67
- relationship to nonlinear Hansch analysis 67
boiling point 22
bond perturbation values 45
bone resorption regulation, CoMFA 172
bootstrapping 104, 105
bovine liver DHFR 118
bradycardic activities, QSAR 134
bradykinin potentiation, QSAR 147
brain capillary permeability coefficients, P_c 128
brain, lipophilicity optimum 128
bromelain inhibition
- comparative QSAR 180
- QSAR 120
bromophenethylamines, QSAR 58*ff*, 63*ff*, 130
bromobutide, QSAR 179
buccal absorption
- acids and bases 81, 82, 127, 176
- pH dependence 81, 82, 126*ff*
buprenorphine 17
butyrylcholinesterase inhibition, QSAR, 120, 147, 150

C (cohesiveness) parameter 22
calcium antagonism 152
- dihydropyridines 51
- verapamil analogs 153
- QSAR 50, 123
calculation of
- MR values, MEDCHEM software 42
- partition coefficients 28, 34*ff*
- - limitations 36
Cambridge database of crystal structures 19, 156, 160
cAMP-phosphodiesterase 134
camphor 17
- binding, to cytochrome P-450 164
cannabinoid receptor binding, CoMFA 171
capacity factor, HPLC 135
carbamates, absorption rate constants 127, 176
Carbo index, R_{AB} 172, 174
carboanhydrase inhibitors 153
- multiple binding modes 163
- QSAR 120
carbonyl group hydration, CoMFA 169
carboxylesterase inhibition, QSAR 147
carboxyribitol diphosphate 17
carcinogenicity 109, 125
- predictions 85
- QSAR 85, 100
carrier protein binding
- CoMFA 171, 175
- similarity matrices 174, 175
CARSO procedure 105
CASE program 86
- application 145
CATALYST program, 3D pharmacophores 161
cavity wall 88
CCK-A antagonist binding, CoMFA, 171
cell culture 15
- data 115
- inhibition, QSAR 124
- models 133
cell wall permeability 117
cells 115
cellular respiration, QSAR 42
central nervous system, see CNS
centrifugal partition chromatographic techniques 32
cephalosporins 10
cerebral vasodilators, QSAR 179
chance correlations 53, 85
- in Hansch analysis 62
- in neural networks 89
- in PLS analysis 167
charge densities 46
charge distributions 40

charge transfer constants 21, 42, 46
charged hydrogen bonds 10
Charton's steric parameter υ 41, 47
chemical similarity 161
chenodesoxycholic acid 80
chicken liver DHFR 118
chiral drugs, QSAR 150
chirality, and biological activity 149
chloramphenicol, receptor site models 152*ff*
chloroform/water f values 34
chloromethyl ketones 10
chlorophenols, QSAR 141
chloroquine resistance 125
chlorpromazine
- CLOGP 38
- other data 39

choice of variables, in Hansch analysis, 59
chromatographic parameters 21, 32*ff*
chymotrypsin
- comparative QSAR 180
- inhibition, QSAR 96*ff*, 120
- ligands, QSAR of enantiomers 150
- QSAR 120

cis/trans isomerism 54
class modeling techniques 100*ff*
class separation, discriminant analysis 100
classical Free Wilson model 63
classification
- multicategorial (ORMUCS) 100
- problem, in SIMCA 101

clearance 15
clique-detection, DISCO 161
CLOGP program 35*ff*
clonidine analogs 134
- CoMFA 169, 175
- QSAR 177, 180
- similarity matrices 174, 175

Clostridium welchii 170
clozapine analogs, CoMFA 171
cluster analysis 107
- substituents 22

cluster significance analysis 107, 110
clustering
- of activity values 96
- of data 16
- methods, in design 113

CNS-active drugs
- lipophilicity optima 128
- QSAR 133

CNS side effects 179
- and lipophilicity 128

COBRA, 3D structure generation 160
cocaine binding site
- CoMFA 171, 175
- similarity matrices 174, 175

cocrystallisation 8
Collander equation 28
Collander-type relationships 28
collinearity 113
  in parameter space 111
colonic absorption, acids 81, 128
CoMFA 1, 6, 106, 159*ff*, 174, 175, 181
- and Hansch analyses 169
- and similarity matrices 174, 175
- applications 167, 169*ff*
- chance correlations 167
- comparison with QSAR 170
- contour maps 168
- critizism 171
- diagram 160
- difference contour maps 168
- electrostatic field 164
- field fit 166
- fields, 164*ff* 172
  - combination with other parameters 167
- history 7, 159
- limitations 172
- methodology 172
- natural logarithms 95
- options, influence on results 172
- patent 169
- PLS analysis 167
  - errors in 168
- practical aspects 172
- predictive ability 168
- program 160
- results, graphical presentation 168
- statistical requirements 172

– steric field 164
– theoretical aspects 172
CoMPA 169
COMPACT program 100
comparative molecular field analysis, see CoMFA
comparative molecular potential analysis, see CoMPA
comparative QSAR analyses 115, 124, 132, 141, 180
comparison of
– biological effects 115
– Hansch and Free Wilson analyses 65*ff*
– QSAR equations 180
compartments 115
complement inhibition, QSAR 121
complexity, drug-receptor interactions 123
components, PLS analysis 102
computer-aided drug design 158, 181
computer-assisted alignment 161
computer-automated structure evaluation, see CASE
computer simulation, drug transport 68*ff*, 74, 75
concanavalin A binding 121
concentration dependence, partition coefficients 32
CONCORD, generation of 3D structures 36, 160
confidence intervals 20, 59, 61, 63, 91, 93, 96, 105
– in Free Wilson analysis 65, 145
– new observations 91
– regression coefficients 53, 94
conformation 182
– at the binding site 161
– in aqueous solution 160
– in the crystal 161
– *in vacuo* 160
– low-energy 157, 158, 161
– maps 19
– minimum energy 89, 182
– of ligands 49
– of proteins 8
– stabilization 9
– superposition 161
conformational
– barriers 182
– changes 31
– degrees of freedom 182
– distortions 20
– energy 17, 50
– entropy 12, 21, 50, 167
– flexibility 3, 89
– preferences 18, 50, 160
– space 160, 161
– – of ligands 157
congeneric drugs 85
– alignment 161
congeneric series 86
congenericity 109
conjunction (and) 88
conjunctions, topological pharmacophore approaches 88
connection layers, neural networks 89
connectivity indices 21, 41, 50*ff*
contour maps
– difference maps 168
– in CoMFA 168
– energy values, GRID 155
CORINA, 3D structure generation 160
correlation between data blocks, in PLS analysis 102
correlation coefficient r 20, 61, 93, 95
– r^2_{PRESS} value 104*ff*
– upper limit 66
correlation matrix, X variables 96*ff*
corticosteroid-binding globulin
– 3D QSAR 159, 171*ff*
– contour maps 168
Coulomb potential 164
covalent bonds 10
Craig plot 109
– σ vs. π diagram 110
critical micelle concentration (CMC) 124
cross product terms, in QSAR 77
cross-validation 88, 99, 102*ff*, 167*ff*, 181

- diagram 103
- in CoMFA PLS analysis 167*ff*, 172
- in Hansch analysis 59
- in PLS analysis 102*ff*
- leave-one-out procedure 103, 104, 167
- r^2_{PRESS} value 104*ff*
- variance of variables 168

crystal structures, Cambridge database 19, 156, 160
crystallographic analyses 8
CSA, see cluster significance analysis 107
cut-off
- biological activities 4
- CoMFA fields 165

cyclandelate 69, 70
cyclic nifedipine lactones, QSAR 50
cyclohexane/water partition coefficients 29, 128
cyclopropylamine MAO inhibitors 152
cysteine proteases 116, 119, 120
- comparative QSAR 180
- inhibitors 10, 119*ff*
- QSAR 180

cytochrome P-450
- induction 129
- inhibitors, multiple binding modes 164
- QSAR 180
- substrates, QSAR 100

cytosine nucleoside deaminase, QSAR 120

D-optimal design 99, 105, 113
D_2 receptor affinity, CoMFA 171
DARC-PELCO approach 65, 88, 145
data
- *in vitro* 15, 95, 121*ff*, 133
- *in vivo* 15, 57, 129*ff*, 133
- modeling, in QSAR 181
- projection 107
- sets, ill-conditioned 76

database
- Cambridge, crystal structures 19, 156, 160
- DESBASE 22
- Hansch equations 115
- searches 36, 161

