

# **QSAR: Hansch Analysis and Related Approaches**

by Hugo Kubinyi



Weinheim · New York  
Basel · Cambridge · Tokyo



Hugo Kubinyi

# **QSAR: Hansch Analysis and Related Approaches**



# **Methods and Principles in Medicinal Chemistry**

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

# **QSAR: Hansch Analysis and Related Approaches**

by Hugo Kubinyi



Weinheim · New York  
Basel · Cambridge · Tokyo

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

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

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

Prof. Hendrik Timmerman  
Faculty of Chemistry  
Dept. of Pharmacochimistry  
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  $\pi$ , 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 75<sup>th</sup> 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.

# 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

<b>6. Design of Test Series in QSAR</b>	<b>109</b>
<b>7. Applications of Hansch Analysis</b>	<b>115</b>
7.1. Enzyme Inhibition	116
7.2. Other <i>in vitro</i> Data	121
7.3. Pharmacokinetic Data	126
7.4. Other Biological Data	129
7.5. Activity-Activity Relationships	133
<b>8. Applications of Free Wilson Analysis and Related Models</b>	<b>139</b>
<b>9. 3D QSAR Approaches</b>	<b>149</b>
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
<b>10. Summary and Conclusions</b>	<b>179</b>
<b>References</b>	<b>183</b>
<b>Index</b>	<b>211</b>

# 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"  $\Phi$  must be a function of the chemical structure  $C$  (eq. 1).

$$\Phi = f(C) \quad (1)$$

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,  $\Delta C$ .

$$\Delta\Phi = f(\Delta C) \quad (2)$$

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 nonspecific 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).

$$\log k_{R-X} - \log k_{R-H} = \rho\sigma \quad (3)$$

$$\log \tau_i - \log \tau_{Et} = \alpha\beta \quad (4)$$

$\tau_i$  in this "biological Hammett equation" stands for the activity value of the  $i^{\text{th}}$  member of a series,  $\tau_{Et}$  is the biological activity value of the ethyl compound of the same series,  $\beta$  is a substituent constant (corresponding to the electronic  $\sigma$  parameter

in the Hammett equation), and  $\alpha$  is a constant characterizing the biological system, which corresponds to the Hammett reaction constant  $\rho$ . 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  $\sigma$  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 " *$\rho$ - $\sigma$ - $\pi$  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  $\pi$  (eq. 6), to be used instead of measured  $\log P$  values (like Hammett  $\sigma$  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].

$$\log 1/C = a \log P + b\sigma + \dots + \text{const.} \quad (5)$$

$$\pi_X = \log P_{R-X} - \log P_{R-H} \quad (6)$$

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

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  $\mu$  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.

$$\log 1/C = \sum a_{ij} + \mu \quad (8)$$

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].

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

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 ( $\alpha$ -glucosidase and  $\beta$ -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

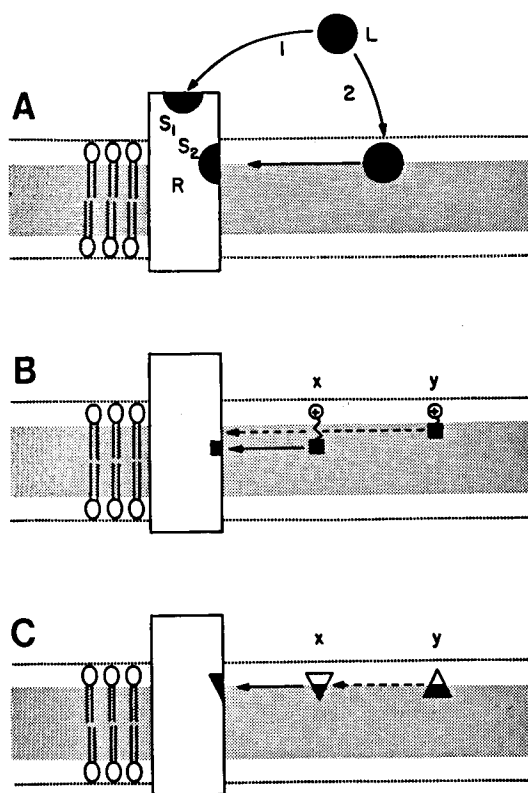
the enzyme lysozyme, which is also found in egg white, milk, blood serum, tears, saliva, some other secretions and tissues of animals, and in plant latices (cited from [99]).