DDT, log P calculation 36
de novo
- approach 62*ff*
- constants 54
- design of ligands, LUDI 155, 156

deazapurines, binding modes 165
decadic logarithms, in QSAR 15, 95
degrees of freedom 66, 94, 102
- in Free Wilson analysis 64, 65
- in pattern recognition 85
- internal 17

density 40
dependent variables 16, 91, 101
deprenyl 10
DESBASE database 22
descriptors, CASE program 86
design 109*ff*
- D-optimal 99, 105, 113
- *de novo*, of ligands 155, 156
- factorial 105, 113
- in QSAR 114
- of congeneric series 109*ff*
- of test series 109*ff*
- rational 99, 109*ff*
- strategies, criteria 114

desolvation 10, 12, 182
- energy 21

determinant 93
- parameter correlation matrix 113

development, of QSAR 4*ff*
DHFR, see dihydrofolate reductase
diacylureas, QSAR 142
diaminodiphenyl sulfone 117
diaminomethyl-pentylquinazoline (DMPQ) 17
diaminophenyl-azobenzenesulfonamide 7
diaminopyrimidines, QSAR 144
diastereomeric mixtures
- of drugs 149, 150
- QSAR 140, 141

diazepam 17
dibenzodioxins, CoMFA 171
dibenzofurans, CoMFA 171
dielectric constant 10, 40
difference contour maps, CoMFA 168
differential molecular connectivity index, 53
diffusion
– control, transport rate constants 73, 74
– layer, aqueous 82
– of drugs
– – into milk and prostatic fluid 129
– – through membranes 31
– rate constants 50, 127
dihydrofolate, binding mode 161*ff*
dihydrofolate reductase 86, 116*ff*, 161*ff*
– active site 155, 162, 163
– different sources 117
– inhibition 116*ff*, 153
– – CoMFA 169
– – distance geometry 157
– – HASL 158
– – ligand design 116
– – MTX resistance, QSAR 124*ff*
– – QSAR 67, 86, 116*ff*, 124, 144
– multiple binding modes 161*ff*
– purine biosynthesis 116
– receptor map 152
dihydropyridines 152
– X-ray structures 51
dihydrotriazines, DHFR inhibition 67
dimethoxyphenethylamines 152
dipole-dipole interactions 11
dipole moment 11, 21, 22, 42, 46
DISCO program, superposition 161
discriminant analysis 100*ff*
discriminant-regression analysis 100
disjunction (exclusive or) 88
dispersion forces 11, 40
dispersive interactions 10
displacement, of water molecules 41
dissociation 15, 77*ff*, 82
– acids and bases 77*ff*, 129
dissolution rate constants 34
distance geometry method 6, 157, 161
distance mapping technique 113
distance of grid points, in CoMFA 163
distances, in multidimensional space 107
distomer 150
distribution
– of drugs 68*ff*, 179, 181
– – QSAR 129
– parameters 15
– rate constants 57, 68*ff*
diuretic activities, CoMFA 172
diyhdropyridines, calcium antagonism 51
DNA
– alkylation, HASL 158
– intercalation 46
– polymerase, QSAR 121
donor-donor interactions 162
dopamine β-hydroxylase substrate properties, QSAR 147
dopamine receptor binding, QSAR 123, 147
dopamine-uptake inhibition, QSAR 123
dose response curves, Escherichia coli, 125
dose response relationships 15
doses, inverse molar 5
drug action, and stereochemistry 149*ff*
drug design 2, 182
– computer-aided 158, 181
– industrial praxis 179
– QSAR in 179
– structure-based 116, 179
drug development 8
– and QSAR 179
– enantiomers 149
drug distribution 2, 68*ff*, 129, 179, 181
drug enantiomers 149
drug-membrane
– interaction 31*ff*, 115
– partition coefficients 8, 31
drug partitioning, standard system 28

drug-receptor
– affinity 21
– binding, negative contributions 12
– complex 10, 12
– interaction 3, 7*ff*, 21, 115, 181
– – complexity 123
– – stereospecificity 7
drug research, in pharmaceutical industry 133
drug resistance 31, 124*ff*
– see also multidrug resistance
drug transport 2, 3, 20, 68*ff*, 179
– computer simulation 68
– in biological systems 68
– rate constants 69, 72, 74
drugs
– acids and bases 79
– activity profiles 137
– amphiphilic 125
dummy variables 21, 54*ff*
DYLOMMS approach 6, 159
dynamic lattice 159

E_s parameter 7
– sign 60
E_s^c values 47, 49
ecotoxic compounds, QSAR 137
EDISFAR program, for design 113
electron
– acceptor substituents 109
– density 172
– – distribution 43, 181
– donor substituents 109
electron-topological approach 158
electronic
– effects 44
– – calculation 47
– parameters 4, 5, 21*ff*, 42*ff*
– – CoMFA 169
– – in QSAR 58
– properties 21, 42*ff*
– similarity matrices 175
electrostatic
– attractions 11
– fields 6, 173
– – CoMFA 164, 167
– interactions 1, 10, 11, 18
– – contour maps 168
– potential 46, 173
– – similarity matrices 173*ff*
electrotopological state index 53
elimination, of drugs 179
– QSAR 129
– rate constants 15, 126
embedded classes of compounds 101
empirical models, nonlinear relationships 68*ff*
emulsin 7
– hydrolysis, CoMFA 169
enantiomers 149*ff*
– alignment 151
– in drug development 149
– of labetalol 149, 150
– pharmacological properties 149, 150
– QSAR 140, 150
end product inhibition 68
endocytosis, in MDR 125
energy balance 10
energy-minimized structures 161
enthalpic contributions, ligand binding, 13, 167
enthalpic interaction, hydrophobic groups 40
enthalpy, ΔH 10, 12
entropic contributions
– hydrophobic groups 40
– ligand binding 13, 27
entropic effects, ligand-protein interaction 167, 182
entropic interactions 11
entropy, ΔS 10
environmental health hazards 85
enzymatic cleavage, glycosides 7
enzymatic reaction 19, 68
enzyme
– inhibitor complexes, 3D structures 8, 116, 161*ff*, 181
– inhibitors 1, 10, 115, 116*ff*, 133
– – binding modes 161*ff*
– – CoMFA 171
– – Free Wilson analysis 147
– – Hansch analyses 116*ff*

equieffective molar doses 15
equilibration, in partitioning 32
equilibrium constants 10, 15
- organic reactions 5
equilibrium model 71, 74, 75
equilibrium systems 74
- pH dependence 82
equivalence (if and only if) 88
equivalence
- of Free Wilson models 63
- of group contributions 66
error, in biological data 96
error terms 91
erythrocyte membrane/buffer partition coefficients 29
Escherichia coli 31, 117, 125, 135
- DHFR 117
- - GRID potentials 155
- - inhibition 117
estradiol 17
estrogen receptor binding, QSAR 123
estrogens, tumor cell response, CoMFA, 172
eudismic index 150, 151
eutomers 150, 151
EVAL method 86
experimental
- error, in regression analysis 91
- partition coefficients 32*ff*, 36
- pK_a values 36
expert systems
- Apex-3D program 158
- carcinogenicity prediction 86
- rule-based 86
explained variance 93
exponential similarity indices, E_{AB} 173
extrapolations
- of R_M values 33
- from QSAR 179
extrathermodynamic approach 5, 57, 115*ff*
- applications 115*ff*
- recommendations 59*ff*
- relationships 16
f values 34*ff*
- calculation of partition coefficients 35
- different scales 34
F test, sequential 59, 61, 94, 96*ff*
F values 94, 95
- table 94
$\mathscr{F}$ values 21*ff*, 42*ff*
- redefinition 44
- validity 44
factor analysis 101
factorial design 105, 113*ff*
- in Free Wilson analysis 114
factorization, of σ values 45
favorable binding sites, GRID 155
ficin inhibition
- comparative QSAR 180
- QSAR 120
field fit, CoMFA 166
field parameter $\mathscr{F}$ 21*ff*, 42*ff*
fields
- contour maps, CoMFA 168
- in CoMFA 160, 164
- hydrophobic 167
filter probe methods 32
Fisher significance ratio, F 20, 94, 96
five-compartment system, drug transport 75
flexibility
- binding site 163
- index 53
- of ligands 163
- of molecules 12
flexible
- drugs, 3D structures 160
- fit 8, 181
- molecules 12
fluidity of membranes 31
folate biosynthesis inhibition 135
force field methods 160
forward selection, stepwise regression 96
fractional factorial design 105
fragment approach 53
fragmental volume constants 41
Franke protein binding model 69

free energy, ΔG 9, 15, 17
- ligand binding 13
- ligand-protein interaction 167, 182

Free Wilson analysis 1, 5, 54, 62*ff*, 139*ff*
- and Hansch analysis, relationships 65*ff*
- applications 139*ff*, 147
- classical version 63
- degrees of freedom 64
- design in 113
- equivalence of different models 63
- graphical presentation 141
- group contributions 66
- interaction terms 67, 142, 144
- interpretation 140
- limitations 139
- nonadditivities 143*ff*
- predictive ability 139
- reduced model 65, 145
- restrictions 143
- scope and limitations 65
- shortcomings 63
- statistical problems 147
- stepwise 86
- variable elimination 144

Free Wilson-type indicator variables 21, 67*ff*, 86
Fujita Ban model 5, 63
- advantages 63

Fujita Ban-type indicator variables 63
fungicides, QSAR 179
furoindolecarboxamides, QSAR 179
future development, of QSAR 180*ff*
fuzzy ALS method, in QSAR 101

G protein-coupled receptors, modeling 8
GABA receptor antagonism, CoMFA 171
GABA-uptake inhibition
- CoMFA 171
- DYLOMMS 159

gastric absorption 126*ff*
- lipophilicity optimum 127
- model 126
- QSAR 176
- rate constants 84, 127

gastric wall 127
gastrointestinal absorption 3, 84
Gaussian function approximations 173, 174
genetic algorithms
- in PLS analysis 167
- variable selection 175

genetic expression, receptor proteins 19
geometric parameters 21, 50
geometry of
- binding site 49
- drug-receptor complex 10

glucocorticoid activities, QSAR 147
glucosidase
- α and β 7
- QSAR 121

glutamate dehydrogenase, QSAR 121
glutamic acids, antitumor activities 62
glutamine synthetase inhibitors, HASL 158
glyceraldehyde phosphate dehydrogenase, QSAR 121
glycine ester amides, QSAR 55, 180
glycocalyx 127
glycolic acid oxidase, QSAR 121
glycoprotein GP 170, MDR 125
glycosides, enzymatic cleavage 7
glyoxalase inhibitors, HASL 158
GOLEM program 86
GOLPE variable selection 105, 167, 174
- similarity matrices 174

GP 170, transport glycoprotein 125
graph theory 50
graphical presentation, Free Wilson analyses 141
GRID program 6, 155
- contour lines 155
- interaction fields 155, 165, 171
  - in CoMFA 155, 163
- water probe 170

grid points, in CoMFA
- distances 163

– variance 165
group contributions 5, 15*ff*, 54, 62*ff*
– additivity 15
– and physicochemical properties 65*ff*
– biological activity values 2, 4, 16, 54
– equivalence 66
– in Free Wilson analysis 62*ff*
growth inhibition, cells 116
guanine deaminase, QSAR 121

H_1 antagonists 123
H_2 antagonists, blood-brain barrier 30, 128
hallucinogens, QSAR 133, 141
Hammett equation 4, 34
– biological 4, 17, 57
Hammett σ parameter 5, 21, 34, 40, 42*ff*
– similarity matrices 174, 175
Hammett-type relationship 5
Hancock E_s^c values 47
Hansch analysis 1, 5, 16, 57*ff*, 115*ff*
– and CoMFA 169
– and Free Wilson analysis, relationships 65*ff*
– applications 115*ff*
– best equation 58*ff*, 61, 96
– chance correlations 62
– group contributions, derived from 63
– outliers 61, 99, 182
– predictions 61
– recommendations 59*ff*
Hansch equations 54, 57*ff*, 115*ff*
– improvement 140
– parabolic 69
– stepwise derivation 139*ff*
– validation 61
hapten-antibody interactions, QSAR 121
HASL, see hypothetical active site lattice model
health hazards, environmental 85
hemoglobin 8
– ligands, QSAR 19, 20
– allosteric modifiers 157
hemolytic activities
– CoMFA 170
– QSAR 123
– regression analysis example 92
herbicides, QSAR 179
hexokinase inhibitors, HASL 158
hexylphenylacetic acid, buccal absorption 81, 82
hidden layers, neural networks 89
high-performance liquid chromatography (HPLC) 33
Higuchi Davis equilibrium model 71
HINT, hydrophobic fields 157, 165
– log P estimation 157
histocompatibility glycoprotein binding, CoMFA 171
history, of QSAR 4*ff*, 159
Hodgkin index, H_{AB} 173, 174
HOMO energies 46
homologous series 5, 171
HPLC capacity factor 135
HSV-thymidine kinase, QSAR 121
human lymphoblastoid DHFR 118
Hyde model 71
hydrated surface areas 29
hydrogen bond
– acceptors 20, 22
– – in LUDI 156
– acceptor solutes 28
– balance 18
– donors 20, 22
– – in LUDI 156
– donor solutes 28
– network 37
– potentials, GRID 155
– properties 22, 54
hydrogen bonds 2, 10*ff*, 21, 89, 161
– binding energies 10*ff*
– intramolecular 37, 54
hydrogen-bonding
– ability 29*ff*, 46, 128
– effects 29
– interactions 179
– parameters 30, 42, 46
hydrolysis rate constants, QSAR 180
hydrophilic barriers 2

hydrophilicity 27
hydrophobic
– atom constants 157
– fields 37, 157
– – CoMFA 167
– – for 3D QSAR studies 157
– – HINT program 157, 165
– fragmental constants 22, 34*ff*, 157
– interactions 1, 11, 12, 13, 157
– properties 21
– surfaces 11
hydrophobicity 22, 27
– maps 157
– profile 157
hydroxyquinoline-carboxylic acids, QSAR 41, 42
hyperbolic regression, in QSAR 77
hyperboxes, in SIMCA 101
hypermolecule 88*ff*, 146
hyperplane 101
hypersaturation phenomena 34
hyperstructure approaches 50, 65, 86*ff*, 146, 151
hypertensive activities, QSAR 134, 180
hypnotic activities, QSAR 142
hypotensive activities, QSAR 134, 180
hypothetical
– active site lattice (HASL) model 158
– binding sites 152*ff*
– pharmacophore, CATALYST program 161
– rate constants 68

IC_{50} values 15
ill-conditioned data sets 76, 109
imidazolines, similarity matrices 174, 175
imidazolyl-phenoxypropanolamines, QSAR 130
immunochemistry, QSAR 121
implication (if then) 88
improvement, Hansch equations 140
independent variables 91, 101
indicator variables 21, 54*ff*, 85, 91, 96
– Free Wilon model 62
– Free Wilson-type 86, 141
– in Hansch analysis 67
– in QSAR 55
induced fit 8
inductive
– contributions 60
– effects 44
– hypotheses, GOLEM program 86
– logic programming 86
industrial practice 115
– drug design 179
inflammatory activities, QSAR 131
information theory approach 113
inhibition
– classes of enzymes 115
– constants 15, 57
inhibitors 1, 20, 116*ff*
input layer, neural networks 89
insecticides, similarity matrices 173
in situ
– absorption models 81, 84
– perfused organs 15
interaction
– charged groups 10
– energies 6, 17, 153, 154, 181, 182
– – calculation 155
– – of functional groups 17
– – of ligands 152
– fields, GRID program 165
– geometry 10
– – of thermolysin inhibitors 163
– ligand-receptor 1
– models 151
– terms, in Free Wilson analysis 67, 142, 144
intercorrelation, π *vs.* σ 43
interface, aqueous/organic 70, 82
intermolecular forces 1, 21
intermolecular interactions 10
internal degrees of freedom 12, 17
internal rotational entropy 12
interpretation, Free Wilson analyses 140
interrelated variables 96, 101, 106
– in pattern recognition 85
intestinal absorption 84, 128, 176
– lipophilicity optimum 127