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 Herbert [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  $\Delta G$  between the free states of both partners

and the drug-receptor complex [DR], which is made up from the enthalpy change  $\Delta H$  and the entropy change  $\Delta S$  (eq. 10). The free energy  $\Delta G$  is related to the equilibrium constant  $K$  for the reaction  $D + R = [DR]$  by eq. 11.

$$\Delta G = \Delta H - T \Delta S \quad (10)$$

$$\Delta G = -2.303 RT \log K \quad (11)$$

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  $170\text{--}600 \text{ kJ} \cdot \text{mol}^{-1}$ . 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  $\Delta H$  considerably, a fact which is much too often neglected in quantitative structure-activity analyses.

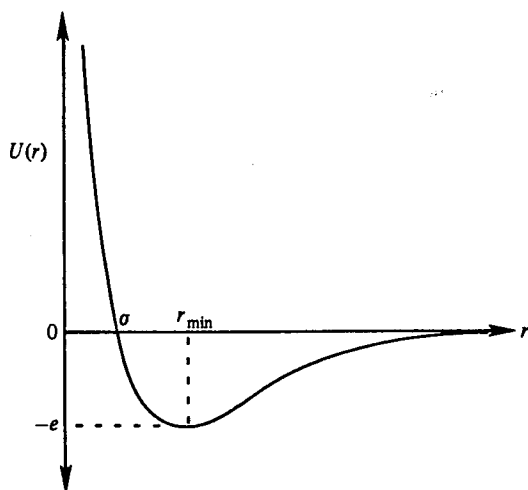
Most interactions between charged groups include hydrogen bonds, *e.g.* between a positive ammonium group and a negative carboxylate, phenolate, phosphate, phosphonate, or sulfate group. Different energy values are given in literature (energies calculated *in vacuo* must not be compared with these values 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  $15\text{--}19 \text{ kJ} \cdot \text{mol}^{-1}$ , while those of neutral hydrogen bonds were estimated to be  $2\text{--}6 \text{ kJ} \cdot \text{mol}^{-1}$  [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



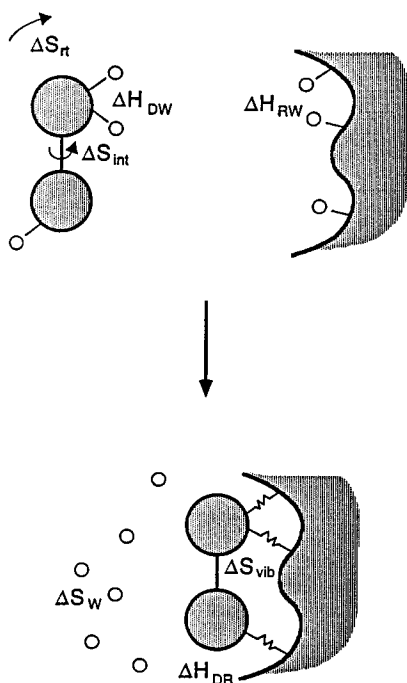
hydrogen atom instead, have been compiled for different enzymes [124]. From a recent comparison of the binding energies of amide-amide hydrogen bonds in aqueous solution and in nonpolar solvents it was concluded that earlier values of neutral hydrogen bond energies may be too small [125].

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;  $\sigma$  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.  $\Delta H_{DW}$  and  $\Delta H_{RW}$  are the enthalpies of hydration of the drug and the receptor, respectively,  $\Delta H_{DR}$  is the enthalpic contribution of the drug-receptor interaction,  $\Delta S_{rt}$  is the overall rotational and translational entropy in solution, and  $\Delta S_{int}$  the internal rotational entropy of the free drug.  $\Delta S_w$  is the increase in entropy due to the release of bound water molecules (small circles) and  $\Delta S_{vib}$  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  $2 \text{ kJ} \cdot \text{mol}^{-1}$ , the contribution of a phenyl ring is about  $8 \text{ kJ} \cdot \text{mol}^{-1}$  [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