- rate constants 84, 127
intestinal wall 127
intraclass correlation matrices 113
intramolecular hydrogen bonds 37, 54
intrinsic activity 1
intrinsic binding energy contributions 17
inverse molar doses 5
inverted matrix 93
invertin 7
in vitro
- activity values 15, 95, 121
- models 1, 115, 133
- - and animal models 180
- *vs. in vivo* data 133
in vivo activities 15, 57, 133
ion pairs, pH partition profile 80
ionization 3, 77*ff*, 181
- acids and bases 77*ff*, 129
- constants, benzoic acids 45
ionized species 82
ions, acids and bases, lipophilicity 77*ff*
irreversible enzyme inhibitors 10
iso-butanol, molecular connectivity values 52
isolated organs 15, 115, 133
isotropic surface areas 29
iterative fitting, HASL model 158

ketanserin 17
KEY program 157
K_i values 15
kinetic control, transport rate constants 73
kinetics of drug transport 68
knowledge base
- Apex-3D program 159
- GOLEM program 86

labetalol, enantiomer properties 149, 150
lack of validity, of a QSAR model 99
lactate dehydrogenase, QSAR 121
Lactobacillus casei
- DHFR inhibition 118
- methotrexate resistance, QSAR 124
latent variables 101
lateral validation, in QSAR 180
lattice, in CoMFA 163
lattice model, CoMFA history 159
LD_{50} test, non-animal methods 137
lead structure optimization 2, 109, 133, 179
learning and behavior, QSAR 147
least squares procedure 57
leave-one-out cross-validation 103, 104, 167
Lennard Jones potential 11, 164
Leo/Hansch method 35*ff*
leukemia
- cell differentiation, CoMFA 172
- tumor cell inhibition, QSAR 125
lidocaine analogs, QSAR 129
ligand
- binding 8*ff*, 115
- - free energy 13
- - modes 49, 161*ff*
- conformations 49
- conformational space 157
- design 3, 8, 179
- flexibility 163
- orientation 9
- points 157
ligand-binding site interactions 8
ligand-enzyme interactions 41
ligand-protein complexes
- 3D structures 8, 116, 161*ff*
- multiple binding modes 161*ff*
ligand-protein interaction 1, 116*ff*, 181
- CoMFA 167, 171
ligand-receptor interactions 8*ff*, 13
- energies 152
limitations
- Collander equation 28
- CoMFA 172
- Free Wilson analysis 139
- of design strategies 114
limited binding site 68, 77
limited solubility 68

linear
- activity-activity relationships 133
- additive structure-activity relationships 65
- dependences, between variables 63
- discriminant analysis 100
- free energy-related approach 5, 16, 57*ff*
- free energy-related parameters 115
- free energy relationships 54, 57*ff*
- Hansch equations 58
- lipophilicity-activity relationships 19, 123
- mixed approach 67
- regression analysis 57, 91*ff*
- similarity indices, L_{AB} 173

lipid membranes 2, 9, 31
lipophilic
- barriers 2
- counter ions 80
- groups, LUDI 156

lipophilicity 2, 26, 27*ff*, 57, 181
- CNS side effects 128
- contributions 34
  - – atom-based 156
- of neutral species 82
- optima 3, 73, 76, 179
  - – antitumor drugs 132
  - – cells and animals 125
  - – CNS-active drugs 128
  - – sensitive and resistant cells 125
- parameters 5, 21, 27*ff*, 34
  - – calculation 34*ff*
  - – CoMFA 167
  - – in QSAR 58
  - – measurement 32*ff*
- pattern 181
- potential, molecular 158
- similarity indices 175
- similarity matrices 77, 171, 175*ff*
- values, HPLC measurements 33

lipophilicity-activity relationships 28, 123
- and CoMFA 170
- nonlinear 5, 16, 29, 68*ff*, 142, 176, 177
  - – models 76

lipoxygenase, QSAR 121
liquid chromatography, high-performance 33
loadings of variables 106
local anesthetics, QSAR 129
lock and key model 7, 8
LOCK program 157
LOCKSMITH program 157
LOCON method 86, 88
log 1/C values 16
log $1/K_i$ values 16
log $1/K_m$ values, papain 119
log D values 78, 81
log k values 15, 16
log k′ values, correlation with log P values 33
log k_{cat} values, papain 119
log P values 5, 27*ff*
- and pK_a values, determination 83
- calculation 34*ff*
- in CoMFA 165
- of alcohols 29

log P_{app} values 77*ff*, 127, 134
- pH dependence, acids and bases 77*ff*

LOGANA method 86, 88
logarithmic scale, activity values 15
logical operators 88
logico-structural approach 158
loss of variance, variables 168
low-energy conformations 157, 158, 161
low-frequency vibration 12
LUDI program, *de novo* design 155, 156
LUMO energies 46
lysozyme 8
- multiple binding modes 161

machine learning program, GOLEM 86
malate dehydrogenase inhibition, QSAR 41, 42, 121
mammalian receptors 8
mandelic acid esters, QSAR 69, 70, 74, 176

MAO, see monoamine oxidase
mapping
- intraclass correlation matrices 113
- receptor 151
mathematical model 62
- pharmacokinetic data 83
matrix
- diagonalization 101
- inversion 93
- of normal equations 92
MCD, Monte Carlo version of MSD method 88
McFarland probability model 70, 71
MDR, see multidrug resistance
mean binding energies, AVERAGE ΔG, 17
mean error, biological data 95
measurement, of partition coefficients 32
mechanism-based enzyme inhibitors 10
mechanism of action 109
mechanisms of toxicicity, and QSAR 85
MEDCHEM software 36, 42
melting point 22, 34
membrane 2, 31
- and drug action 31
- channels 31
- NMR measurements 31
- partitioning, drugs 8, 21
- permeation 27
- perturbation 2, 125
- potential 31
- transport, neutral forms 79
membrane-bound
- proteins, 3D structures 8
- receptors 31, 115
membrane/buffer partition coefficients 31
membrane-embedded receptors 8
membrane/water partition coefficients, measurement 30
MEP, see molecular electrostatic potential
mesylglycine phenyl esters
- papain inhibiton 119
- QSAR 54
metabolic pathways 20
metabolism 3, 179
- of enantiomers 149
- of lipophilic drugs 68
- QSAR 129
- rate constants 126
- reactive species formation 129
metamitron, QSAR 179
methodology
- CoMFA 172
- QSAR 5
methotrexate 17, 116
- analogs 153
- binding mode 161*ff*
- resistance, QSAR 124*ff*
methylene group, fragment values 35
mg/kg values 16
micelle formation 68
micelles 80
Michealis Menten constants 15
MIMUMBA, 3D structure generation 160
minimal steric difference method 50, 88, 146, 152
minimal topological difference method 50, 88, 146, 152
- and CoMFA 169
minimum analog peptide sets, for QSAR 113
minimum-energy conformations 89, 182
minimum receptor occupation 68
mixed approach 63, 65*ff*, 88, 146
- applications 67*ff*
mixed Hansch/Free Wilson model 5, 65*ff*
MO calculations 151
model
- hypotheses, from Hansch analysis 61
- refinement, in CoMFA 167
- simulations, in pharmacokinetics 126
- systems, kinetically controlled 74
models

- *in vitro* 1, 115, 133
- *in vivo* 15, 57, 133
- nonlinear 2, 68*ff*
- quantitative 2
- sensitivity 136

molar
- doses, inverse 5
- polarization, P_M 0
- refractivity 21*ff*, 40*ff*, 49
  - and connectivity values 53
  - calculation 42
  - contributions, atom-based 42, 156
  - correlation with other properties 40*ff*
  - in QSAR studies 40*ff*, 57
  - MEDCHEM software 42
  - nonlinear relationships 77
  - scaling of values 41
- volume 21, 40, 49

molecular
- alignment, in CoMFA 166
- connectivity indices 50*ff*
- dynamics 18
- electrostatic field 10
- electrostatic potential 155
- field analysis, comparative, see CoMFA
- graphics 116
- lipophilicity potential 156, 158
- mechanisms, of drug resistance 124*ff*
- modeling 3, 179, 182
- orbital energies 46, 151
- shape analysis (MSA) 50, 89
- shape descriptors 53
- similarity indices 172*ff*, 181
- weight 21, 22, 40, 49, 74

monoamine oxidase
- inhibition 10, 152
  - QSAR 82, 83, 121, 176
- ligands
  - CoMFA 171, 175
  - similarity matrices 174, 175

Monte Carlo version, of MSD method 88

morphinans, QSAR predictions 140

morphine, SMILES code 35

mosquito larvae 43

MPTP toxic activation
- CoMFA 171, 175
- similarity matrices 174, 175

MR values 21, 22, 40*ff*
- scaling of 41

MSA, see molecular shape analysis

MSD, see minimal steric difference

MTD, see minimal topological difference

MTX, see methotrexate

MULTICASE program 86, 87

multicategorial classification (ORMUCS) 100

multicollinearity 113

multidimensional
- activity data 137
- mapping 113
- parameter space 22

multidrug resistance 31, 50, 125

multi-layer network 89

multiparameter equations 57, 96

multiple binding modes 3, 20, 109, 150, 161*ff*, 182

multiple regression equations
- from PLS analysis 167
- in QSAR 181

multivariate statistical methods 85, 100*ff*

muscarinic receptor
- agonists and antagonists, QSAR 122
- affinities 151
- ligands
  - CoMFA 171
  - QSAR 122, 147

mutagenesis, ligand binding 10, 19

mutagenesis test, *Salmonella* 85

mutagenicity 46
- CoMFA 169
- QSAR 85, 109, 125

myclobutanil, QSAR 179

Mycobacterium lufu 117

Mycobacterium smegmatis 135

naphthylamidines, trypsin inhibition 120

narcotic activities 109

natural logarithms, in CoMFA 15, 95
negation (not) 88
negative r^2_{PRESS} values, PLS analysis 104
net atomic charges 46
net free energy of interaction 182
neural networks 89
- problems 89
- similarity matrices 174
neuroleptics
- and membranes, NMR 31
- QSAR 147
- spectral mapping 136
neutral
- form, acids and bases 79, 82
- hydrogen bonds 10
neutron diffraction, small-angle 30
Nicod's function (neither nor) 88
nifedipine analogs, QSAR 50
nimodipine 50
- log P values 31
NIPALS algorithm 101, 102
nipecotic acids, DYLOMMS 159
nisoldipine, log P values 31
nitromethylene insecticides, similarity matrices 173
NMR measurements, neuroleptics 31
noise variables, in PLS analysis 105
nonadditivity
- in Free Wilson analysis 143*ff*
- partition coefficients 37
- of group contributions 20
non-animal methods, for LD_{50} test 137
noncongeneric series 86
noncovalent interactions 6, 11
non-designed data sets 105
non-equilibrium model system, drug transport 75
nonlinear
- activity-activity relationships 135
- CoMFA models, critizism 171
- iterative partial least squares 101
- lipophilicity-activity relationships 5, 16, 20, 68*ff*, 142, 176, 177
-- and CoMFA 170
-- similarity matrices 175
- mapping 107
- mixed approach 67
- models 2, 68*ff*
-- applicability 76, 77
-- calculation 76
-- comparison 131
-- in pharmacokinetics 129
-- other 76
-- transport and distribution 5
- molar refractivity-activity relationships 77
- parameter β, bilinear model 73
- structure-activity relationships 66, 68*ff*
-- and CoMFA 167
-- empirical models 68*ff*
-- enantiomers 150
-- in Free Wilson analysis 142
-- theoretical models 70*ff*
- terms, in Free Wilson analysis 141
- version of PLS analysis 105
non-NMDA antagonist binding, CoMFA 171
nonspecific
- binding, to proteins 121
- biological activities 2, 5, 17, 29, 109
-- QSAR 123
- structure-activity relationships 57
- toxicities 115
norepinephrine-uptake inhibition 152
- QSAR 123, 140, 147
norfloxacin, development 179
normal distribution, experimental error 16
normal equations, regression analysis 91
nucleoside transporter protein 158
number of
- components, in PLS analyses 103
- degrees of freedom, in Free Wilson analysis 141
- terms, in Hansch analysis 59
- variables, in regression equations 96
numerical equivalence, group contributions 66

Occam's razor 59
octanol/water partition coefficients 5, 27*ff*
– advantages 28
– standard system 29
opiate receptors 13
optically active compounds 140, 149*ff*
optimal set of analogs 114
optimization, lead structures 2, 133, 179
optimized map, in MTD 88, 146
optimum lipophilicity 3, 73, 76, 179
– antitumor drugs 132
– blood-brain barrier 128
– cells and animals 125
optimum number of components, in PLS analysis 103
orbital energies 42
ordered water molecules 11, 41
orientation
– of a ligand 9
– of drugs, in a membrane 31
– rules, in 3D QSAR 161
ORMUCS 100
ortho effects 54
ortho-substituents 45
ouabain 17
outliers
– in Hansch analyses 61
– in QSAR 99, 182
output layer, neural networks 89
overall variance 93, 104
overlap of properties, REMOTEDISK 158
overlap volumes 89
overprediction 96
– in Hansch analysis 62
– in PLS analysis 102*ff*
overtraining, neural networks 89
oxathiolanes, QSAR of enantiomers 150, 151
oxygen affinities, hemoglobin 19, 20
oxytocin contraction inhibition, QSAR 147

π values 21, 22, 34
– different scales 34*ff*, 59
– lipophilicity calculation 34*ff*
$\pi_{benzene}$ values 34
P values (resonance π bond perturbation) 45
P values, see partition coefficients
P_{app}, P_i and P_u values 77*ff*
papain
– comparative QSAR 180
– inhibition, QSAR 119
– ligands
– – CoMFA 169
– – QSAR 54, 55
parabolic
– lipophilicity-activity relationships 68, 123
– lipophilicity term 57
– model 5, 58, 68*ff*, 74, 81, 96
– – comparison with bilinear model 76, 131
– – systematic deviations 50
parachor 21, 40, 41
parameter
– correlation matrix 22, 113
– focusing 107 110
– interrelation 109
– range, and predictions 61
– space 3
– – collinearity 111
– – multidimensional 22
parameters 21
– aliphatic 22
– aromatic 21, 23*ff*
– BC(DEF) 22
– collections 21*ff*
– connectivity indices 21, 41, 50*ff*
– electronic 21*ff*, 42*ff*, 58
– field 21*ff*, 42*ff*
– for QSAR, from CoMFA 169, 181
– geometric 21, 50
– hydrogen-bonding 30, 42, 46
– indicator variables 21, 54*ff*, 85, 91, 96
– linear free energy-related 115
– lipophilicity 21, 27*ff*, 58

- molar refractivity 21*ff*, 40*ff*, 49
- other 49*ff*
- physicochemical 21*ff*
- polarizability 21, 40*ff*, 58
- quantum-chemical 21, 42, 46, 85, 126
- resonance 21*ff*, 42*ff*
- spectroscopic 42, 45
- steric 21, 47*ff*, 58
- STERIMOL 22, 48*ff*
- table 23*ff*
- topological 21, 50*ff*

parametrization, of enantiomers 151
pargylin 10
parsimony, principle of 59, 61
partial charges 42
partial F test 94*ff*
partial least squares method, see PLS
partition coefficients 5, 21, 22, 27*ff*
- acids and bases 77*ff*
- additive constitutive parameter 34
- and connectivity values 53
- apparent P_{app} 77*ff*
- atom-based calculation 37
- calculation 28, 34*ff*
- CLOGP program 35*ff*
- different solvent systems 28
- experimental 36
- measurement 27, 32
- nonadditivities 37
- pH dependence 77*ff*
- principal component analysis 29
- step by step calculations 35

partitioning 27*ff*
- alkylpyridines 28
- equilibrium, of an acid 77
- of drugs 15
  - into a membrane 31
- of ion pairs 80

passive diffusion, in MDR 125
pattern recognition 85, 100
- by neural networks 89
- in QSAR 6

pattern searching, pharmacophoric 151
PCA, see principal component analysis
PDE inhibitors, activity-activity relationships 134
penetration, blood-brain barrier 128
penfluridol 17
penicillin 7, 10
peptides, receptor selectivity 32
perfused organs 15
permeability, cell wall 117
perturbation method 18
Pfeiffer's rule 149, 150, 151
PGDP/water system
- f values 34
- partition coefficients 30

pH absorption profiles 82
pH dependence
- absorption rate constants 84
- log P_{app} values 77*ff*
- partition coefficients of amino acids 79
- transfer rate constants 84

pH partition
- hypothesis 82
- profiles 78*ff*

pH shift 82
pH values 77*ff*
pharmacodynamic effects 15, 115
pharmacokinetic data 15, 115, 126
- acids and bases, pH dependence 83
- mathematical models 83

pharmacokinetics, QSAR 126, 129
pharmacological models 137
pharmacological profiles, of enantiomers 149
pharmacophore 50, 151
- CATALYST program 161
- CoMFA 159*ff*
- conformation 12
- hypotheses 6, 161
  - in 3D QSAR 161
- identification 155
- recognition, Apex-3D program 158

pharmacophoric groups 3
pharmacophoric pattern searching 151
Phe-X-Phe conformations 18
phenanthreneaminoalkylcarbinols, QSAR 130, 139

phenethylamines
– binding site model 152
– QSAR 88, 140
phenol 17
phenol coefficient 145
phenol sulfotransferase, QSAR 121
phenols, QSAR 141
phenoxyacetic acids 5, 28
phenoxypropionic acids, QSAR of enantiomers 149
phenylalkylamines, QSAR 141
phenylbutazone 16
phenyldialkyltriazenes 125
phenylenediamineplatinum dichlorides, CoMFA 169
phenylethanolamine-N-methyltransferase, QSAR 147
phenyl ethers, QSAR 66
phenyl glucosides
– binding to concanavalin A 121
– CoMFA 169
phenylimidazoles, multiple binding modes 163, 164
phenylmethylcarbamates 146
phenylquinolinylmethanols, QSAR 139
phenyltriazines, QSAR 124
phenyltropane carboxylic acids
– CoMFA 171, 175
– similarity matrices 174, 175
phorbol esters, QSAR 131*ff*, 177
phosphinates 17*ff*
phosphodiesterase inhibitors 134
phospholipids 31
phosphonamidates 17*ff*, 20
phosphonates 17*ff*
phosphonic acids, chiral, QSAR 150
phosphorylase inhibitors, multiple binding modes 163
photosynthetic reaction center 8
physicochemical
– model 96
– parameters 21*ff*, 85, 91
– – range 96
– properties, and group contributions 65*ff*
pI_{50} values 16
pK_a values 42, 45, 77*ff*
– and log P values, determination 83
– CoMFA 169
– 3D QSAR 175
– experimental 36
– of benzoic acids, CoMFA 169
– similarity matrices 174
pK_a–pH approximation 81
pK_{HB} values 46
placenta barrier 129
placental transfer ratios 129, 177
plasmin inhibition, QSAR 141
PLS analysis 6, 26, 27, 85, 100, 101*ff*, 141, 166*ff*
– advantages 106
– algorithm 102
– bootstrapping 104, 105
– components 102
– – selection, in CoMFA 172
– cross-validation 102*ff*
– diagram 102
– genetic algorithms 167
– GOLPE variable selection 167
– in CoMFA 166*ff*
– in DYLOMMS 159
– in HASL 158
– in QSAR 106, 181
– nonlinear version 105
– in pattern recognition 85
– predictions 103*ff*
– predictive ability 102, 105
– results 102
– robustness 105
– similarity matrices 174, 175
– statistical parameters 103
– validity 102
– variable selection 105, 167, 174, 175
– variance of variables 168
– vectors 101*ff*
polar interactions 1
polar properties 21
polarity 26
polarizability 21, 40*ff*
– parameters 21, 40*ff*
– – in QSAR 58

– pattern 181
POMONA database 36, 39
porin, 8
positive inotropic responses, PDE inhibitors 134
potential maps, DYLOMMS 159
prazosin 134
predictions
– carcinogenicity 85*ff*
– from Free Wilson analysis 64, 65
– from Hansch analysis 61
– from PLS models 104
– in QSAR 3, 122, 140, 179
– prospective 85
– toxicity, criticism 85
– value of 86
predictive ability
– CoMFA models 167, 168
– Free Wilson analysis 139
– HASL method 158
– in PLS analysis 102, 105
– of a model 100
– of CoMFA models 167
– of DYLOMMS models 159
– of Hansch equations 61
predictive value, CoMFA models 167
preselection, of variables 85
PRESS value, cross-validation 103, 104
principal component analysis 6, 22, 26, 85, 101, 137
– DYLOMMS 159
– in design 113
– partition coefficients 29
– scores 27
principal component regression analysis
– Free Wilson model 65
– stepwise 141
principal components 101
– in PLS analysis 102
– of properties 113
principal properties 101
– of amino acids 113
– of substituents 114
principle of parsimony 59, 61
probability
– model, see McFarland model
– of success, predictions 86
– to arrive at a receptor site 57
probe atoms 6
– GRID 155
– in CoMFA 164
problems
– in QSAR 181
– neural networks 89
progestational activities, QSAR 147
progesterone
– hypermolecule 146
– receptor binding, QSAR 146
prolyl endopeptidase inhibition, CoMFA 171
prontosil rubrum 7
properties of proteins, in the crystal 8
property-property relationships 57
propicillin absorption, pH dependence 84
propranolol
– buccal absorption 81, 82
– log P calculation 37
– log P values 31
– pH partition profile 78
propylene glycol dipelargonate, see PGDP
prostaglandin analogs, CoMFA 171, 172
proteases, comparative QSAR 180
protein binding
– model 69
– QSAR 129
– specific and unspecific 121
protein-protein surface interactions 19
proteins
– conformations 8
– crystallography 3, 6, 17, 19, 179, 182
– crystals, functional properties, 8
– 3D structures 8, 116, 161*ff*
pseudoequilibrium systems 74
pseudosubstituents 63, 114
purine biosynthesis 116
purine nucleoside phosphorylase, *multiple binding modes 163, 165*

purine-sulfonamides, antitumor activities 158
purines, binding modes 165
purinones, QSAR 179
pyrenamines, QSAR 179
pyrimidinone H_2-antihistaminics 152

QSAR 1*ff*, 15, 16
– abstracts 6, 7
– analyses 115*ff*, 139*ff*, 169*ff*
– and multiple binding modes 165
– best equation 58*ff*, 61, 96
– comparative 132, 141, 180
– development 4
– drug design and 179
– enzyme inhibitors 121
– equations, significance and validity 95
– future development 180, 181
– history 4*ff*, 159
– introduction 1*ff*
– lateral validation 180
– methodology 5
– models, validity 99
– multiple binding modes 163
– paradigm 179
– parameters, from CoMFA 169, 181
– philosophy 3
– predictions 122, 140
– problems 181
– publications 6
– relevance 61
– strategy 3
– studies 15, 115*ff*, 139*ff*, 169*ff*
– – reviews 133
– success stories 179
– three-dimensional 149*ff*
quadrants, of Craig plots 109
qualitative data 100
qualitative model, in Hansch analysis 60
quality of fit 93, 95
– PLS analysis 102
quantitative
– models 2, 57
– structure-activity relationships, see QSAR
– structure-pharmacokinetics relationships 126, 129
quantum-chemical parameters 21, 42, 46, 85, 165
– in QSAR 46, 126
– rapid calculation 47
quaternary ammonium compounds 4
– transport rate constants 72, 74
– QSAR 122, 143, 144
quinazolines, QSAR 179

ϱ-σ-π analysis 5
r value, see correlation coefficient
r^2_{PRESS}value 104*ff*
$\mathcal{R}$ values 21*ff*, 42*ff*
– redefinition 44
– validity 44
$\mathcal{R}^+$ and $\mathcal{R}^-$ values 44*ff*
R_f values 32
R_M values 32*ff*
– advantages 33
– correlation with log P values 33
racemates 149
random number simulations, Hansch analysis 62
random walk 2
rank-ordered biological data 107
rat, gastric and intestinal absorption 84
rat, *in situ* absorption model 81
rate constants 15, 115
– of drug transport 69, 72*ff*, 126
– – experimental 73
– of metabolic degradation 15
– pH dependence 82
– pharmacokinetic 126
rational
– alignment 161
– design, in QSAR 99, 109*ff*
– drug design 2, 3, 7
– selection of substituents 22
reaction constant, ϱ 5
receptor 7*ff*, 31, 115, 151
– 3D structures 8

– affinities 15, 151, 153
– agonist 1
– antagonist 1
– binding 15, 115, 123, 133
– – CoMFA 171, 175
– – similarity matrices 174, 175
– binding site models 152*ff*, 155
– cavity 88
– conformation 13
– function 19
– ligands, QSAR 122*ff*
– map, of DHFR 152
– mapping 151*ff*
– modeling 8, 157
– – chloramphenicol binding 54
– – REMOTEDISK 158
– occupation 68
– regulation 19
– response 13
– selectivity 19
– sites, hypothetical 152*ff*, 155
– space 163
– subtype selectivity 32
– surface 161
– – mapping 6
recommendations, for Hansch analysis 59*ff*
reduced Free Wilson model 65, 145
redundant data sets 104
reference compound 5
– in Free Wilson analysis 62
reference substituent selection, in Free Wilson analysis 145
refractive index 40, 41
regression analysis 6, 16, 57, 85, 91*ff*, 100, 101
– and PLS analysis 102
– comparison of different models 94
– example 92
– linear multiple 92, 181
– ordinary 100
– procedure 91*ff*
– proper application 99
– stability and validity 99
– stepwise 96*ff*, 158
regression coefficients 20, 96
– from PLS analysis 167
– significance 59
Rekker method 34*ff*
relationships
– between Hansch and Free Wilson analyses 65*ff*
– nonlinear 66, 68*ff*
relevance, of QSAR analyses 61
REMOTEDISC 158
renin inhibition
– CoMFA 171
– QSAR 147
replication (so if) 88
repulsive forces, ligand binding 11
residual variance 93
residuals, in regression analysis 99
resistance
– drug 31, 124*ff*
– methotrexate 124*ff*
– multidrug 31, 50, 125
– trimethoprim 31, 125
resonance
– contributions 45, 60
– effects 44
– parameter $\mathcal{R}$ 21*ff*, 42*ff*
respiration inhibition, tumor cells, QSAR 41
response surfaces 105
restricted rotations 182
restrictions, of Free Wilson analyis 143
retention values, RPTLC 32
retro-thiorphan, binding mode 163, 166
reversed-phase thin-layer chromatography 32
Rhinocladium beurmanni 49, 74
ribonucleosides, antiviral 158
ribosomal protein 153
rigid analogs 12, 88, 89, 151
risk assessment, and QSAR 85
robustness, of statistical methods 101
rodent carcinogenicity, prediction 85*ff*
rotational entropy 12, 17
rule-based algorithms, LUDI 155
rule-based expert system, GOLEM 86

σ bond perturbation parameter S 45
σ constants 4, 5, 21*ff*, 42*ff*
– 3D QSAR 175
– aliphatic σ^* values 45
– compilations 44
– different scales 44
– in QSAR 43
– similarity matrices 174
σ receptor ligands, CoMFA 171
s value, see standard deviation
s_{PRESS} value 103
S values (field-inductive σ bond perturbation) 45
salicylamides, CoMFA 171
salicylic acid, pH partition profile 80
tetrabutyl-ammonium ion pair 80
Salmonella typhimurium 46
– mutagenesis test 85
Sartorius absorption simulator® 126, 127
scaling, of MR values 41
scores, principal components 27
SDEP value 103
searches, in 3D databases 36, 161
selection
– of substituents, in drug design 107
– of variables
– – GOLPE 105
– – in Hansch analysis 59
selectivity, 12
– and QSAR 179
semiempirical methods 19, 40, 47, 160
semiquantitative data 100
sensitivity, of test models 136
sequence specificity of DNA alkylation, HASL 158
sequential F test 59, 61, 94*ff*
sequential simplex technique 110*ff*
serendipity 7
serine proteases 116, 120
– comparative QSAR 180
– inhibitors 10
– QSAR 116, 120, 141, 180
– selectivities 141
serotonin
– 5-HT_2 agonism 152
– antagonists, QSAR 123
– ligands, CoMFA 171
– receptor binding, QSAR 147
serotonin-uptake inhibition, QSAR 123
serum albumin binding 121
shake flask method 32
shape similarity, of molecules 22, 173*ff*
Sheffer's function (incompatible with) 88
SIBIS method 88
side effects, sedative 128
sigmoidal curves 79
significance
– level 95
– of cross-validation results 104
– QSAR equations 62, 95
– of regression coefficients 59
– of regression equations 94
– of terms, in Free Wilson analysis 65
– of variables 96
– probability value 107
significant components, in PLS analysis 103
SIMCA model 85, 101, 107
similarity
– ASP program 173
– chemical 161
– indices 50, 172, 173
– matrices 77, 171, 172*ff*, 181
– – and CoMFA studies 174, 175
– – lipophilicity 171, 175*ff*
– QSAR analyses 172*ff*
– searches 36
simplex technique 100, 110
simulations
– bilinear model 73
– nonlinear models 71
single class discrimination 101
single-point determination 64, 114, 141
singularity problems 63
sinus terms, in QSAR 77
size, of substituents 26
skin permeation, phorbol esters 131
small-angle neutron diffraction 30
SMILES code 35

soaking, protein crystals 8
soft modeling techniques 6, 85
solubility 4, 27, 33, 40, 57
– ions 79
– limited 13, 68
solvation-desolvation balance 10, 182
solvation of drugs, in membranes 31
solvent-accessible surface 40, 49
solvent polarity, scales 28
solvent/water systems, different 28
spasmolytic activities, QSAR 69, 70, 74
Spearman rank correlation coefficient 174
specific binding, to proteins 121
specific biological activities 109
spectral mapping 136, 137
– neuroleptic drugs, diagram 136
spectroscopic data 21, 42, 45
stability
– of a model 99, 100
– of a PLS model 105
– of a regression model 99
stabilization, of conformations 9
standard deviation of error of predictions, SDEP 103
standard deviation s 20, 55, 61, 66, 93*ff*, 96
– in Free Wilson analysis 141
– of regression coefficients 53
– s_{PRESS} value 103
standard system, drug partitioning 28*ff*
Staphylococcus aureus 141
start map, MTD 88
statistical
– design 113, 181
– methods 91*ff*
– – multivariate 100
– – proper application 109
– parameters, PLS analysis 103*ff*
– problems, in Free Wilson analysis 147
– requirements, CoMFA studies 172
– significance 94
– unicorns, in QSAR 77
stepwise
– approaches, Topliss scheme 113
– derivation, Hansch equations 139*ff*
– discriminant analysis 100
– Free Wilson analysis 86
– iteration of β, bilinear model 73
– principal component regression 65, 141
– regression analysis 59, 85, 96*ff*, 106, 158
– – in Free Wilson analysis 145
– selection of variables 96*ff*
stereochemistry, and drug action 149*ff*
stereospecificity, drug-receptor interactions 7
steric
– bulk 22, 59
– constraints, binding site 49
– crowding of ligand groups 20
– field, CoMFA 6, 164, 167
– fit 1
– interactions 1, 181
– – contour maps 168
– overlap, CoMFA steric fields 165
– parameters 21, 41, 47*ff*
– – CoMFA 169
– – compilations 49
– – in QSAR 57*ff*
– properties 21
– – and connectivity values 53
STERIMOL parameters 22, 48*ff*
steroid-binding globulins, 3D QSAR 159, 171, 173*ff*
steroids
– CoMFA 159, 160, 168, 171*ff*
– contour maps 168
– QSAR 133
– similarity matrices 173*ff*
structural variation, in data sets 109
structure-activity relationships 5, 15, 28, 57*ff*, 115*ff*, 139*ff*
– and multiple binding modes 163
– comparative 115, 124, 132, 141, 180
– nonlinear 66, 68*ff*
– nonspecific 57, 123
structure and function, of receptors 8

structure-based drug design 116, 179
structure-pharmacokinetics relationships 126*ff*
Student t values 94
– table 95
substituent group contributions 5, 63*ff*
substituent parameters, for correlation analysis 21*ff*
substituents
– cluster analysis 22
– principal properties 114
– rational selection 22
– size 26
substrates 1, 20
substructure searches 36
subtilisin, comparative QSAR 180
success stories, in QSAR 179
succinate dehydrogenase inhibitors, QSAR 88
suizide enzyme inhibitors 10
sulfa drugs, antibacterial 124
sulfanilamide 7
sulfonamide carboanhydrase inhibitors 153
sulfonamides 45
– binding 152
– CoMFA 172
suloctodil 16
sum of squared errors of predictions, PRESS 103*ff*
superdelocalizability 46
superposition
– atom-by-atom 88
– of molecules 161
– of ring systems 88
superstructure 86
– searches 36
surface
– area, CoMFA 169
– complementarity 1
– solvent-accessible 49
– tension 40
symmetrical curves, McFarland model 71
symmetrical matrix 93
symmetry equations 63
t values, Student 94*ff*
table
– F values 94
– substituent parameters 23*ff*
– t values 95
tangens terms, in QSAR 77
TCDD, see tetrachlorodibenzodioxin
template field, CoMFA field fit 166
teratogenicity 85, 109
test models 137
– sensitivity 136
– *in vitro* 115
test series, design 109*ff*
test set 85, 86, 100, 104
– in pattern recognition 85
tetrachlorodibenzodioxin receptor binding 123
– CoMFA 171
tetracyanoethylene 46
tetrahydrocannabinol analogs, CoMFA 171
theoretical aspects, CoMFA 172
theoretical models, nonlinear relationships 70*ff*
thermodynamic
– analysis 13
– control, transport rate constants 73
– perturbation method 18
– relationships 20
thermolysin
– active site 166
– inhibitors 17*ff*, 20
– – CoMFA 171
– – binding mode 166
– – interaction geometries 163
– – multiple binding modes 161
thin-layer chromatography, reversed-phase 32
thiophosponates, QSAR 96, 98
thiopurine methyltransferase, QSAR 121
thiorphan, binding mode 163, 166
three-compartment system 72*ff*
– equilibrium model 75
– water/n-octanol/water 73
three-dimensional

– grid, in CoMFA 159
– QSAR, see 3D QSAR
thrombin inhibitors
– QSAR 141
– multiple binding modes 163
thromboxane synthase, QSAR 121
thymidilate synthase, QSAR 121
thymidine kinase, QSAR 121
thyroid hormones
– multiple binding modes 163
– receptor affinity, QSAR 123
– QSAR 4, 123
thyromimetic activities, QSAR 147
thyroxine, 17
– binding protein affinity 123
timolol, log P values 31
tissue levels 126
Topliss operational schemes 110*ff*
– applications 113
– diagram 112
topological
– indices 21, 50, 53, 85
– – critizism 53
– – in QSAR 53
– pharmacophore methods 86, 88
– state 53
topology-based electronic descriptors 53
topology of drugs 182
total topological index 53
toxic activation of MPTP analogs
– CoMFA 171, 175
– similarity matrices 174, 175
toxicity 4, 15, 109
– alkanes, CoMFA 172
– benzoic acids 5, 43
– CoMFA 170
– QSAR 43, 85, 100, 147, 179
– predictions, criticism 85
training, of neural networks 89
training set 85, 86, 100, 103
– in pattern recognition 85
– in QSAR 181
tranquillizing activities, QSAR 147
trans-membrane domains 8
transfer rate constants 126, 176
transformation, of biological data 15
transition state 149, 161
– stabilization energies 19
translational entropy 12, 17
transport glycoprotein GP-170 125
transport of drugs 2, 3, 5*ff*, 20*ff*, 179, 181
– and CoMFA 167
– computer simulation 68*ff*
– rate constants 72*ff*, 84
– – experimental 73, 84
triangularisation of a matrix 93
triazines
– CoMFA 172, 175
– similarity matrices 174, 175
trifluoperazine, pH partition profile 79
trimethoprim 116*ff*
– analogs 31
– – ligand design 116
– – multiple binding modes 163
– diphenylsulfone analog 117
– in DHFR active site 155
– resistance 31, 125
trinitrobenzene 46
trypsin inhibition
– comparative QSAR 180
– QSAR 120, 141
tumor cells
– drug resistance, QSAR 125
– inhibition, QSAR 42, 124
– multidrug resistance, QSAR 31, 50, 125
– response to estrogens, CoMFA 172
tumor promoting activities, phorbol esters 131
tyrosine kinase inhibition, CoMFA 171
tyrosyl-tRNA synthetase, 19

unexplained variance 93, 104
unicorns, statistical 77
unionized form, correction for 82
unsymmetrical curves, bilinear model 75
upper limit of correlation coefficient r 66

uptake inhibitors, QSAR 123
uridine phosphorylase, QSAR 121
U. S. National Toxicology Program 85

valence connectivity values 53
validation
- of Hansch equations 61
- of QSAR studies 180
- procedure, in QSAR 181
validity
- of a PLS model 102
- of a regression model 99
- of QSAR equations 62, 95, 99, 181
valproate 17
van der Waals
- potentials 164
- radii r_v 47
- spheres 165
  - overlap 11
- volumes 41, 49, 128
  - CoMFA 169
vaporization enthalpy 22
variable elimination, Free Wilson analysis 144
variable selection
- GOLPE 105, 167
- in Hansch analysis 59
- in PLS analysis 105, 172
- stepwise 96*ff*
variables
- dummy 21, 54*ff*
- indicator 21, 54, 85, 91, 96
- interrelated 101
- significance 96
- stepwise selection 96*ff*
variance 93
- biological data 95
- of variables, loss of 168
- unexplained 93, 104
variance-covariance matrix 113
vector addition, dipole moments 46
vectors, orthogonal 101
verapamil analogs, QSAR 123, 152*ff*
vertebrate DHFR 118
viral coat protein ligands
- CoMFA 171
- multiple binding modes 163
volume 22
Voronoi binding site models, 157

Walter Reed Army Institute for Medical Research 130
water dragging effect 29, 30
water/gas phase equilibrium constants 30
water molecules, ordered 41
water/n-octanol/water three-compartment system 72*ff*
water solubility 4
weighting of variables 107, 166, 171
Wiener index 53
WIZARD, 3D structure generation 160
wrong predictions, in QSAR 122

X block, in PLS analysis 101*ff*
X-ray structures
- conformations 161
- dihydropyridines 51
- enzyme inhibitor complexes 8, 116, 161*ff*
- in CoMFA 167
xanthine oxidase, QSAR 121

Y block, in PLS analysis 101*ff*

z scales, amino acids 26, 113